MolGuide: 2D Molecular Optimization with Preserved Structural Motifs Guidance

Xinyang Liu[1], Bowei Fang[1], Ping Zhang[2], Bo Chen[1], Huaxiu Yao[3]
[1]Xidian University
[2]The Ohio State University
[3]UNC-Chapel Hill
{xinyangatk, fbw}@stu.xidian.edu.cn, zhang.10631@osu.edu,
bchen@mail.xidian.edu.cn, huaxiu@cs.unc.edu

Abstract

Molecular optimization is a very important practical problem in drug discovery. Deep learning based methods has been raised increasingly and provided useful solutions. However, existing methods for 2D molecular optimization mainly adopt the auto-regressive paradigm which reduce the diversity of generated structural motifs due to the limitation during training. Recently, various diffusion model has achieve promising results in image transfer and editing, which motivate us adapt it in a one-shot generation mode to molecular optimization. In this paper, we present a conditional discrete diffusion model by incorporating features of source molecules into the modeling of target molecules via cross-attention module. Besides, to address the challenge of controlling generation process, we propose a new method that optimizing the molecular with preserved structural motifs guidance with the assistance of ChatGPT. We conduct a toy experiment and our model achieves a promising results compared with other auto-regressive based methods.

Introduction

The goal of 2D Molecular Optimization is to obtain target molecules that not only have the desired properties but are also similar to the source molecule. Work to data using deep learning models has provided effective methods for this task by directly generating desirable molecules with various generative models. Specifically, these methods have been primarily categorized into two groups based on their generation mode: atom-based methods and fragments-based methods. Existing methods have been achieved promising performance due to their auto-regressive generation paradigm which focus only on generated part while predicting the current state (Jin et al. 2018; Jin, Barzilay, and Jaakkola 2020, 2018; Fu, Xiao, and Sun 2020; You et al. 2018; Liu et al. 2018; Du et al. 2022; Zhang et al. 2023). Under a careful training strategy, molecular can be optimized well by constraining next prediction state at every step. However, there are still limitation to the complex training strategy, difficulty in parallelization and the diversity of generated structural motifs, which is usually find in fragment-based methods (Jin et al. 2018; Jin, Barzilay, and Jaakkola 2020) that require collecting vocabulary list in advance.

Motivated by the wild success of applying diffusion models to image transfer and editing (Song et al. 2020; Ho, Jain, and Abbeel 2020; Austin et al. 2021), the methods employing diffusion model on 3D molecular optimization task have been raised increasingly (Xu et al. 2022; Hua et al. 2023). But one important difference with 3D molecular optimization is that 2D molecular lack detailed spatial coordinates so that learning a good distribution of the target molecules while keeping similarity conditional on the source molecules is intractable. Specifically, too much diversity during one-shot based generation process makes the similarity of molecular before and after optimization hard to guarantee, which is also a common challenge in one-shot generation (Liu et al. 2021). Besides, Recent advancements in large language models (LLMs), such as ChatGPT, have demonstrated remarkable promise in various domains, including drug discovery (OpenAI 2023; Liu et al. 2023b; Guo et al. 2023; Li et al. 2023a; Qian et al. 2023). However, existing works mainly focus on investigating the capabilities of conversational LLMs on chemical reaction and retrosynthesis. While drug editing, a critical task in the drug discovery pipeline, remains largely unexplored.

In this paper, we propose **MolGuide**, a conditional diffusion model under novel guidance framework with the assistance of ChatGPT for molecular optimization. Specifically, we first design a graph condition mechanism that is adapted to diffusion model (Vignac et al. 2022). Then we construct potential structural motifs with ChatGPT to guide the generation process from multiple perspectives. Moreover, we incorporate the Optimal Transport (OT) Distance between molecular and structural motifs into energy-function to correct the generation process. Empirically, MolGuide, with a one-shot based generation mode, achieves a promising results compared with methods in the auto-regressive generation manner.

The Proposed MolGuide

In this section, we present the **MolGuide** with preserved structural motifs guidance for molecular optimization. Atoms and bonds in molecular can be naturally represented as nodes and edges in graphs $\mathcal{G} = (X, E)$, where $X \in R^{n \times a}$ is the node feature matrix and $E \in R^{n \times n \times b}$ is the edge attributes tensor. Let n be the maximum number of nodes, a and b denote the number of possible types

of nodes and edges, respectively. Then we have $x_i \in R^a$ to denote its one-hot encoding and $E_{ij} \in R^b$ to denote one-hot encoding of each egde. In particular, we consider the virtual edge as an additional node type. Given the molecular in source domain, our goal is to generate a set of molecular in target domain. Specifically, The generated target molecules should have better properties while being structurally similar to the source molecule. For a pair of data (M_X, M_Y) denoted molecular in source domain and target domain respectively, we construct their graph $(\mathcal{G}_{M_X}, \mathcal{G}_{M_Y})$ The goal of 2D Molecular Optimization is to obtain target molecules that not only have the desired properties but are also similar to the source molecule. Motivated by the promising results from applying diffusion models to image transfer and inpainting, we adapt such a one-shot generative model to the 2D molecular optimization and design a new framework with preserved structural motifs to guide the optimization process. We will give a full overview of our model in the following subsections.

