SculptDrug : A Spatial Condition-Aware Bayesian Flow Model for Structure-based Drug Design

Qingsong Zhong[1], Haomin Yu[3], Yan Lin[4], Wangmeng Shen[1], Long Zeng[1], Jilin Hu[*] [1, 2]

[1]School of Data Science and Engineering, East China Normal University, China
[2]School of Chemistry and Molecular Engineering, East China Normal University, China
[3]School of Science, Engineering and Environment, University of Salford, UK
[4]Department of Computer Science, Aalborg University, Denmark
{xxrelax, wmshen, longzeng}@stu.ecnu.edu.cn, h.yu6@salford.ac.uk, lyan@cs.aau.dk, jlhu@dase.ecnu.edu.cn

Abstract

Structure-Based Drug Design (SBDD) has emerged as a popular approach in drug discovery, leveraging three-dimensional protein structures to generate drug ligands. However, existing generative models encounter several key challenges: (1) Incorporating boundary condition constraints, (2) Integrating hierarchical structural conditions and (3) Ensuring spatial modeling fidelity. To overcome these limitations, we propose SculptDrug, a spatial condition-aware generative model based on Bayesian Flow Networks (BFNs). First, SculptDrug follows a BFNs-based framework and employs a progressive denoising strategy to ensure spatial modeling fidelity, iteratively refining atom positions while enhancing local interactions for precise spatial alignment. Second, we introduce the Boundary Awareness Block, which incorporates protein surface constraints into the generative process to ensure that the generated ligands are geometrically compatible with the target protein. Finally, we design a Hierarchical Encoder that captures global structural context while preserving fine-grained molecular interactions, ensuring overall consistency and accurate ligand-protein conformations. We evaluate SculptDrug on the CrossDocked dataset, and experimental results demonstrate that SculptDrug outperforms state-of-the-art baselines, proving the efficacy of spatial condition-aware modeling.

Code —
https://github.com/decisionintelligence/SculptDrug.git

Introduction

Drug discovery is the process of identifying potential therapeutic molecules to treat diseases, playing a crucial role in improving health and addressing unmet medical needs (Segler et al. 2018). Yet, this process is highly complex and time-consuming. Structure-Based Drug Design (SBDD) has emerged as a powerful approach to streamline this process (Anderson 2003), leveraging the structural information of biological targets to rationally design molecules with improved specificity and efficacy (Isert, Atz, and Schneider 2023).

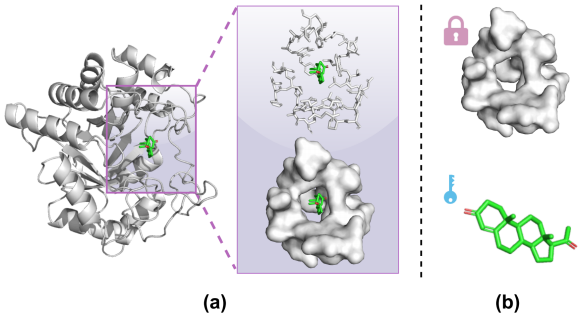

Figure 1: Protein-ligand representation: (a) The protein ribbon model highlights the binding pocket, with close-ups showing the pocket in stick (top) and surface (bottom) representations, emphasizing spatial complementarity. (b) A symbolic "lock-and-key" analogy illustrates the specificity of protein-ligand binding.

The efficacy of SBDD lies in its ability to leverage the lock-and-key principle of protein-ligand interactions (Jorgensen 1991), providing a framework for designing molecules that precisely target biological systems, as shown in Figure 1. Specifically, the protein surface functions as a unique "lock", while the ligand serves as the complementary "key". This structural compatibility ensures precise protein-ligand interactions, highlighting the effectiveness of SBDD in developing targeted and efficient drugs. Traditionally, SBDD relies on virtual screening, which aims to select ligands from an immense chemical space estimated to contain approximately 10^{60} potential molecules (Reymond et al. 2012). virtual screening focuses on predefined compound libraries based on molecular properties and structural features (Lionta et al. 2014), which cover only a small fraction of the space and significantly limit ligand discovery (Reymond et al. 2012). Recent deep generative models (Kingma and Welling 2014; Goodfellow et al. 2020; Ho, Jain, and Abbeel 2020; Graves et al. 2023) enable ligand generation beyond predefined libraries, supporting broader exploration of chemical space (Cheng et al. 2021).

