Supplementary Materials for *Sculpting Molecules in Text-3D Space: A Flexible Substructure Aware Framework for Text-Oriented Molecular Optimization*

S1. More Related Works

In recent years, the scientific discourse surrounding molecule discovery has burgeoned, driven by a collective effort to confront 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)[1–4] and generative adversarial networks (GANs)[5–8], which leverage deep learning[9–14] techniques to generate new 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,[15,16] 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[17–22] have shown potential in optimizing drug-like properties and exploring chemical space efficiently. However, the diversity of molecules is still a significant challenge in the realm of molecular generation, and both VAE and GAN based models encounter limitations in effectively addressing this issue. The inherent nature of these models often results in generated molecules lacking the desired diversity and complexity due to the vast chemical space and intricate nature of molecular structures. However, a promising solution emerges through the adoption of diffusion model-based molecular generation. Unlike traditional approaches, diffusion models excel at capturing the stochastic nature of molecular transformations, allowing for the generation of diverse molecular structures while maintaining chemical validity. By embracing the principles of diffusion processes, these models offer a more nuanced and accurate representation of molecular diversity, thereby presenting a potential breakthrough in advancing the capabilities of molecular generation methodologies. Through manipulating the diffusion process, it becomes possible to expand more diverse molecules with desired properties, which offers valuable insights and inspiration for scientific research.

Along with the extensive data accessible, researchers have turned their attention to leveraging large-scale, unlabeled molecular datasets to pretrain deep neural networks. The paradigm of unsupervised pretraining on a vast amount of molecular data has emerged as a cornerstone in the quest for enhanced representation learning. Nevertheless, this paradigm encounters prominent challenges. 1) Unsupervised pretraining, while advantageous in acquiring knowledge of chemical structures devoid of supervised annotations, achieves this through the reconstruction of masked topological[23] or geometric substructures[24]. In contrast to the familiar territory of supervised learning with labeled data, a significant challenge arises in generalizing to unseen categories and tasks without the luxury of such labeled instances or the refinement afforded by fine-tuning. 2) The success of multimodal pretraining in various domains such as image captioning[25,26], video analysis[27,28], and natural language processing[29] has shown that the synergy between different modalities enables the model to capture intricate relationships and patterns that may be overlooked in unimodal settings. By providing models with a rich tapestry of information, multimodal pretraining not only enhances their ability to comprehend and generate content across modalities but also fosters a more nuanced understanding of the underlying semantic connections between different types of data. Previous studies primarily concentrated on string representations of molecules[1,2,30,31], such as SMILES (Simplified Molecular Input Line Entry System)[32], without effectively incorporating the structural information inherent in molecules. And there are also many works centering on two-dimensional molecular graphs[33–36]. Unsupervised approaches that incorporate both two-dimensional and three-dimensional geometric structures in both pretraining and downstream tasks remain less explored. It is noteworthy that a concurrent study[37] also explores the integration of 3D and textual data for molecular properties and caption learning. However, a significant distinction lies in our model's development of a robust diffusion model as a versatile decoder to enhance the optimization of molecular structures. From a broader perspective, we assess our text-3D approach in a zero-shot optimization scenario and propose future investigation into extending zero-shot to few-shot optimization, as suggested by[38], to further enhance performance.

S2. Background for baseline models and optimization tasks

Representative state-of-art machine learning baselines. The baselines we used in this work include:

- **MoleculeSTM:** A multi-modal model, which enhances molecule representation learning through the integration of textual descriptions. On zero-shot text-based molecule optimization tasks, the effectiveness of MoleculeSTM has been confirmed compared to the existing methods. However, this approach has overlooked the 3D geometric information. Moreover, this method suffers from less diverse generation results as it lacks flexibility during the generation process.

- **GPT3.5:** With its immense language processing capabilities and a broad understanding of chemistry concepts, has the potential to revolutionize molecule optimization. GPT3.5 can learn the patterns and relationships inherent in molecular structures by training on vast amounts of chemical data. By interacting with the model, we could input specific molecular structures or describe desired modifications, and GPT3.5 can generate novel molecules with desired properties. However, it is important to note that GPT3.5, as powerful as it is, should be used in conjunction with human expertise and validation. The generated molecules may not follow chemical rules or lack of sophisticated modifications.

- **Galactica:** A versatile scientific language model, extensively trained on a vast repository of scientific text and data. Its capabilities extend beyond standard scientific NLP tasks, excelling in sophisticated endeavors like predicting citations, conducting mathematical reasoning, predicting molecular properties, and annotating proteins at a remarkable level of proficiency.

The 18 optimization tasks in Table 2 in the main article. The highest occupied molecular orbital (HOMO) energy and the lowest unoccupied molecular orbital (LUMO) energy are two important measures of the reaction capability and activity of a molecule. Common chemical reactions usually involve the electron transfer from the HOMO of one molecule to the LUMO of another molecule with approximate energy value. Hence, a higher HOMO and/or a lower LUMO correspond to a larger reaction tendency, and vice versa. The HOMO-LUMO gap accounts for the light absorption of a molecule, as photons with energy equal to the gap can excite electrons from the HOMO to the LUMO.

Polarity determines the intermolecular interaction strength and is quantified by the dipole moment, which is a measure of the separation of the charge centers of electrons and atom nuclei. The total polarity of a molecule, determined by the sum of polar vectors of each heteroatomic bond, can be tuned by adjusting individual bond polarity or modifying the spatial displacement of these polar bonds. A larger polarity leads to stronger interactions and hence higher freezing and boiling points of a matter. Polarity also governs the solubility, as a molecule tends to dissolve into solvents with similar polarity (the like dissolves like rule). Water solubility and permeability of a molecule are respectively determined by its interaction strength with water, which is polar, and lipid membranes, which are non-polar. Therefore, both can be tuned via modifications of the polarity. Specifically, water solubility is also determined by the hydrogen-bond forming capability of the molecule. A hydrogen bond is formed between a highly electronegative atom (fluorine (F), oxygen (O), and sometimes nitrogen (N)), termed *hydrogen bond acceptor*, and a hydrogen atom bonded with another highly electronegative atom, termed *hydrogen bond donor*. Other than water solubility, hydrogen bonds play a significant role in various biochemical interactions.

Lastly, drug-likeness prediction assesses the bioactivity of a molecule qualitatively and is important for the virtual screening of drug candidates.

Evaluation of optimization performance. The evaluation metric is the satisfactory hit ratio. Given the input molecule M_0 and a text prompt y, optimized molecule M_y is outputted by 3DToMolo. Then we use the hit ratio to measure if the output molecule can align with the conditions in the text prompt

$$hit(M_0, y) = \begin{cases} 1, valid(M_y) \cap satisfy(M_0, M_y, y) \\ 0, otherwise \end{cases}, hit(y) = \frac{\sum_{i=1}^{N} hit(M_0^i, y)}{N}, \tag{S1}$$

where N is the total number of molecules to be optimized, $valid()$ is the chemical test for measuring whether a molecule is valid or not, and $satisfy()$ is the condition calculation function. It is task-specific, the logarithm of partition coefficient (LogP), quantitative estimate of drug-likeness (QED), and topological polar surface area (tPSA) are used as the proxies to evaluate the molecule solubility[39], drug-likeness[40], and permeability[41], respectively. The count of hydrogen bond acceptors (HBA) and hydrogen bond donors (HBD) are calculated explicitly. The value of Homo, Lumo, Homo-Lumo gap, and polarity are obtained through DFT. For single-objective optimization, a successful hit represents the measurement difference between the input molecule and output molecule above a certain threshold (we take 0). For multiple-objective property-based optimization, a text prompt should compose multiple properties' descriptions and a successful hit needs to satisfy all the properties simultaneously.

Binding-affinity-based molecule optimization. We conducted binding-affinity-based molecule optimization to validate that the ligands generated by 3DToMolo could bind with the receptor in lower docking scores. Note that different from traditional pocket-ligand joint geometry modeling methods[42], 3DToMolo doesn't require training another neural network model on the pocket side, which also demonstrates the flexibility of the text-prompt driven molecule optimization. The assay description serves as the basis for binding-affinity-based optimization, with each assay representing a distinct binding affinity task. A notable instance is exemplified by ChEMBL 1613777[43] ("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.").