Graph Conditional Diffusion Model for Molecular Optimization

Discrete Denoising Diffusion Model for Graph Generation (**DiGress**) (Vignac et al. 2022) are a variant of diffusion models for 2D molecule data. DiGress diffuse separately on each node and edge feature via transition probabilities defined by the matrices $Q^t = (Q_X^t, Q_E^t)$, e.g., $[Q_X^t]_{ij} = q(x^t = j|x^{t-1} = i)$ and $[Q_E^t]_{ij} = q(a^t = j|a^{t-1} = i)$. At time t, form $\mathcal{G}^t = (X^t, E^t)$ by sampling each node and edge type from a categorical distribution which are formulated by:

$$\begin{aligned} q(\mathcal{G}^t|\mathcal{G}^{t-1}) &= (X^{t-1}Q_X^t, E^{t-1}Q_E^t) \\ q(\mathcal{G}^t|\mathcal{G}) &= (X\bar{Q}_X^t, E\bar{Q}_E^t) \end{aligned} \tag{1}$$

for $\bar{Q}_X^t = Q_X^1 ... Q_X^t$ and $\bar{Q}_E^t = Q_E^1 ... Q_E^t$. To generate samples, the forward process is reversed using a Markov chain:

$$q(\mathcal{G}^{t-1}|\mathcal{G}^t, \mathcal{G}^0) \propto \mathcal{G}^{t-1}(\mathrm{Q^t})^\top \odot \mathcal{G}^0 \bar{Q}^{t-1} \tag{2}$$

The predicted clean graph $\mathcal{G}^0$ is parameterized by a denoising neural network $\epsilon_\theta(\mathcal{G}^t, t)$, and is trained using a cross-entropy loss l between the predicted probabilities $\hat{p}_G^0 = (\hat{p}_X^0, \hat{p}_E^0)$ for each node and edge and the true graph G, with their contributions weighed by a hyperparameter $\lambda \in R^+$:

$$\begin{aligned} l(\hat{p}_G^0, \mathcal{G}) = &\sum_{i=1}^{n} \text{cross-entropy}(\hat{p}_{X_i}^0, X_i) \\ &+ \lambda \sum_{1 \le i,j \le n} \text{cross-entropy}(\hat{p}_{E_i}^0, E_i) \end{aligned} \tag{3}$$

DiGress show that the distribution of generated samples $p(\mathcal{G}^0)$ is exchangeable to any permutation if the noise prediction network is permutation equivariant and the transition probabilities is equivariant to joint permutations of $\hat{p}_\theta(\mathcal{G})$ and $\mathcal{G}^t$. Our work is based on their implementation, but we adapted it to work on 2D molecular optimization.

To better guide the generation process of target molecular conditional on the structural information of the source molecular, we carefully design a conditional mechanism (Rombach et al. 2022) that matching the target molecular with the source molecular by incorporating the $\mathcal{G}_{M_X}$ into the denoising neural network $\epsilon_\theta(\mathcal{G}_{M_Y}^t, t, \mathcal{G}_{M_X})$. More specifically, we first employ a pre-trained graph encoder $\tau_{M_X}(\cdot)$ that project $\mathcal{G}_{M_X}$ to the intermediate representation $\tau_{M_X}(\mathcal{G}_{M_X})$. Then we map it to the intermediate layers of the graph transformer network via a Node & Edge cross-attention layer implementing: $\mathrm{CrossAttention}(Q, K, V) = \mathrm{softmax}(\frac{QK^\top}{\sqrt{d}} \cdot V)$, with:

$$\begin{aligned} Q &= W_Q^{(i)} \cdot f_i(\mathcal{G}_{M_Y}^t), \\ K &= W_K^{(i)} \cdot \tau_{M_X}(\mathcal{G}_{M_X}), \\ V &= W_V^{(i)} \cdot \tau_{M_X}(\mathcal{G}_{M_X}) \end{aligned} \tag{4}$$

where $f_i(\mathcal{G}_{M_Y}^t)$ denote a intermediate representation of the graph transformer implementing ϵ_θ and $W_Q^{(i)}, W_K^{(i)}, W_V^{(i)}$ are learnable projection matrices. After aligning the feature between Nodes & Nodes and Edges & Edges in two domains, we can easily match the their relation between two graphs.

