Zhang *et al. BMC Bioinformatics* (2025) 26:123
https://doi.org/10.1186/s12859-025-06072-w

BMC Bioinformatics

RESEARCH **Open Access**

Sculpting molecules in text-3D space: a flexible substructure aware framework for text-oriented molecular optimization

Kaiwei Zhang[1†], Yange Lin[2†], Guangcheng Wu[3], Yuxiang Ren[2*], Xuecang Zhang[2], Bo Wang[2,4], Xiao-Yu Zhang[1] and Weitao Du[2,5*]

†K. Zhang, Y. Lin have contributed equally to this work.

*Correspondence:
renyuxiang1@huawei.com;
duweitao@mail.ustc.edu.cn

[1] Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100085, China
[2] Huawei Technologies, Shenzhen, China
[3] Department of Chemistry, The University of Hong Kong, Hong Kong SAR 999077, China
[4] School of Chemistry and Chemical Engineering, Harbin Institute of Technology, Harbin 150001, China
[5] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China

Abstract

The integration of deep learning, particularly AI-Generated Content, with high-quality data derived from ab initio calculations has emerged as a promising avenue for transforming the landscape of scientific research. However, the challenge of designing molecular drugs or materials that incorporate multi-modality prior knowledge remains a critical and complex undertaking. Specifically, achieving a practical molecular design necessitates not only meeting the diversity requirements but also addressing structural and textural constraints with various symmetries outlined by domain experts. In this article, we present an innovative approach to tackle this inverse design problem by formulating it as a multi-modality guidance optimization task. Our proposed solution involves a textural-structure alignment symmetric diffusion framework for the implementation of molecular optimization tasks, namely 3DToMolo. 3DToMolo aims to harmonize diverse modalities including textual description features and graph structural features, aligning them seamlessly to produce molecular structures adhere to specified symmetric structural and textural constraints by experts in the field. Experimental trials across three guidance optimization settings have shown a superior hit optimization performance compared to state-of-the-art methodologies. Moreover, 3DToMolo demonstrates the capability to discover potential novel molecules, incorporating specified target substructures, without the need for prior knowledge. This work not only holds general significance for the advancement of deep learning methodologies but also paves the way for a transformative shift in molecular design strategies. 3DToMolo creates opportunities for a more nuanced and effective exploration of the vast chemical space, opening new frontiers in the development of molecular entities with tailored properties and functionalities.

Keywords: Molecule optimization, Diffusion model, Drug discovery, Multi-modality, AI for biology

Introduction

In the realm of molecule optimization, a pivotal undertaking in drug discovery, and chemical engineering including catalyst and polymer material designs, the imperative lies in enhancing the desired properties of candidate molecules through strategic chemical modifications. This pivotal process revolves around the core objective of generating molecules that not only meet stringent structural, physical, and electrochemical criteria, but also retain essential structural features (beneficial for relatively straightforward and economic synthesis [1]). At the center of this inverse design issue is the fact that the targeted properties are diverse, spanning the spectrum from qualitative to quantitative aspects. These encompass properties dependent on electronic structures to overarching global descriptions, intricately tied to the two-dimensional (2D) bond topology of the molecule. A concomitant challenge emerges from the multifaceted nature of the goals, necessitating the manipulation of scales and modalities within the molecule. This spans the gamut from fine-tuning atom types and topological structure of atoms to orchestrating alterations in the three-dimensional (3D) conformer structures, reflecting the varied and nuanced nature of the optimization objectives. As a result, traditional solutions rely on the knowledge and expertise of medicinal chemists, often executed through fragment-based screening or synthesis [2–4]. However, such approaches are inherently limited by their lack of scalability and automation.

In recent years, the landscape of computational lead generation has witnessed the emergence of *in silico* methodologies. These methodologies prominently feature deep learning techniques such as latent-space-based generation and Monte-Carlo tree searching (MCTS) algorithms, which trade explicit mechanistic interpretability to model more complex biological relationships learned directly from data such as SMILES (Simplified Molecular Input Line Entry System) [5–8] and two-dimensional molecular graphs [9–12]. The ensuing consequence is the flourishing advancement in the field of molecular discovery, driven by the intricate challenges inherent in identifying novel compounds endowed with specific and desired properties. Within this expansive domain, one prominent line of research focuses on generative models, such as variational autoencoders (VAEs) [5, 6, 13, 14] and generative adversarial networks (GANs) [15–18], which leverage deep learning [19–24] techniques to generate novel molecules. These models have demonstrated promising results in generating diverse and chemically valid molecules. By formulating the molecule optimization problem as a sequence-to-sequence or graph-to-graph translation problem, [25, 26] also utilizes molecular autoencoders as the backbone model for purely 2D molecule optimization. Another approach entails the employment of Reinforcement Learning (RL) algorithms to iteratively optimize molecular structures guided by predefined objectives. RL-based methods [27–32] have shown potential in optimizing drug-like properties and exploring chemical space efficiently.

However, a notable gap persists in the utilization of traditional encoder-decoder-based *de novo* molecule generation methods for molecule optimization tasks. Lead optimization [33] focuses on improving the properties of existing lead compounds, leveraging experimental data and medicinal chemistry expertise to systematically refine molecular structures. This approach tends to retain the major scaffold of molecules for yielding drug candidates with better-defined pharmacological profiles and higher likelihoods of success in clinical trials. Unlike unconditional generative models that generate molecules in a zero-shot manner from informationless noise, effective automatic molecule optimization demands the learning of the distribution differences between molecules before and after the optimization, aligning with preferred properties. MoleculeSTM [11] addresses this challenge by introducing a latent optimization block that guides property-directed transformations through vector movements in the latent space. Since this occurs in the latent space rather than the real 3D molecular space, such approaches grapple with diversity collapsing issues, potentially leading to the loss of crucial molecular structure information. On the other hand, implicit searching-based methods, such as reinforcement learning and MCTS, necessitate expert-designed optimization paths. These paths are instrumental in training the reward function, ensuring it aligns with fixed properties, and formulating policies for molecular modifications. In practice, this entails identifying disconnection sites for optimization, such as optimal side chains, at each step. A learned policy network then selects the best actions from a pre-fixed set of valid molecule modifications. However, this approach may suffer from inflexibility, as the predefined optimization path data and the modification set may not capture the diverse and nuanced possibilities inherent in molecule optimization.

Despite significant progress in molecular optimization, there remains a critical challenge: existing methods often lack the ability to simultaneously address both 2D molecular features (e.g., atom types and bond topology) and 3D conformer structures while accommodating diverse and complex optimization goals. Traditional approaches rely on predefined pathways or single-modality optimization, which limits their flexibility and adaptability to real-world challenges. This challenge highlights the need for a multi-modality-driven framework capable of unifying molecular property and structure descriptions to enable robust and versatile optimization.

In general, it is highly desirable to develop a methodology that is purposefully tailored for optimizing both the 2D aspects (atom types and chemical bond topology) and the 3D conformer structure of molecules, while maintaining compatibility with a broad spectrum of complex goals to enable multi-goal guidance optimization. Capitalizing on the remarkable capabilities exhibited by large language models (LLMs), there is a natural inclination to explore the feasibility of consolidating property and structure descriptions into a unified text format. Then, we are able to leverage the prowess of LLMs to extract a unified representation from such textual amalgamation. In pursuit of this, we advocate the training of a joint molecule diffusion model designed to capture the fine-grained distributions of 2D+3D molecule structures. The crux of an ideal molecule optimization lies in achieving alignment within the representation spaces of both the text side and the molecule structure side. To this end, we introduce the 3D-based Text-oriented Molecular Optimization (3DToMolo), wherein this specific cross-modality alignment is realized through contrastive training. This involves training the representation pair obtained from a lightweight LLM and an (SE(3)) equivariant graph transformer specifically tailored for molecules. The intermediary steps introduced during the forward diffusion process play a crucial role as a medium connecting the initial molecules with those possessing target properties. In contrast to generative approaches of sampling from white noise, the intermediate molecule representations retain essential structural information from the original molecules. Moreover, control over text descriptions is meticulously exerted at each step during the subsequent backward optimization process. Beyond the flexibility inherent in optimizing entire regions of molecules, 3DToMolo showcases its prowess in two practical scenarios where substructures are preserved. In these instances, specific three-dimensional structures are pre-fixed, and optimization exclusively occurs within the remaining inpainting areas. This unique capability underscores the versatility and effectiveness of 3DToMolo in addressing the diverse challenges of molecular optimization, marking a significant step forward in multi-modality-driven molecular design.

Results

Definition of text - structural optimization

Natural-language texts provide a cohesive framework for articulating intricate details regarding the structural and property characteristics of molecules. We follow the approach presented by MoleculeSTM [11] for optimizing structures of molecules, guided by textual prompts. These prompts may encompass qualitative and quantitative descriptions, addressing single or multiple goals. Nevertheless, a notable limitation of the latent space optimization approach proposed in MoleculeSTM lies in its lack of 3D structure encoding. It is imperative to recognize that the spatial arrangement information of 2D chemical bonds, included in 3D conformer structures of molecules, can contribute significantly to their chemical and physical properties. Consequently, successful optimization of molecules or well-known scaffolds with precisely tuned properties requires the integration of 3D structures.

Task definition. Given a molecule or molecular fragment M_0 with known 2D and 3D structures, molecule optimization aims to modify atom types, 3D positions and associated bond relations [34] to produce another molecule M_1. This transformation is guided by the prompt-text y, ensuring that M_1 aligns better with the given text than the original molecule M_0.

To establish a connection between the original molecule structure and the optimized molecule, we introduce a series of noised states M_t. Coarsening fine-grained details, M_t is a blurred version preserving essential semantic information. With a well-selected time horizon T, h_T serves as a common representation bridging M_0 and M_1. Utilizing diffusion-based generative models, known for their efficacy in generating molecule graphs [35] and 3D conformers [36], we propose that parameterizing and controlling the denoising process, which reverses h_T to M_0, provides a flexible and grounded method for optimizing molecules.

Suppose M_t is generated by a Markov chain defined as:

$$dM_t = f(M_t, t)dt + g(t) \cdot dW_t, \tag{1}$$

where W_t denotes Brownian motion, and f and g are smooth functions depending on the current molecules and time t. Let $p_t(M_t)$ be the marginal distribution of the noised molecule M_t. A θ parameterized $SE(3)$-equivariant graph transformer S_θ is employed to learn the gradient of the log-likelihood $p_t(M_t)$: $\nabla \log p_t(M_t)$. The optimizing process with prompt y follows the formula:

$$dM = [f(M,t) - g^2(t) \cdot \nabla \log p_t(M,y)]dt + g(t) \cdot dW_t, \tag{2}$$

where $\nabla \log p_t(M,y) = \nabla \log p_t(M) + \nabla \log p_t(y|M)$. Fitting the conditional probability $p_t(y|M)$ involves using the latent molecular embedding extracted from another graph transformer, trained independently with S_θ (by pairing with the text embedding of the prompt y). We will outline the overall workflow of 3DToMolo in the next section.