Redox potentials. Redox potential, synonymous with reduction-oxidation potential, quantifies a substance's propensity to lose or gain electrons in a chemical redox reaction. This property holds significance in diverse fields such as chemical synthesis, catalysis, and energy storage. In the context of energy storage, enhancing the energy density of batteries necessitates elevated voltage, requiring electrolyte molecules with an expansive electrochemical window.

Penicillin and triptolide optimization. The β-lactam ring is crucial for the antibiotic function of penicillin[44,45]. However, this ring is vulnerable to hydrolysis in acidic environments[46]. A proposed hydrolysis mechanism suggests the initial nucleophilic attack by the oxygen atom from another amide group on the carbonyl group in the β-lactam ring[47]. To enhance penicillin's acid resistance, reducing the electron density on the attacking oxygen atom and weakening its nucleophilicity are crucial. Additionally, the β-lactam serves as the binding site for β-lactamases, enzymes that confer resistance to β-lactam drugs by hydrolyzing the amide bond in the β-lactam ring[48]. Thus, an effective strategy is replacing the benzyl group by a more electron-withdrawing functional group which exhibits substantial steric hindrance to impede lactamase binding.

Triptolide has been extensively investigated for its mechanism of action and pharmacological activities. It exhibits significant benefits, including anti-cancer, anti-rheumatoid, anti-inflammatory, anti-Alzheimer's effects, and more[49]. Despite these promising attributes, neither triptolide nor its analogs have gained approval as drugs, primarily due to poor water solubility[50]. Therefore, our objective is to enhance the water solubility of triptolide.

S3. SE(3) equivariant graph transformer

From the reconstruction training objective Eq. (6) in the main article, the neural network F_θ is responsible for mapping the noised molecule M_t to the original molecule M_0. We adopt a graph transformer architecture due to its capacity for capturing complex graph information[51–53]. The inputs consist of the nuclear charges H_i, a dense edge matrix consisting of bond types (where the 'null' bond is also denoted as an edge type), and the positions $P_i \in \mathbf{R}^3$. The invariant inputs H and E are fed into an embedding layer to create a high-dimensional representation of atoms and bonds, denoted by A and W. The embedding of equivariant P will be introduced in the following paragraph.

SE(3) Embedding. A notable property of molecular 3D conformers data distributions is the SE(3) symmetry[54] and the reflection anti-symmetry (Figure S7). Therefore, our embedding of equivariant P should also be rotationally and translationally invariant and reflect the difference between chiral molecules. Furthermore, the 3D output $F_\theta^P(M)$ should transform equivalently with respect to the $SE(3)$ group. A unified framework for encoding and decoding equivariant quantities is through scalarization and tensorization[53], which depends on a set of equivariant frames. Two common choices of equivariant frames are vector frames and spherical harmonics. In this work, we apply edge-based vector frames that are easily merged with 2D invariant features from A and W.

Let $\mathbf{p}_i$ and $\mathbf{p}_j$ be the two atom positions connected by bond e_{ij}. Then, the SE(3) symmetric, reflection anti-symmetric frame is defined by:

$$\mathcal{F}_{ij} = (\mathcal{F}_{ij}^1, \mathcal{F}_{ij}^2, \mathcal{F}_{ij}^3) := \left(\frac{\mathbf{p}_i - \mathbf{p}_j}{||\mathbf{p}_i - \mathbf{p}_j||}, \frac{\mathbf{p}_i \times \mathbf{p}_j}{||\mathbf{p}_i \times \mathbf{p}_j||}, \frac{\mathbf{p}_i - \mathbf{p}_j}{||\mathbf{p}_i - \mathbf{p}_j||} \times \frac{\mathbf{p}_i \times \mathbf{p}_j}{||\mathbf{p}_i \times \mathbf{p}_j||} \right). \tag{S2}$$

Then, fixing an equivariant frame $\mathcal{F}$, we define the **Scalarization** operation with respect to an equivariant vector $\mathbf{v}$ as:

$$S(\mathbf{v}, \mathcal{F}) := (\mathcal{F}^1 \cdot \mathbf{v}, \mathcal{F}^2 \cdot \mathbf{v}, \mathcal{F}^3 \cdot \mathbf{v}),$$

such that the output is a tuple of invariant scalars. Reversely, we define the **Tensorization** operation with respect to a tuple of scalars $a = (a_1, a_2, a_3)$ as:

$$T(a, \mathcal{F}) := (a_1 \mathcal{F}^1, a_2 \mathcal{F}^2, a_3 \mathcal{F}^3),$$

such that the output is an equivariant vector. By iterating these two operations through different layers of the neural network, we can reform standard graph transformers into equivariant architectures.

Attention-Based Update Module. Following the standard transformer architecture[55], we implement a residual structure and layer normalization between different update blocks to prevent gradient exploration and vanishing. Within each layer, we utilize the feature-wise attention mechanism that aggregates 2D and 3D information between atoms. Unlike the standard method of calculating attention coefficients, 3DToMolo leverages the equivariant 3D features from the position matrix P, thereby merging 3D information and 2D information during updating.

Let $A^{l-1} \in \mathbf{R}^{N \times d}$, $W^{l-1} \in \mathbf{R}^{N \times N \times d}$ be the atom-wise features and edge-wise features at layer $l-1$, and let $P^{l-1} \in \mathbf{R}^{N \times 3}$ denote the equivariant vector features matrix at layer $l-1$ (P^0 represents the initial position matrix P). Building on the previous section, we have edge-wise equivariant frames $\mathcal{F}_{ij}$ for each edge $e_{ij} \in E$. We then scalarize the equivariant positions p_i^{l-1} and

p_j^{l-1} to obtain two scalar tuples: $S(\mathbf{p}_i^{l-1}, \mathcal{F}_{ij})$ and $S(\mathbf{p}_j^{l-1}, \mathcal{F}_{ij})$. The self-attention message passing is given by:

$$\alpha_{ij}^{l-1} = \text{Softmax}\{\frac{(W_Q a_i^{l-1}) \cdot (W_K a_j^{l-1})}{\sqrt{d}} + \phi^{l-1}(S(\mathbf{p}_i^{l-1}, \mathcal{F}_{ij}) || S(\mathbf{p}_j^{l-1}, \mathcal{F}_{ij}) || w_{ij}^{l-1})\}, \tag{S3}$$

$$\Delta x^{l-1} = \sum_j \alpha_{ij}^{l-1} \cdot (W_V a_j^{l-1}), \tag{S4}$$

$$a_i^l = \phi_a(\Delta x^{l-1}) + a_i^{l-1}, \tag{S5}$$

$$w_{ij}^l = \phi_w(\Delta x_i^{l-1} || \Delta x_j^{l-1}) + a w_{ij}^{l-1}, \tag{S6}$$

where $||$ denotes the concatenation operation between tensors. On the other hand, the equivariant P^l is updated through tensorization:

$$\mathbf{p}_i^l = \mathbf{p}_i^{l-1} + \sum_j T(\phi_{\mathbf{p}}(\alpha_{ij}^{l-1}), \mathcal{F}_{ij}).$$

For the normalization update module, we apply standard layer normalization on scalar A and W, and apply equivariant E_3 normalization[52] to the position feature P.

Our updating formulas mix different types of features through self-attention, and we can extend the capacity of 3DToMolo by incorporating SMILES representation through cross attention. As SMILES and molecule graphs are merely different representations of molecules, this flexibility allows us to take advantage of powerful pretrained SMILES encoders like REINVENT4[56].

S4. Multi-Modality Optimization of 3DTomolo

In our 3DToMolo framework (Figure S6), besides the target molecule probability density $p_0(\text{mol})$, we introduce a family of noised distributions $p_t(\text{mol})$ whose stationary distribution is $p_0(\text{mol})$. Therefore, $p_t(\text{mol})$ can be considered as a blurred encoding of the raw molecule for optimization. Essentially, if we initiate the denoising process from a $p_t(\text{mol})$ with a value of t close to zero, the resultant molecule will resemble the raw molecule. Formally, let X_t denote the noised molecule derived from X_0, and let df represent the differentiable guidance from a specific prompt (see Eq. (7) and Eq. (8) in the main article). The optimization process can be illustrated as:

$$X_0 \xrightarrow{\text{Noising}} X_t \xrightarrow{df\text{guided Denoising}} X_0^f.$$

Although the noising process blurs high-resolution information, it preserves the skeleton information. Furthermore, we can parameterize the noising process[57] to control which parts of the information are preserved.

