ECloudGen: Access to Broader Chemical Space for Structure-based Molecule Generation

Odin Zhang[*] Jieyu Jin[*] Haitao Lin[†] Jintu Zhang[*] Chenqing Hua[‡]
Yufei Huang[*] Huifeng Zhao[*] Chang-Yu Hsieh[*] Tingjun Hou[*]

Abstract

AI-aided drug design has facilitated structure-based molecule generation strategies. However, despite significant success, the restriction of the scarcity of protein-ligand data prevents the models from fully exploiting the learning chemical space and discovering unexplored potential drugs. The limited chemical space sampling contrasts with the original intention of generation models to explore a broader chemical space, leading to what we term the Chemical Space Generation Paradox. To address the proposed paradox, we developed ECloudGen with the following attributes: (1) Fundamental Physical Representation: We introduce the electron cloud representation, unifying all biological forces under one representation, offering a compact and continuous learning space. (2) Broad and Structurally Ordered Chemical Space: Utilizing electron clouds as generative agents, ECloudGen leverages data without binding structure to access a broader chemical space. In implementation, ECloudDiff as a latent ECloud-based diffusion model is established to sample high-fidelity electron clouds conditioned on pockets's structure; and CEMP as a novel contrastive learning strategy is proposed to structurally organize the chemical space, thus enabling controllable generation. Subsequent experiments confirm ECloudGen's state-of-the-art performance, in generating chemically feasible molecules with high binding efficacy, drug-likeness, and other chemical properties. Besides, ECloudGen proves to encompass a broader chemical space and also demonstrates superiority in controllable generation in extensive experiments.

1 Introduction

The pursuit of new pharmaceuticals is one of the critical strategies in combating diseases [1]. Within the space of Artificial Intelligence-Driven Drug Discovery (AIDD), molecule generation is core work tasked with the discovery of innovative, effective medicinal compounds [2]. Contemporary research has shifted gears to a systematic approach in this area referred to as structure-based molecule generation. This method carefully 'designs' molecules targeted to specific protein structures, thereby differing from the traditional ligand-based methods that 'randomly generate' molecules by replicating the graph patterns found in the training data [3]. It thus heralds a change toward a more proactive and focused approach to exploiting the potential of AI in changing the field of drug discovery.

The field of Structure-Based Molecule Generation (SBMG) has witnessed significant advancements with a variety of modeling techniques emerging, ranging from auto-regressive methods [4, 5, 6, 7] to diffusion-based methods [8, 9, 10, 11]. However, a fundamental challenge remains largely unaddressed: *the scarcity of 3D protein-ligand data severely limits the chemical space accessible for SBMG*. To illustrate, datasets like PDBBind provide only a limited snapshot of this space [12],

[*]Innovation Institute for AI in Medicine of Zhejiang University, Zhejiang University.
[†]AI Lab, Research Center for Industries of the Future, Westlake University
[‡]Montreal Institute for Learning Algorithms; McGill University, Montreal, QC, Canada
[‡]Odin Zhang and Jieyu Jin are co-first authors.

Preprint. Under review.

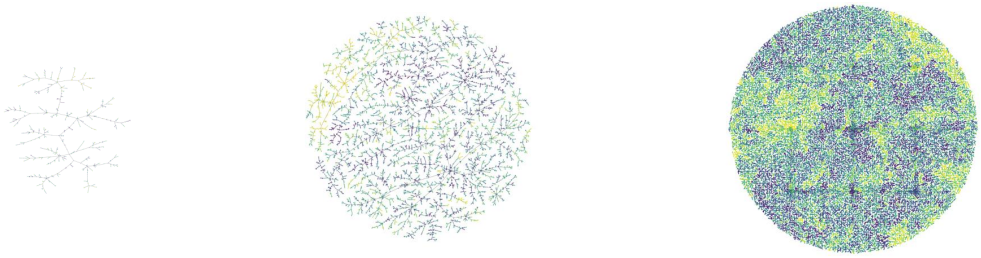

Figure 1: Illustration of the **Chemical Generation Space Paradox**, contrasting three scales of chemical spaces. The PDBBind dataset encompasses roughly 10^5 protein-ligand pairs, the chemical compound library contains about 10^9 molecules, and the vast estimated total chemical space holds approximately 10^{60} compounds.

containing only tens of thousands of protein-ligand pairs ($\sim 10^5$), a mere fraction compared to the billions of 2D molecules ($\sim 10^9$) cataloged in compound libraries such as ZINC [13]. This discrepancy is even more severe when considering the estimated total chemical space, which encompasses roughly 10^{60} compounds [14]. The demonstration of this data scarcity is visualized in Figure 1. Although generation models aim to navigate and explore wider chemical spaces, the sparseness of available data limits SBMG models to a much narrower chemical domain. This dilemma is what we term the **Chemical Generation Space Paradox**, underscoring the discrepancy between the vast potential of chemical diversity and the relatively narrow range that current SBMG models have explored.

To address this widely neglected paradox, we introduce ECloudGen, *a novel approach that emphasizes a more physical and continuous molecular representation through* ***electron clouds***. Although the ball-and-stick model is an early and very familiar concept in chemistry, it is important to note that it is only an approximation based upon classical mechanics [15]. The electron cloud model, on the other hand, is a quantum-based, foundational representation, as it describes the probability density of electrons [16]. This is more in tune with the dynamical and probabilistic nature of atoms [17, 18]. Learning using electron clouds offers two important advantages:

(1) Firstly, from a physics perspective, all the intermolecular forces reduce to only one force, i.e., electromagnetic force. Examples include hydrogen bonds, van der Waals forces, and $\pi - \pi$ stacking, which are basically variants of dipole-dipole interactions [19]. By using an electron cloud-based learning model, the model bypasses the redundant classical approximation and directly imbibes the unified laws of interaction for superior generalization capabilities.

(2) Secondly, compared with learning on discrete distribution, learning on continuous fields is an easier task. In the ball-and-stick representation, the difficulty of training is from both the atom positions, which discretize the spatial domain with existence and non-existence, and atom element types, which are treated as discrete variables (see Figure 5). In contrast, electron clouds are continuous in the spatial domain, enabling efficient utilization of the valuable protein-ligand data. Moreover, electron clouds provide a denser learning space. As shown in Appendix D, the molecules which appear distinct in the chemical space tend to merge in electron cloud space, thus providing a more generalization capability of the model toward new, unseen molecules, which is a very important advantage, considering the scarcity of 3D data. Therefore, we propose that generating electron clouds from protein-binding sites is a more effective strategy than directly sampling atom element types and positions.

In this way, ECloudGen is designed with two primary modules: the first generates accurate electron clouds from protein pockets using the CrossDock dataset [20]; the second translates these electron clouds into molecular structures, leveraging extensive chemical libraries like ZINC [13]. This operation on the electron clouds provides an easy, physical transformation from proteins into ligands, allowing the use of data without binding structures to expand the chemical space accessible to SBMG process. By this means, ECloudGen overcomes the 'Chemical Generation Space Paradox'. Overall, a visualization of the workflow of ECloudGen is illustrated in Figure 2. In implementation, we integrate the state-of-the-art to better refine the chemical space captured through the large datasets [21, 22]. In this respect, we proposed a **3D Conditional Latent Diffusion** for effective generation of high-fidelity electron clouds (Section 2.2), together with the Roformer architecture to decode the electron clouds into molecules [23]. We also develop the **Contrastive ECloud-Molecule Pre-training** (CEMP) to

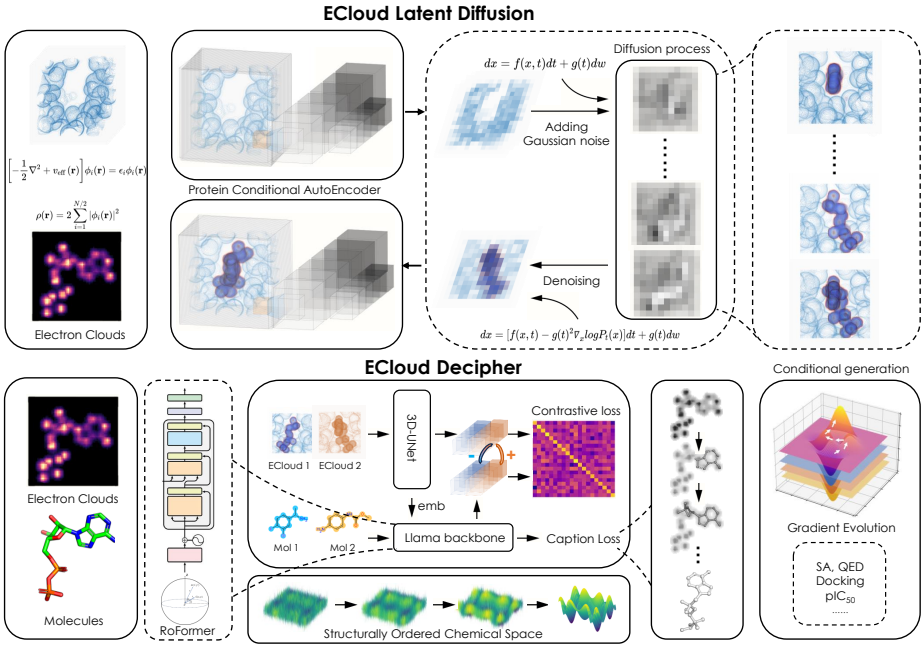

Figure 2: Demonstration of ECloudGen. The model consists of two primary modules of which the first module generates accurate electron clouds from protein pockets using the CrossDock dataset [20], and the second module translates these electron clouds into molecular structures, leveraging extensive chemical libraries like ZINC [13].

arrange the chemical space structurally (Section 2.3), thereby improving generalization and controlled generation. For a thorough understanding of the model design and the Chemical Space Generation Paradox, detailed discussions are provided in Section 2 and Appendix C.