Development of a text-structural diffusion model

3DToMolo unfolds in two phases: pretraining and the subsequent application of pretrained models to three types of downstream optimization tasks, as illustrated in Fig. 1. During the pretraining phase, two key objectives are pursued. First, the alignment of textual descriptions and chemical structures is undertaken. Second, an unconditional 2D+3D molecular generation model is initiated. For both objectives, we employ an encoder-decoder-based equivariant graph transformer that takes the 2D molecular graph and the 3D coordinates of each atom as input. However, for the first goal, we exclusively utilize the encoder component to extract the latent representation of molecules. This decoupled workflow enables the utilization of extensive structural data lacking accompanying text descriptions for training S. This aspect is crucial for the generation of diverse optimized structures.

On the prompt-text embedding side, we leverage the widely acclaimed large language model, LLAMA [37], as the text encoder, tapping into its ability to capture nuanced semantic representations from textual descriptions. Then, the alignment is achieved through contrastive learning of the two latent representations: the molecule structure encoding and its paired text embedding. As a possible extension, the text-structure alignment can be independently fine-tuned for domain-specific texts, e,g., materials [38]. To validate the effectiveness of our learned molecule latent embedding, we conduct tests on retrieval and property prediction tasks. The experimental results are provided in Table 1. In line with prior research on molecule pretraining [11, 39], we adopt the MoleculeNet benchmark [40], which encompasses eight single-modal binary classification datasets aimed at evaluating the efficacy of pretrained molecule representation approaches. We consider nine pretraining-based methods as baselines: AttrMask [39], ContextPred [39], InfoGraph [41], MolCLR [42], GraphMVP [10], MoleculeSTM [11], GEM [43], and Grover [44]. We adopt the area under the receiver operating characteristic curve (ROC-AUC) [45] as the evaluation metric. As delineated in Table 1, our observations reveal that methods based on pretraining markedly enhance overall classification accuracy compared to randomly initialized counterparts. Additionally, 3DToMolo demonstrates superior performance on five out of eight tasks, while achieving comparable results to the leading baselines in the remaining three tasks. Since 3DToMolo lies in its ability to leverage pretrained chemical structure representations that incorporate external domain knowledge, which potentially provides a beneficial implicit bias for property prediction tasks. The key hyperparameters of molecule encoder are layers of Graph Transformer 5, learning rate $\{1e-4, 1e-5\}$, hidden states dimension (X: 256, E: 128, pos: 64).

Additionally, we pretrain an unconditional diffusion model designed as the backbone to capture the vast and complex data distribution and generate new structures within

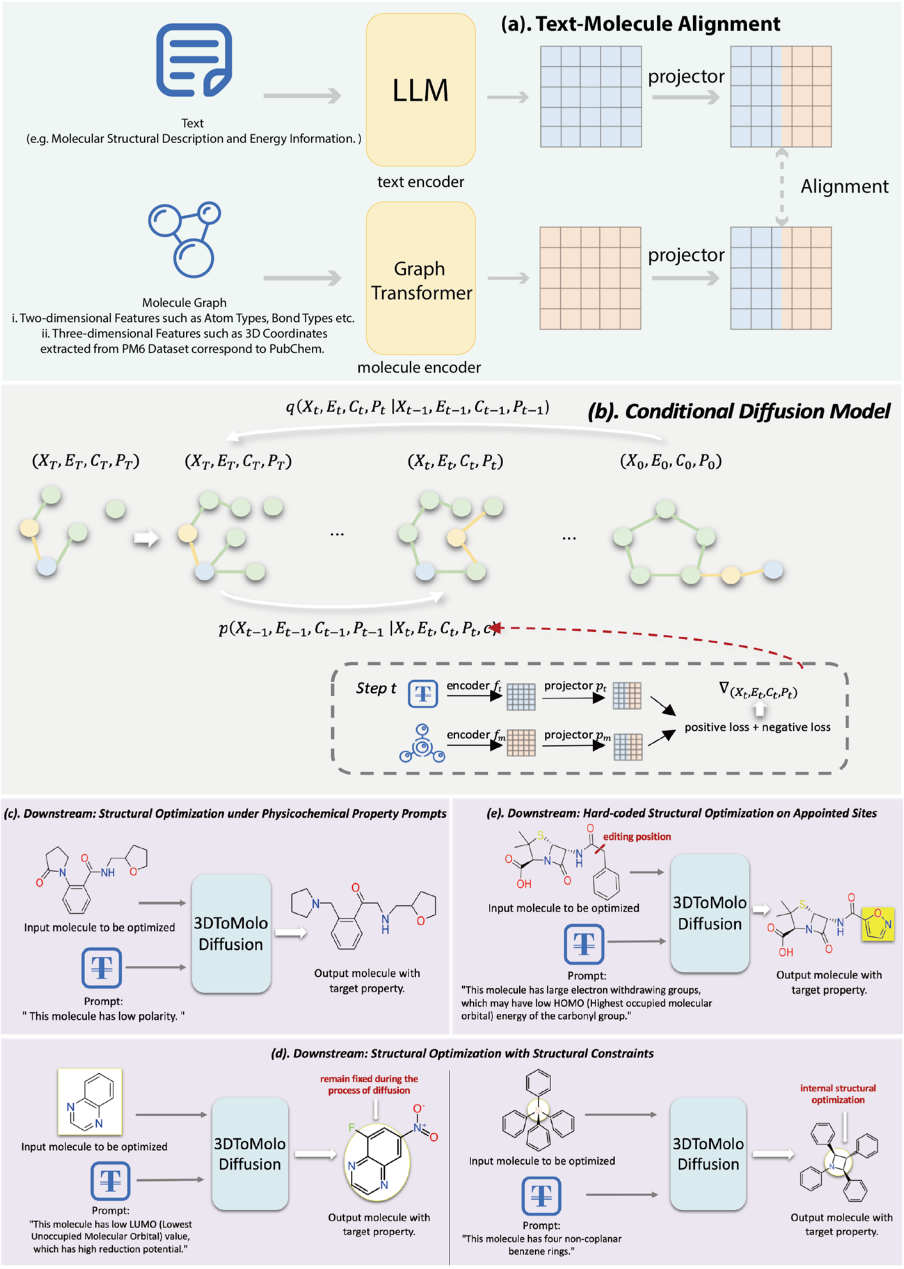


Fig. 1 Overview of 3DToMolo. **a** The alignment of textual description and chemical structures of molecules, which is realized through contrastive learning of the two latent representations: molecule structure encoding with its paired text embedding. **b** Conditional diffusion model. In order to maintain molecule optimization in alignment with the prompt, conditional diffusion model incorporates text prompts at each step during the subsequent backward optimization process. **c** The zero-shot prompt-driven molecule optimization task involves modifying the input molecule in response to a given text prompt related to physicochemical properties. 3DToMolo necessitates the overall optimization of both 2D and 3D features of molecules, ensuring a balanced alignment with the input molecule and the text prompt which is achieved by the conditional diffusion model, shown as (b). **d** Molecule optimization under structural constraints. This task further enhances the similarity to the input molecule by retaining essential structural features. **e** Molecule optimization under appointed sites. Given the precise position within the input molecule, 3DToMolo aims to optimize molecule by offering strategies for atoms and the bonds connected with the site

Table 1 Downstream prediction results conducted on eight binary classification datasets sourced from MoleculeNet. The mark '-' represents the randomly initialized method. The best overall results are highlighted in bold text

methods	BBBP	Tox21	Bace	ToxCast	Sider	ClinTox	MUV	HIV
-	68.67	73.09	77.89	63.40	59.03	68.97	70.94	77.28
AttrMask	67.79	75.00	80.28	63.57	58.05	75.44	73.76	75.44
ContextPred	63.13	74.29	78.75	61.58	60.26	80.34	71.36	70.67
InfoGraph	64.84	76.24	77.64	62.68	59.15	76.51	71.97	70.20
MolCLR	67.79	75.55	71.14	64.58	58.66	84.22	72.76	75.88
GraphMVP	68.11	77.06	80.48	65.11	60.64	84.46	74.38	77.74
MoleculeSTM	69.98	76.91	80.77	65.06	60.96	92.53	73.40	76.93
GEM	72.40	**78.10**	**85.60**	**69.20**	67.20	90.10	**81.70**	80.60
Grover	68.00	76.30	79.50	63.40	60.70	76.90	75.80	77.80
3DToMolo	**72.50**	77.16	85.53	67.11	**67.88**	**93.18**	77.65	**80.66**

the chosen chemical space. The diffusion model samples Gaussian noise and undergoes iterative denoising, resembling standard diffusion sampling. The validity of our pre-trained generative models is verified through standard chemical validity tests, as detailed in Appendix N Table N6. The introduction of the datasets we used is provided in the method section.

By integrating the alignment objective into the denoising process (with the noising step t iteratively tuned to retain the similarity ratio between M_1 and M_0), as detailed in the method section, the model is capable of optimizing molecules with desired properties in a seamless, end-to-end manner. Notably, 3DToMolo is zero-shot, signifying that throughout the optimization process. We refrain from introducing any feedback for multi-stage correction, and the denoising process is executed in a single run. We propose three types of downstream tasks in the following sections and systematically verify the robust effectiveness of 3DToMolo for text-structural optimization. All downstream tasks conducted are based on zero-shot optimization, which refers to a scenario where a model is required to optimize or adapt to new tasks or data without having been explicitly trained on those specific tasks or data beforehand. For instance, in molecular optimization, a model might be able to modify a molecule to achieve a desired biological activity, even if it has never been trained on molecules with that specific task, by leveraging its understanding of chemical space learned during pretraining. Zero-shot optimization is particularly valuable because it allows for the exploration of new chemical spaces or the design of molecules with novel properties, without requiring a model to be retrained or fine-tuned on every new task.

Flexible molecule optimization under physicochemical property prompts

According to the degree of human knowledge involved in the molecule optimization process, we may classify them into two categories:

1. **Flexible optimization:** This category encompasses processes that do not explicitly specify the sites for optimizing atoms and the bonds connected with them.

2. **Hard-coded optimization:** In contrast, this category involves processes that precisely indicate the locations for optimization or substructures to be retained, providing hard constraints regarding the targeted atoms and their geometry.

Details of both optimization algorithms are available in Appendix D. Here, we prioritize the first category, wherein the prompt exerts a global influence on the entire molecular structure without specifically identifying optimization sites. Formally, given a molecule M_0 to be optimized, we adopt the following pipeline:

$$M_0 \rightarrow h_T \xrightarrow{y} M_1,$$

where y denotes the text-prompt. We deliberately choose a small value for T, ensuring that the Tanimoto similarity coefficient [46] between h_T and M_0 approaches unity. While our text prompts encompass constraints ranging from 2D structure considerations (e.g., the number of hydrogen bond donors or acceptors) to properties determined by 3D structure (e.g., polarity), we primarily focus on analyzing how 3DToMolo effectively utilizes 3D structure information to enhance the alignment of the optimized molecule with the given text prompt. It is important to note that, in addition to energies directly calculated from 3D electronic configurations, we are equally intrigued by properties that, while validated through the generated SMILES, might demonstrate indirect connections with 3D structures throughout the optimization process. Our goal is to investigate whether our optimization process, which involves 3D structures, takes advantage of such properties. Consequently, we have designed multi-objective prompts, such as "soluble in water and having high polarity," to assess whether the 3D structure constraint, specifically the requirement for polarity, aids in guiding our optimization process through the vast chemical space.