Molecular Optimization with Preserved Structural Motifs (PSM)

We notice that there are many structural similarities between the target molecule and the source molecule. Inspired by the prompt learning using in natural language process and image generation (Liu et al. 2023a; Li and Liang 2021; Wang et al. 2023; Zhou et al. 2022), the molecular optimization can be view as a generation process with some fixed structural motifs as prompt. For a pair of data (M_X, M_Y), we aim to optimize the Eq. 3 with preserved common structural motifs between the molecules M_X and M_Y at each time-step t. Thanks to the method proposed by (Jin, Barzilay, and Jaakkola 2020), we can easily decomposing a molecule M into disconnected fragments by breaking all the bridge bonds that will not violate chemical validity. Specifically, we extract two motifs set $S^{M_X} = \{S_i^{M_X}\}_{i=1}^{K_{M_X}}$ and $S^{M_Y} = \{S_i^{M_Y}\}_{i=1}^{K_{M_Y}}$ from the source molecule and the target molecule during training, where K_{M_X} and K_{M_Y} indicate the number of structural motifs belonging to them respectively. Then we filter the common structural motifs from two set:

$$S = S^{M_X} \cap S^{M_Y} \tag{5}$$

where $S = \{S_i\}_{i=1}^{K}$ denote the common structural motifs set having K motifs and we define a set $n_L = \{n_i\}_{i=1}^{i=L}$ that contains all L nodes in the common substructure motifs without repetition. Therefore, we can construct a subgraph $\mathcal{G}_L = (X_L, E_L)$ with nodes set n_L and mask the input graph $\mathcal{G}^t = (X^t, E^t)$ with this subgraph at each time step t:

$$\mathcal{G}_{\text{PSM}}^t = (X_{\text{PSM}}^t, E_{\text{PSM}}^t) \tag{6}$$

$$\begin{aligned} X_{\text{PSM}}^t &= \text{MASK}_X \odot X_L + (1 - \text{MASK}_X) \odot X^t \\ E_{\text{PSM}}^t &= \text{MASK}_E \odot E_L + (1 - \text{MASK}_E) \odot E^t \end{aligned} \tag{7}$$

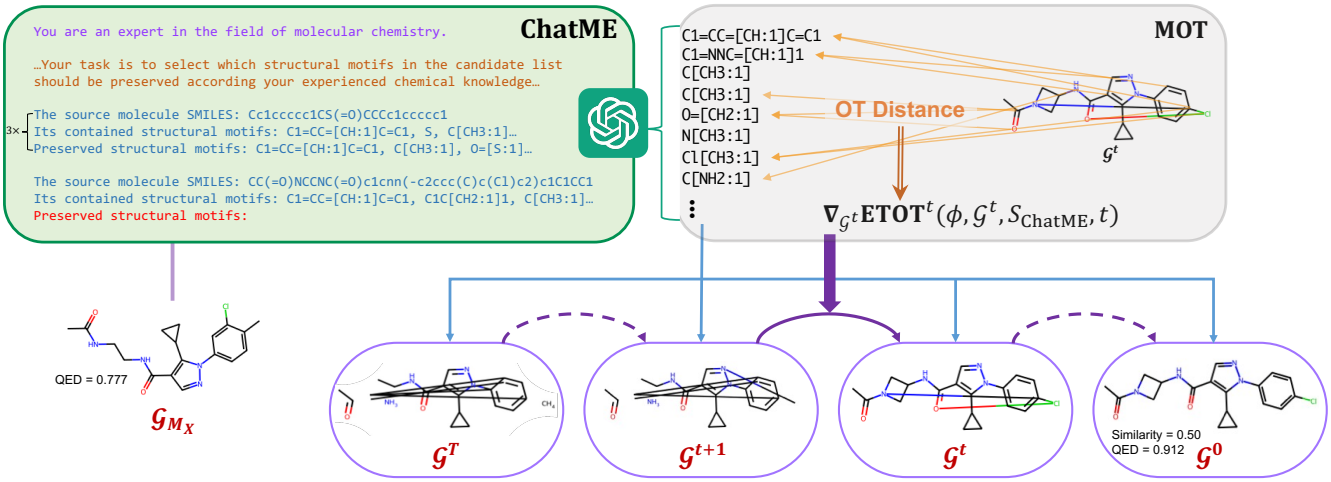

Figure 1: The framework of proposed **MolGuide** for molecular optimization with preserved structural guidance.

where $\text{MASK}_X \in R^{n \cdot a}$ and $\text{MASK}_E \in R^{n \cdot n \cdot b}$ are masks indicating the n_L first nodes (Lugmayr et al. 2022). Then we can replace $\mathcal{G}^t$ in Eq. 3 with $\mathcal{G}^t_{\text{PSM}}$ and train a preserved structural motifs condition denoising neural network $\epsilon_\theta(\mathcal{G}^t_{\text{PSM}}, t)$. During test, the most chanllengable obstacle is obtaining a set of common structural motifs with only source molecules in test data. One possible way is to manually pick out all appropriate common structural motifs using expert knowledge and experience. However, when test data is enormous, it becomes nearly intractable and not conducive to industrial applications. Fortunately, with the capabilities and expert knowledge of large language models (LLMs), such as ChatGPT, on molecular analysis, we can obtain the potential common structural motifs of source molecular by resorting to ChatGPT.