2 ECloudGen: Addressing the Chemical Space Generation Paradox

2.1 Problem Formulation

Electron Cloud Electron clouds, also known as electron density, are the foundation of quantum chemistry. As we have discussed in Section 1, the electron cloud is fundamental because most physical properties of molecular systems are fundamentally linked with electron density. This provides a description not only of bonded interactions, like chemical bonds, but also of non-bonded interactions like intermolecular forces, giving a much more complete description [19]. Essentially, electron clouds are based on the solutions to the Schrödinger equation, which is a central concept in quantum mechanics [24], as follows:

$$\left[-\frac{1}{2}\nabla^2 + v_{\text{eff}}(\mathbf{r})\right]\phi_i(\mathbf{r}) = \epsilon_i\phi_i(\mathbf{r}), \qquad \rho(\mathbf{r}) = \sum_i |\phi_i(\mathbf{r})|^2. \tag{1}$$

Here, $\phi_i(\mathbf{r})$ stands for the electron's wave function, ϵ_i denotes the energy eigenvalue, $-\frac{1}{2}\nabla^2$ represents the kinetic energy operator, and $v_{\text{eff}}(\mathbf{r})$ is the effective potential. $\rho(\mathbf{r})$ refers to the electron cloud, indicating the probability density of the presence of an electron at a specific position $\mathbf{r}$. However, directly solving the Schrödinger equation to obtain the electron clouds is computationally expensive. To make the problem tractable, we apply an efficient semi-empirical GFN2-xTB method [25] for ECloudGen. GFN2-xTB, an adaptation of the tight-binding approach, is specifically designed for chemical systems and has been validated as sufficiently precise for calculating electron clouds [26].

In this work, we leverage this unique representation to address the **Chemical Generation Space Paradox** (see Section 1), thereby significantly improving the quality of molecule generation. We use the grid data structure to store the electron clouds, and formulate the problem as follows: given pocket electron clouds $\rho_p \in \mathbb{R}^{n_e \times n_e \times n_e}$, where n_e denotes the grid size, ECloudGen aims to sample suitable ligand electron clouds $\tilde{\rho}_l \sim P(\rho_l|\rho_p)$ that fit within the same grid space $\tilde{\rho}_l \in \mathbb{R}^{n_e \times n_e \times n_e}$. These clouds are then translated into a series of molecules $\{m_i\}_{i=0}^{N}$. In the context of conditional

generation, with the optimization goal of $\max_m f(m)$ subject to constraints $h_i(m) < 0$, ECloudGen seeks to identify a diverse set of molecules $\{m_i\}_{i=0}^N$ that achieve satisfactory optimization objective.

2.2 ECloud Latent Diffusion

The initial phase of ECloudGen involves generating electron clouds from specified protein pockets, framing it as a conditional 3D generation problem. Inspired by the success of diffusion models in image generation [27, 28, 21], we leverage this framework to generate high-fidelity electron clouds adhering to protein constraints. To this end, we develope the **ECloudLDiff** module, an adaptation of the diffusion model that functions within latent space. This module is not only memory-efficient but also adept at capturing semantically significant patterns in latent space, thereby boosting both the efficiency and the efficacy of the generation process.

Protein-Conditional AutoEncoder To construct the latent space for the diffusion process, we initiate by training a protein-conditional autoencoder **(P-CAE)**. In this setup, the protein electron cloud ρ_p conditions the mapping of the ligand electron cloud ρ_l into a low-dimensional latent space:

$$\rho_{p_z} = \texttt{PktEncoder}(\rho_p), \quad \rho_{l_z} = \texttt{LigEncoder}(\rho_l), \quad \tilde{\rho}_l = \texttt{Decoder}([\rho_{l_z} \| \rho_{p_z}]). \tag{2}$$

Here, ρ_{p_z} denotes the embedding of ρ_p, ρ_{l_z} represents the embedding of ρ_l, and $\tilde{\rho}_l$ is the reconstructed version of the ligand electron cloud. The `PktEncoder`, `LigEncoder`, and `Decoder` are based on a similar 3D CNN framework, incorporating both sub-sampling and up-sampling techniques to facilitate the encoding and decoding processes.

Conditional Latent Diffusion In contrast to the vanilla diffusion model, which operates directly in the standard space, our approach involves applying noising and denoising processes within the P-CAE encoded latent space. During the noising process, Gaussian noise with variances $\{\beta_0, \ldots, \beta_T\}$ is added to $\rho_{l_z}^0$ within the latent space, where the posterior distribution q can be represented as:

$$q(\rho_{l_z}^{1:T} | \rho_{l_z}^0) = \prod_{t=1}^{T} q(\rho_{l_z}^t | \rho_{l_z}^{t-1}), \quad q(\rho_{l_z}^t | \rho_{l_z}^{t-1}) = \mathcal{N}(\rho_{l_z}^t; \sqrt{1-\beta_t}\rho_{l_z}^{t-1}, \beta_t I). \tag{3}$$

Further, we can derive a closed-form expression for the probability distribution at any time step t as:

$$q(\rho_{l_z}^t | \rho_{l_z}^0) = \mathcal{N}(\rho_{l_z}^t; \sqrt{\overline{\alpha}_t}\rho_{l_z}^0, (1-\overline{\alpha}_t)I) \tag{4}$$

where $\alpha_t = 1 - \beta_t$ and $\overline{\alpha}_t = \prod_{s=1}^t \alpha_s$. This deterministic noise addition process aims to evolve the original molecular electron cloud $P(\rho_{l_z}^t)$ into a Gaussian distribution $P(\rho_{l_z}^T)$. To generate a new electron cloud $\tilde{\rho}_{l_z}^0$ within the latent space, we first sample from the noise distribution $P(\rho_{l_z}^T)$. Then, we condition on ρ_{p_z} and reverse the noising process, transforming the sampled noise back into an actual electron cloud $\tilde{\rho}_{l_z}^0$. The denoising process can be expressed as:

$$p_\theta(\rho_{l_z}^{0:T-1} | \rho_{p_z}, \rho_{l_z}^T) = \prod_{t=1}^{T} p_\theta(\rho_{l_z}^{t-1} | \rho_{p_z}, \rho_{l_z}^t), \tag{5}$$

$$p_\theta(\rho_{l_z}^{t-1} | \rho_{p_z}, \rho_{l_z}^t) = \mathcal{N}(\rho_{l_z}^{t-1}; \mu_\theta(\rho_{p_z}, \rho_{l_z}^t, t), \sigma_t^2 I). \tag{6}$$

Here, p_θ denotes the distribution whose parameters are subject to estimation. The neural network μ_θ aims to estimate the mean of distribution p_θ, while σ_t^2 is a user-defined variance. As formulated in Equation 7, the transformation of $\rho_{l_z}^T$ is sequentially navigated through $p_\theta(\rho_{l_z}^{t-1} | \rho_{p_z}, \rho_{l_z}^t)$, resulting into the initial distribution sample $\tilde{\rho}_{l_z}^0$. Once obtaining $\tilde{\rho}_{l_z}^0$, the P-CAE's decoder is applied to reconstruct the electron cloud within the latent space,

$$\tilde{\rho}_l = \texttt{Decoder}([\tilde{\rho}_{l_z}^0 \| \rho_{p_z}]). \tag{7}$$

Training Objective The training objectives for optimizing the 3D Conditional Latent Diffusion module can be summarized as follows. First, we consider the auto-encoder loss associated with compressing molecular electron clouds:

$$\mathcal{L}_{\text{recon}} = \mathbb{E}_{(\rho_p, \rho_l) \sim \text{data}} \left[\|\rho_l - \tilde{\rho}_l\|^2 \right], \tag{8}$$

where ρ_l and $\tilde{\rho}_l$ represent the input and reconstructed ligand electron clouds, respectively. Second, we consider the pocket-conditional latent diffusion loss which can be expressed as:

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{\epsilon \sim \mathcal{N}(0,I), t \sim \text{Uniform}(1,T)} \left[\|\epsilon - \mu_\theta(\rho_{p_z}, \rho_{l_z}^t, t)\|^2 \right]. \tag{9}$$

Here, $\mathcal{L}_{\text{diff}}$ denotes the standard diffusion loss, instructing the model on the denoising of electron clouds during generation. Furthermore, we consider two regularization losses, $\mathcal{L}_{t=0}$ and $\mathcal{L}_{t=T}$, to modulate the diffusion process at starting $t = 0$ and termination $t = T$ steps, enhancing stability and sampling quality:

$$\mathcal{L}_{t=0} = \mathbb{E}_{(\rho_{p_z}, \rho_{l_z}) \sim \text{data}} \left[\|\rho_{l_z}^0 - \rho_{l_z}\|^2 \right], \tag{10}$$

$$\mathcal{L}_{t=T} = \mathbb{E}_{(\rho_{p_z}, \rho_{l_z}) \sim \text{data}} \left[\|\mu_\theta(\rho_{p_z}, \rho_{l_z}^T, T)\|^2 \right]. \tag{11}$$

The goal of $\mathcal{L}_{t=0}$ is to ensure the model's initial output closely aligns with the original latent representation, while $\mathcal{L}_{t=T}$ aims to control the noise introduction, preventing excessive magnitudes in the latent space encoder output. Additionally, we have a clash loss $\mathcal{L}_{\text{clash}}$ to augment protein-awareness and mitigate steric clashes:

$$\mathcal{L}_{\text{clash}} = \rho_p \odot \tilde{\rho}_l. \tag{12}$$

Here, ρ_p denotes the input protein electron cloud, $\tilde{\rho}_l$ is the reconstructed ligand electron cloud, and $\odot$ is the element-wise multiplication.

2.3 ECloud Decipher

The next challenge involves translating sampled electron clouds into their corresponding molecules. This task can be framed as a 3D image captioning problem. To tackle it, we develop the ECloud Decipher module by leveraging 3D-Unet [29] and RoFormer [23] to construct the ECloud-based chemical space. Furthermore, to facilitate effective conditional generation within this chemical space, we propose Contrastive ECloud-Molecule Pre-training (CEMP), which aims to organize the chemical space, making it amenable to gradient-based conditional generation. Utilizing CEMP, we ensure that the latent representations of electron clouds align with their corresponding molecular structures.

RoFormer To preserve the spatial relationships including 3D structural information of electron clouds and the topological relationships within chemical structures, we adopt RoFormer [23] as a variant of Transformers that have been widely used in large language models[30, 31]. The core of RoFormer is Rotational Position Encoding (RoPE), which introduces absolute position encoding at the input stage and inherently derives relative position encoding between inputs n and m during the attention computation. For clarity, we omit the multi-head dimension and present the basic structure of the RoFormer layer:

$$q = W_q x, \quad k = W_k x, \quad v = W_k x, \tag{13}$$

$$z = \texttt{RoAttn}(q, k, v), \quad z' = \texttt{FFN}(z). \tag{14}$$

Here, W_q, W_k, W_v denote the embeddings for query, key, and value, respectively. `RoAttn` denotes the rotational attention computation module, and `FFN` is the fully-connected network. The expansion of the `RoAttn` module can be written as follows:

$$q'_n = q_n e^{in\theta}, \quad k'_m = k_m e^{im\theta}; \tag{15}$$

$$A_{nm} = \texttt{Re}[q'_n (k'_m)^*], \quad z_n = \sum_m \texttt{softmax}(A_{\cdot m}) v_m. \tag{16}$$

Here, q'_n and k'_m are the complex features of the n-th query and the m-th key, respectively. θ denotes a predefined base angle, d_k is the embedding dimension, and `Re` represents the real-part operation. After the `softmax` on the attention map A, z_n is obtained as the output feature.

ECloud Caption To caption the electron cloud, we use the 3D-Unet backbone to encode the electron clouds into the chemical space, followed by decoding them into the corresponding molecules using RoFormer. Omitting the model details, we express the caption training objective as follows:

$$h_\rho = \texttt{ECloudNet}(\rho_l), \quad s_i = \texttt{RoFormer}_\text{d}(s_{<j} | h_\rho), \tag{17}$$

$$\mathcal{L}_{\text{Caption}} = -\sum_{i=1}^{t} \log p(s_i | s_{0:i-1}, \rho_{l_d}), \tag{18}$$

where h_ρ represents the latent chemical features of the electron cloud ρ_l, s_i is the i-th token of the decoded SMILES, $\texttt{RoFormer}_\text{d}$ denotes the RoFormer acting as the decoder, and $\mathcal{L}_{\text{Caption}}$ is the negative log-likelihood loss function.

CEMP for Structuring Chemical Space Controllable molecule design is one of the most important elements in practical drug discovery, including the optimization of the binding affinity between molecules and proteins to enhance drug efficacy, as well as the improvement of ADMET properties to enhance safety and bioavailability [32, 33]. On the other hand, the current challenges to structure-based models of molecule generation rely mostly on the size of the accessible chemical space being fairly limited. ECloudGen, in this regard, offers a very promising solution by accessing a far broader chemical space. To fully utilize this advantage, the chemical space must be structured so that structurally similar molecules are grouped together. A common approach is to employ a variational autoencoder (VAE) architecture [34], which naturally clusters data in latent space through its Gaussian prior constraint. However, the simplicity of the Gaussian constraint makes it inadequate to capture the complexity of the chemical space. One possible modification is to relax this constraint through the adoption of a mixture of Gaussian distributions [35]; nevertheless, this still does not solve the problem fundamentally. We then explore another approach—contrastive learning. The basic idea behind contrastive learning is to bring positive samples closer in the representation space while pushing negative samples away. Although it is primarily used in the pre-training to improve performance in downstream tasks [36, 37]—seen in Contrastive Language-Image Pre-training (CLIP) [22], which attains superior zero-shot classification performance for embedding images and text into a shared space—its principles are found to be ideally suited for the organization of the chemical space. Therefore, we propose Contrastive ECloud-Molecule Pre-training (CEMP) to structure the complex chemical space of ECloudGen, thus enabling efficient gradient-based conditional generation.

Based on the aforementioned ECloud Caption, we introduce an additional RoFormer encoder for SMILES to align the molecular modality with the electron cloud modality in the chemical space:

$$h_\rho = \texttt{ECloudNet}(\rho_l), \quad h_s = \texttt{RoFormer}_e(s), \tag{19}$$

Here, s is the molecular SMILES, h_ρ and h_s denote the respective embeddings for electron cloud and molecule in the chemical space. The electron cloud of a given molecule is considered a positive sample, whereas electron clouds from different molecules are considered negative samples. To perform contrastive learning on these samples, the InfoNCE loss [38] can be used:

$$\mathcal{L}_{\text{Contra}} = -\frac{1}{2b}\sum_i \left[\log \frac{\exp(h_{s,i}^\top h_{\rho,i})}{\sum_{j=0}^K \exp(h_{s,i}^\top h_{\rho,j})} + \log \frac{\exp(h_{\rho,i}^\top h_{s,i})}{\sum_{j=0}^K \exp(h_{\rho,i}^\top h_{s,j})}\right] \tag{20}$$

where $\mathcal{L}_{\text{Contra}}$ aims to maximize the similarity between positive pairs while minimizing the similarity between negative pairs. CEMP facilitates flexible clustering in the latent space, helping the model distill hidden physicochemical knowledge in a self-supervised way.

Conditional Generation Conditional generation is more sophisticated in molecule generation that not only broadens theoretical understanding but is also more closely aligned with real-world applications. A common approach in deep learning models for controlled purposes is to treat property scores as input tokens for conditioning, like MolGPT [39] and PrefixMol [40]. However, the constraints in drug design are not about achieving a fixed value (*e.g.*, QED=0.8), but rather about optimizing for higher values (*e.g.*, *maximize* QED). Therefore, framing molecular conditional generation as an optimization problem aligns better with real-world scenarios:

$$\begin{aligned} &\textit{maximize} && f(m), \\ &\textit{subject to} && h_i(m) < 0, \quad i = 1, \ldots, k, \end{aligned} \tag{21}$$

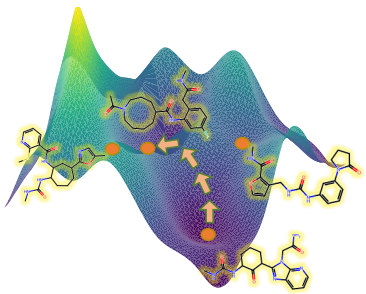

Figure 3: Illustration of metadynamics.

where m is the molecule, $f(m)$ denotes the given molecular property that chemists want to optimize, and $\{h_i\}_{i=0}^k$ is the set of constraints.