However, existing generative models still face some challenges in learning patterns from protein-ligand complexes.

*Corresponding author

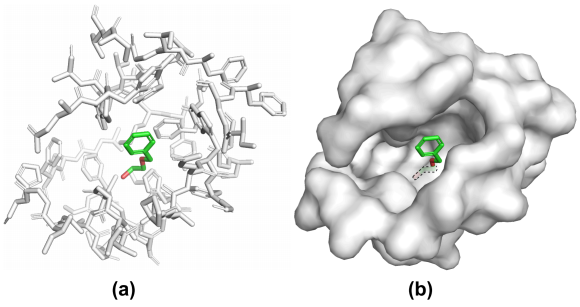

Figure 2: (a) The generated ligand exhibits reasonable atomic distances when evaluated against protein atoms. (b) However, it erroneously penetrates the solvent-excluded surface, violating spatial plausibility.

These challenges can be categorized into the following aspects. (1) *Boundary Condition Constraints.* Designing effective ligands is challenging due to boundary conditions, as ligands are often encapsulated by the protein surface. Due to the lock-and-key nature of protein-ligand interactions, effective ligands must fit within the protein surface boundary. Existing generative models often ignore these constraints, leading to misaligned or buried ligands (Figure 2). (2) *Hierarchical Structure Condition Integration.* Integrating hierarchical structural conditions remains a challenge in ligand design. The local structure is crucial for determining the precise alignment of the key's teeth within a lock, ensuring a better fit based on the lock-and-key principle. Meanwhile, the global structure aims to capture the overall shape of the lock, offering a broader perspective, as the lock's shape and integrity guide the key's design to achieve proper alignment and effective functionality. It is essential to seamlessly integrate both local and global structural information to design ligands that are both functionally effective and structurally stable.

Diffusion models and Bayesian Flow Networks (BFNs) (Graves et al. 2023) are powerful tools for molecular generation, but often struggle with (3) *Spatial Modeling Fidelity*. Unlike images with fixed spatial layouts, molecules rely on flexible interatomic distances to define structure. Noise added to atomic coordinates can distort these distances, causing atoms to fall outside interaction thresholds and be excluded from local computations, ultimately compromising chemical validity.

To address the above challenges, we propose SculptDrug, a spatial condition-aware generative model based on BFNs, designed to generate geometrically accurate and chemically plausible drug-like ligands. First, to improve spatial modeling fidelity, SculptDrug adopts a progressive denoising strategy that gradually refines atom positions and types, enabling more accurate modeling of protein–ligand interactions. Second, we introduce the Boundary Awareness Block, which encodes protein surface information to guide ligand placement and ensure geometric compatibility, effectively avoiding steric clashes. Finally, we incorporate a Hierarchical Encoder that integrates both global structural constraints, ensuring the ligand fits the overall protein pocket, and fine-grained local interactions, which resemble the alignment of a key's teeth within a lock. This design enables the model to maintain structural context integrity while effectively capturing conformational complexity across multiple structural levels.

Overall, the main contributions of this work are summarized as follows:

- We propose a structure-based drug design framework, SculptDrug, which leverages a progressive denoising strategy to achieve high-fidelity spatial interaction modeling.
- We introduce the Boundary Awareness Block to incorporate protein surface geometry into the generative process, encouraging the generated ligands to align with structural constraints and reducing the likelihood of steric clashes.
- We design a Hierarchical Encoder that incorporates both global and local structural contexts, supporting multi-scale alignment.
- We validate the effectiveness of SculptDrug through extensive experiments, demonstrating its superior performance in generating drug-like ligands compared to existing methods.

Related Work

Ligand generation has evolved from 1D representations like SMILES (Bjerrum and Threlfall 2017; Segler et al. 2018), to 2D molecular graphs (Liu et al. 2018; Jin, Barzilay, and Jaakkola 2018), and more recently to 3D structure-based approaches. While 1D/2D methods capture some chemical constraints, they lack the spatial detail essential for modeling protein-ligand interactions. With the rise of deep learning in 3D modeling, structure-aware ligand generation has become a central research focus. Existing methods can be broadly categorized into: (1) *Voxel-based* and (2) *Euclidean space-based* models.