During generation process, we carefully design a **Chat**GPT for **M**otifs **E**xtraction (**ChatME**) module aims to generate potential common structural motifs while given a source molecule and its contained all structural motifs as candidate list. As Fig 1 shows, we first define the role ChatGPT plays, such as a chemical expert, and the description of current task. Then, we construct the question prompt template P consisting of several examples. Based on these, **ChatME** will return a set S_{ChatME} denote the potential common structural motifs in candidate list:

$$S_{\text{ChatME}} = \text{ChatME}(P, M_X, S_{M_X}) \tag{8}$$

where S_{ChatME} can be view as preserved structural motifs that considers expert knowledge and the trade-off between similarity and optimization properties. After that, we get the nodes set n_L^{ChatME} and build the subgraph $\mathcal{G}_L^{\text{ChatME}} = (X_L^{\text{ChatME}}, E_L^{\text{ChatME}})$. So far, we can update the X^t and E^t with the similar way above to mask the generated graph at each reverse iteration step after sampling $\mathcal{G}^t$.

Molecular-Motifs Aligned Energy-Guidance

Given two discrete probability distribution P and Q on space $\mathcal{X} \in R^d$: $\text{P} = \sum_{i=1}^n \theta_i \delta_{h_i}$, and $\text{Q} = \sum_{j=1}^m \beta_j \delta_{u_j}$, where h_i and u_j are two points in the arbitrary same space $\mathcal{X}$. $\theta \in \Sigma^n$ and $\beta \in \Sigma^m$, the simplex of R^n and R^m, denote two probability values of the discrete states satisfying $\sum_{i=1}^n \theta_i = 1$ and $\sum_{j=1}^m \beta_i = 1$. δ_h refers to a point mass located at coordinate $h \in R^d$. To measure such two discrete distributions, we formulate the OT distance between P and Q as the optimization problem: $\text{OT}(\text{P}, \text{Q}) = \min_{\text{T} \in \Pi(\theta, \beta)} \sum_{i,j} t_{ij} c_{ij}$, with $\text{T}1^m = \theta$, $\text{T}^\top 1^n = \beta$. 1^m is the m dimensional vector of ones. $c_{ij} = c(h_i, u_j) \geq 0$ is the transport cost defined by an arbitrary cost $c(\cdot)$. The optimal transport plan T is often trained by minimizing the OT cost with the iterative Sinkhorn algorithm (Cuturi 2013; Liu et al. 2023c; Li et al. 2023b).

Now we present the details of our proposed module MOT guidance, which aligns **Mo**lecule atoms & bonds in and common structural **Mo**tifs under **O**ptimal **T**ransport framework (**MOT**) for energy-guided molecular optimization. As shown in Fig. 1, MOT consists of four components, the atoms&edges $\text{P} = (\text{P}^X, \text{P}^E)$ set of the target molecular, the preserved structural motifs $\text{Q} = (\text{Q}^X, \text{Q}^E)$ set of the source molecular. and the energy-guidance with OT distance between P and Q.

Time-Independent P set over node & edge embeddings For a target molecular graph $\mathcal{G} = (X, E)$, MOT first apply noise to it at time-step t, and feed the noised graph $\mathcal{G}^t$ to a graph transformer network $g_\phi(\mathcal{G}^t_{M_Y}, t)$ to obtain its nodes features $H^{X,t} = \{h_i^{X,t}\}_{i=1}^n$ and edge features $H^{E,t} = \{h_j^{E,t}\}_{j=1}^{n^2}$, where $h_i^{X,t} \in R^d$ denotes i-th node embedding, $h_j^{E,t} \in R^d$ denotes j-th edge embedding and d is the embedding dimension. For convenience, MOT view each node and edge equally that the weight $\theta^X = \frac{1}{n}$ is a uniform distribution over n nodes and the weight $\theta^E = \frac{1}{n^2}$ is a uniform distribution over n^2 edges. After giving the node embedding matrix $H^{X,t}$, edge embedding matrix $H^{E,t}$ and their weights θ^X and θ^E respectively, MOT obtains the time-independent discrete distribution $\text{P}^t = (\text{P}^{X,t}, \text{P}^{E,t})$ of $\mathcal{G}^t$

Algorithm 1: Training MolGuide.