Due to the sparsity of protein-ligand data, previous structure-based molecule generation models have struggled to model a sufficiently continuous chemical space for conditional generation. In contrast, ECloudGen, through prior preparation, not only possesses a rich chemical space but also structures this space in an orderly manner using CEMP. As a result, ECloudGen is well-equipped to perform conditional molecular generation. Specifically, gradient-based optimization is utilized to optimize molecular structures in the chemical space. To obtain diverse optimized molecules, as illustrated in

Figure 3, the meta-dynamics methodology [36, 41, 42] is leveraged to bump the model out of local minima when structural changes are minimal, where the adding potential is defined as (for simplicity, we describe this in 2D data):

$$\phi(H, c_i, r) = H \cdot \sum_{c_i \in \{c\}} \exp\left(-\frac{1}{2r}(h_\rho - c_i)^2 + \frac{1}{2}(\log(2\pi) + \log(r))\right). \tag{22}$$

Here, $\phi(H, c_i, r)$ denotes the Gaussian potential function to lift the optimization target out of local minima. c_i is the position of the Gaussian potential, r is its variance, and H is its height coefficient. The algorithm details can be found in Appendix C.

3 Related Work

Thanks to the rapid development of graph machine learning and graph neural networks [43, 44, 45, 46, 47, 48, 49], numerous structure-based-drug-design models have been proposed, ranging from auto-regressive models [50, 4, 5, 51, 52], to diffusion-based approaches [8, 10, 9, 53, 54]. Diverging from the graph-based generative models, ECloudGen is inspired by voxel-conditioned molecule generation to utilize voxel grids to store electron clouds, which further represent molecules. Voxel computations are based on Gaussian distributions centered on atoms, describing molecule shapes, *e.g.,* ShapeVAE [55], SQUID [56], VoxMol [57], ShapeMol [58] in ligand-based design, and LiGAN [59] in structure-based design. It is important to note that these voxel methods typically require separate channels for different atomic types, whereas ECloudGen uses only one channel, which is electron density. Additionally, unlike methods that compute atomic positions in isolation, electron cloud computation considers inter-atomic interactions, offering a more comprehensive physical description. A detailed review of related works can be found in Appendix A.

4 Experiment and Analysis

4.1 Binding Behaviour and Drug-likeness

The potential of a molecule to become a viable drug is determined by its binding affinity to the target and its molecular properties, commonly referred to as drug-likeness [60, 61]. Therefore, the quality of generated molecules is primarily evaluated based on these two criteria. Previous structure-based molecule generation models have been effective in designing molecules with strong binding affinities. However, they still exhibit considerable limitations in optimizing drug-likeness properties. While some methods have improved molecular properties through fragment-wise generation approaches [52, 62], these strategies have not fully addressed the Chemical Space Generation Paradox. In this study, we demonstrate how accessing a larger chemical space, facilitated by ECloudGen, can significantly enhance the quality of generated molecules.

We empirically compare ECloudGen with tenn baselines including Pocket2Mol [4], ResGen [5], GraphBP [51], FLAG [52], FragGen [62], DiffBP [9], DiffSBDD [8], TargetDiff [10], Lingo3dMol [6], PrefixMol [40]. The experiments are conducted using the CrossDock dataset with a fixed data split method [20]. To assess the quality of generated samples, including their binding affinity and drug-likeness, we use Vina score [63] to measure the affinities between the generated molecules and their corresponding protein targets. We use QED [64], SA [65], Lipinski [66], and LogP [67] to assess the drug-likeness of the generated molecules. It is important to note that the Vina score is derived from the summation of individual atom/fragment energy contributions, which inherently tends to favor larger molecules. To mitigate this bias in our evaluation, we adopt Ligand Binding Efficiency (LBE), a metric commonly used in computer-aided drug discovery to normalize the Vina score based on the number of atoms, *i.e.,* LBE $=$ Vina$/N_{\text{atoms}}$, where N_{atoms} is the number of atoms. Previous studies have shown that LBE is a more realistic metric than the unnormalized Vina score [68, 69], indicating that LBE is a better alternative for evaluating binding efficiency. Further details on the baseline models and the related metrics can be found in Appendix B.

As drug design research typically focuses on a few molecules with the highest potential, we compute the average results for the top five molecules of each target in the test set. We summarize the empirical results in Table 1. We can observe that ECloudGen the best results on the LBE metric, indicating its ability to design molecules that effectively bind to protein targets. Notably, compared to the Pkt2Mol model, which also generates tightly bound molecules, ECloudGen can generate molecules with significant improvements of 22.8%, 132.3%, and 1.4% on the QED, SA, and Lipinski metrics, respectively, demonstrating the model's ability to generate high-fidelity drug-like molecules. The

Table 1: Evaluation of ligand binding efficiency (**LBE**) and other molecular properties for baselines and ECloudGen. ↓ means the smaller the value, the better the performance, and ↑ means the opposite. Values in **bold** refer to the best metrics. AR-A/F denotes the autoregressive atom-wise/fragment-wise strategy, Chem-L denotes the chemical language strategy.

Class	Method	LBE (↓)	QED (↑)	SA (↑)	Lipinski(↑)	LogP (-2 - 5)
Ref.	Test	-0.34	0.56	0.73	4.89	0.10
AR-A	Pkt2Mol [4]	**-0.42**	0.57	0.34	4.93	0.76
AR-A	ResGen [5]	-0.41	0.55	0.32	4.95	1.64
AR-A	GraphBP [51]	-0.33	0.56	0.48	4.77	1.29
AR-F	FLAG [52]	-0.31	0.52	0.58	4.87	0.45
AR-F	FragGen [62]	-0.36	0.57	0.72	4.86	1.27
Diff	DiffBP [9]	-0.33	0.49	0.43	4.78	0.11
Diff	DiffSBDD [8]	-0.35	0.55	0.50	4.85	0.11
Diff	TargetDiff [10]	-0.33	0.57	0.48	4.78	0.62
Chem-L	Lingo3dMol [6]	-0.28	0.47	0.65	4.50	0.74
Chem-L	PrefixMol [40]	-0.36	0.51	0.69	4.83	0.08
Our	**ECloudGen**	**-0.42**	**0.70**	**0.79**	**5.00**	0.09

significant improvements in drug-likeness by ECloudGen primarily stems from its access to a broader chemical space than the baselines can access. Empowered by this extensive chemical knowledge, ECloudGen generates molecules that not only bind tightly to target proteins but also possess rational chemical structures, advancing structure-based molecule generation models closer to real-world applications. Moreover, we provide visualization of generated molecules by different models in Appendix F.

4.2 Generated Chemical Space Measurement

In addition to its outstanding binding performance, a key claim of ECloudGen is its ability to address the 'Chemical Space Generation Paradox'. This suggests that ECloudGen encompasses a broader chemical space compared to other baseline models. This subsection aims to measure and compare the chemical space spanned by the generated molecules of different models. To measure the scale/size of the chemical space of the generated molecules, one can use internal diversity (IntDiv), previously defined in Pocket2Mol, as follows:

$$\mathrm{IntDiv}_p(\mathcal{G}) = 1 - \sqrt[p]{\frac{1}{|\mathcal{G}|^2} \sum_{G_1, G_2 \in \mathcal{G}} T\left(G_1, G_2\right)^p}. \tag{23}$$

Here, $\mathcal{G}$ denotes the set of generated molecules, T represents Tanimoto similarity [70], and p is a positive value that adjusts the contribution of similarity between different molecular pairs. While IntDiv has been widely used across various tasks, it may not be the most optimal metric for measuring the scale/size of the chemical space. From the perspective of point-set topology, a robust measure of chemical space should satisfy subadditivity and dissimilarity preference axioms [71], whereas IntDiv only fulfills the latter, making it a suboptimal choice. To address this and provide a better and more complete measurement, we employ the Maximum Exclusion Circle (Circles) [72] and Bottleneck (BN) metrics [73]. Circles adhere to both axioms, while Bottleneck metric respects dissimilarity preference, offering a more comprehensive assessment of the chemical space. These two metrics can be formulated as follows:

$$\mathrm{Circles}(\mathcal{G}) := \max_{\mathcal{C} \subseteq \mathcal{G}} |\mathcal{C}| \quad \text{s.t. } T(G_1, G_2) > t, \quad \forall G_1, G_2 \in \mathcal{C}, \tag{24}$$

$$\text{Bottleneck}\ (\mathcal{G}) := \sum_{G_1 \in \mathcal{G}} \min_{\substack{G_2 \in \mathcal{G} \\ G_2 \neq g_1}} T(G_1, G_2). \tag{25}$$

Intuitively, the 'Circles' metric represents the number of mutually exclusive circles required to cover the generated molecules within the chemical space. On the other hand, the 'Bottleneck' metric measures the sum of distances between each molecule and its closest neighbor.