Another method for achieving fine-grained optimization of the molecule involves noising only a sub-structure S_0 of the molecule, ensuring that the de-noising process does not disrupt the fixed sub-structure. Formally, we decompose $X_0 = M_0 \coprod S_0$, and we only optimize the M_0 part:

$$X_t \xrightarrow{Noising} M_t \coprod S_0 \xrightarrow{df\text{guided Denoising}} X_0^f := M_0^f \coprod S_0.$$

Note that we don't assume that the position of S_0 is near the center of the molecule. Thus we need to find a way for delineating S_0 during the optimization process. To remedy this issue, we leverage the rich 3D geometry information embedded in 3DToMolo, providing a fine-grained ability to precisely locate the substructure S_0. Formally, let $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ denote the positions of the designated optimization sites. In this structural manipulation process, we employ a meticulous spatial translation approach. Around each $\mathbf{x}_i$, a specified number of atoms are randomly initialized. The objective is to ensure that the center of mass (COM) of this newly initialized set of atoms matches the COM of the designated optimization sites: $\bar{\mathbf{X}} := \sum_i \mathbf{x}_i / n$, which is realized by Euclidean translation. Subsequently, a similar strategy is employed as in the previous section. The atoms situated specifically at the optimization sites and their associated chemical bonds are selectively noised, introducing controlled perturbations to the molecular structure. Meanwhile, all other aspects of the 2D and 3D structures within the molecule are held constant. This precise and localized perturbation strategy ensures that the optimization is targeted and confined solely to the designated sites while preserving the overall structural integrity of the molecule.

It is worth noting that we can even modify the atom numbers once we obtain M_t. For example, let M_0 denote the atom set of hydrogens and their related chemical bonds. Then, M_t contains the noised atom set with random dropout or supplemental new atoms. During the denoising process, the 2D part inside the subgraph S_0 (H_t and E_t) is adjusted, while the 3D positions are fine-tuned.

3D-Aware Substituent Group Optimization. A common template-based molecule synthesis pipeline involves two steps: 1. Finding the optimization site[58]; 2. Composing the selected fragments at the optimization site. This scenario imposes a highly nontrivial challenge on latent-space based representation methods for molecules[35], as locating the precise location of the optimization site while keeping the surrounding structure unchanged is difficult. However, leveraging **3D structure-based information** provides direct guidance for optimization by noising only the optimization area and generating the substituent group by moving random atoms to the optimization site. Formally, let $\mathbf{v}_e$ denote the position of the optimization site. We initialize a set of atoms with 3D positions denoted by $\mathbf{G}_T$ according to the stationary distribution of the noising process. We then translate $\mathbf{G}_0$ to the optimization site position and start denoising:

$$\mathbf{G}_T \rightarrow \mathbf{G}_T + \mathbf{v}_e \xrightarrow{df\text{guided Denoising}} \mathbf{G}_0^f. \tag{S7}$$

The algorithms of 3DToMolo for three guidance optimization settings are provided in Section S15..

S5. Calculation of Redox Potentials

To rapidly assess the oxidation and reduction potentials of generated molecules, we employ machine learning techniques for qualitative structure-property relationship (QSPR) predictions based on straightforward structural descriptors. These descriptors, inspired by prior research[59], encompass factors such as the number of atoms of each element with the same coordination number, the count of rings with various sizes, the presence of aromatic rings, and the quantity of hydroxyl groups. A comprehensive list of descriptors is provided in Table S1. The dataset used comprises DFT-calculated redox potentials, combining 149 electrolyte additives for lithium-ion batteries[59] and 41,801 molecules from the Organic Materials for Energy Application Database (OMEAD)[60]. We use the gradient boosting regression (GBR) method. The model is trained on three-quarters of the dataset and validated on the remaining quarter. In configuring the hyperparameters for the scikit-learn [1] GBR module, we set the learning rate to 0.001, the number of estimators to 100,000, the maximum depth to 6, the minimum sample split to 3, and the maximum features to 2. On the test dataset, 3DToMolo yields a root mean square error (RMSE) of 0.455 V for the reduction potential and 0.447 V for the oxidation potential, as illustrated in Figure S1.

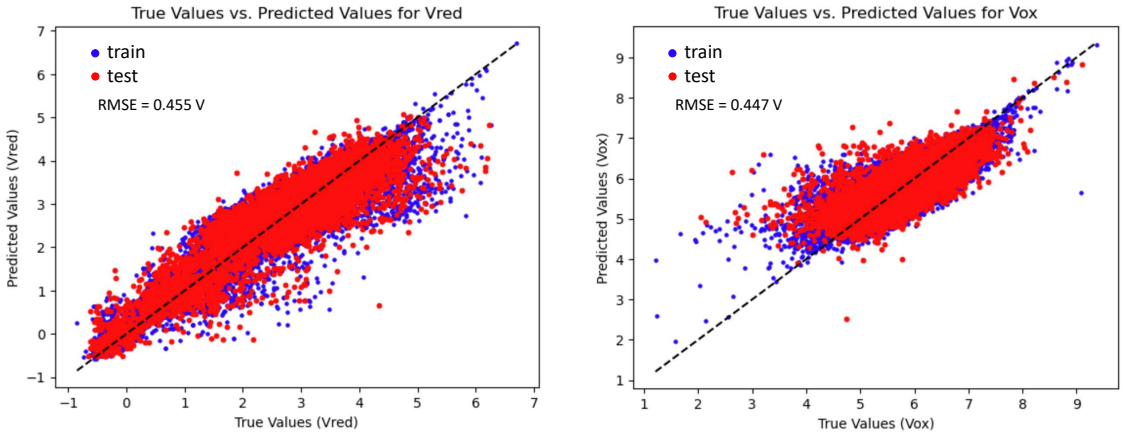

Figure S1. Gradient boosting regression on the DFT calculated redox potentials. The dataset consists of 149 electrolyte additives for lithium-ion batteries[59] and 41,801 molecules in the organic materials for energy application database (OMEAD)[60]. Three quarters of the dataset are randomly sampled to be the train set while the rest are used as the test set. The root mean square errors (RMSE) on the test set are 0.455 V and 0.447 V respectively for the reduction potential (left) and the oxidation potential (right).

Table S1. Descriptors used in the redox potential regressor.

Class	Descriptors
Atomic coordination number	B3, C4, C3, C2, N3, N2, N1, O2, O1, F1, Si4, P4, P3, S4, S3, S2, S1, Cl1
Functional group & substructure	Three- or four-member ring, five-member ring, six- or more-member ring, aromatic ring, hydroxyl group

As in the referenced work[59], in the *atomic coordination number* class, the letter stands for the element and the digit stands for the coordination number. For example, we count how many carbon atoms have a coordination number of 4 and use this number as the value for the descriptor "C4".

[1] https://scikit-learn.org/stable/

S6. Similarity vs. denoising steps

The similarity between two molecules is quantified by the Tanimoto coefficient based on the Morgan extended-connectivity fingerprints (ECFP) for the molecules[61]. As references, given by Tanimoto coefficient (ECFP), the similarity between benzene and thiophene is $2/9 \approx 0.222$, and the similarity between naphalene and quinoxaline is 0.3125.