Input: A pair of graph $(\mathcal{G}_{M_X}, \mathcal{G}_{M_Y})$ and $\mathcal{G}_{M_Y} = (X, E)$
Get common structural motifs S and construct subgraph $\mathcal{G}_{\text{PSM}}$
Sample $t \sim \mathcal{U}(1, ..., T)$
Sample $G^t_{M_Y} \sim X\bar{Q}^t_X \times E\bar{Q}^t_E$
Get new $G^t_{M_Y} = \text{MASK}(G^t_{M_Y}, \mathcal{G}_{\text{PSM}})$ with Eq. (7)
Encode graph condition $\tau(\mathcal{G}_{M_X})$
$(\hat{p}^0_X, \hat{p}^0_E) \leftarrow \epsilon_\theta(\mathcal{G}^t_{M_Y}, \tau(\mathcal{G}_{M_X}), t)$
optimizer.step $(l_{\text{CE}}(\hat{p}^0_X, X) + \lambda l_{\text{CE}}(\hat{p}^0_E, E))$

Algorithm 2: Training MOT.

Input: A pair of graph $(\mathcal{G}_{M_X}, \mathcal{G}_{M_Y})$ and $\mathcal{G}_{M_Y} = (X, E)$
Get common structural motifs S.
Sample $t \sim \mathcal{U}(1, ..., T)$
Sample $\mathcal{G}^t_{M_X} \sim X\bar{Q}^t_X \times E\bar{Q}^t_E$
optimizer.step $(\text{TEOT}^t(\phi, \mathcal{G}^t_{M_X}, S, t))$

over the nodes & edges set at time-step t.

Q **set over motifs embeddings** In addition to the set P, MOT also represents the set of the structural motifs in a source molecular graph as subgraph-level embeddings. Specifically, MOT first feed the graph to a pre-trained graph encoder to get its node embeddings $V^X = \{v^X_i\}^n_{i=1}$ and edge embedding $V^E = \{v^E_j\}^{n^2}_{j=1}$, then pooling corresponding node and edge embeddings in each structural motifs to get motif embeddings as follow:

$$U = \{U_k\}^K_{k=1}, U_k = \text{concat}[\text{Pooling}(V^X_{S_k}), \text{Pooling}(V^E_{S_k})], \quad S = \{S_k\}^K_{k=1} \tag{9}$$

where K denotes the number of motifs in a graph. $V^X_{S_k}$ and $V^E_{S_k}$ are two sets of node embeddings and edge embeddings in the S_k motif, respectively. We choose $\text{sum}(\cdot)$ as $\text{Pooling}(\cdot)$ operation and define the weight $\beta = \frac{1}{K}$ in Q is a uniform distribution over K motifs. After giving the motifs embedding matrix U and its weights β, MOT obtains the discrete distribution Q of $\mathcal{G}$ over the motifs set.

Energy-Guidance with Time-Independent OT Distance For a molecular graph $\mathcal{G}$ and a set of potential preserved structural motifs S, the two discrete distributions P and Q can be viewed as semantic representations from two different domains. MOT bridges the domain semantic gap by minimizing the OT distance of $\text{P} = (\text{P}^X, \text{P}^E)$ and Q, $e.g.$, $\text{OT}(\text{P}, \text{Q})$. Furthermore, for better aligning molecular generation and potential structural motifs, we develop a **T**ime-independent **E**nergy-function based on **OT** Distance **(TEOT)**:

$$\text{TEOT}^t(\phi, \mathcal{G}^t_{M_Y}, S_{M_X}, t) = \text{OT}(\text{P}^t, \text{Q}) \tag{10}$$

where t is he current time-step and $g_\phi(\cdot)$ can be parameterized by equivariant model graph transformer.

Inspired by (Ho and Salimans 2022; Yu et al. 2023; Bao et al. 2022), now we describe preserved structural motifs-guided diffusion model, which guides the optimized molecules of Eq. 2 towards desired potential structural motifs S_{M_X} by leveraging a time-independent energy-function based on OT Distance $\text{TEOT}^t(\phi, \mathcal{G}^t_{M_Y}, S, t)$:

$$p_\phi(\mathcal{G}^t, S_{M_X}) \propto \exp\left(-\lambda \nabla_{\mathcal{G}^t} \text{TEOT}^t(\phi, \mathcal{G}^t, S_{\text{ChatME}}, t)\right) \tag{11}$$

$$\mathcal{G}^{t-1} \sim p_\theta(\mathcal{G}^{t-1}|\mathcal{G}^t) p_\phi(\mathcal{G}^t, S_{\text{ChatME}}) \tag{12}$$

Besides, we can also introduce the molecular fingerprint of the source molecule as a global feature into the denoising neural network $\epsilon(\theta, \mathcal{G}^t, c)$v for option strategy. Then we can easily train a conditional diffusion model with the inherent structural information.