Table 2: Measurement of chemical spaces generated by different methods on the CrossDock test set.

Measure	Pkt2mol	ResGen	GraphBP	FLAG	FragGen	DiffBP	DiffSBDD	TargetDiff	Lingo3dMol	PrefixMol	ECloudGen
Circles (↑)	28	20	29	10	63	12	19	61	9	4	**85**
BN (↑)	20.5	14.5	22.3	8.4	45.0	9.9	15.7	48.1	5.7	2.8	**53.4**

In Table 2, we summarize the results of the measurement of the generated chemical space using 'Circles' and 'BN'. We can observe that ECloudGen achieves the highest values in both Circle

and Bottleneck metrics. Specifically, the chemical compounds generated by ECloudGen not only encompass more mutually exclusive circles, indicating a broader coverage of chemical space, but also exhibit greater minimum distances between molecules. This suggests that the compounds generated by ECloudGen are structurally more unique and diverse. These strong improvments confirm that ECloudGen indeed encompasses a broader chemical space, and further addresses the Chemical Space Generation Paradox.

4.3 Conditional Generation

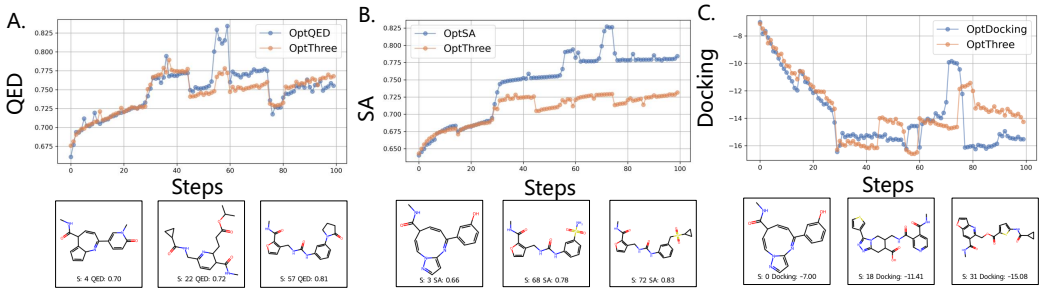

Figure 4: Objective function values for single- and multi-condition generation, with the horizontal axis representing optimization steps. The example image is from single-condition generation.

Another noteworthy feature of ECloudGen is its capability for conditional generation, an area that has been relatively underexplored in structure-based molecule generation methods. It is achieved by employing gradient-based evolution directly on the molecules within the chemical space that are to be optimized. Specifically, given a target, we first construct a prediction model for the desired pharmacological property within the chemical space. Starting with molecules generated for that target, we guide their evolution within the chemical space using the gradients of the prediction model to optimize the molecules. Additionally, the metadynamics-inspired approach is utilized to enable molecules to escape from basins when they become trapped in local minima.

Single-condition optimization We select the protein with PBD ID of 4bel as our case study, optimizing for two molecular properties, QED and SA, in addition to binding affinity estimation via docking. We first conduct single-condition optimization (see illustration of blue lines in Figure 4(A)(B)(C)). The sudden jumps observed during convergence are due to the additional potential (metadynamics) applied by ECloudGen when the molecules become trapped in valleys of the chemical space. Although the optimization objective may temporarily worsen after these jumps, the molecules rapidly evolve to better values in new regions.

Multi-condition optimization In addition to the single-condition optimization, we also perform multi-condition optimization across three conditions once (see illustration of orange lines in Figure 4(A)(B)(C)). We can see that the optimization outcomes for Docking and QED in multi-condition optimization closely align with those observed in single-condition optimization. However, the performance in optimizing for SA is slightly less effective, which may be attributed to the inherently lower SA values of molecules in the newly accessed regions. Illustrative examples of molecules from single-condition optimization are presented in Figure 4, while the trajectories for multi-condition optimization can be found in Appendix E.

5 Conclusion and Limitation

In this work, we first identify the Chemical Space Generation Paradox within the field of structure-based molecule generation. To address this paradox, we propose ECloudGen, which leverages electron clouds as the generative agent, enabling the training process to utilize unlabeled data from chemical libraries. Combining the model with the proposed pretaining method-CEMP, ECloudGen structurally orders the vast learned chemical space, facilitating the conditional generation. Future work could explore several elements presented in this paper, such as investigating electron clouds as a fundamental physical representation for unified molecular modeling, and incorporating graph completion techniques into ECloud Latent Diffusion to achieve structural modification tasks. However, limitations still exist. Current ECloudGen has been trained on 2 billion molecules randomly sampled from the CHEMBL library [74] due to the limited time and resource budget. Additionally, the data preparation relies on the accuracy of GFN2-xTB method. While it is highly efficient, the software is still not very friendly to certain complex systems [26] in accuracy. These limitations remain our future focus to address.

References

[1] Rita M Meganck and Ralph S Baric. Developing therapeutic approaches for twenty-first-century emerging infectious viral diseases. *Nature medicine*, 27(3):401–410, 2021.

[2] Youjun Xu, Kangjie Lin, Shiwei Wang, Lei Wang, Chenjing Cai, Chen Song, Luhua Lai, and Jianfeng Pei. Deep learning for molecular generation. *Future medicinal chemistry*, 11(6):567–597, 2019.

[3] Kenneth M Merz Jr, Dagmar Ringe, and Charles H Reynolds. *Drug design: structure-and ligand-based approaches*. Cambridge University Press, 2010.

[4] Xingang Peng, Shitong Luo, Jiaqi Guan, Qi Xie, Jian Peng, and Jianzhu Ma. Pocket2mol: Efficient molecular sampling based on 3d protein pockets. In *International Conference on Machine Learning*, pages 17644–17655. PMLR, 2022.

[5] Odin Zhang, Jintu Zhang, Jieyu Jin, Xujun Zhang, RenLing Hu, Chao Shen, Hanqun Cao, Hongyan Du, Yu Kang, Yafeng Deng, et al. Resgen is a pocket-aware 3d molecular generation model based on parallel multiscale modelling. *Nature Machine Intelligence*, 5(9):1020–1030, 2023.

[6] Wei Feng, Lvwei Wang, Zaiyun Lin, Yanhao Zhu, Han Wang, Jianqiang Dong, Rong Bai, Huting Wang, Jielong Zhou, Wei Peng, et al. Generation of 3d molecules in pockets via a language model. *Nature Machine Intelligence*, pages 1–12, 2024.

[7] Odin Zhang, Tianyue Wang, Gaoqi Weng, Dejun Jiang, Ning Wang, Xiaorui Wang, Huifeng Zhao, Jialu Wu, Ercheng Wang, Guangyong Chen, et al. Learning on topological surface and geometric structure for 3d molecular generation. *Nature Computational Science*, 3(10):849–859, 2023.

[8] Arne Schneuing, Yuanqi Du, Charles Harris, Arian Jamasb, Ilia Igashov, Weitao Du, Tom Blundell, Pietro Lió, Carla Gomes, Max Welling, et al. Structure-based drug design with equivariant diffusion models. *arXiv preprint arXiv:2210.13695*, 2022.

[9] Haitao Lin, Yufei Huang, Meng Liu, Xuanjing Li, Shuiwang Ji, and Stan Z. Li. Diffbp: Generative diffusion of 3d molecules for target protein binding, 2022.

[10] Jiaqi Guan, Wesley Wei Qian, Xingang Peng, Yufeng Su, Jian Peng, and Jianzhu Ma. 3d equivariant diffusion for target-aware molecule generation and affinity prediction. *arXiv preprint arXiv:2303.03543*, 2023.

[11] Chenqing Hua, Sitao Luan, Minkai Xu, Zhitao Ying, Jie Fu, Stefano Ermon, and Doina Precup. Mudiff: Unified diffusion for complete molecule generation. In *Learning on Graphs Conference*, pages 33–1. PMLR, 2024.

[12] Renxiao Wang, Xueliang Fang, Yipin Lu, Chao-Yie Yang, and Shaomeng Wang. The pdbbind database: methodologies and updates. *Journal of medicinal chemistry*, 48(12):4111–4119, 2005.

[13] John J Irwin and Brian K Shoichet. Zinc- a free database of commercially available compounds for virtual screening. *Journal of chemical information and modeling*, 45(1):177–182, 2005.