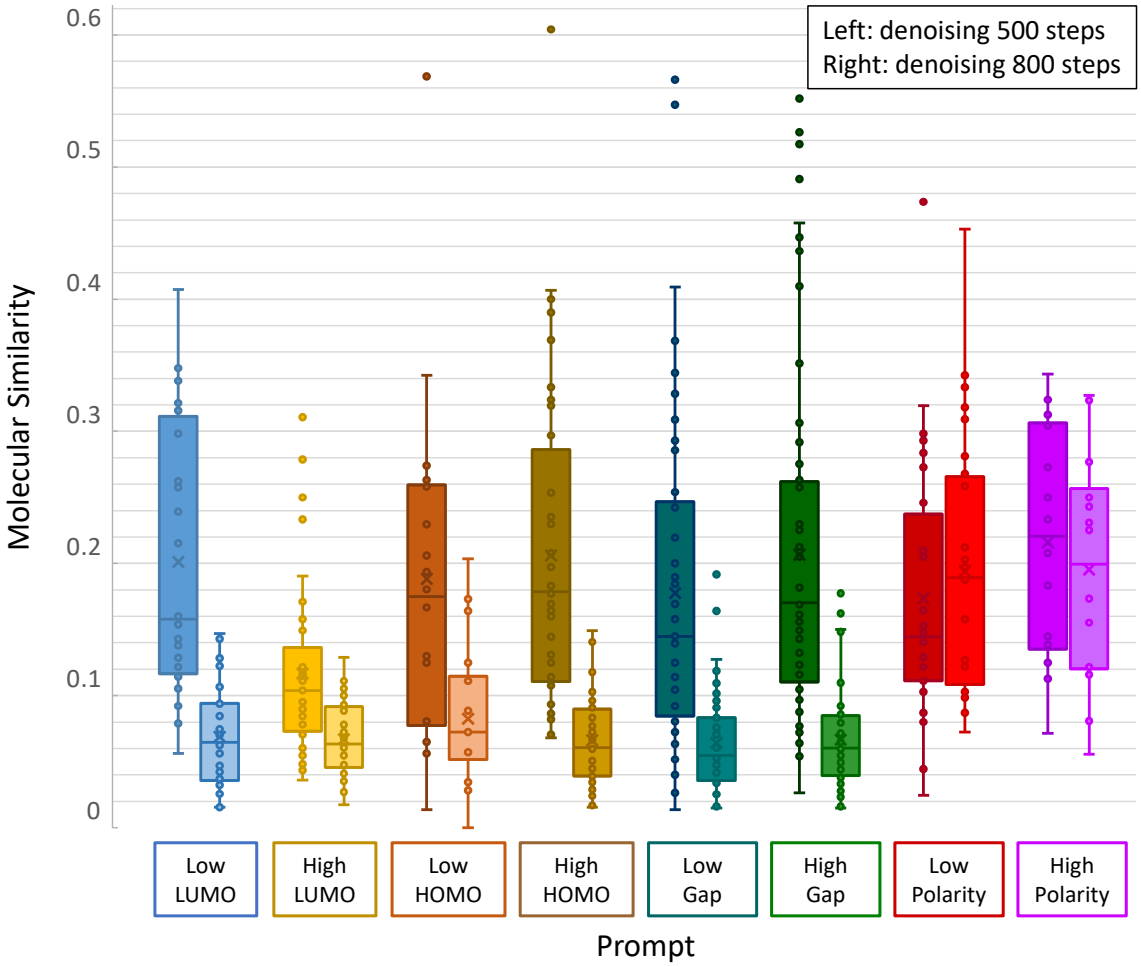

Figure S2. Box plots of molecular similarity between the optimized molecule and the original molecule in two settings: denoising 500 steps and denoising 800 steps. The average is marked by the symbol "×" and the median is marked by the horizontal line.

S7. More optimized results with non-coplanar benzene rings

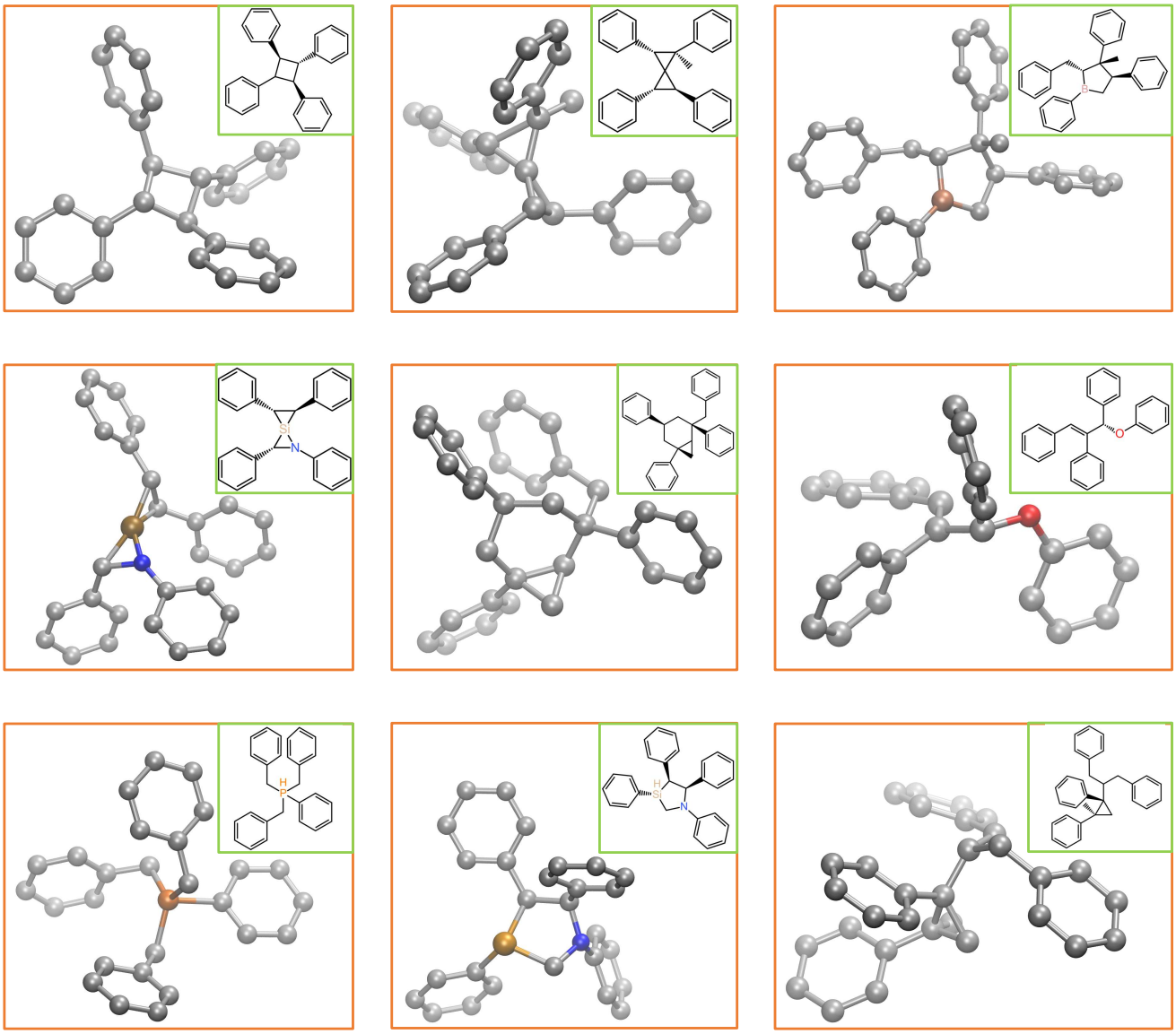

Figure S3. Selected optimized structures with non-coplanar benzene rings using the tetraphenylsilane molecule as the template.

S8. Tetraphenylsilane optimization task using GPT3.5.

Table S2. GPT3.5 optimization on the tetraphenylsilane molecule.