Voxel-based methods discretize 3D space into regular grids and model ligand density distributions within this voxelized space. Representative works such as LIGAN (Masuda, Ragoza, and Koes 2020) and VoxBind (Pinheiro et al. 2024) generate ligands aligned to protein binding pockets using voxel-based representations. However, these methods suffer from high computational cost at fine resolutions and limited spatial precision due to the loss of structural detail during voxelization.

Euclidean space-based methods generate atomic types and positions directly in continuous 3D coordinates, typically representing proteins and ligands as point clouds or graphs. Early models, including GraphBP (Liu et al. 2022), Pocket2Mol (Peng et al. 2022), and FLAG (Zhang et al. 2023c), adopt autoregressive strategies at either the atom or fragment level. SurfGen (Zhang et al. 2023a) and ResGen (Zhang et al. 2023b) incorporate protein context through surface features or residue-level graphs. However, the autoregressive nature of these methods can lead to error accumulation and limited global awareness.

Diffusion-based approaches, such as DIFFBP (Lin et al. 2025a), TARGETDIFF (Guan et al. 2023a), and DIFFSBDD (Schneuing et al. 2024), leverage 3D-equivariant de-

noising to iteratively refine atom placements, improving spatial consistency and structural integrity. Extensions like D3FG (Lin et al. 2023) and DECOMPDIFF (Guan et al. 2023b) introduce fragment priors and scaffold decomposition to enhance diversity and controllability. Nonetheless, modeling discrete atom types alongside continuous coordinates remains challenging. The introduction of Bayesian Flow Networks adds a new dimension to molecular generation (Song et al. 2024a). MOLCRAFT (Qu et al. 2024) generates ligands in continuous parameter spaces. However, their limited ability to perceive protein spatial geometry constrains performance in structure-guided ligand design.

Preliminaries

Problem Definition

Definition 1 (Protein structure). The protein structure refers to the three-dimensional arrangement of atoms within a protein molecule. It is represented as $\mathcal{P} = \{(\mathbf{x}_p^n, \mathbf{a}_p^n)\}_{n=1}^{N_p}$, where $\mathbf{x}_p^n$ denotes the 3D coordinates of the n-th atom in the protein, $\mathbf{a}_p^n$ represents the atom type, and N_P is the total number of atoms in the protein.

Definition 2 (Protein surface). The protein surface refers to the solvent-excluded surface that defines the geometric boundary of a protein. It is represented as a surface graph $\mathcal{S} = \left(\{(\mathbf{x}_s^n, \mathbf{a}_s^n)\}_{n=1}^{N_s}, \mathcal{E}_s\right)$, where $\mathbf{x}_s^n \in \mathbb{R}^3$ is the coordinate of the n-th surface vertex, $\mathbf{a}_s^n \in \mathbb{R}^d$ is the associated feature vector describing geometric and biochemical properties (e.g., shape index, hydrophobicity, polarity, electrostatic charge), N_s is the number of surface points, and $\mathcal{E}_s \subseteq \{(i,j)\}$ defines edges connecting nearby surface points.

Definition 3 (Ligand). The generated ligand molecule, denoted as $\mathcal{M} = \{(\mathbf{x}_m^n, \mathbf{a}_m^n)\}_{n=1}^{N_m}$, consists of atom position $\mathbf{x}_m^n$ and type $\mathbf{a}_m^n$, where N_m represents the total number of atoms in the ligand.

The task can be formally defined as follows: Given the protein structure $\mathcal{P}$ and its surface representation $\mathcal{S}$, the goal is to generate a ligand molecule $\mathcal{M}$ that exhibits high binding affinity, favorable drug-likeness, and a well-formed 3D conformation. This process can be expressed as:

$$\mathcal{F}(\mathcal{S}, \mathcal{P}) = \mathcal{M}, \tag{1}$$

where $\mathcal{F}$ denotes the generative function that maps the protein structure and surface to a chemically and geometrically plausible ligand.

Bayesian Flow Networks

Bayesian Flow Networks (BFNs) represent a novel generative modeling framework that combines Bayesian inference with neural networks to iteratively refine distribution parameters for generating complex data distributions.