[14] Pavel G Polishchuk, Timur I Madzhidov, and Alexandre Varnek. Estimation of the size of drug-like chemical space based on gdb-17 data. *Journal of computer-aided molecular design*, 27:675–679, 2013.

[15] Scott R Hinze, David N Rapp, Vickie M Williamson, Mary Jane Shultz, Ghislain Deslongchamps, and Kenneth C Williamson. Beyond ball-and-stick: Students' processing of novel stem visualizations. *Learning and instruction*, 26:12–21, 2013.

[16] Charles T Sebens. Electron charge density: a clue from quantum chemistry for quantum foundations. *Foundations of Physics*, 51(4):75, 2021.

[17] Mårten Skogh, Werner Dobrautz, Phalgun Lolur, Christopher Warren, Janka Biznárová, Amr Osman, Giovanna Tancredi, Jonas Bylander, and Martin Rahm. The electron density: a fidelity witness for quantum computation. *Chemical Science*, 15(6):2257–2265, 2024.

[18] Hiroshi Nakatsuji. Electron-cloud following and preceding and the shapes of molecules. *Journal of the American Chemical Society*, 96(1):30–37, 1974.

[19] Deborah Leckband and Jacob Israelachvili. Intermolecular forces in biology. *Quarterly reviews of biophysics*, 34(2):105–267, 2001.

[20] Paul G Francoeur, Tomohide Masuda, Jocelyn Sunseri, Andrew Jia, Richard B Iovanisci, Ian Snyder, and David R Koes. Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. *Journal of chemical information and modeling*, 60(9):4200–4215, 2020.

[21] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[22] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

[23] Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568:127063, 2024.

[24] Erwin Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Physical review*, 28(6):1049, 1926.

[25] Christoph Bannwarth, Sebastian Ehlert, and Stefan Grimme. Gfn2-xtb—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *Journal of chemical theory and computation*, 15(3):1652–1671, 2019.

[26] Christoph Bannwarth, Eike Caldeweyher, Sebastian Ehlert, Andreas Hansen, Philipp Pracht, Jakob Seibert, Sebastian Spicher, and Stefan Grimme. Extended tight-binding quantum chemistry methods. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 11(2):e1493, 2021.

[27] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[28] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

[29] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 424–432. Springer, 2016.

[30] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models (2023). *arXiv preprint arXiv:2302.13971*, 2023.

[31] Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. Palm: Scaling language modeling with pathways. *Journal of Machine Learning Research*, 24(240):1–113, 2023.

[32] William L Jorgensen. Efficient drug lead discovery and optimization. *Accounts of chemical research*, 42(6):724–733, 2009.

[33] Martin Morgenthaler, Eliane Schweizer, Anja Hoffmann-Röder, Fausta Benini, Rainer E Martin, Georg Jaeschke, Björn Wagner, Holger Fischer, Stefanie Bendels, Daniel Zimmerli, et al. Predicting and tuning physicochemical properties in lead optimization: amine basicities. *ChemMedChem: Chemistry Enabling Drug Discovery*, 2(8):1100–1115, 2007.

[34] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[35] Nat Dilokthanakul, Pedro AM Mediano, Marta Garnelo, Matthew CH Lee, Hugh Salimbeni, Kai Arulkumaran, and Murray Shanahan. Deep unsupervised clustering with gaussian mixture variational autoencoders. *arXiv preprint arXiv:1611.02648*, 2016.

[36] Benjamin Kaufman, Edward C Williams, Carl Underkoffler, Ryan Pederson, Narbe Mardirossian, Ian Watson, and John Parkhill. Coati: Multimodal contrastive pretraining for representing and traversing chemical space. *Journal of Chemical Information and Modeling*, 64(4):1145–1157, 2024.

[37] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. *Advances in neural information processing systems*, 33:5812–5823, 2020.

[38] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

[39] Viraj Bagal, Rishal Aggarwal, PK Vinod, and U Deva Priyakumar. Molgpt: molecular generation using a transformer-decoder model. *Journal of Chemical Information and Modeling*, 62(9):2064–2076, 2021.

[40] Zhangyang Gao, Yuqi Hu, Cheng Tan, and Stan Z Li. Prefixmol: Target-and chemistry-aware molecule design via prefix embedding. *arXiv preprint arXiv:2302.07120*, 2023.

[41] Alessandro Barducci, Massimiliano Bonomi, and Michele Parrinello. Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(5):826–843, 2011.

[42] Stefan Grimme. Exploration of chemical compound, conformer, and reaction space with meta-dynamics simulations based on tight-binding quantum chemical calculations. *Journal of chemical theory and computation*, 15(5):2847–2862, 2019.

[43] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

[44] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1263–1272. JMLR. org, 2017.

[45] William L Hamilton. *Graph representation learning*. Morgan & Claypool Publishers, 2020.

[46] Sitao Luan, Chenqing Hua, Qincheng Lu, Jiaqi Zhu, Mingde Zhao, Shuyuan Zhang, Xiao-Wen Chang, and Doina Precup. Is heterophily a real nightmare for graph neural networks to do node classification? *arXiv preprint arXiv:2109.05641*, 2021.

[47] Chenqing Hua, Guillaume Rabusseau, and Jian Tang. High-order pooling for graph neural networks with tensor decomposition. *Advances in Neural Information Processing Systems*, 35:6021–6033, 2022.

[48] Sitao Luan, Chenqing Hua, Qincheng Lu, Jiaqi Zhu, Mingde Zhao, Shuyuan Zhang, Xiao-Wen Chang, and Doina Precup. Revisiting heterophily for graph neural networks. *Advances in neural information processing systems*, 35:1362–1375, 2022.

[49] Chenqing Hua, Sitao Luan, Qian Zhang, and Jie Fu. Graph neural networks intersect probabilistic graphical models: A survey. *arXiv preprint arXiv:2206.06089*, 2022.

[50] Shitong Luo, Jiaqi Guan, Jianzhu Ma, and Jian Peng. A 3D generative model for structure-based drug design. In *Thirty-Fifth Conference on Neural Information Processing Systems*, 2021.

[51] Meng Liu, Youzhi Luo, Kanji Uchino, Koji Maruhashi, and Shuiwang Ji. Generating 3d molecules for target protein binding. In *International Conference on Machine Learning*, 2022.

[52] ZAIXI ZHANG, Shuxin Zheng, Yaosen Min, and Qi Liu. Molecule generation for target protein binding with structural motifs. In *International Conference on Learning Representations*, 2023.

[53] Haitao Lin, Yufei Huang, Odin Zhang, Yunfan Liu, Lirong Wu, Siyuan Li, Zhiyuan Chen, and Stan Z. Li. Functional-group-based diffusion for pocket-specific molecule generation and elaboration. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 34603–34626. Curran Associates, Inc., 2023.

[54] Jiaqi Guan, Xiangxin Zhou, Yuwei Yang, Yu Bao, Jian Peng, Jianzhu Ma, Qiang Liu, Liang Wang, and Quanquan Gu. DecompDiff: Diffusion models with decomposed priors for structure-based drug design. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 11827–11846. PMLR, 23–29 Jul 2023.

[55] Miha Skalic, José Jiménez, Davide Sabbadin, and Gianni De Fabritiis. Shape-based generative modeling for de novo drug design. *Journal of chemical information and modeling*, 59(3):1205–1214, 2019.

[56] Keir Adams and Connor W Coley. Equivariant shape-conditioned generation of 3d molecules for ligand-based drug design. *arXiv preprint arXiv:2210.04893*, 2022.

[57] Pedro O O Pinheiro, Joshua Rackers, Joseph Kleinhenz, Michael Maser, Omar Mahmood, Andrew Watkins, Stephen Ra, Vishnu Sresht, and Saeed Saremi. 3d molecule generation by denoising voxel grids. *Advances in Neural Information Processing Systems*, 36, 2024.

[58] Ziqi Chen, Bo Peng, Srinivasan Parthasarathy, and Xia Ning. Shape-conditioned 3d molecule generation via equivariant diffusion models. *arXiv preprint arXiv:2308.11890*, 2023.

[59] Yibo Li, Jianfeng Pei, and Luhua Lai. Structure-based de novo drug design using 3d deep generative models. *Chemical science*, 12(41):13664–13675, 2021.

[60] Oleg Ursu, Anwar Rayan, Amiram Goldblum, and Tudor I Oprea. Understanding drug-likeness. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(5):760–781, 2011.

[61] Sheng Tian, Junmei Wang, Youyong Li, Dan Li, Lei Xu, and Tingjun Hou. The application of in silico drug-likeness predictions in pharmaceutical research. *Advanced drug delivery reviews*, 86:2–10, 2015.