In order to provide a reasonable and comprehensive evaluation of the molecule optimization ability of 3DToMolo, we first benchmarked representative state-of-art machine learning baselines, including:

- **MoleculeSTM:** A multi-modal model, which enhances molecule representation learning through the integration of textual descriptions.
- **GPT-3.5:** With its immense language processing capabilities and a broad understanding of chemistry concepts, has the potential to revolutionize molecule optimization.
- **Galactica:** A versatile scientific language model, extensively trained on a vast repository of scientific text and data.

Note that while the training data for Large Language Model (LLM)-based models encompasses significantly more scientific texts than domain-specific models, such as ours, it is limited in its ability to assimilate information from modalities other than textual, such as 3D structures. The effectiveness of models is assessed through a satisfactory hit ratio, indicating whether the output molecule generated by the model aligns with the conditions specified in the text prompt when given both a text prompt and a molecule for optimization.

Table 2 Results on 18 text prompts oriented towards diverse physicochemical objectives. The inputs consist of 200 molecules randomly sampled from ZINC, with the evaluation measured by the hit ratio (%) of the property changes in each experiment. Best baseline results are highlighted with underlined text. Best overall results are marked by *. Statistically significant improvement (t-test over 5 different dataset splits, p-value< 0.05) is highlighted with bold text

Prompts	**Baselines**			**3DToMolo**
	MoleculeSTM	**GPT-3.5**	**Galactica**	
This molecule has low HOMO (Highest occupied molecular orbital) value, which is more stable.	35.00	37.50	00.00	**46.50***
This molecule has high HOMO (Highest occupied molecular orbital) value, which is more reactive and susceptible to electron acceptance or participation in chemical reactions.	28.00	40.50	00.00	**58.50***
This molecule has low LUMO (Lowest Unoccupied Molecular Orbital) value.	09.50	27.50	00.00	**30.00***
This molecule has high LUMO (Lowest Unoccupied Molecular Orbital) value.	53.50	30.50	00.00	**88.00***
This molecule has low HOMO-LUMO gap value, which has enhanced light absorption properties. The small energy difference allows the molecule to absorb photons in the visible or ultraviolet range, resulting in a higher likelihood of exhibiting color or being used as a dye or pigment.	09.00	29.50	00.00	**87.50***
This molecule has high HOMO-LUMO gap value, which is insulating or non-conductive. The large energy difference between the HOMO and LUMO orbitals makes it less likely for electrons to be excited across the gap, resulting in low electrical conductivity.	56.50	40.00	00.00	**89.00***
This molecule has high polarity.	41.50	44.50	00.00	**58.00***
This molecule has low polarity.	45.00	35.50	00.00	**73.50***
This molecule is soluble in water, which may have high polarity.	28.50	25.00	04.00	**46.00***
This molecule is insoluble in water, which may have low polarity.	52.00	74.00	03.50	**81.00***
This molecule is soluble in water.	29.50*	24.50	04.00	26.50
This molecule is insoluble in water.	52.00	66.50	03.50	**88.00***
This molecule has high permeability.	34.50	23.00	05.50	**89.00***
This molecule has low permeability.	24.50	39.00*	00.50	21.00
This molecule has more hydrogen bond acceptors.	07.50	20.00*	00.00	10.50
This molecule has more hydrogen bond donors.	05.00	12.50	00.00	**34.50***
This molecule is like a drug.	42.00	59.00*	00.00	41.50
This molecule is not like a drug.	39.00	31.00	00.00	**56.50***

Table 2 summarizes the molecule optimization performance of 3DToMolo and existing approaches on a randomly selected subset of 200 molecules from the Zinc dataset [47]. To establish a zero-shot generalization stage for testing 3DToMolo, these 200 molecules were intentionally excluded from the model's pre-training data (although they may be present in other baseline models depending on their respective training datasets). We delve into 18 optimization tasks encompassing both 2D and 3D-related optimization. The tasks cover a wide range of energetic and structural properties of molecules (scientific background in Appendix B). It is evident that 3DToMolo consistently achieves exemplary hit ratios across the majority of the 18 tasks. This observation underscores the validity and benefits of incorporating 3D structures of molecules into the diffusion model and aligning chemical space with semantic space, thereby facilitating the exploration of output molecules satisfied with the desired properties. Incorporation of 3D structures also provides an additional navigation for exploring the accessible chemical space. Our optimization results diversify from different input molecules, different prompts, and different parallel runs (Appendix K Figure K6). This feature enables 3DToMolo to effectively improve the hit ratio by conducting multi-run optimization (Appendix K Table K4).

Visual analysis on single-objective molecule optimizations. We conduct a detailed visual analysis of disparities between original and optimized molecules, focusing on the single-objective tasks. Common modifications involve the addition, removal, and replacement of functional groups or molecular cores, with frequent occurrences of molecular skeleton rearrangements due to our ability to manipulate three-dimensional structures. For instance, atoms with high electronegativity have large electron affinities and thus can lower overall electron energy levels, whereas atoms with low electronegativity do the opposite. Therefore, to heighten electron energy levels in response to tasks like increasing the HOMO energy (Fig. 2a1) and the LUMO energy (Fig. 2b1), 3DToMolo removes highly electronegative atoms and functional groups in the input molecule, such as fluorine and chlorine atoms, as well as sulfone and isoxazole groups. Conversely, in tasks requiring the reduction of electron energy levels, electron-withdrawing

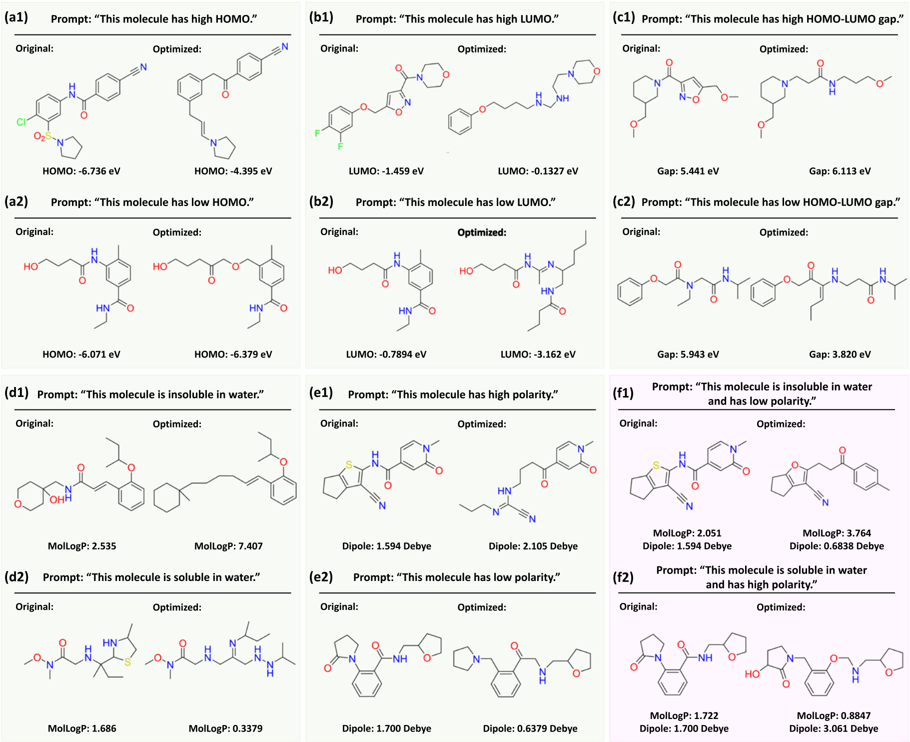
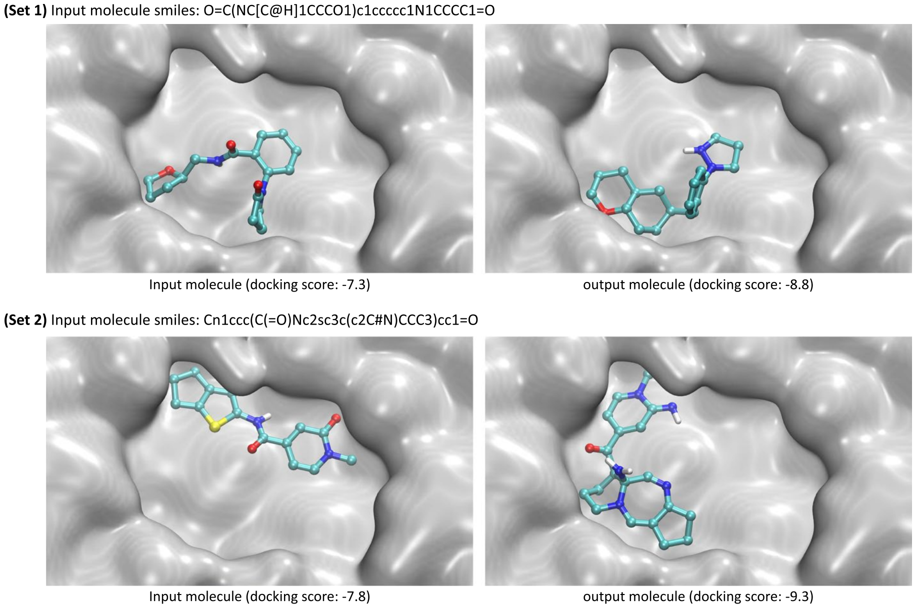


Fig. 2 Exemplary prompt-driven molecule optimizations: **a** the highest occupied molecular orbital (HOMO) energy optimization, **b** the lowest unoccupied molecular orbital (LUMO) energy optimization, **c** the HOMO-LUMO energy gap optimization, **d** the water solubility optimization, **(e)** the polarity optimization, **f** the water solubility and polarity multi-objective optimization. Prompts shown in the figure are simplified and exact prompts used in experiments can be found in Table 2

functional groups or atoms with high electronegativity are introduced (Fig. 2a2) and 2b2). In Fig. 2c1), the widening of the HOMO-LUMO gap is achieved by replacing the isoxazole group with a saturated chain. In contrast, Fig. 2(c2) showcases the narrowing of the HOMO-LUMO gap through the introduction of a double bond conjugated with the carbonyl group. This is because the introduction/removal of conjugated structure can result in denser/sparser electron energy levels and hence a wider/narrower HOMO-LUMO gap. Figure 2d1) and 2d2) illustrate that the addition and removal of hydrogen bond-forming groups, like hydroxyl groups and amines, modulate aqueous solubility by increasing and decreasing it, respectively. Concerning molecular polarity, optimizations such as changing a sulfur atom to a nitrogen atom increase bond polarity, enhancing overall polarity (Fig. 2e1), while the removal of a polar carbonyl group decreases polarity (Fig. 2e2). Additionally, we conduct binding-affinity-based molecule optimization. As shown in Fig. 3, two sets of output molecules have lower docking scores, validating that the ligands generated by 3DToMolo could bind the receptor with higher affinity.