In each iteration, BFNs define a sender distribution $p_S(\mathbf{y}|\mathbf{x}, \alpha(t))$ that describes how the noisy observations $\mathbf{y}$ arises from data $\mathbf{x}$ under noise level $\alpha(t)$. The sender distribution can be expressed as:

$$p_S(\mathbf{y}|\mathbf{x}, \alpha(t)) = \prod_{d=1}^{D} p_S(y^{(d)}|x^{(d)}; \alpha(t)), \tag{2}$$

Through Bayesian updating, we iteratively refine the parameter $\boldsymbol{\theta}$ based on noisy observations $\mathbf{y}$. Over time, the randomness introduced by the noise causes $\boldsymbol{\theta}$ to evolve stochastically. By applying multiple Bayesian updates, we eventually obtain the Bayesian flow distribution, which represents the marginal distribution of the parameters $\boldsymbol{\theta}$ over all updates up to time step t. It is defined as:

$$p_F(\boldsymbol{\theta}|\mathbf{x}; t) = p_U(\boldsymbol{\theta}|\boldsymbol{\theta}_0, \mathbf{x}; \beta(t)), \tag{3}$$

where $\beta(t)$ represents the total noise intensity over all previous updates.

Bayesian updates only perform independent inference for each dimension, and BFNs leverage neural networks to incorporate contextual information, deriving the output distribution $p_O(\mathbf{x}|\boldsymbol{\theta}; t)$ by feeding the Bayesian-updated parameter θ and the time step t into the network for more accurate predictions.

The loss function in BFNs aims to minimize the divergence between the sender distribution $p_S(\mathbf{y}|\mathbf{x}, \alpha(t))$ and the receiver distribution $p_R(\mathbf{y}|\boldsymbol{\theta}; t, \alpha(t))$. Thus, the total loss $\mathcal{L}(\mathbf{x})$ consists of two components: n-step loss $\mathcal{L}_n(\mathbf{x})$ and reconstruction loss

$$\mathcal{L}(\mathbf{x}) = \mathcal{L}_n(\mathbf{x}) + \mathcal{L}_r(\mathbf{x}), \tag{4}$$

$$\mathcal{L}_n(\mathbf{x}) = \mathbb{E}_{p(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{n-1})} \sum_{i=1}^{n} D_{KL}\left(p_S^i \,\|\, p_R^i\right), \tag{5}$$

$$\mathcal{L}_r(\mathbf{x}) = -\mathbb{E}_{p_F(\boldsymbol{\theta}|\mathbf{x}, 1)} \ln p_O(\mathbf{x}|\boldsymbol{\theta}; 1). \tag{6}$$

By introducing sender and receiver distributions and minimizing their divergence, BFNs provide a flexible and effective framework for generative modeling.

Spatial Condition-Aware Model

To advance SBDD, we propose a novel model, SculptDrug , based on BFNs. SculptDrug enhances the ligand generation process by integrating both the surface and structural information of the target protein. At each step, the model progressively captures the conditional information of the protein, refining it from coarse to fine, ensuring that the generated ligands adhere to fundamental biological and chemical principles.

As shown in Figure 3, during the BFNs process, at the initial step ($i = 0$), a prior input distribution is provided. In each subsequent transmission step i, the parameters $\boldsymbol{\theta_{i-1}}$ of the previously learned distribution are input into SculptDrug . The model then reconstructs the ligand structure $\mathcal{M}'$ before noise is added and generates an output distribution through a Spatial Condition-Aware (SCA) neural network. The output distribution is given by:

$$p_O(\mathcal{M}' \mid \boldsymbol{\theta}_{i-1}, t_{i-1}) = \text{SCA}(\mathcal{S}, \mathcal{P}, \boldsymbol{\theta}_{i-1}, t_{i-1}). \tag{7}$$

Subsequently, the sender modifies the precision of the ligand data based on a predefined schedule to obtain the sender distribution $p_S(\mathcal{M}_\alpha|\mathcal{M}; \alpha_i)$. Simultaneously, the receiver distribution $p_R(\mathcal{M}'_\alpha|\boldsymbol{\theta}_{i-1}; t_{i-1}, \alpha_i)$ is calculated by applying the same precision to the output distribution. A sample is then drawn from the sender distribution, and through Bayesian updates, the input distribution parameters $\boldsymbol{\theta}_i$ are