Training and sampling procedures Given a pair of graph data $(\mathcal{G}_{M_X}, \mathcal{G}_{M_Y})$ denoting the source molecular and the target molecular, respectively. We first extract the common structural motifs between molecules in two domain. Then obtain a noised graph with preserved structural motifs and introduce the condition mechanism via encoding the source molecular into $\tau(\mathcal{G}_{M_X})$. After that, all parameters in MolGuide are optimized according to Eq. (3). We summarize the training algorithm of MolGuide at algorithm. 1. Besides, similarly, we train the guidance module given Given a pair of graph data $(\mathcal{G}_{M_X}, \mathcal{G}_{M_Y})$ denoting the source molecular and the target molecular, respectively. After obtaining the common structural motifs S and noised graph $\mathcal{G}_{M_Y}$, MOT represents them as motif embeddings and node & edge embeddings. Parameters of graph transformer model in MOT is optimized under OT framework according to Eq. (10).. We summarize the training algorithm of MOT at algorithm 2. After training the conditional discrete diffusion model and its guidance module, we can sample from MolGuide with algorithm. 3.

Algorithm 3: Sampling from MolGuide.

Input: A source graph $\mathcal{G}_{M_X}$
Get potential common structural motifs S_{ChatME} with Eq. (8) and construct subgraph $\mathcal{G}^{\text{ChatME}}_L$ with L nodes.
Sample $G^T \sim q_X(n) \times q_E(n)$
Encode graph condition $\tau(\mathcal{G}_{M_X})$
for $t = T$ **to** 1 **do**
  $(\hat{p}^0_X, \hat{p}^0_E) \leftarrow \epsilon_\theta(\mathcal{G}^t, \tau(\mathcal{G}_{M_X}), t)$
  $p_\phi(\mathcal{G}^t, S_{\text{ChatME}}) \propto \exp\left(-\lambda \nabla_{\mathcal{G}^t} \text{TEOT}^t(\phi, \mathcal{G}^t, S_{\text{ChatME}}, t)\right)$
  Sample $\mathcal{G}^{t-1} \sim p_\theta(\mathcal{G}^{t-1}|G^t) p_\phi(\mathcal{G}^t, S_{\text{ChatME}})$
  Get new $\mathcal{G}^t = \text{MASK}(\mathcal{G}^t, \mathcal{G}_{\text{ChatME}})$ with Eq. (7)
end for

Experiments

We follow the experimental design by (Jin et al. 2018) and evaluate our model on their molecular optimization task. At test time, the molecular similarity between M_X and output M_Y must be over a threshold $\text{sim}(M_X, M_Y) \geq \delta$.

Dataset We train and evaluate our model on **Drug-likeness (QED)** task. The QED score measures a compound's drug-likenes. In this task, the model aims to optimize molecules with QED scores from the lower range $[0.7, 0.8]$ into the higher range $[0.9, 1.0]$ and the similarity constraint is $\text{sim}(M_X, M_Y) \geq \delta$ and δ is 0.4 here. The