Prompt	Validity ratio	Hit ratio
1) Please edit the central silicon atom of the tetraphenylsilane molecule while maintaining the non-coplanar structure of the four benzene rings. Please generate (another) 10 edited molecules in SMILES.	0.64	0
2) Tetraphenylsilane is a molecule that has four non-coplanar benzene rings. Its SMILES is **smiles**. Please edit the central silicon atom of the tetraphenylsilane molecule while the edited molecules maintain four non-coplanar benzene rings. Please generate (another) 10 edited molecules in SMILES.	1.00	0
3) Please generate (another) 10 molecules that have four non-coplanar benzene rings in SMILES.	1.00	0
4) Tetraphenylsilane is a molecule that has four non-coplanar benzene rings. Its SMILES is **smiles**. Please edit the internal region of the tetraphenylsilane molecule. The edited molecules must have four benzene rings that are non-coplanar. Please generate (another) 10 edited molecules in SMILES.	1.00	0
5) Tetraphenylsilane is a molecule that has four non-coplanar benzene rings. Its SMILES is **smiles**. Please edit the internal region of the tetraphenylsilane molecule. The edited molecules must have four benzene rings. These four benzene must be non-coplanar. Please generate (another) 10 edited molecules in SMILES.	1.00	0
6) Tetraphenylsilane is a molecule that has four non-coplanar benzene rings. Its SMILES is **smiles**. Please replace the central silicon atom by other substructure such that the edited molecules have four non-coplanar benzene rings. Please generate (another) 10 edited molecules in SMILES.	1.00	0.70
7) Tetraphenylsilane is a molecule that has four non-coplanar benzene rings. Its SMILES is **smiles**. Please replace the central silicon atom by other substructure with more than one atom such that the edited molecules have four non-coplanar benzene rings. Please generate (another) 10 edited molecules in SMILES	1.00	0

We asked GPT3.5 to generate 10 molecules everytime and repeated several times. From the second time and after, we added "another" to the prompt. In most cases, GPT3.5 fails because the generated molecule does not have four benzene rings or put them in linear connection. In those satisfactory cases, GPT3.5 is simply replacing the silicon atom by other atoms, including Ge, Sn, As, P, Pb, Bi and so on. The abbreviated SMILES string in the instructions is as follows: **smiles** = "c1ccc(c(c1)[Si](c2ccccc2)(c3ccccc3)c4ccccc4)".

S9. Multi-site optimization task using GPT3.5.

Table S3. GPT3.5 editing on the penicillin molecule.

Prompt	Validity ratio	Hit ratio
1) Please edit the penicillin molecule **smiles** without modifying the sub-structure **sub-smiles** so that it has a large electron withdrawing group, which may have low HOMO (Highest occupied molecular orbital) energy of the carbonyl group. Please generate (another) 10 edited smiles.	0	0
2) Please edit the penicillin molecule **smiles** without modifying the sub-structure **sub-smiles** so that it has a large electron withdrawing group, which may have low HOMO (Highest occupied molecular orbital) energy of the carbonyl group. Please generate (another) 10 valid edited smiles.	0	0
3) Please edit the penicillin molecule **smiles** so that it has a large electron withdrawing group next to the amide group, which may have low HOMO (Highest occupied molecular orbital) energy of the carbonyl group. Please generate (another) 10 valid edited smiles.	0	0
4) Please edit the penicillin molecule **smiles** so that it has a better resistance to acids and lactamases. Please generate (another) 10 edited smiles.	0	0
5) Please edit the penicillin molecule so that it has a better resistance to acids and lactamases. Please generate (another) 10 edited smiles.	0	0
6) Please modify the benzyl group in the penicillin molecule so that the molecule has a large electron withdrawing group. Please generate (another) 10 edited molecules in SMILES.	0	0
7) Please modify the benzyl group in the penicillin molecule while maintain the rest of the structure so that the molecule has a large electron withdrawing group. Please generate (another) 10 edited molecules in SMILES.	0	0
8) Please modify the benzyl group in the penicillin molecule while maintain the rest of the structure so that the molecule has a large electron withdrawing group. An example of satisfactory editing is **example-smiles**. Please generate (another) 10 edited molecules in SMILES.	0.20	0
9) Please modify the benzyl group in the penicillin molecule while maintain the rest of the structure so that the molecule has a large electron withdrawing group. An example of satisfactory editing is **example-smiles**. Please generate (another) 10 edited molecules in SMILES that you had never provided.	0	0

We asked GPT3.5 to generate 10 molecules everytime and repeated several times. From the second time and after, we added "another" to the prompt. GPT3.5 failed to generate valid SMILES strings in most cases. None of valid cases has preserved the core structure and thus could not be regarded as satisfactory. The abbreviated SMILES strings in the instructions are as follows:

smiles = "O=C(N[C@@H]1C(=O)N2[C@H](C(S[C@]12[H])(C)C)C(=O)O)CC1C=CC=CC=1".

sub-smiles = "O=C(N[C@@H]1C(=O)N2[C@H](C(S[C@]12[H])(C)C)C(=O)O)".

example-smiles = "CC1=C(C(=NO1)C2=CC=CC=C2)C(=O)N[C@H]3[C@@H]4N(C3=O)[C@H](C(S4)(C)C)C(=O)O".

S10. Diversity of optimization results and multi-run performance

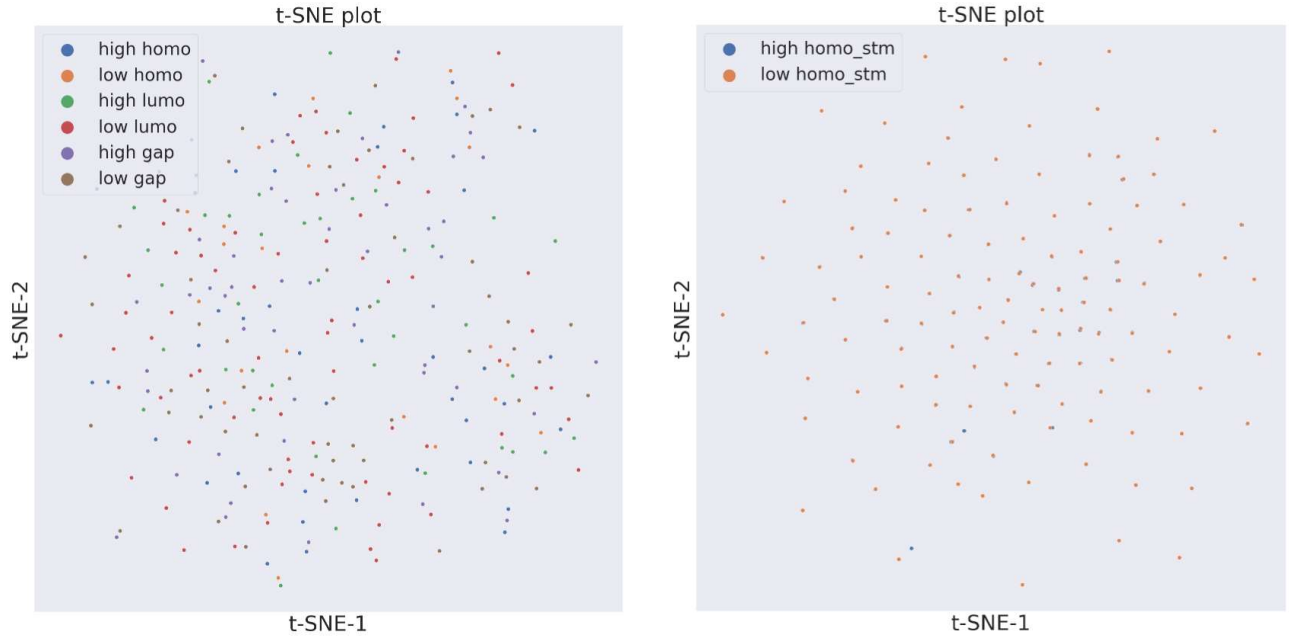

Figure S4. T-SNE plots of valid optimization results of 200 ZINC molecules under various prompts. In a t-SNE plot, each dot represents a molecule. The short distance between two dots means the corresponding molecules have similar structures. On the left is the plot for our 3DToMolo model. On the right is the plot for MoleculeSTM results that we reproduce. The overlaps of blue and orange dots in the right plot reflect that for most of input molecules, the output results by MoleculeSTM are in fact the same, even though the prompts are opposite.