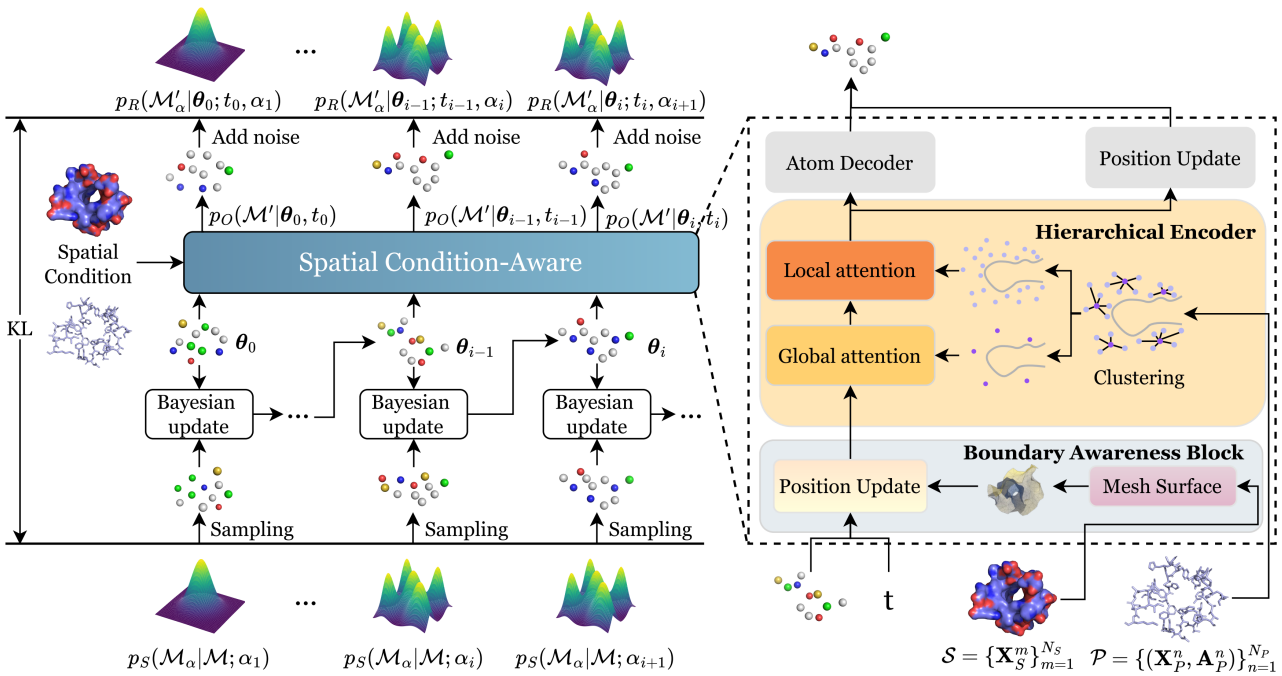

Figure 3: Overview of the SculptDrug framework for ligand generation.

refined for the next round of transmission. After multiple iterations, the data's distribution progressively transitions from the prior distribution to a posterior distribution that more accurately approximates the true distribution of ligand structure.

Importantly, Bayesian updates are performed independently for each data dimension, while the Spatial Condition-Aware neural network plays a crucial role in integrating contextual and conditional information across dimensions, ensuring a more accurate and coherent reconstruction of the ligand structure.

To enhance the reconstruction of denoised ligands under spatial and conditional constraints, we introduce two key modules. The **Boundary Awareness Block** incorporates protein surface information to guide ligand placement within chemically and structurally plausible regions. The **Hierarchical Encoder** captures multi-scale protein context by jointly modeling global pocket geometry and local atomic interactions, enabling precise structural conditioning during ligand generation.

Boundary Awareness Block

In drug design, the binding of ligand to target proteins relies heavily on their precise spatial matching. This relationship is analogous to a key fitting into a lock, where the ligand, as the "key", must align perfectly with the "lock" represented by the protein. To address this, we propose the Boundary Awareness Block, which enhances the model's ability to understand the geometry of protein surfaces and incorporate this essential spatial information during ligand generation.