Visual analysis on multi-objective molecule optimizations. We further analyze the multi-objective molecule optimization. Water solubility and polarity are two positively correlated properties. Consequently, 3DToMolo turns the 2-oxo-1-pyrindinyl group into a benzene group, which reduces the solubility as well as the polarity of the input molecule (Fig. 2f1)). In contrast, 3DToMolo adds a hydroxyl group to the input molecule when given the opposite prompt, which increases the solubility and the polarity (Fig. 2f2). More results on multi-objective optimization are presented in Appendix L Table L5. We observe that in the multi-objective task of improving both the solubility and the polarity, 46% of the input molecules have been observed a solubility improvement after optimization, while 31.5% have improvements in both properties,



(Set 1) Input molecule smiles: O=C(NC[C@H]1CCCO1)c1ccccc1N1CCCC1=O

Input molecule (docking score: -7.3)

output molecule (docking score: -8.8)

(Set 2) Input molecule smiles: Cn1ccc(C(=O)Nc2sc3c(c2C#N)CCC3)cc1=O

Input molecule (docking score: -7.8)

output molecule (docking score: -9.3)

Fig. 3 The visualization of binding-affinity-based molecule optimization. The text prompt is from ChEMBL 1613777 ("This molecule is tested positive in an assay that are inhibitors and substrates of an enzyme protein. It uses molecular oxygen inserting one oxygen atom into a substrate, and reducing the second into a water molecule." [48])

higher than the hit ratio of the single-objective solubility improvement task (Table 2). It hints that coupling with the polarity in the prompt helps us better tune the solubility. A possible reason could be that 3DToMolo tunes the polarity more flexibly, as discussed in the next paragraph.

Case study for 3D structural manipulation. In addressing prompts related to molecular conformation, 3DToMolo adeptly achieves the goal by manipulating 3D structures beyond functional-group-wise modifications. For instance, when instructed to decrease the polarity of the input molecule, 3DToMolo strategically adds a polar hydroxyl group. The added hydroxyl group spatially cancels out the dipole moment of another existing C-O bond (Fig. 4a), resulting in a decreased total dipole moment. In another example, when tasked with increasing the polarity of a molecule with six heteroatoms, including two fluorine atoms, 3DToMolo removes highly polar C-F bonds and outputs a molecule with four heteroatoms (Fig. 4b). This decision is based on the understanding that, in a stable conformation, the two C-F bonds contradict the dipole of pyridine ring. Thus, the replacement of C-F bonds by a hydroxyl group more aligned with the pyridine dipole in fact increases polarity. These examples underscore 3DToMolo's ability to comprehend entire molecules, including transient conformational information, a crucial aspect for precise task execution.

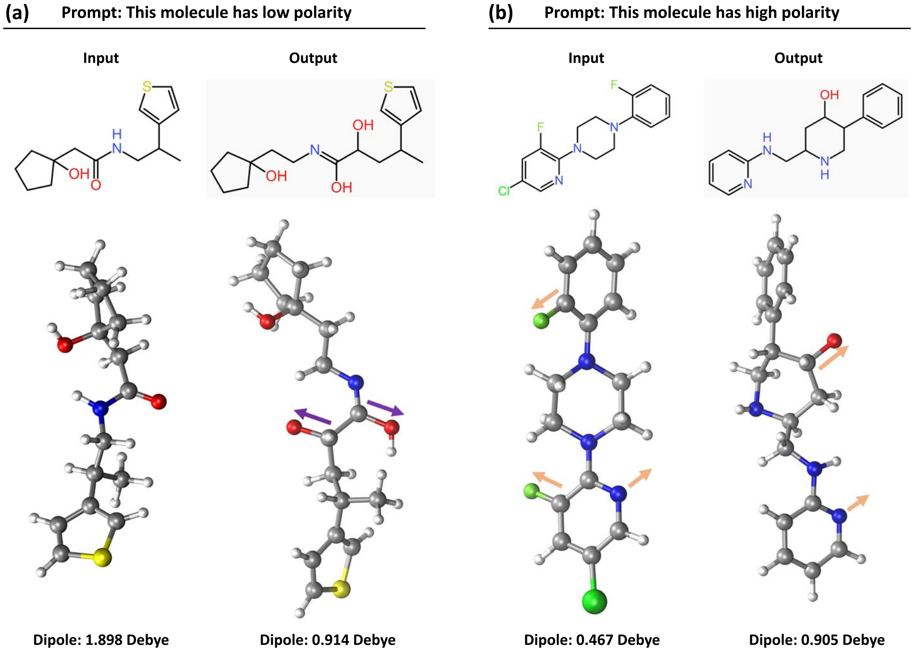


Fig. 4 Exemplary optimizations involving spatial information in the polarity-related prompts. Above are the 2D graphs of molecules and below are their corresponding 3D conformations. **a** Under the prompt "This molecule has low polarity", a hydroxyl group is added to the molecule, which neutralizes the dipole of the neighboring hydroxyl group due to the opposite alignment, as illustrated by the arrows. Consequently, the dipole moment of the molecule is reduced from 1.898 Debye to 0.914 Debye. **b** Under the prompt "This molecule has high polarity", the output molecule discards two C-F bonds which counteract the dipole of the pyridine ring and hence do not contribute much to the polarity. The removal of C-F bonds and the introduction of an aligned hydroxyl group raise the dipole moment of the molecule from 0.467 Debye to 0.905 Debye

Prompt-driven molecule optimization with structural constraints

While the flexible optimization scenario offers maximal optimization diversity within the chemical space, there are situations where specific substructures, as designated by experts, must be preserved. Formally, a molecule is decomposed into two disjoint parts: $M_0 \coprod S_0$, with S_0 representing the substructure to be protected. Consequently, we transit from recovering $p_t(M \coprod S)$ to the conditional density $p_t(M|S)$. Given that S is fixed, the gradient required by the optimization process in Eq. 2 becomes:

$$\nabla p_t(M|S) \rightarrow \nabla_M p_t(M \coprod S).$$

While both the variational-based prompt molecule optimization MoleculeSTM and our denoising-based approach share a common step of encoding molecules into a latent space, it is methodologically impossible to decompose the molecule into two parts in MoleculeSTM. This is because the optimization process in MoleculeSTM occurs in the latent space, which is different from our formulation.

In the following experiments, we have carefully chosen specific physicochemical tasks to conduct a thorough examination. Specifically, we protect all core structures excluding the hydrogen atoms or other removable atoms. By only optimizing the pre-defined removable atoms, we aim to maintain the integrity of the molecular backbone, emphasizing the impact on non-hydrogen constituents, which play key roles in defining the original molecule's chemical properties and functionality.

Redox potential related prompt-driven molecule optimization. In the context of energy storage, enhancing the energy density of batteries necessitates elevated voltage, requiring electrolyte molecules with an expansive electrochemical window. As case studies, we aim to increase the oxidation potential and decrease the reduction potential of exemplary electrolyte molecules. Generally, the oxidation and reduction potentials of a molecule can be influenced by the introduction of substituent function groups. Thus in the following experiment, we protect the skeleton of original electrolyte molecules and allow modifications and substitutions only to the hydrogen atoms.

Thiophene is a common structure used in electrolyte additives for lithium-ion batteries. To augment thiophene's resistance to high voltage, we apply the prompt "This molecule has low HOMO (Highest occupied molecular orbital) value, which has high oxidation potential" while constraining all atoms except hydrogens. A comparative experiment without the prompt serves as the baseline. With the prompt, 19.5% of the generated derivatives exhibit an increased oxidation potential, compared to 12.3% without the prompt. Selected successful examples are presented in Fig. 5a. A similar experiment is conducted on phosphate, frequently employed in electrolytes to enhance battery stability at elevated temperatures. The success rate is 8.12% with the prompt and 5.66% without. To illustrate the modification of reduction potential, we use quinoxaline, pertinent to redox flow batteries. By constraining all atoms except hydrogens and employing corresponding prompts, we successfully modify the reduction potential in two directions (Fig. 5a). In the more desirable direction of lowering the reduction potential, we achieve a success rate of 69.1%.

Additionally, we conducted quantitative experiments a dateset of 143 common electrolyte additives [49]. We focus on properties people generally concern about in liquid electrolytes, including the redox potential, the polarity, and the water solubility.

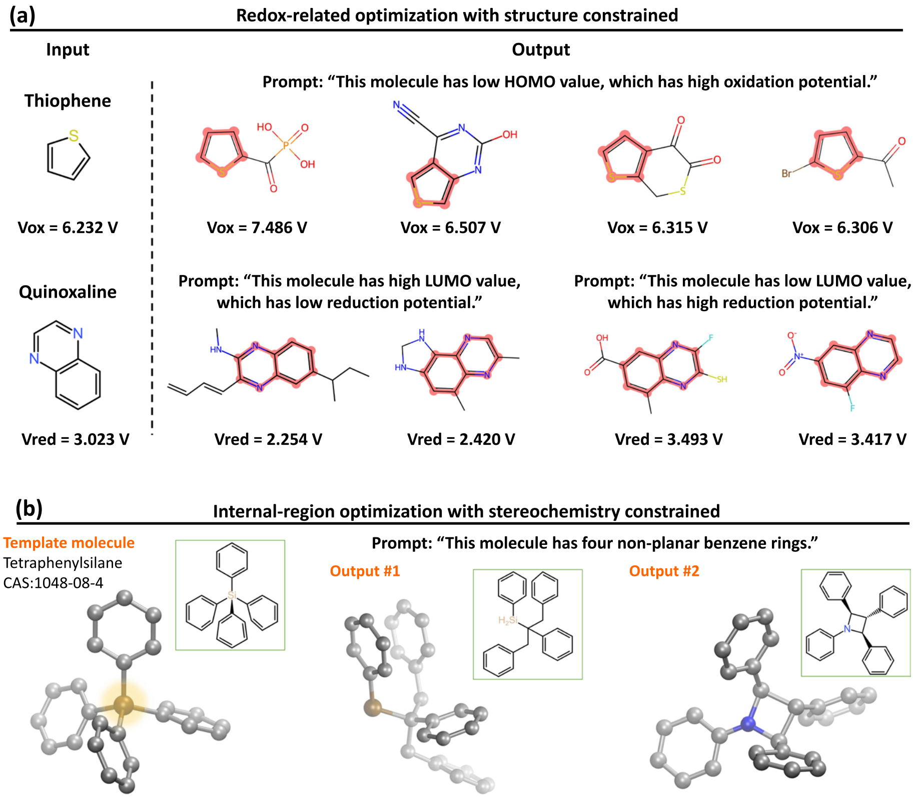


Fig. 5 Molecule optimization with structural constraints. **a** Redox-related prompt-driven molecule optimization with preserved skeletons. **b** A case study for optimization on the internal region. On the left is a reported spiro-linked π –conjugated molecule, tetraphenylsilane, used as the template. The center silicon atom (yellow-shaded area) is diffused while the four benzene rings remain fixed until final steps of the denoising process. On the right are two selected output structures whose benzene rings remain non-coplanar as desired. Hydrogen atoms are not shown for the sake of clear visualization

As shown in Table 3, 3DToMolo significantly outperforms the state-of-the-art baseline GPT–3.5. The major challenge for GPT–3.5 lies in the validity of output molecules, which contains problems such as incomplete rings and chemically incorrect bondings. 3DToMolo circumvents this defect, as the incorporation of 3D information makes it less likely to forget those hanging rings or bonding condition of atoms.