[62] Odin Zhang, Yufei Huang, Shichen Cheng, Mengyao Yu, Xujun Zhang, Haitao Lin, Yundian Zeng, Mingyang Wang, Zhenxing Wu, Huifeng Zhao, et al. Deep geometry handling and fragment-wise molecular 3d graph generation. *arXiv preprint arXiv:2404.00014*, 2024.

[63] Amr Alhossary, Stephanus Daniel Handoko, Yuguang Mu, and Chee-Keong Kwoh. Fast, accurate, and reliable molecular docking with QuickVina 2. *Bioinformatics*, 31(13):2214–2216, 02 2015.

[64] Richard Bickerton, Gaia Paolini, Jérémy Besnard, Sorel Muresan, and Andrew Hopkins. Quantifying the chemical beauty of drugs. *Nature chemistry*, 4:90–8, 02 2012.

[65] P Ertl and Ansgar Schuffenhauer. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of cheminformatics*, 1:8, 06 2009.

[66] Christopher A Lipinski, Franco Lombardo, Beryl W Dominy, and Paul J Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, 23(1-3):3–25, 1997.

[67] Arup K. Ghose, Vellarkad N. Viswanadhan, and John J. Wendoloski. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. a qualitative and quantitative characterization of known drug databases. *Journal of Combinatorial Chemistry*, 1(1):55–68, December 1998.

[68] Cele Abad-Zapatero and James T Metz. Ligand efficiency indices as guideposts for drug discovery. *Drug discovery today*, 10(7):464–469, 2005.

[69] Andrew L Hopkins, György M Keserü, Paul D Leeson, David C Rees, and Charles H Reynolds. The role of ligand efficiency metrics in drug discovery. *Nature reviews Drug discovery*, 13(2):105–121, 2014.

[70] Taffee T Tanimoto. Elementary mathematical theory of classification and prediction. 1958.

[71] Yutong Xie, Ziqiao Xu, Jiaqi Ma, and Qiaozhu Mei. How much space has been explored? measuring the chemical space covered by databases and machine-generated molecules. *arXiv preprint arXiv:2112.12542*, 2021.

[72] Yutong Xie, Ziqiao Xu, Jiaqi Ma, and Qiaozhu Mei. How much of the chemical space has been explored? selecting the right exploration measure for drug discovery. In *ICML 2022 2nd AI for Science Workshop*, 2022.

[73] Herbert Edelsbrunner and John L Harer. *Computational topology: an introduction*. American Mathematical Society, 2022.

[74] Anna Gaulton, Louisa J Bellis, A Patricia Bento, Jon Chambers, Mark Davies, Anne Hersey, Yvonne Light, Shaun McGlinchey, David Michalovich, Bissan Al-Lazikani, et al. Chembl: a large-scale bioactivity database for drug discovery. *Nucleic acids research*, 40(D1):D1100–D1107, 2012.

[75] Rafael Gómez-Bombarelli, Jennifer N Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D Hirzel, Ryan P Adams, and Alán Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *ACS central science*, 4(2):268–276, 2018.

[76] Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Junction tree variational autoencoder for molecular graph generation. In *International conference on machine learning*, pages 2323–2332. PMLR, 2018.

[77] Ryan-Rhys Griffiths and José Miguel Hernández-Lobato. Constrained bayesian optimization for automatic chemical design using variational autoencoders. *Chemical science*, 11(2):577–586, 2020.

[78] Qi Liu, Miltiadis Allamanis, Marc Brockschmidt, and Alexander Gaunt. Constrained graph variational autoencoders for molecule design. *Advances in neural information processing systems*, 31, 2018.

A Other Related Work

Structure-based drug design Current advancement in structure prediction methods boosted interest in structure-based drug design. The front-runners like Pkt2Mol[4], GraphBP[51], and ResGen[5], generate molecules atom-by-atom in an autoregressive manner, a strategy which we refer to as AR-A. However, some researchers have noted that these atom-wise methods often produce molecules with multi-fused rings, which, while fostering strong interactions with proteins, are unrealistic to be drug candidates. In response, models such as FLAG[52] and FragGen[62] have opted to generate molecules fragment-by-fragment, a strategy we refer to as AR-F, alleviating issues associated with unusual local structures. Another storyline is diffusion-based methods, which challenge the realism of the autoregressive approach by suggesting that molecules do not have a naturally defined growth path. Models like DiffSBDD[8], DiffBP[9], and TargetDiff[10] utilize diffusion models to grow molecules in one-shot, moving away from the traditional step-by-step assembly. Currently, there's growing enthusiasm for applying language model techniques to structure-based molecule design. Given the natural alignment of autoregressive methods with language generation, Lingo3DMol[6] integrates protein data as an embedding vector within a Transformer architecture for autoregressive molecular generation. Another model, PrefixMol[40], uses protein embeddings and properties such as SA and QED as prefixes, serving as conditions in the molecule generation model. Despite the field's rapid development and the introduction of novel approaches, most methods are constrained by the limited chemical space spanned by the training data, primarily from sources like CrossDock (the docking-enhanced version of PDBBind).

Chemical Space Exploration Expanding the accessible chemical space for models not only enhances the quality of generated molecules but also facilitates exploration and optimization within this continuous space. Rafael[75] et al. pioneered the application of VAE to condense the latent chemical space, subsequently employing Bayesian Optimization to optimize molecular structures towards better drug-likeness. Similar methodologies are employed in JTVAE[76] and CBO-VAE[77], where Bayesian Optimization is executed within the latent spaces constructured by different VAEs. Beyond leveraging Bayesian Optimization for latent space exploration, another approach is using gradient decent way to evolve molecules, potentially contributing to shape the latent chemical space[78]. The challenge within this direction is the propensity of molecules to become trapped in local minima. In addition to selecting more robust optimizers, utilizing the concept of metadynamics[41, 36] to fill the local minima where molecules are trapped is also an effective approach.

B Metrics Description

Quantitative Estimate of Drug-likeness (QED) Developed as described in Ref [64], QED is a metric that provides a quantitative estimate of drug-likeness based on a molecule's desirability across a range of properties. It is a composite score derived from molecular properties that are typically considered favorable for drug candidates.

Synthetic Accessibility (SA) The Synthetic Accessibility Score, as outlined in Ref [65], measures the ease with which a molecule can be synthesized. The calculation of the SA score is based on several factors, including molecular complexity (like containing unusual rings) and fragment contributions.

Lipinski's Rule of Five (Linpinski): Introduced in Ref [66], Lipinski's Rule of Five predicts the drug-likeness of compounds based on four key physical properties: no more than 5 hydrogen bond donors, no more than 10 hydrogen bond acceptors, a molecular weight under 500 Dalton, and an octanol-water partition coefficient (LogP) not greater than 5. These rules are used to evaluate the likelihood of success in the later stages of drug development.

LogP The partition coefficient between octanol and water (LogP) [67] is a measure of a molecule's hydrophobicity. It is crucial in determining a compound's absorption and distribution characteristics, influencing its pharmacokinetic profile. Generally, it is considered that a LogP value between -2 and 5 represents a reasonable range for potential drug candidates.

Ligand Binding Efficiency (LBE) To address the inherent bias of the Vina score favoring larger molecules due to its calculation method — summing up the individual atom or fragment energy contributions — we adopt the Ligand Binding Efficiency (LBE). LBE normalizes the Vina score by the number of atoms (N_{atoms}) in the molecule, calculated as $\text{LBE} = \frac{\text{Vina}}{N_{\text{atoms}}}$.