To achieve this, we follow a systematic approach for understanding and simplifying the protein "lock" and encoding the ligand "key". First, we decode the lock by extracting and simplifying the protein surface structure. Once the surface geometry and spatial features of the lock are identified, we encode the ligand key, ensuring both components are optimized for precise spatial matching. The final stage involves adapting the "key" to fit the "lock" by updating the atomic positions of the ligand.

Extracting the Protein Surface. Inspired by SurfGen (Zhang et al. 2023a) and SurfPro (Song et al. 2024b), we extract the surface structure of the binding pocket by selecting residues within 10 Å of any ligand atom. The solvent-excluded surface is computed using MSMS (Sanner, Olson, and Spehner 1996), which outputs a triangular mesh of surface vertices and faces. We retain only the inward-facing vertices and their adjacent edges that are spatially close to the ligand. Each vertex is annotated with geometric and biochemical features: the shape index is computed from local curvature, while hydrophobicity, polarity, and electrostatic charge are assigned based on the nearest residue. The resulting surface graph aligns with the formal definition of $\mathcal{S}$.

Fitting the Key to the Lock. We construct a unified spatial graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ by merging protein surface points and ligand atoms. The edge set $\mathcal{E}$ consists of both the original mesh connectivity $\mathcal{E}_S$ from the protein surface and additional k-nearest neighbor (k-NN) edges constructed in Euclidean space. Each edge $(i, j) \in \mathcal{E}$ is annotated with a 6-dimensional one-hot vector $\mathbf{t}_{ij} \in \mathbb{R}^6$, indicating both the types of the connected nodes and the source of the edge. The Euclidean distance between nodes is further encoded using Gaussian radial basis functions:

$$\boldsymbol{\phi}_{ij} = \phi(\|\mathbf{x}_i - \mathbf{x}_j\|) \in \mathbb{R}^g, \tag{8}$$

We construct edge features via outer product:

$$\mathbf{e}_{ij} = \mathbf{t}_{ij} \otimes \boldsymbol{\phi}_{ij} \in \mathbb{R}^{6\times g}, \tag{9}$$

which is then flattened and concatenated with the source and target node features:

$$\tilde{\mathbf{e}}_{ij} = \text{concat}(\mathbf{h}_i, \mathbf{h}_j, \text{flatten}(\mathbf{e}_{ij})) \in \mathbb{R}^{2d+6g}. \tag{10}$$

Given these representations, the attention weights and vector messages are defined as:

$$\alpha_{mj} = \text{softmax}_j \left(\frac{f_q(\mathbf{h}_m)^\top f_k(\tilde{\mathbf{e}}_{mj})}{\sqrt{d}} \right), \tag{11}$$

$$\mathbf{v}_{mj} = f_v(\tilde{\mathbf{e}}_{mj}) \cdot (\mathbf{x}_j - \mathbf{x}_m), \tag{12}$$

where f_q, f_k, f_v are learnable multi-layer perceptrons (MLPs), and d denotes the hidden dimension.

The ligand atom positions $\mathbf{x}_m$ are updated via message aggregation:

$$\mathbf{x}_m \leftarrow \mathbf{x}_m + \Delta\mathbf{x}_m, \quad \text{where} \quad \Delta\mathbf{x}_m = \sum_{j\in\mathcal{N}(m)} \alpha_{mj} \cdot \mathbf{v}_{mj},$$

with $\mathcal{N}(m)$ denoting the spatial neighbors of node m in graph $\mathcal{G}$.

Hierarchical Encoder

To better balance the local and global structural information of the protein, we propose a layered encoder that extracts features from multiple structural levels, enhancing the model's understanding of protein context during ligand generation.

While atomic point clouds provide fine-grained local details, they often fail to capture higher-order structural patterns and are sensitive to noise. To address this, we first generate *virtual atoms* by clustering protein atoms, producing a coarse-grained representation that aggregates within-cluster features and highlights global structure. Ligands then interact with these virtual atoms to integrate global context, while fine-grained interactions are captured via multi-type edges that model detailed biochemical cues. To support both levels of interaction, we adopt *equivariant attention mechanisms* in the global and local modules, drawing inspiration from the attention design in DecompDiff (Guan et al. 2023b).