Internal-region molecule optimization. Modification on the internal region of a molecule is challenging [50], as it necessitates rational linkages of fragments. This task becomes even more difficult when specific stereochemistry requirements are imposed. Here, we demonstrate the competence of 3DToMolo for such tasks via a case study on tetraphenylsilane. Non-coplanar benzene rings in tetraphenylsilane (Fig. 5b, left) is a desired structural feature for optical materials that have high refractive index and low light double-refraction (birefringence) [51, 52]. Benzene rings contribute to the strong refractive ability. The non-coplanar configuration hinders the $\pi - \pi$ stacking, preventing the formation of layering structure and thus reducing the double-refraction. To generate more candidate molecules with satisfactory configuration, we protect the benzene rings in the tetraphenylsilane molecule and diffuse the center silicon atom under the prompt "This molecule has four non-planar benzene

Table 3 The results of molecule optimization with structural constraints on the electrolyte additive dataset [49] were evaluated using the hit ratio (%) of property changes in each objective-oriented prompt. Statistically significant improvement (t-test over 5 different dataset splits, p-value< 0.05) is highlighted with bold text

Prompts	**Validity (%)**		**Hit Ratio (%)**	
	GPT-3.5	**3DToMolo**	**GPT-3.5**	**3DToMolo**
This molecule has low HOMO (Highest occupied molecular orbital) value, which has high oxidation potential.	2.867	**41.54**	0.2797	**10.91**
This molecule has high HOMO (Highest occupied molecular orbital) value, which has low oxidation potential.	0.4196	**42.73**	0.06993	**29.65**
This molecule has low LUMO (Lowest Unoccupied Molecular Orbital) value, which has high reduction potential.	0	**44.55**	0	**25.38**
This molecule has high LUMO (Lowest Unoccupied Molecular Orbital) value, which has low reduction potential.	0.4895	**42.45**	0.4895	**18.81**
This molecule has high polarity.	1.399	**42.03**	1.119	**30.35**
This molecule has low polarity.	0.1399	**43.15**	00.00	**12.17**
This molecule is soluble in water.	2.448	**55.24**	1.678	**36.50**
This molecule is insoluble in water.	1.399	**57.27**	1.119	**20.63**

Note that the prompts we input into GPT—3.5 have subtle differences from those listed in the above table. To ensure a fair comparison, we formulated our request to GPT—3.5 as follows: "Optimize this molecule: <SMILES> to achieve <specified properties>. Optimization is restricted to hydrogen atoms only, while the rest of the input molecule remains fixed. The number of atoms (including hydrogen) added to the optimized molecule must be within the range of 1 to 10. Provide 10 different and valid results." For the compound prompts, the hit ratio is calculated based on the satisfaction of the redox potential part. In most cases, GPT—3.5 fails to meet the requirements, as the generated molecules do not conform to either the specified structural criteria or the desired properties

rings". The protection is removed at final steps during the denoising process. Valid generated structures are examined by Density Functional Theory (DFT) [53, 54] computation. Most of the structures have connected four benzene rings via the generated central motif and maintain the rings non-coplanar. Two optimized results are exemplified in Fig. 5b and more can be found in Appendix G Figure G4. As a comparison, GPT—3.5's performance on the same task is poor either because it fails to generate required structures or because it merely conducts single-atom replacement on the silicon atom without the capability of providing more sophisticated internal structures (detailed in Appendix I).

We focus on the prompt-driven experiments in this section. However, we note that 3DToMolo is able to handle prompt-free unconditional generation (examples in Appendix H and Figure H5), no different from other generative models that do not incorporate LLM models.

Hard-coded molecule optimization on appointed sites

Precisely optimizing on pre-appointed optimization sites is notoriously difficult for latent space-based molecule representations, primarily due to the missing of exact spatial decoding. On the other hand, several machine-learning based optimization site identifiers have been proposed, specifically tailored for domain-specific tasks. Since many of these identifiers are trained on datasets where the goals are explicitly defined, and such detailed objectives may be scarce in textual representations of molecular structures. Consequently, 3DToMolo faces a hurdle in automatically identifying the desired optimization sites solely from textual prompts. In light of this, we embark on an

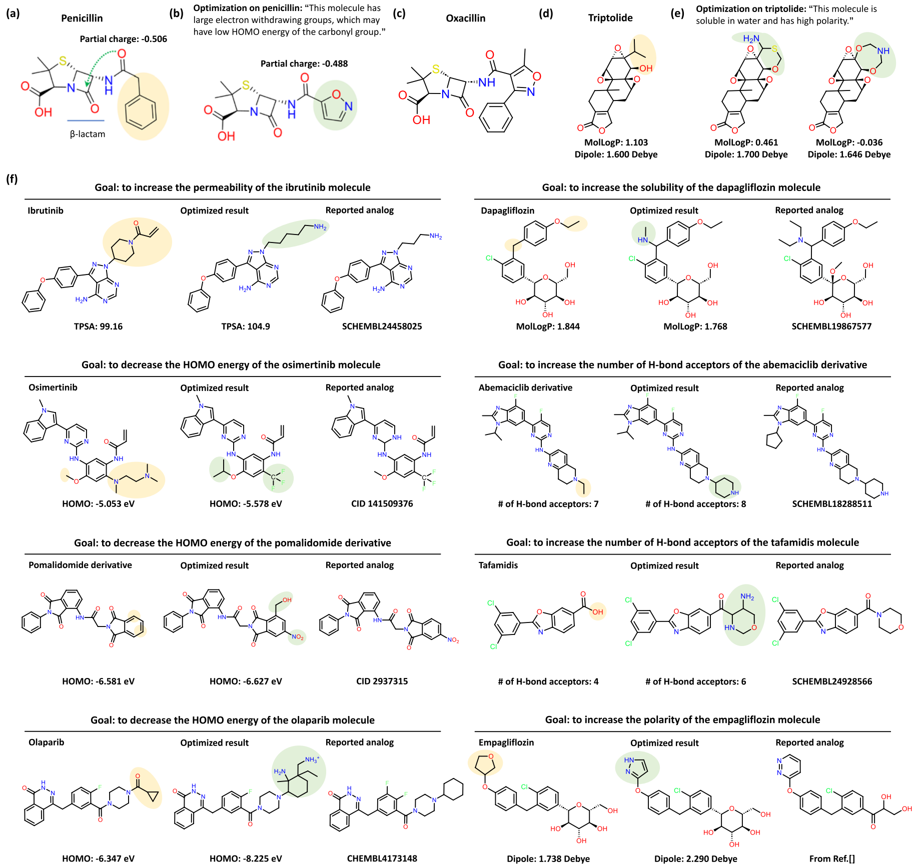


Fig. 6 Drug molecule optimization on appointed sites. **a** The molecular structure of penicillin. The blue line marks the β-lactam ring. The dashed arrow shows the nucleophilic attack from the oxygen atom on the side chain to the β-lactam ring. The yellow region marks the functional group to be optimized. The partial charge on the nucleophilic-attacking oxygen atom is −0.506. **b** One representative example of our optimization results under the given prompt is shown in the figure. The green region marks the modification where the benzyl group is replaced by a more electron withdrawing isoxazole group. With the modification, the partial charge on the nucleophilic-attacking oxygen atom is successfully reduced to −0.488. **c** The molecular structure of oxacillin, an existing analogue to our modification. **d** The molecular structure of triptolide. The yellow regions mark the optimization sites. **e** Two exemplary optimizations where the two modified side chains have connected. Both exhibit increases in solubility as well as in polarity, as the given prompt orders. **f** Selected examples in diverse drug molecule optimization tasks that have reported analogs

exploration to determine the adaptability of 3DToMolo to hard-code optimization on pre-appointed sites, establishing a comprehensive optimization pipeline.

From a methodological perspective, the 3D positions of the appointed sites are utilized during the hard-coded optimization process (see Appendix D for the algorithm details). We showcase 3DToMolo's capability on two exemplary drug-related molecules, penicillin and triptolide. For penicillin, the crucial β-lactam ring [55, 56] is vulnerable to β-lactamase binding [57] and acidic hydrolysis [58] (Fig. 6a). One proposed mechanism suggests the initial nucleophilic attack by the oxygen atom from another amide group on the carbonyl group in the β-lactam ring [57]. Thus, an effective strategy is replacing the benzyl group by a more electron-withdrawing functional group with substantial steric volume to impede lactamase binding and to weaken the nucleophilicity of the attacking oxygen atom. We optimize the penicillin molecule under the prompt "This molecule has large electron-withdrawing groups" while maintaining structural constraints on the entire molecule except for the benzyl group. Of the optimized structures, 23% successfully exhibit a decrease in the electron density on the attacking oxygen atom revealed by DFT computation, with 43% of these structures featuring a ring of at least five members, indicative of substantial steric effects. An exemplary result demonstrates the replacement of the benzyl group with an isoxazole group (Figs. 6b). Notably, the isoxazolyl series of semi-synthetic penicillins, such as oxacillin and (Fig. 6c), has been recognized for superior resistance to acids and β-lactamases [59]. A comparative test utilizing GPT—3.5 as a baseline reveals its inability to generate valid SMILES strings or preserve the core structure under varying instructions (Appendix J).

We further demonstrate the capability of 3DToMolo beyond simple side chain substitution in the optimization of Triptolide [60, 61] with the goal of enhancing its water solubility. Referencing to reported modified derivatives [62], we choose two adjacent sites for optimization, as depicted in Fig. 6d. The remaining structure is constrained. We employ the prompt "This molecule is soluble in water and has high polarity" that has been proven to be more effective in the previous section. Within successful optimizations,

Table 4 The results of molecule optimization based on specified editing positions on the small molecule drug dataset, with the evaluation measured by the hit ratio (%) of the property changes in each objective-oriented prompt. Statistically significant improvement (t-test over 5 different dataset splits, p-value< 0.05) is highlighted with bold text

Drug Molecule	Objectives	Prompts	Validity (%)		Hit Ratio (%)	
			GPT-3.5	Ours	GPT-3.5	Ours
Apixaban	Electron-donating substitution	High HOMO energy	55	**73**	1	**18**
Empagliflozin	More hydrophilic	Large polarity	0	**15**	0	**8**
Ibrutinib	Higher permeability	High permeability	19	**33**	0	**2**
Dapagliflozin	Large aqueous solubility	More soluble	0	**35**	0	**13**
Osimertinib	Electron-withdrawing substitution	Low HOMO energy	30	**52**	0	**39**
Olaparib	Electron-withdrawing substitution	Low HOMO energy	**34**	30	1	**2**
Abemaciclib	More hydrogen bonding interactions	More hydrogen bond acceptors	20	**42**	0	**21**
		More hydrogen bond donors	0	**45**	0	**25**
Pomalidomide	Electron-withdrawing substitution	Low HOMO energy	24	**64**	**11**	10
Tafamidis	More hydrogen bonding interactions	More hydrogen bond acceptors	46	**61**	0	**13**
		More hydrogen bond donors	10	**48**	0	**5**
Lisdexamfetamine	To control the hydrolysis rate	High HOMO energy	0	**11**	0	**8**
		Low HOMO energy	0	**19**	0	**5**

Note that the prompts shown in the table have been abbreviated. The prompts actually used in 3DToMolo can be found in Table 2, while we formulated our request to GPT—3.5 as follows: "Optimize the following molecule: <SMILES> to achieve the desired properties: <specified properties>, while ensuring that the specified substructure <substructure> remains unchanged. Provide 10 different and valid results. Note in this experimental setting, we employed a multi-run approach to facilitate the generation of a total of 100 optimized results for each drug molecule. Specifically, each iteration obtained 10 distinct outputs from GPT—3.5, with a total of 10 iterations conducted. In most cases, GPT—3.5 fails to meet the requirements, as the generated molecules do not conform to either the specified structural criteria or the desired properties we observe some connected structures between these two optimization sites (Fig. 6e), which is not achievable through sequential side chain substitution.