Table S4. Multi-run performance evaluation compared with Single-run

Prompts	MoleculeSTM	GPT3.5	3DToMolo
This molecule has low HOMO (Highest occupied molecular orbital) value, which is more stable.	35.00 -> 35.00	37.50 -> 72.00	46.50 -> 75.00
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 -> 28.00	40.50 -> 82.50	58.50 -> 97.50
This molecule has more hydrogen bond donors.	05.00 -> 05.00	12.50 -> 74.00	34.50 -> 78.00
This molecule has high permeability and 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.	01.00 -> 01.00	18.00 -> 44.00	33.50 -> 94.00

S11. Summary of multi-objective molecule optimization results

Table S5. Multi-objective molecule optimization

Prompts	MoleculeSTM	GPT3.5	3DToMolo
This molecule is soluble in water and has high polarity.	08.00	68.00	31.50
This molecule is insoluble in water and has low polarity.	06.00	59.00	95.00
This molecule has high permeability and 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.	01.00	44.00	94.00
This molecule is soluble in water and has more hydrogen bond acceptors.	01.00	66.50	55.00
This molecule is insoluble in water and has more hydrogen bond acceptors.	01.00	24.00	15.00

S12. Synthetic accessibility of optimization results

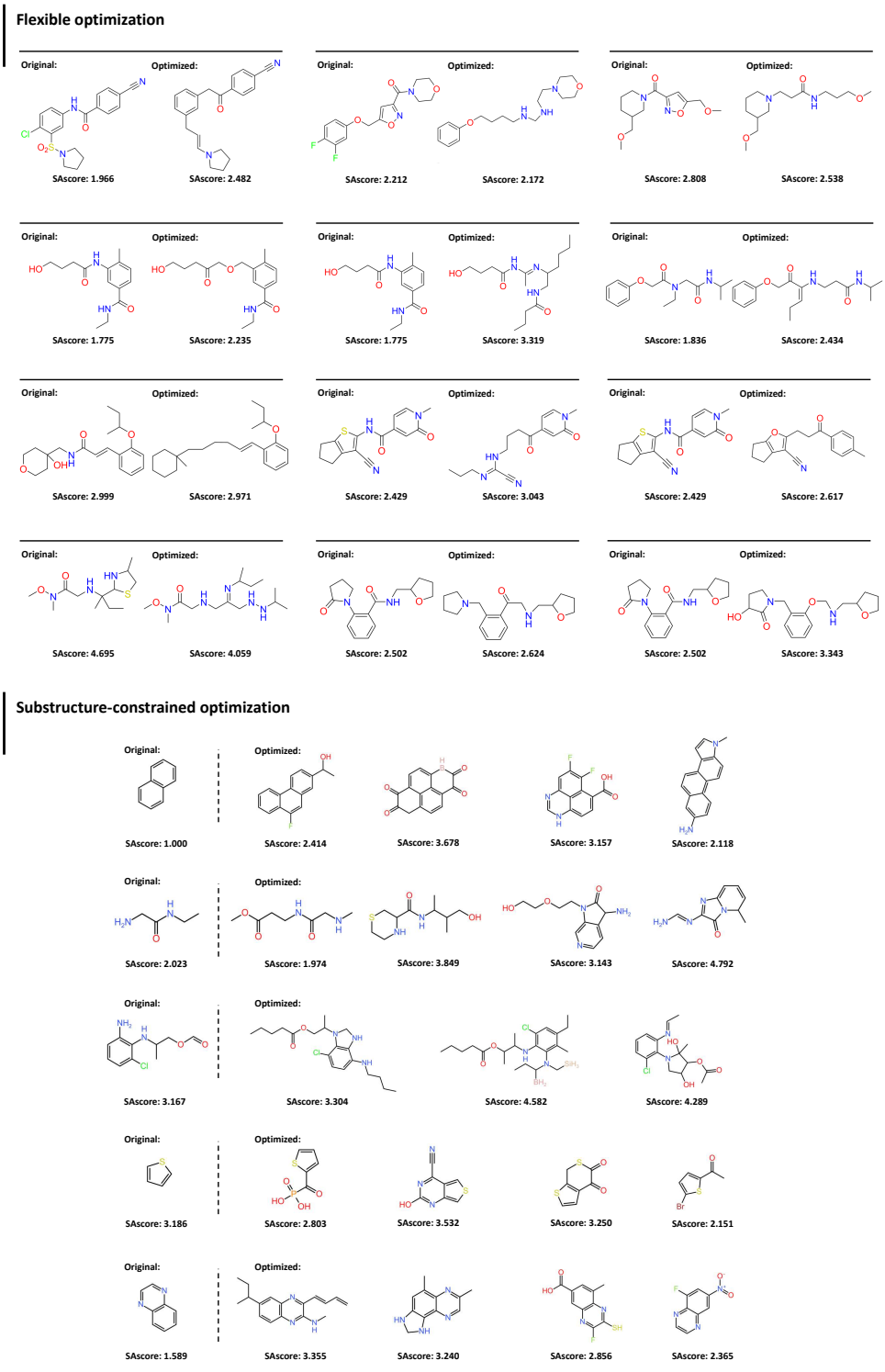

Figure S5. A summary of the synthetic accessibility score (SAscore) for presented molecules in the main article. The SAscore ranges from 1 to 10, where 1 means "easy to make" and 10 means "very difficult to make". As references, SAscores for natural products mainly range from 5 ~ 6 and SAscores for bioactive molecules mainly range from 3 ~ 4.[62]

S13. Validity of our pretrained generative model

Table S6. Validation results on PCQM4Mv2.

Model	Mol stable ↑	Atom stable ↑	Validity ↑	Unique ↑	AtomTV ↓	BondTV ↓	ValW1 ↓	Bond Lengths W1 ↓	Bond Angles W1 ↓
EDM	55.0	92.9	34.8	100.0	0.212	0.049	0.112	0.002	6.23
3DToMolo[63]	**84.5**	**99.6**	**80.0**	**100.0**	**0.059**	**0.021**	**0.008**	**0.003**	**2.16**

We provide standard metrics for testing the performance of 3DToMolo's diffusion model. The molecule's stability and validity are measured by default Rdkit[64] algorithms. We also provide W_1 and total variation distances between our generated molecules' 2D topology and the training set, proposed by the referenced work[52]. The hyper-parameters for training the baseline EDM[63] follow the same setting as the referenced work[52]. As we can see from the table, 3DToMolo achieves significantly better performance than EDM, which demonstrates the effectiveness of our proposed equivariant transformer for generating valid and diverse molecules in the chemical space.

S14. Validation of backbone model on chirality distinction

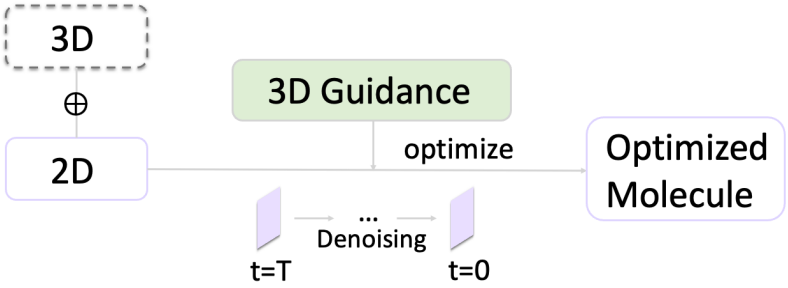

Figure S6. Method Sketch.

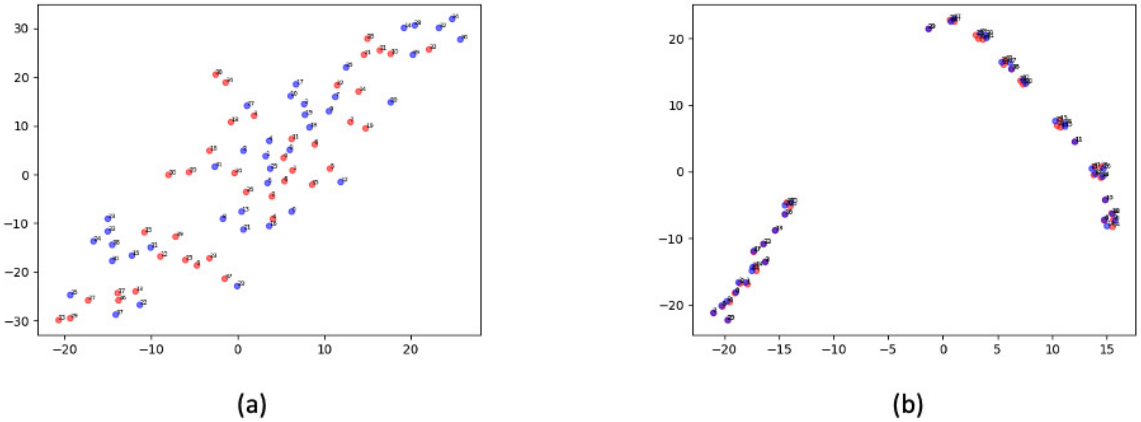

Figure S7. Comparison between 3DToMolo and EDM[63] on the chirality discrimination. T-SNE projects the representations of molecules with different noise-adding steps (we take 0 to 200 at an interval of 5) into a 2D representation. In the scatter plot, the red and blue dots represent a pair of left and right-handed chiral molecules, respectively. Figure (a) is the scatter plot for our 3DToMolo and (b) is for EDM. The dispersion or overlap of scatter points implies the capability of model for distinguishing left and right-handed chirality molecules. We could observe that the backbone model of 3DToMolo could distinguish between left and right-handed molecules with high precision, which is not achievable for EDM.