C Optimization Algorithm

Algorithm 1 Metadynamics for Molecular Optimization

Input: Initial molecule and its embedding vector m_0, v_0, objective function $f(v)$, constraint functions $\{h_i(v)\}_{i=1}^{n}$ and the corresponding Lagrange multipliers $\{\lambda_i\}_{i=1}^{n}$, optimization steps n, and the decoder of ECloud Decipher `Decoder`, the Gaussian potential $\phi(H, c_i, r)$ with the given H and r
for $i = 0$ **to** n_{steps} **do**
  Calculate objective function $f(v_i)$
  Aggregate constraints and compute total constraint penalty $\sum_i \lambda_i h_i(v)$
  if $\frac{v_i \cdot v_{i-1}}{\|v_i\| \|v_{i-1}\|} < \delta$ **and** $i > 0$ **then**
    Calculate adding potential $\phi(H, v_i, r)$
  else
    $\phi(H, v_i, r)$=0
  end if
  Calculate the Lagrangian function $\mathcal{L} = f(v_i) + \sum_i \lambda_i h_i(v) + \phi(H, v_i, r)$
  Calculate the gradient $\nabla \mathcal{L}$
  Update $v_{i+1} = v_i + \nabla \mathcal{L}$
  Decode the vector $m_i = \texttt{Decoder}(v_i)$
end for
Output: $\{m_i\}_{i=0}^{n_{steps}}$

D Electron Cloud-based Space

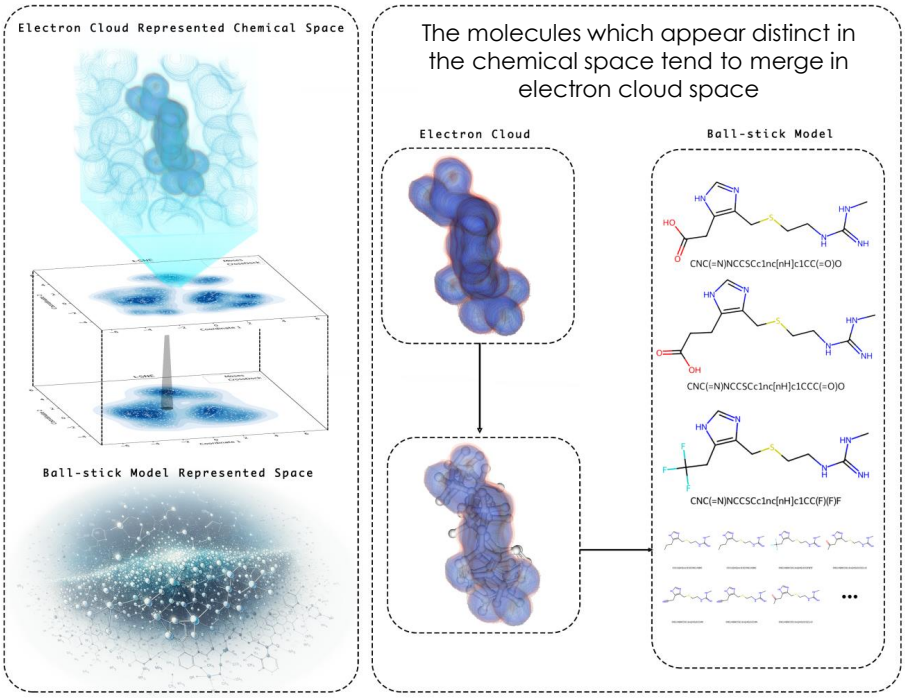

Figure 5: Illustration of the Electron-cloud Represented Space versus the Ball-stick Model Represented Space

Because the electron-cloud representation offers a more compact learning space, learning from one electron cloud can correspond to multiple molecules. This reduces the model's overall data demand and shifts the focus of data requirements to the second phase—ECloud Decipher—from electron cloud to molecular decoding. In this phase, a large number of molecules without binding structures can be utilized.

E Trajectories for Multi-condition optimization

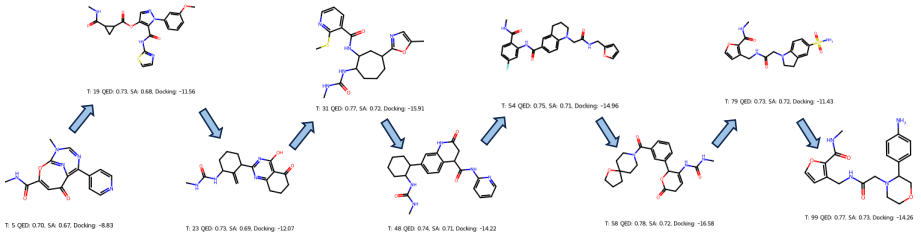

Figure 6: Trajectories for Multi-condition optimization

F Methods Visualization

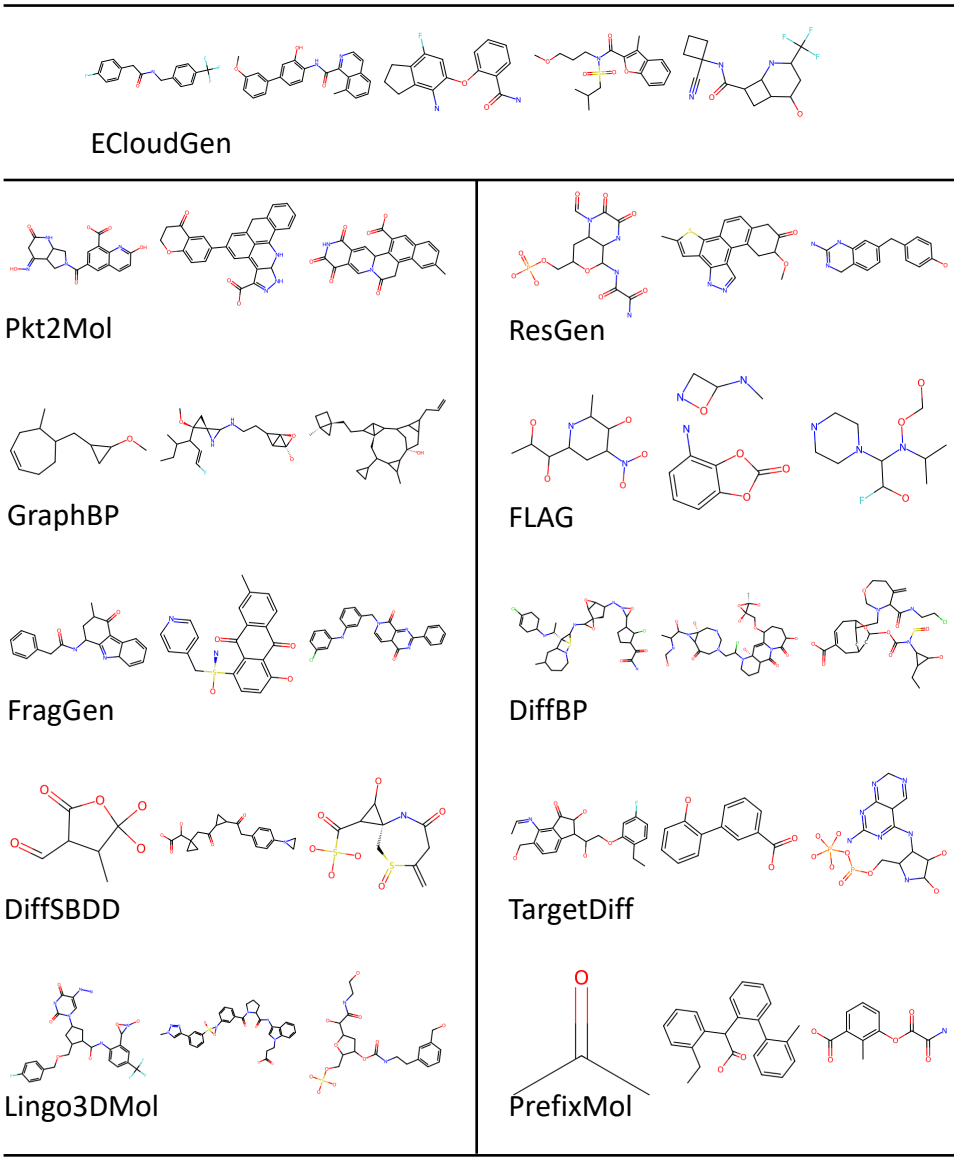

Figure 7: Visualization of molecules generated by different methods.

G Experimental Details

Platform We use three NVIDIA A100 (81920MiB) GPUs for a trial. The codes are implemented in Python 3.9 mainly with Pytorch 1.12, and run on Ubuntu Linux.

Dataset The ECloud Latent Diffusion module is trained on the CrossDock dataset, following the well-acknowledged train-test split used in other models. The ECloud Decipher model is trained on 2 billion molecules randomly sampled from the CHEMBEL library.

H Method Details

The electron clouds of ligands and proteins are stored within 32-point grid boxes. The `PktEncoder`, `LigEncoder`, and the `Decoder` utilize two-layer 3D-CNNs, and both $\texttt{Roformer}_{\text{d}}$ and $\texttt{Roformer}_{\text{e}}$ consist of four layers. The hyperparameters for the 3D-Unet are detailed in Table 3. The training process employs an Adam optimizer in conjunction with a WarmupLinearLR scheduler.

Table 3: Hyperparameters of the 3D-Unet in ECloudGen

Parameter	**Value**
In Channels	128
Model Channels	128
Out Channels	64
Number of Residual Blocks	1
Attention Resolutions	[4, 2, 1]
Dropout	0
Channel Multiplier	[1, 2, 2]
Convolution Resample	True
Dimensions	3
Number of Heads	8

I Broader Impacts

In this work, we develop structure-based molecule generation models that have the potential to accelerate the process of drug development, thereby making contributions to human health. However, it is important to note that there is a potential, albeit currently very low, for these models to be misused in the development of chemical weapons. The methodology itself is neutral; what is crucial is the implementation of stringent regulations to prevent misuse.