To provide a quantitative comparison, we construct a small task set of optimization on appointed sites in common drug molecules under given prompts. Akin to penicillin and triptolide above, modification sites and optimization strategies of each molecule are carefully chosen based on relevant studies (details in Appendix B and Figure B1). The performance of 3DToMolo is summarized in Table 4, whereas GPT—3.5 is used as a baseline. Compared to the failure of GPT—3.5 shown in the table, 3DToMolo is more powerful in adapting the domain knowledge it has learned in diverse drug discovery tasks. Several optimized results by 3DToMolo not only meet the structural and property requirements, but also are chemically reasonable, as they resemble known molecules in existing database (Fig. 6f). These findings underscore the remarkable proficiency of 3DToMolo in selectively appointing and modifying substructures based on natural language guidance, particularly in scenarios involving complex isomeric structures and three-dimensional considerations.

Discussion

From a broad perspective, our text-structural optimization strategy falls within the category of multi-modality controlled molecule structural modification approaches. Given the inherent significance of 3D structures in shaping molecular properties, we employed SE(3)-equivariant graph transformers for the intricate task of encoding and decoding molecule representations. In summary, we integrated three types of modalities of a molecule: molecule graph, 3D conformers, and text descriptions. Combined with the noising-denoising 2D+3D diffusion models, 3DToMolo proves instrumental in achieving highly promising optimization outcomes. It allows us to optimize molecular structures not only at internal regions, enhancing flexibility, but also in pre-assigned periphery through hard-coded implementations. Notably, the achievement of such comprehensive optimization results would be unattainable without the incorporation of fine-grained 3D position imputing manipulations.

Considering the crucial aspect of data efficiency, our design involves the decoupling of the molecule structure generation model and the text-structural alignment guidance. This separation enables us to leverage vast amounts of unlabeled structure data during the training of the structure generator. 3DToMolo proves particularly advantageous when labeled data for guiding text-structural alignment is limited. By tapping into the abundance of unlabeled data, 3DToMolo gains a robust understanding of diverse molecular structures. Remarkably, we are able to selectively fine-tune specific aspects of the model in a low-rank manner [63], particularly when dealing with intricate molecular geometry configurations such as binding scenarios or conformers exhibiting high energy states. An additional rationale for favoring diffusion models in molecule optimization lies in their global optimization approach. This stands in contrast to the local reinforced, and often greedy approach of optimizing one disconnection site at a time. The global methodology employed by diffusion models contributes to the generation of more diverse and comprehensive optimization results. Finally, it is noteworthy that 3DToMolo distinguishes itself by not necessitating optimization trajectories as part of our training data, in contrast to traditional MCTS methods. This is attributed to the fixed nature of the noising process equation for every molecule, and only the denoising process is learned. Conversely, the manipulation of the noising process to reinforce intermediate states for synthesizability (as an example) is a plausible avenue. This adjustment holds the potential to be beneficial for generating retrosynthetic pathways for our optimized molecules.

As the field of multi-modality large neural networks advances swiftly, our research is only a preliminary attempt on harnessing the potential of multimodality information to guide the optimization of molecular structures. As an illustrative example, we utilized paired text-molecule data to train the alignment between the molecular representation X and the corresponding text representation Y. What remains unexplored is the potential of adversarial matching between X and Y, wherein the mapping function G learns to map the distribution of X to that of Y. This approach has the capacity to leverage unpaired text and molecule data, offering a promising avenue for further investigation. On the text side, 3DToMolo involves utilizing a pretrained Large Language Model (LLM) for extracting text embeddings. This LLM model is specifically trained by predicting the next word token based on context. While effective, a more intricate strategy involves directly training a text-molecule equivariant large model in a similar way as Emu2 [64]. This advanced approach allows for the generation of multi-modality outputs. Unlike our current approach, where the model is trained to predict molecules given the text, this more involved method also operates in reverse. It trains the model not only to predict molecules based on textual descriptions but also to predict text from molecular information. This bidirectional training scheme contributes to a more versatile and expressive representation. Moreover, extending beyond text, the inclusion of illustrations from academic papers adds another layer of informative guidance for optimizing molecules. Such image representations can also serve as valuable cues for refining and enhancing the structural optimization process.

Lastly, the uncertainty in synthesizability of generated molecules remains an unsolved problem for many deep learning models. To ensure the optimization is not only towards better properties but also better synthesizability, the model should be trained with a considerable amount of data and human expert knowledge in the synthesis domain. This is challenging because the data and the knowledge are constantly updating as new synthesis methods are discovered. Notably, although 3DToMolo does not incorporate synthesis-related data, its optimization results maintain synthetic accessibility scores [66] (SAscore) comparable with their inputs (Figure M7 in Appendix M).

Methods

Datasets

We use PCQM4Mv2 dataset[1] to pretrain an unconditional diffusion model for modeling complex data distribution to generate new structures within the chosen chemical space. PCQM4Mv2 is a quantum chemistry dataset including 3,746,619 molecules originating from the PubChemQC project [67]. MoleculeSTM dataset [11] with over 280K chemical structure-text pairs is used to train a text-molecule model. To better align chemical space

---

[1] https://ogb.stanford.edu/docs/lsc/pcqm4mv2/

with semantic space, we effectively incorporate 3D structure information of molecules to enhance the alignment with textual description. However, MoleculeSTM lacks 3D coordinates of molecules, thus we extract 3D information and energy-related values from PubchemQC according to the PIDs in MoleculeSTM. Regarding downstream tasks, a novel molecule could be generated from Gaussian noise, or a molecule selected from Zinc dataset [47] could be optimized through applying noise and recursively denoising.

Training details

All methods are implemented in Python 3.9.13. PyTorch Lightning is utilized to implement a framework that maximizes flexibility without sacrificing performance at scale. All experiments were conducted on Ubuntu 20.04.6 LTS with AMD EPYC 7742 64-Core Processor, 512GB of memory, and 80GB NVIDIA Tesla V100.

Structural optimization through diffusion

Drawing inspiration from the demonstrated effectiveness of diffusion models in generating data from noisy inputs, such as [68], and image-editing applications [69], we propose a 2D-3D joint diffusion model. We aim to introduce fine-grained prompt control through the gradients derived from the contrastive loss, aligning the text-based prompt with our 2D-3D joint representation. To achieve this, our method employs a two-stage strategy. The first stage utilizes a relatively large database of molecules with diverse physicochemical properties. In the second stage, we leverage a text-molecule structure pair database for cross-modality alignment and text-guided molecule structural optimization.

Denoising diffusion process

In the first stage, we conduct pretraining of a generative 2D-3D molecular diffusion model. Following the structure of typical diffusion models, 3DToMolo encompasses two key processes: the forward process and the reverse process. To adapt these processes to 2D-3D molecular graphs, we represent the molecule M as a combination of node features (atom types) H, an adjacency matrix E (representing chemical bond edges), and 3D positions P, denoted as (H, E, P). Specifically, suppose the molecule is composed of n atoms, then $H = [H^1, \dots, H^n]$ represents the one-hot embedding in the periodic table for the n atoms. In the forward process, the original structure of the molecule undergoes a joint Markovian transition step by step. We denote the intermediate structure as $M_t := (H_t, E_t, P_t)$, where the initial structure is denoted as M_0. For the 3D point cloud P_t, a Markov chain is implemented by incrementally introducing Gaussian noise over T steps. The transition from P_{t-1} to P_t is described as follows:

$$q(P_t|P_{t-1}) = \mathcal{N}(P_t; \sqrt{1-\beta_t}P_{t-1}, \beta_t I),$$

where $t \in \{1, \dots, T\}$ denotes the diffusion step. Here, $\mathcal{N}$ represents the Gaussian distribution, and the hyperparameter $\beta_t \in (0, 1)$ controls the scale of the Gaussian noise added at each step. By leveraging the additivity property of two independent Gaussian noises, we can directly express the state P_t in terms of the initial P_0:

$$P_t = \sqrt{\bar{\alpha}_t}P_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, \tag{3}$$

where $\epsilon \sim \mathcal{N}(0, I)$, and $\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$. On the other hand, we treat $z_t = (E_t, H_t)$ as discrete random variables, and thus, we subject them to discrete Markov chains Q_t:

$$q(z_t|z_{t-1}) = \mathcal{C}(z_{t-1}Q_t),$$

where the transition matrix Q_t represents the probability of jumping between states at diffusion step t, and $\mathcal{C}$ denotes the corresponding categorical distribution. Specifically, $\{Q_t\}_0^T$ comprises two components: $\{Q_t^H, Q_t^E\}_0^T$, with $Q_t^H = \alpha_t^z \cdot \mathbf{I} + \beta_t^z \cdot \mathbf{I_a} m_h$, where $m_h \in \mathbf{R^a}$ represents the marginal distribution of atom types in the training set. The Q_t^E for edge diffusion is defined similarly. Finally, α_t^z and β_t^z govern the noise schedule of the discrete part.

Stationary Distribution.

A key ingredient for the noising process is that the stationary distribution of $q(M_t)$ is known. For example, $q(P_t)$ approaches $\mathcal{N}(0, I)$ as $t \to \infty$. Similarly, $q(H_t)$ and $q(E_t)$ follow the categorical distribution defined by m_h and m_e as $t \to \infty$. Since these stationary distributions are simple, we know how to sample them in a trivial way.

Next, the diffusion model learns to remove the added noises from M_t to recover M_{t-1} using neural networks. Starting from h_T, the reverse process gradually reconstructs the relations within the 2D and 3D representation of the molecules through the denoising transition step. Note that although the most straightforward parameterization is to directly predict M_t given M_{t-1}, DDIM [70] has demonstrated that predicting the raw molecule M_0 is equivalent to predicting M_{t-1} with the additional advantage of accelerating the generative process during inference.

Joint Denoising Process.

We implement an equivariant graph transformer architecture F_θ inspired by [50, 71] for predicting M_0 from M_t:

$$F_\theta(M_t) = (F_\theta^P(M_t), F_\theta^E(M_t), F_\theta^H(M_t)).$$

For the 3D part, utilizing Eq. 3 and the Bayes formula, the posterior probability is given by

$$q(P_{t-1}|P_t, F_\theta^P(M_t)) \sim \mathcal{N}(\mu_t \cdot F_\theta^P(M_t) + \nu_t \cdot P_t, \sigma_t \cdot I), \tag{4}$$

where μ_t, ν_t, and σ_t are parameters that don't depend on the neural network.

For the 2D part, similarly, we have the discrete denoising Markov chain with the transition probability given by:

$$q(H_{t-1}|H_t, F_\theta^H(M_t)) \sim H_t(\bar{Q}_t)^T \odot F_\theta^H(M_t)\bar{Q}_{t-1}, \tag{5}$$

where $\bar{Q}_t := Q_1 \cdots Q_t$. Following the transition states of the joint denoising process step by step, the reverse process gradually reconstructs the relations within a molecule graph and its corresponding 3D structures.

Optimization of the Diffusion Process. As demonstrated earlier, the optimization objective for learning F_θ is to reconstruct M_0 from a noised M_t. In practice, we found it beneficial to include a regularization term dependent on the sampled diffusion step t:

$$\mathbf{E}_t \mathbf{E}_{q(M_t|M_{t-1})}[\lambda_t |M_0 - F_\theta(M_t)|^2], \tag{6}$$

where λ_t is a set of parameters depending on the noise schedule α_t. From the posterior distribution (Eq. 4 and Eq. 5) perspective, the reconstruction loss is equivalent to optimizing an Evidence Lower Bound of the likelihood [72] of the original molecular distribution $p(M_0)$.