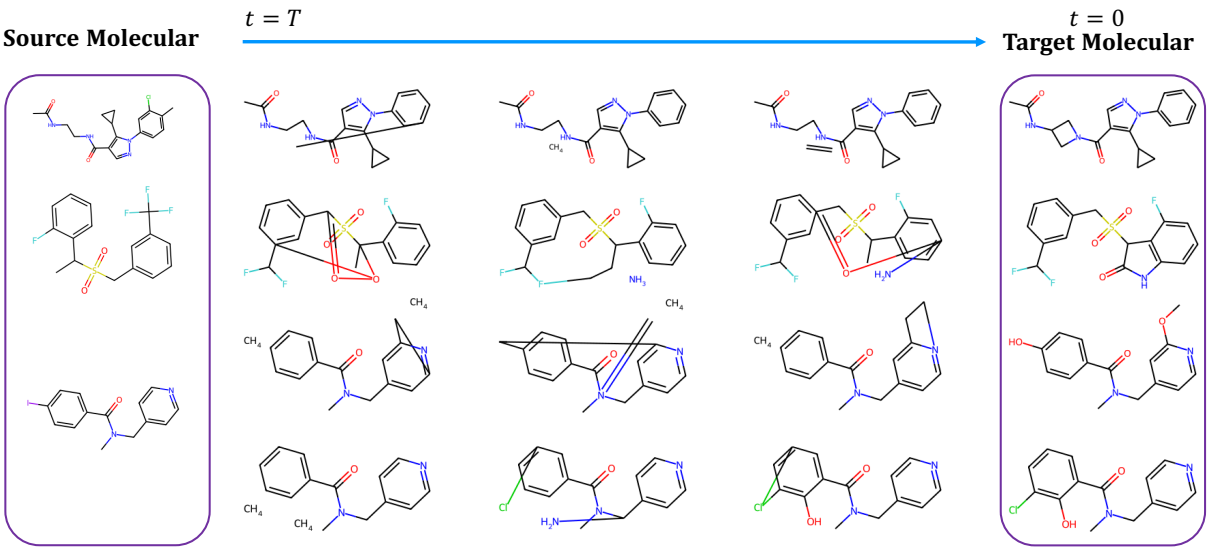

Figure 2: The visualization of generation process with preserved structural motifs.

test set contains 800 molecules. Besides, we train a graph auto-encoder using the ZINC250K dataset, which has 250K drug-like molecules.All datasets above have the maximum number of atoms is 38 and include 9 atom types and 3 edge type.

Baselines We compare our approach with the baselines including MMPN (Dalke, Hert, and Kramer 2018), the method learning rules for generating "molecular paraphrases"; GCPN (You et al. 2018), a reinforcement learning based model; CG-VAE (Liu et al. 2018), a graph-based VAE and translation based method JT-VAE, sSeq2Seq, VJTNN, AtomG2G and HierG2G (Jin et al. 2018; Jin, Barzilay, and Jaakkola 2020, 2018). Except for MMPN, all other baselines follow auto-regressive generation paradigms for 2D molecular optimization.

Table 1: Optimization performance on QED task. The best results are **highlighted**.

Method	Generation Mode	Success	QED Diversity	Novelty
MMPA	Rules-based	32.9%	0.236	99.9%
CG-VAE	Atom-based	4.8%	-	-
JT-VAE	Fragment-based	8.8%	-	-
GCPN	Atom-based	9.4%	0.216	100%
VSeq2Seq	Fragment-based	58.5%	0.331	99.6%
VJTNN	Fragment-based	59.9%	0.373	98.3%
AtomG2G	Atom-based	73.6%	0.421	-
HierG2G	Fragment-based	76.9%	0.477	-
MolGuide	One-shot	75.4%	0.510	98.8%

Training Details We set the training hyper-parameters of discrete diffusion model to be the same as DiGress (Vignac et al. 2022) and train a graph auto-encoder with the same architecture of unconditional graph transformer in DiGress. During training with preserved structural motifs, to enhance the generalization of the source-to-target molecule mapping, we randomly select 70% of the structural motifs in the common structural motifs of each pair of molecules and repeat 3 times. Besides, we train the graph transformer in MOT module with the same architecture of unconditional graph transformer in DiGress.

We present translation Accuracy, Diversity and Novelty as evaluation metric in Table 1. From the results, we can find that our proposed method consistently outperforms a serious of Atom-based and Fragment-based methods, all of which follow the auto-regressive generation paradigm (Dalke, Hert, and Kramer 2018; You et al. 2018; Liu et al. 2018; Jin et al. 2018; Jin, Barzilay, and Jaakkola 2018, 2020). Moreover, we visualize the generation process with preserved structural motifs as shown in Fig. 2. It is worth noting that there exist totally different reserve processes corresponding to a same source molecule with most preserved structural motifes and finally our purposed method can still generate two different target molecules.

Conclusion and Limitation

In this paper, we have proposed **MolGuide**, a one-shot generative model for molecular optimization, and carefully design a guidance framework aligning structural motifs and generated molecules by preserving structural motifs with the assistance of ChatGPT and energy-guidance under OT Distance. However, the controllability of one-shot generative model still has a crucial limitation in molecular optimization. For further exploration, we will pay more attention to the methods having a good ability of control during generation process. Solving 2D molecule optimization or even graph generation with one-shot generative model has always been our primary research focus. Thus, We hope one-shot generative model, including our purposed method MolGuide, can bring a novel insight of molecular optimization to community.

References

Austin, J.; Johnson, D. D.; Ho, J.; Tarlow, D.; and Van Den Berg, R. 2021. Structured denoising diffusion models in discrete state-spaces. *Advances in Neural Information Processing Systems*, 34: 17981–17993.

Bao, F.; Zhao, M.; Hao, Z.; Li, P.; Li, C.; and Zhu, J. 2022. Equivariant energy-guided sde for inverse molecular design. *arXiv preprint arXiv:2209.15408*.

Cuturi, M. 2013. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26.

Dalke, A.; Hert, J.; and Kramer, C. 2018. mmpdb: An open-source matched molecular pair platform for large multiproperty data sets. *Journal of chemical information and modeling*, 58(5): 902–910.