S15. Algorithms of 3DToMolo for three guidance optimization settings

Algorithm 1 Flexible molecule optimization under text prompts

Input: Given a molecule to be optimized $M_0 := (H_0, E_0, P_0)$ and text-prompt y implies the orientation of optimization.

Sampling atoms to input: $H_0, E_0, P_0 \leftarrow H_0 \cup H_{ex}, E_0 \cup E_{ex}, P_0 \cup P_{ex}$
for $t := 1$ to T **do**
  $z_{t-1} \leftarrow (E_{t-1}, H_{t-1})$
  $H_t, E_t \leftarrow C(z_{t-1} Q_t)$
  $P_t \leftarrow \sqrt{\bar{\alpha}_t} P_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$ ▷ Forward process
end for
$M'_T \leftarrow M_T$
for $t := T$ to 1 **do** ▷ Conditional denoising process
  $q_\theta(M'_{t-1}|M'_t, y) \propto q_\theta(M'_{t-1}|M'_t) \cdot e^{-\lambda < \nabla_{M'_t} ||y - f(M'_t)||^2, M'_{t-1} >}$
  $M'_{t-1} \sim q_\theta(M'_{t-1}|M'_t, y)$ ▷ Sample based on the differential of CLIP loss
end for

Output: Output optimized molecule M'_0 with desired property y.

Algorithm 2 Molecule optimization with Structural constraints

Input: Given a molecule to be optimized $M_0^c := (H_0^c, E_0^c, P_0^c)$ with substructure S_0 to be protected and text-prompt y implies the orientation of optimization.

$M_0^c \leftarrow M_0 \cup S_0, M_0 := (H_0, E_0, P_0)$
Sampling atoms to M_0: $H_0, E_0, P_0 \leftarrow H_0 \cup H_{ex}, E_0 \cup E_{ex}, P_0 \cup P_{ex}$
for $t := 1$ to T **do**
  $z_{t-1} \leftarrow (E_{t-1}, H_{t-1})$
  $H_t, E_t \leftarrow C(z_{t-1} Q_t)$
  $P_t \leftarrow \sqrt{\bar{\alpha}_t} P_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$ ▷ Forward process
end for
$M_T := (H_T \cup S_0, E_T \cup S_0, P_T \cup S_0)$ ▷ Substructure remain fixed
$M'_T \leftarrow M_T$
for $t := T$ to 1 **do** ▷ Conditional denoising process
  $q_\theta(M'_{t-1}|M'_t, y) \propto q_\theta(M'_{t-1}|M'_t) \cdot e^{-\lambda < \nabla_{M'_t} ||y - f(M'_t)||^2, M'_{t-1} >}$
  $M'_{t-1} \sim q_\theta(M'_{t-1}|M'_t, y)$ ▷ Sample based on the differential of CLIP loss
  $M'_{t-1}[S_0] \leftarrow M_T[S_0]$
end for

Output: Output optimized molecule M'_0 with desired property y.

Algorithm 3 Hard-coded molecule optimization on appointed sites

Input: Given a molecule to be optimized $M_0^c := (H_0^c, E_0^c, P_0^c)$ with editing site X and text-prompt y.

Extract a subgraph of input surrounding X: $M_0^s \leftarrow Subgraph(M_0^c|X, k)$ ▷ K-hop subgraph
$M_0 \leftarrow M_0^s[X], M_0 := (H_0, E_0, P_0) = (H_0^s[X], E_0^s[X], E_0^s[X])$
Sampling atoms to M_0: $H_0, E_0, P_0 \leftarrow H_0 \cup H_{ex}, E_0 \cup E_{ex}, P_0 \cup P_{ex}$
Match the center of mass (COM) of (H_{ex}, E_{ex}, P_{ex}) with $\bar{\mathbf{X}} := \sum_i \mathbf{x}_i / n$
for $t := 1$ to T **do**
    $z_{t-1} \leftarrow (E_{t-1}, H_{t-1})$
    $H_t, E_t \leftarrow C(z_{t-1} Q_t)$
    $P_t \leftarrow \sqrt{\bar{\alpha}_t} P_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon$ ▷ Forward process
end for
$M_T := (H_T \cup M_0^s, E_T \cup M_0^s, P_T \cup M_0^s)$
$M_T' \leftarrow M_T$
for $t := T$ to 1 **do** ▷ Conditional denoising process
    $q_\theta(M_{t-1}'|M_t', y) \propto q_\theta(M_{t-1}'|M_t') \cdot e^{-\lambda < \nabla_{M_t'} ||y - f(M_t')||^2, M_{t-1}' >}$
    $M_{t-1}' \sim q_\theta(M_{t-1}'|M_t', y)$ ▷ Sample based on the differential of CLIP loss
    $M_{t-1}'[M_0^s \backslash X] \leftarrow M_T[M_0^s \backslash X]$
end for

Output: Output optimized molecule M_0' with desired property y.

References

1. Gómez-Bombarelli, R. & et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* 10.1021/acscentsci.7b00572 (2018).
2. Kusner, M. J., Paige, B. & Hernández-Lobato, J. M. Grammar variational autoencoder. In *International conference on machine learning*, 1945–1954 (PMLR, 2017).
3. Nakata, Y. & et al. Molecular generation for organic electrolyte molecule discovery using conditional variational autoencoders. *The J. Phys. Chem. Lett.* 10.1021/acs.jpclett.8b02011 (2018).
4. Simonovsky, M. & Komodakis, N. Constrained graph variational autoencoders for molecule design. *arXiv preprint arXiv:1805.09076* (2018). 1805.09076.
5. Goodfellow, I. *et al.* Generative adversarial networks. *Commun. ACM* **63**, 139–144 (2020).
6. Prykhodko, O. *et al.* A de novo molecular generation method using latent vector based generative adversarial network. *J. Cheminformatics* **11**, 1–13 (2019).
7. Gomez-Bombarelli, R. & et al. Chemgan challenge for drug discovery: can ai reproduce natural chemical diversity? *ChemRxiv* 10.26434/chemrxiv.5309669.v1 (2018).
8. De Cao, N. & Kipf, T. Molgan: An implicit generative model for small molecular graphs. *arXiv preprint arXiv:1805.11973* (2018). 1805.11973.
9. Krishnan, S. R. *et al.* De novo structure-based drug design using deep learning. *J. Chem. Inf. Model.* **62**, 5100–5109 (2021).
10. Arús-Pous, J. *et al.* Randomized smiles strings improve the quality of molecular generative models. *J. cheminformatics* **11**, 1–13 (2019).
11. Bagal, V., Aggarwal, R., Vinod, P. & Priyakumar, U. D. Molgpt: molecular generation using a transformer-decoder model. *J. Chem. Inf. Model.* **62**, 2064–2076 (2021).
12. Mahmood, O., Mansimov, E., Bonneau, R. & Cho, K. Masked graph modeling for molecule generation. *Nat. communications* **12**, 3156 (2021).
13. Gupta, A. *et al.* Generative recurrent networks for de novo drug design. *Mol. informatics* **37**, 1700111 (2018).
14. Li, Y., Pei, J. & Lai, L. Structure-based de novo drug design using 3d deep generative models. *Chem. science* **12**, 13664–13675 (2021).
15. He, J. *et al.* Molecular optimization by capturing chemist's intuition using deep neural networks. *J. Cheminformatics* **13**, 26, 10.1186/s13321-021-00497-0 (2021).