Text-structure alignment

Prompt guidance. In the previous section, we demonstrated how to reconstruct a molecule from a denoising process, laying the groundwork for meaningful molecule optimization. However, the challenge lies in guiding the denoising process to ensure that the final optimization result aligns with a given prompt. Formally, the prompt guidance denoted by y is expected to influence the transition probability of the denoising process:

$$q_\theta(M_t|M_{t-1}) \rightarrow q_\theta(M_t|M_{t-1}, y) = p(y, t) \cdot q_\theta(M_t|M_{t-1}).$$

To address this, we introduce the Clip mapping [73], previously used in text-image alignment, to establish a connection between the text prompt and our molecular structure. The contrastive-based CLIP loss minimizes the cosine distance in the latent space between the molecule representation X and a given prompt text y:

$$f(x, y) = \mathrm{Clip}(x, y),$$

where Clip returns the cosine distance between their encoded vectors. We utilize a pretrained molecular embedding model from [50] that maps M to its vector embedding X. On the text side, we extract a latent embedding from a light version of pretrained large language model LLAMA-7B [37]. During the optimization of Clip, the parameters of the two encoders are efficiently fine-tuned in a stop-gradient way.

Then, the amplitude of f directly measures the alignment between a given molecule and its prompt text. In other words, for an original molecule embedding X_0, we aim for the optimized molecule $X_{\mathrm{optimized}}$ to satisfy the condition:

$$df \cdot (X_{\mathrm{optimized}} - X_0) > 0.$$

Thanks to the auto-differentiation technique developed by the deep learning community, obtaining the gradient df with respect to the parameters of the molecular embedding model, is straightforward.

Now, we design $q_\theta(M_{t-1}|M_t, y)$ based on the differential df. Assuming Clip is robustly trained, let $p(y|M_t) = \mathcal{N}(f(M_t), \sigma_y \cdot I)$, where σ_t is a hyperparameter. Then, using the Taylor expansion:

$$q_\theta(M_{t-1}|M_t, y) = q_\theta(M_{t-1}|M_t) \cdot p(y|M_{t-1}) \tag{7}$$

$$\approx q_\theta(M_{t-1}|M_t) \cdot e^{<\nabla \log p(y|X(M_t)), M_{t-1} - M_t>}, \tag{8}$$

where $\nabla \log p(y|X(M_t)) \propto -\nabla||y - f(M_t)||^2$. Combining the above, we set $q_\theta(M_{t-1}|M_t, y)$ to be:

$$q_\theta(M_{t-1}|M_t, y) \propto q_\theta(M_{t-1}|M_t) \cdot e^{-\lambda < \nabla_{M_t} ||y - f(M_t)||^2, M_{t-1}>}, \tag{9}$$

and the parameter λ is introduced to control the strength of the prompt guidance.

It is worth emphasizing that all the aforementioned optimization experiments were conducted with a focus on adapting the de-noising steps. Our findings from extra experiments (detailed in Appendix F) and inherent nature of diffusion model indicate that the choice of de-noising steps is correlated with the resemblance between the optimized molecule and the initial input. In cases where minimal alterations to the input were anticipated, opting for smaller de-noising steps proved to be effective.

Multi-identity alignment. To alleviate the mode-collapse issues of the global CLIP loss, we propose to utilize the method in [38] to enhance the original text embedding y_0 with its identity-wise embedding $y_1, \ldots, y_N$ automatically extracted from the **grammar-parse tree** of the text. Note that we empirically find this technique to be beneficial for multi-objective prompt tasks. Let's take "*This molecule is soluble in water, which has lower HOMO value.*" Then, the extracted identity-wise embedding is $y_1 =$ "*soluble in water*$''$ and $y_2 =$ "*lower HOMO value*$''$. The concatenated embedding $y = (y_0, y_1, y_2)$ is fed into the sampling formula Eq. 7 during inference.

Manifold constraint. Drawing inspiration from the geometric explanation of the diffusion process proposed in [72], the 3D score function $\nabla_\theta p_\theta(P_t)$ points towards the normal direction of the data manifold defined by the probability density $q(P_0)$. In the molecular scenario, this data manifold corresponds to valid molecules, constituting a low-dimensional sub-manifold within the space of all chemical graphs. For instance, the valence rule of atoms imposes a strict constraint on the topology of the graph.

However, the gradient df (as seen in guidance-sampling: Eq. 7) may have negative components along the direction of $\nabla p_\theta(P_t)$, potentially leading to a deviation from the data manifold defined by $q(P_0)$. To address this concern, we propose subtracting the negative component from df to enhance the validity of the final denoising result:

$$df \to df - s(df, \nabla p_\theta(P_t)) \cdot \frac{\nabla p_\theta(P_t)}{||\nabla p_\theta(P_t)||},$$

where $s(df, \nabla p_\theta(P_t)) := df \cdot \frac{\nabla p_\theta(P_t)}{||\nabla p_\theta(P_t)||}$. Empirical findings suggest that incorporating the manifold constraint during sampling improves the validity of optimized results for both single-objective and multi-objective optimization tasks.

Evaluation of physiochemical properties

The energy minimization of optimized structures and the computation of charge distribution as well as other physiochemical properties are done by Gaussian16 [54] using B3LYP functional and 6-31 G* basis set. The evaluation metric for optimized results under text prompts is the satisfactory hit ratio, which gauges whether the output molecule can fulfill the conditions specified in the text prompt (detailed in Appendix B). The high-throughput computation of redox potentials is done by a structural-descriptor-based machine learning regressor (detailed in Appendix E).

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-025-06072-w.

Acknowledgements
We appreciated Chunngai Hui for the insightful discussions in organic chemistry and small-molecule drug related tasks.

Author contributions

K. Z. and W. D. conceptualized the research. K. Z. wrote the code. K. Z., W. D., and Y. L. designed the downstream tasks. K. Z., Y. L., and G. W. analyzed the data and results. K. Z., W. D., and Y. L. drafted the manuscript. B. W., Y. R., and X.-Y. Z. supervised the experimental process.

Funding
This work was supported by the National Natural Science Foundation of China (NSFC) (Grant 62376265).

Availability of data and materials

All datasets used in this document are publicly available. The PCQM4MV2 dataset including optimized structures, various properties and 3D information of all the 3,378,606 training molecules is available at acs.jcim.7b00083: https://doi.org/10.1021/acs.jcim.7b00083. The MoleculeNet dataset is available at C7SC02664A: https://doi.org/10.1039/C7SC02664A. The MoleculeSTM dataset is available at s42256-023-00759-6: https://doi.org/10.1038/s42256-023-00759-6. The PubChemQC dataset is available from https://nakatamaho.riken.jp/pubchemqc.riken.jp/. We used the HOMO, LUMO, and HOMO-LUMO gap from PubChemQC dataset, which was obtained at B3LYP/6-31 G* level. The redox dataset is available in the Supplementary of acsomega.8b00576: https://doi.org/10.1021/acsomega.8b00576.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no Conflict of interest.

Received: 22 November 2024 Accepted: 30 January 2025

References

1. Chen Z, Min MR, Parthasarathy S, Ning X. A deep generative model for molecule optimization via one fragment modification. Nat Mach Intell. 2021;3(12):1040–9. https://doi.org/10.1038/s42256-021-00410-2.
2. Gerry CJ, Schreiber SL. Chemical probes and drug leads from advances in synthetic planning and methodology. Nat Rev Drug Discov. 2018;17(5):333–52.
3. Hoffer L, Voitovich YV, Raux B, Carrasco K, Muller C, Fedorov AY, Derviaux C, Amouric A, Betzi S, Horvath D, et al. Integrated strategy for lead optimization based on fragment growing: the diversity-oriented-target-focused-synthesis approach. J Med Chemi. 2018;61(13):5719–32.
4. Souza Neto LR, Moreira-Filho JT, Neves BJ, Maidana RR, Guimarães ACR, Furnham N, Andrade CH, Silva FP. In silico strategies to support fragment-to-lead optimization in drug discovery. Front Chem. 2020;8:93.
5. Kusner MJ, Paige B, Hernández-Lobato JM. Grammar variational autoencoder. In: International Conference on Machine Learning, pp. 1945–1954;2017. PMLR
6. Gómez-Bombarelli R, et al. Automatic chemical design using a data-driven continuous representation of molecules. ACS Central Sci. 2018. https://doi.org/10.1021/acscentsci.7b00572.
7. Sanchez-Lengeling B, Aspuru-Guzik A. Inverse molecular design using machine learning: Generative models for matter engineering. Science. 2018;361(6400):360–5.
8. Segler MH, Kogej T, Tyrchan C, Waller MP. Generating focused molecule libraries for drug discovery with recurrent neural networks. ACS Central Sci. 2018;4(1):120–31.