Du, Y.; Fu, T.; Sun, J.; and Liu, S. 2022. Molgensurvey: A systematic survey in machine learning models for molecule design. *arXiv preprint arXiv:2203.14500*.

Fu, T.; Xiao, C.; and Sun, J. 2020. Core: Automatic molecule optimization using copy & refine strategy. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 638–645.

Guo, T.; Guo, K.; Liang, Z.; Guo, Z.; Chawla, N. V.; Wiest, O.; Zhang, X.; et al. 2023. What indeed can GPT models do in chemistry? A comprehensive benchmark on eight tasks. *arXiv preprint arXiv:2305.18365*.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.

Ho, J.; and Salimans, T. 2022. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*.

Hua, C.; Luan, S.; Xu, M.; Ying, R.; Fu, J.; Ermon, S.; and Precup, D. 2023. MUDiff: Unified Diffusion for Complete Molecule Generation. *arXiv preprint arXiv:2304.14621*.

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

Jin, W.; Barzilay, R.; and Jaakkola, T. 2020. Hierarchical generation of molecular graphs using structural motifs. In *International conference on machine learning*, 4839–4848. PMLR.

Jin, W.; Yang, K.; Barzilay, R.; and Jaakkola, T. 2018. Learning multimodal graph-to-graph translation for molecular optimization. *arXiv preprint arXiv:1812.01070*.

Li, J.; Liu, Y.; Fan, W.; Wei, X.-Y.; Liu, H.; Tang, J.; and Li, Q. 2023a. Empowering Molecule Discovery for Molecule-Caption Translation with Large Language Models: A ChatGPT Perspective. *arXiv preprint arXiv:2306.06615*.

Li, M.; Wang, D.; Liu, X.; Zeng, Z.; Lu, R.; Chen, B.; and Zhou, M. 2023b. PatchCT: Aligning Patch Set and Label Set with Conditional Transport for Multi-Label Image Classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.

Li, X. L.; and Liang, P. 2021. Prefix-tuning: Optimizing continuous prompts for generation. *arXiv preprint arXiv:2101.00190*.

Liu, M.; Yan, K.; Oztekin, B.; and Ji, S. 2021. Graphebm: Molecular graph generation with energy-based models. *arXiv preprint arXiv:2102.00546*.

Liu, P.; Yuan, W.; Fu, J.; Jiang, Z.; Hayashi, H.; and Neubig, G. 2023a. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9): 1–35.

Liu, Q.; Allamanis, M.; Brockschmidt, M.; and Gaunt, A. 2018. Constrained graph variational autoencoders for molecule design. *Advances in neural information processing systems*, 31.

Liu, S.; Wang, J.; Yang, Y.; Wang, C.; Liu, L.; Guo, H.; and Xiao, C. 2023b. ChatGPT-powered Conversational Drug Editing Using Retrieval and Domain Feedback. *arXiv preprint arXiv:2305.18090*.

Liu, X.; Wang, D.; Li, M.; Duan, Z.; Xu, Y.; Chen, B.; and Zhou, M. 2023c. Patch-Token Aligned Bayesian Prompt Learning for Vision-Language Models. *arXiv preprint arXiv:2303.09100*.

Lugmayr, A.; Danelljan, M.; Romero, A.; Yu, F.; Timofte, R.; and Van Gool, L. 2022. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11461–11471.

OpenAI, R. 2023. GPT-4 technical report. *arXiv*, 2303–08774.

Qian, C.; Tang, H.; Yang, Z.; Liang, H.; and Liu, Y. 2023. Can Large Language Models Empower Molecular Property Prediction? *arXiv preprint arXiv:2307.07443*.

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

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

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

Wang, Z.; Jiang, Y.; Lu, Y.; Shen, Y.; He, P.; Chen, W.; Wang, Z.; and Zhou, M. 2023. In-context learning unlocked for diffusion models. *arXiv preprint arXiv:2305.01115*.

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

You, J.; Liu, B.; Ying, Z.; Pande, V.; and Leskovec, J. 2018. Graph convolutional policy network for goal-directed molecular graph generation. *Advances in neural information processing systems*, 31.

Yu, J.; Wang, Y.; Zhao, C.; Ghanem, B.; and Zhang, J. 2023. Freedom: Training-free energy-guided conditional diffusion model. *arXiv preprint arXiv:2303.09833*.

Zhang, M.; Qamar, M.; Kang, T.; Jung, Y.; Zhang, C.; Bae, S.-H.; and Zhang, C. 2023. A survey on graph diffusion models: Generative ai in science for molecule, protein and material. *arXiv preprint arXiv:2304.01565*.

Zhou, K.; Yang, J.; Loy, C. C.; and Liu, Z. 2022. Learning to prompt for vision-language models. *International Journal of Computer Vision*, 130(9): 2337–2348.