16. Hoffman, S. C., Chenthamarakshan, V., Wadhawan, K., Chen, P.-Y. & Das, P. Optimizing molecules using efficient queries from property evaluations. *Nat. Mach. Intell.* **4**, 21–31, 10.1038/s42256-021-00422-y (2022).

17. Atance, S. R., Diez, J. V., Engkvist, O., Olsson, S. & Mercado, R. De novo drug design using reinforcement learning with graph-based deep generative models. *J. Chem. Inf. Model.* **62**, 4863–4872, 10.1021/acs.jcim.2c00838 (2022). PMID: 36219571, https://doi.org/10.1021/acs.jcim.2c00838.

18. Popova, M., Isayev, O. & Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci. advances* **4**, eaap7885 (2018).

19. Olivecrona, M., Blaschke, T., Engkvist, O. & Chen, H. Molecular de novo design through deep reinforcement learning. *J. cheminformatics* **9**, 48 (2017).

20. Putin, E. *et al.* Reinforcement learning for molecular de novo design. *J. cheminformatics* **10**, 1–11 (2018).

21. You, J., Liu, B., Ying, R., Pande, V. & Leskovec, J. Graph convolutional policy network for goal-directed molecular graph generation. In *Advances in Neural Information Processing Systems*, 6410–6421 (2018).

22. Segler, M. H., Preuss, M. & Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic ai. *Nature* **555**, 604–610 (2018).

23. Hu, W. *et al.* Strategies for pre-training graph neural networks. *arXiv preprint arXiv:1905.12265* (2019).

24. Liu, S., Guo, H. & Tang, J. Molecular geometry pretraining with se(3)-invariant denoising distance matching. *ArXiv* **abs/2206.13602** (2022).

25. Karpathy, A. & Fei-Fei, L. Deep visual-semantic alignments for generating image descriptions. *IEEE Transactions on Pattern Analysis Mach. Intell.* **39**, 664–676, 10.1109/TPAMI.2015.2437384 (2015).

26. Lu, J., Batra, D., Parikh, D. & Lee, S. Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 13–23, 10.5555/3321432.3321512 (2019).

27. Wang, L. *et al.* Temporal segment networks: Towards good practices for deep action recognition. *arXiv preprint arXiv:1608.00859* (2016).

28. Miech, A. *et al.* Howto100m: Learning a text-video embedding by watching hundred million narrated video clips. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2652–2661, 10.1109/ICCV.2019.00276 (2019).

29. Liu, Z. & Yang, J. Advancements in multimodal pretraining for natural language processing. *J. Artif. Intell. Res.* **56**, 567–580, 10.1614/JAIR.2023.01234 (2023).

30. Sanchez-Lengeling, B. & Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **361**, 360–365 (2018).

31. Segler, M. H., Kogej, T., Tyrchan, C. & Waller, M. P. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS central science* **4**, 120–131 (2018).

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

33. Duvenaud, D. K. *et al.* Convolutional networks on graphs for learning molecular fingerprints. *Adv. neural information processing systems* **28** (2015).

34. Liu, S., Demirel, M. F. & Liang, Y. N-gram graph: Simple unsupervised representation for graphs, with applications to molecules. *Adv. neural information processing systems* **32** (2019).

35. Liu, S. *et al.* Multi-modal molecule structure-text model for text-based retrieval and editing. *Nat. Mach. Intell.* **5**, 1447–1457, 10.1038/s42256-023-00759-6 (2023).

36. 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. communications* **13**, 862 (2022).

37. Li, S. *et al.* Towards 3d molecule-text interpretation in language models. *ArXiv* **abs/2401.13923** (2024).

38. Liu, S. *et al.* Conversational drug editing using retrieval and domain feedback. In *The Twelfth International Conference on Learning Representations* (2024).

39. Leo, A., Hansch, C. & Elkins, D. Partition coefficients and their uses. *Chem. reviews* **71**, 525–616 (1971).

40. Bickerton, G. R., Paolini, G. V., Besnard, J., Muresan, S. & Hopkins, A. L. Quantifying the chemical beauty of drugs. *Nat. chemistry* **4**, 90–98 (2012).

41. Ertl, P., Rohde, B. & Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Medicinal Chem.* **43**, 3714–3717, 10.1021/jm000942e (2000). PMID: 11020286, https://doi.org/10.1021/jm000942e.

42. Caumes, G., Borrel, A., Abi Hussein, H., Camproux, A.-C. & Regad, L. Investigating the importance of the pocket-estimation method in pocket-based approaches: An illustration using pocket-ligand classification. *Mol. informatics* **36**, 1700025 (2017).

43. Mendez, D. *et al.* Chembl: towards direct deposition of bioassay data. *Nucleic acids research* **47**, D930–D940 (2019).

44. Kardos, N. & Demain, A. L. Penicillin: the medicine with the greatest impact on therapeutic outcomes. *Appl. microbiology biotechnology* **92**, 677–687 (2011).

45. Waxman, D. J. & Strominger, J. L. Penicillin-binding proteins and the mechanism of action of beta-lactam antibiotics. *Annu. Rev. Biochem.* **52**, 825–869 (1983).

46. Klein, A. R. *et al.* Probing the fate of different structures of beta-lactam antibiotics: Hydrolysis, mineral capture, and influence of organic matter. *ACS Earth Space Chem.* **5,6**, 1511–1524 (2021).

47. Lima, L. M., da Silva, B. N. M., Barbosa, G. & Barreiro, E. J. β-lactam antibiotics: An overview from a medicinal chemistry perspective. *Eur. journal medicinal chemistry* **208**, 112829 (2020).

48. Tooke, C. L. *et al.* β-lactamases and β-lactamase inhibitors in the 21st century. *J. molecular biology* **431**, 3472–3500 (2019).

49. 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. product reports* **29**, 457–475 (2012).

50. Tong, L. *et al.* Triptolide: reflections on two decades of research and prospects for the future. *Nat. product reports* **38**, 843–860 (2021).

51. Du, W., Chen, J., Zhang, X., Ma, Z.-M. & Liu, S. Molecule joint auto-encoding: Trajectory pretraining with 2d and 3d diffusion. In *Thirty-seventh Conference on Neural Information Processing Systems* (2023).

52. Vignac, C., Osman, N., Toni, L. & Frossard, P. Midi: Mixed graph and 3d denoising diffusion for molecule generation. *arXiv preprint arXiv:2302.09048* (2023).

53. Du, W. *et al.* SE(3) equivariant graph neural networks with complete local frames. In Chaudhuri, K. *et al.* (eds.) *Proceedings of the 39th International Conference on Machine Learning*, vol. 162 of *Proceedings of Machine Learning Research*, 5583–5608 (PMLR, 2022).

54. Batzner, S. *et al.* E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. communications* **13**, 2453 (2022).

55. Vaswani, A. *et al.* Attention is all you need. In Guyon, I. *et al.* (eds.) *Advances in Neural Information Processing Systems*, vol. 30 (Curran Associates, Inc., 2017).

56. Loeffler, H. H. *et al.* Reinvent 4: Modern ai–driven generative molecule design. *J. Cheminformatics* **16**, 20, 10.1186/s13321-024-00812-5 (2024).

57. Du, W., Zhang, H., Yang, T. & Du, Y. A flexible diffusion model. In *International Conference on Machine Learning*, 8678–8696 (PMLR, 2023).

58. Powers, A. S. *et al.* Geometric deep learning for structure-based ligand design. *ACS Cent. Sci.* **9**, 2257–2267, 10.1021/acscentsci.3c00572 (2023). https://doi.org/10.1021/acscentsci.3c00572.

59. 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* **3**, 7868–7874 (2018).

60. Carvalho, R. P., Brandell, D. & Araujo, C. M. An evolutionary-driven ai model discovering redox-stable organic electrode materials for alkali-ion batteries. *Energy Storage Mater.* **61**, 102865 (2023).

61. Bajusz, D., Rácz, A. & Héberger, K. Why is tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. cheminformatics* **7**, 1–13 (2015).

62. Ertl, P. & Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. cheminformatics* **1**, 1–11 (2009).

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

64. Landrum, G. *et al. rdkit/rdkit: 2020_03_1 (Q1 2020) Release* (Zenodo, 2020).