9. Duvenaud DK, Maclaurin D, Iparraguirre J, Bombarell R, Hirzel T, Aspuru-Guzik A, Adams RP. Convolutional networks on graphs for learning molecular fingerprints. Adv Neural Inf Process Syst. 2015;28:96.
10. Liu S, Demirel MF, Liang Y. N-gram graph: simple unsupervised representation for graphs, with applications to molecules. Adv Neural Inf process Syst. 2019;32:45.
11. Liu S, Nie W, Wang C, Lu J, Qiao Z, Liu L, Tang J, Xiao C, Anandkumar A. Multi-modal molecule structure-text model for text-based retrieval and editing. Nat Mach Intell. 2023;5(12):1447–57. https://doi.org/10.1038/s42256-023-00759-6.
12. Zeng Z, Yao Y, Liu Z, Sun M. A deep-learning system bridging molecule structure and biomedical text with comprehension comparable to human professionals. Nat Commun. 2022;13(1):862.
13. Nakata Y, et al. Molecular generation for organic electrolyte molecule discovery using conditional variational autoencoders. J Phys Chem Lett. 2018. https://doi.org/10.1021/acs.jpclett.8b02011.
14. Simonovsky M, Komodakis N. Constrained graph variational autoencoders for molecule design. arXiv preprint arXiv:1805.09076 [cs.LG]. 2018.
15. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial networks. Commun ACM. 2020;63(11):139–44.
16. Prykhodko O, Johansson SV, Kotsias P-C, Arús-Pous J, Bjerrum EJ, Engkvist O, Chen H. A de novo molecular generation method using latent vector based generative adversarial network. J Cheminform. 2019;11(1):1–13.
17. Gomez-Bombarelli R, al. Chemgan challenge for drug discovery: can ai reproduce natural chemical diversity? ChemRxiv 2018. https://doi.org/10.26434/chemrxiv.5309669.v1
18. De Cao N, Kipf T. Molgan: An implicit generative model for small molecular graphs. arXiv preprint arXiv:1805.11973 [stat.ML] (2018)
19. Krishnan SR, Bung N, Vangala SR, Srinivasan R, Bulusu G, Roy A. De novo structure-based drug design using deep learning. J Chem Inf Model. 2021;62(21):5100–9.
20. Arús-Pous J, Johansson SV, Prykhodko O, Bjerrum EJ, Tyrchan C, Reymond J-L, Chen H, Engkvist O. Randomized smiles strings improve the quality of molecular generative models. J Cheminform. 2019;11(1):1–13.
21. Bagal V, Aggarwal R, Vinod P, Priyakumar UD. Molgpt: molecular generation using a transformer-decoder model. J Chem Inf Model. 2021;62(9):2064–76.
22. Mahmood O, Mansimov E, Bonneau R, Cho K. Masked graph modeling for molecule generation. Nat Commun. 2021;12(1):3156.
23. Gupta A, Müller AT, Huisman BJ, Fuchs JA, Schneider P, Schneider G. Generative recurrent networks for de novo drug design. Mol Inform. 2018;37(1–2):1700111.
24. Li Y, Pei J, Lai L. Structure-based de novo drug design using 3d deep generative models. Chem Sci. 2021;12(41):13664–75.
25. He J, You H, Sandström E, Nittinger E, Bjerrum EJ, Tyrchan C, Czechtizky W, Engkvist O. Molecular optimization by capturing chemist's intuition using deep neural networks. J Cheminform. 2021;13(1):26. https://doi.org/10.1186/s13321-021-00497-0.
26. Hoffman SC, Chenthamarakshan V, Wadhawan K, Chen P-Y, Das P. Optimizing molecules using efficient queries from property evaluations. Nat Mach Intell. 2022;4(1):21–31. https://doi.org/10.1038/s42256-021-00422-y.
27. Atance SR, Diez JV, Engkvist O, Olsson S, Mercado R. De novo drug design using reinforcement learning with graph-based deep generative models. J Chem Inf Model. 2022;62(20):4863–72 (**(PMID: 36219571)**).
28. Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. Sci Adv. 2018;4(7):7885.
29. Olivecrona M, Blaschke T, Engkvist O, Chen H. Molecular de novo design through deep reinforcement learning. J Cheminform. 2017;9(1):48.
30. Putin E, Asadulaev A, Ivanenkov Y, Aladinskiy V, Sidorov P, Majorov K. Reinforcement learning for molecular de novo design. J Cheminform. 2018;10(1):1–11.
31. You J, Liu B, Ying R, Pande V, Leskovec J. Graph convolutional policy network for goal-directed molecular graph generation. Adv Neural Inf Process Syst. 2018;31:6410–21.
32. Segler MH, Preuss M, Waller MP. Planning chemical syntheses with deep neural networks and symbolic ai. Nature. 2018;555(7698):604–10.
33. Jorgensen WL. Efficient drug lead discovery and optimization. Acc Chem Res. 2009;42(6):724–33.
34. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open babel: an open chemical toolbox. J Cheminform. 2011;3(1):1–14.
35. Jo J, Lee S, Hwang SJ. Score-based generative modeling of graphs via the system of stochastic differential equations. In: International Conference on Machine Learning, 10362–10383;2022. PMLR
36. Liu S, Du W, Ma Z-M, Guo H, Tang J. A group symmetric stochastic differential equation model for molecule multimodal pretraining. In: International Conference on Machine Learning, 21497–21526;2023. PMLR
37. Touvron H, Lavril T, Izacard G, Martinet X, Lachaux M-A, Lacroix T, Rozière B, Goyal N, Hambro E, Azhar, F, et al. Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971, 2023.
38. Rubungo AN, Arnold C, Rand BP, Dieng AB. Llm-prop: Predicting physical and electronic properties of crystalline solids from their text descriptions. arXiv:2310.14029, 2023.
39. Hu W, Liu B, Gomes J, Zitnik M, Liang P, Pande V, Leskovec J. Strategies for pre-training graph neural networks. arXiv preprint arXiv:1905.12265, 2019.
40. Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, Leswing K, Pande V. Moleculenet: a benchmark for molecular machine learning. Chem Sci. 2018;9(2):513–30.
41. Sun F-Y, Hoffmann J, Tang J. Infograph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization. arXiv:1908.01000 (2019)
42. Wang Y, Wang J, Cao Z, Farimani AB. Molclr: Molecular contrastive learning of representations via graph neural networks. arXiv:2102.10056, 2021.
43. Fang X, Liu L, Lei J, He D, Zhang S, Zhou J, Wang F, Wu H, Wang H. Geometry-enhanced molecular representation learning for property prediction. Nat Mach Intell. 2022;4(2):127–34.

44. Rong Y, Bian Y, Xu T, Xie W, Wei Y, Huang W, Huang J. Self-supervised graph transformer on large-scale molecular data. Adv Neural Inf Process Syst. 2020;33:12559–71.
45. Bradley AP. The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern Recognit. 1997;30(7):1145–59.
46. Bajusz D, Rácz A, Héberger K. Why is tanimoto index an appropriate choice for fingerprint-based similarity calculations? J Cheminform. 2015;7(1):1–13.
47. Irwin JJ, Tang KG, Young J, Dandarchuluun C, Wong BR, Khurelbaatar M, Moroz YS, Mayfield J, Sayle RA. Zinc20-a free ultralarge-scale chemical database for ligand discovery. J Chem Inf Model. 2020;60(12):6065–73. https://doi.org/10.1021/acs.jcim.0c00675. ((**PMID: 33118813**)).
48. Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, Magariños MP, Mosquera JF, Mutowo P, Nowotka M, et al. Chembl: towards direct deposition of bioassay data. Nucl Acids Res. 2019;47(D1):930–40.
49. Okamoto Y, Kubo Y. Ab initio calculations of the redox potentials of additives for lithium-ion batteries and their prediction through machine learning. ACS Omega. 2018;3(7):7868–74.
50. Du, W, Chen J, Zhang X, Ma Z, Liu S. Molecule Joint Auto-Encoding: Trajectory Pretraining with 2D and 3D Diffusion, 2023.
51. Chen Y, Xu J, Gao P. A route to carbon-sp3 bridging spiro-molecules: synthetic methods and optoelectronic applications. Org Chem Front. 2024;11:508.
52. Seto R, Sato T, Kojima T, Hosokawa K, Koyama Y, Konishi G-I, Takata T. 9,9′-spirobifluorene-containing polycarbonates: transparent polymers with high refractive index and low birefringence. J Polym Sci Part A: Polym Chem. 2010;48(16):3658–67. https://doi.org/10.1002/pola.24150.
53. Smith DG, Burns LA, Simmonett AC, Parrish RM, Schieber MC, Galvelis R, Kraus P, Kruse H, Di Remigio R, Alenaizan A, et al. PSI4 1.4: Open-source software for high-throughput quantum chemistry. J Chem Phys. 2020;152:48.
54. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Petersson GA, Nakatsuji H, Li X, Caricato M, Marenich AV, Bloino J, Janesko BG, Gomperts R, Mennucci B, Hratchian HP, Ortiz JV, Izmaylov AF, Sonnenberg JL, Williams-Young D, Ding F, Lipparini F, Egidi F, Goings J, Peng B, Petrone A, Henderson T, Ranasinghe D, Zakrzewski VG, Gao J, Rega N, Zheng G, Liang W, Hada, M, Ehara M, Toyota K, Fukuda R, Hasegawa, J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Throssell K, Montgomery, JA Jr, Peralta, J.E, Ogliaro F, Bearpark MJ, Heyd JJ, Brothers EN, Kudin KN, Staroverov VN, Keith TA, Kobayashi R, Normand, J, Raghavachari K, Rendell AP, Burant JC, Iyengar SS, Tomasi J, Cossi M, Millam JM, Klene M, Adamo C, Cammi, R, Ochterski JW, Martin RL, Morokuma K, Farkas, O, Foresman JB, Fox DJ. Gaussian16 Revision C.01. Gaussian Inc. Wallingford CT, 2016.
55. Kardos N, Demain AL. Penicillin: the medicine with the greatest impact on therapeutic outcomes. Appl Microbiol Biotechnol. 2011;92:677–87.
56. Waxman DJ, Strominger JL. Penicillin-binding proteins and the mechanism of action of beta-lactam antibiotics. Annu Rev Biochem. 1983;52:825–69.
57. Lima LM, Silva BNM, Barbosa G, Barreiro EJ. β-lactam antibiotics: an overview from a medicinal chemistry perspective. Eur J Med Chem. 2020;208: 112829.
58. Klein AR, Sarri E, Kelch SE, Basinski JJ, Vaidya S, Aristilde L. Probing the fate of different structures of beta-lactam antibiotics: hydrolysis, mineral capture, and influence of organic matter. ACS Earth Space Chem. 2021;56:1511–24.
59. Rolinson GN. Forty years of beta-lactam research. J Antimicrob Chemother. 1998;41(6):589–603.
60. Zhou Z-L, Yang Y-X, Ding J, Li Y-C, Miao Z-H. Triptolide: structural modifications, structure-activity relationships, bioactivities, clinical development and mechanisms. Nat Prod Report. 2012;29(4):457–75.
61. Tong L, Zhao Q, Datan E, Lin G-Q, Minn I, Pomper MG, Yu B, Romo D, He Q-L, Liu JO. Triptolide: reflections on two decades of research and prospects for the future. Nat Prod Report. 2021;38(4):843–60.
62. Hou W, Liu B, Xu H. Triptolide: medicinal chemistry, chemical biology and clinical progress. Eur J Med Chem. 2019;176:378–92.
63. Hu EJ, shen Wallis P, Allen-Zhu Z, Li Y, Wang S, Wang L, Chen W. LoRA: Low-rank adaptation of large language models. In: International Conference on Learning Representations, 2022. https://openreview.net/forum?id=nZeVKeeFYf9
64. Sun Q, Cui Y, Zhang X, Zhang F, Yu Q, Luo Z, Wang Y, Rao Y, Liu J, Huang T, Wang X. Generative multimodal models are in-context learners. arXiv:2312.13286, 2023.
65. Gao W, Coley CW. The synthesizability of molecules proposed by generative models. J Chem Inf Model. 2020;60(12):5714–23. https://doi.org/10.1021/acs.jcim.0c00174. (**PMID: 32250616**).
66. Ertl P, Schuffenhauer A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. J Cheminform. 2009;1(8):1–11.
67. Nakata M, Shimazaki T. Pubchemqc project: a large-scale first-principles electronic structure database for data-driven chemistry. J Chem Inf Model. 2017;57(6):1300–8. https://doi.org/10.1021/acs.jcim.7b00083. ((**PMID: 28481528**)).
68. Vahdat A, Kreis K, Kautz J. Score-based generative modeling in latent space. Adv Neural Inf Process Syst. 2021;34:11287–302.
69. Meng C, He Y, Song Y, Song J, Wu J, Zhu J-Y, Ermon S. Sdedit: Guided image synthesis and editing with stochastic differential equations. arXiv preprint arXiv:2108.01073, 2021.
70. Song J, Meng C, Ermon S. Denoising diffusion implicit models. In: International Conference on Learning Representations 2021. https://openreview.net/forum?id=St1giarCHLP
71. Vignac C, Osman N, Toni L, Frossard P. Midi: Mixed graph and 3d denoising diffusion for molecule generation. arXiv preprint arXiv:2302.09048, 2023.
72. Du W, Zhang H, Yang T, Du Y. A flexible diffusion model. In: International Conference on Machine Learning, 2023;8678–8696. PMLR
73. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, et al: Learning transferable visual models from natural language supervision. In: International Conference on Machine Learning, 2021;8748–8763. PMLR

Publisher's Note