Generating Virtual Atoms for Protein Representation. To obtain a coarse-grained protein representation, we apply k-means++ (Arthur and Vassilvitskii 2007) clustering to protein atoms based on 3D coordinates, generating virtual atoms at cluster centroids. Let C denote the atoms in a cluster. The virtual atom's position is defined as:

$$\mathbf{x}_v = \frac{1}{|C|} \sum_{i\in C} \mathbf{x}_i^P. \tag{13}$$

To compute its feature, we perform a distance-aware aggregation from all atoms in the cluster to the virtual atom:

$$\mathbf{h}_v = \sum_{i\in C(v)} \alpha_{iv}\, \text{MLP}\Big([\mathbf{h}_i^P,\ \phi(\|\mathbf{x}_i^P - \mathbf{x}_v\|)] \Big), \tag{14}$$

where $\boldsymbol{\phi}(\cdot)$ is a radial basis expansion of the interatomic distance, and $\alpha_{iv} \in \mathbb{R}$ is a learned scalar aggregation weight produced by an MLP over the same input, normalized with a softmax over atoms $i \in C(v)$. . Unlike individual atoms, each virtual atom encodes a higher-level abstraction of local structure, offering a coarser yet semantically rich representation. This abstraction allows the global attention mechanism to operate over a reduced node set, significantly improving efficiency while retaining key spatial information.

Global Attention with Adaptive Edge Selection. To capture the global protein pocket context that influences ligand topology and scaffold formation, we apply a unified attention mechanism over ligand atoms and virtual atoms.

Given features and coordinates $\{\mathbf{h}_a, \mathbf{x}_a\}_{a\in\mathcal{V}}$, where $\mathcal{V}$ denotes the union of ligand atoms and virtual atoms, each node updates its representation by attending to all others:

$$\mathbf{h}_a^{(l+1)} = \mathbf{h}_a^{(l)} + \sum_{v\in\mathcal{V}, v\neq a} \psi_h \left(\mathbf{h}_a^{(l)}, \mathbf{h}_v, \mathbf{e}_{av}, \boldsymbol{\phi}_{av} \right) \tag{15}$$

where $\psi_h(\cdot)$ is a graph attention layer and $\boldsymbol{\phi}_{av}$ encodes the inter-node distance.

To incorporate geometric influence, node coordinates are refined through a second attention stream:

$$\begin{aligned} \mathbf{x}_a^{(l+1)} = \mathbf{x}_a^{(l)} + \sum_{v\in\mathcal{V}, v\neq a} (\mathbf{x}_a^{(l)} - \mathbf{x}_v^{(l)}) \cdot \\ \psi_x \left(\mathbf{h}_a^{(l+1)}, \mathbf{h}_v^{(l+1)}, \mathbf{e}_{av}, \boldsymbol{\phi}_{av} \right), \end{aligned} \tag{16}$$

where ψ_x is computed similarly to ψ_h but with separate parameters. This dual-stream mechanism allows the model to separately capture feature-level and geometry-level interactions.

To suppress weak or noisy interactions, we apply an adaptive edge selection strategy, retaining only edges whose mean attention score exceeds a threshold τ:

$$\mathcal{E}' = \left\{ (a, v) \in \mathcal{E} \;\middle|\; \frac{1}{H} \sum_{h=1}^{H} \alpha_{av}^{(h)} > \tau \right\}. \tag{17}$$

Local Interaction Refinement with Distance-Aware Edges. We designed fine-grained edge connection rules grounded in established domain knowledge, specifically inspired by the affinity calculations in AutoDock Vina (Trott and Olson 2010). This approach ensures that our modeling of molecular interactions is both accurate and biologically meaningful. Specifically, leveraging the interaction formulas from VinaDock, three types of edge with distance thresholds set to 2.7 Å, 3.4 Å, and 5 Å are established. The first two thresholds are intended to capture steric and short-range interactions, while the 5 Å threshold is specifically tailored to describe long-range interactions, such as van der Waals forces. These thresholds ensure that the model accurately captures different types of interactions based on spatial proximity, enhancing the precision of ligand generation conditioned on protein structures. We apply the same graph attention modules as in the global stage, but restrict attention to explicitly connected node pairs, enabling fine-grained interaction modeling between ligand atoms and their spatially proximal protein neighbors.

Methods \ Metrics	Vina Score			Vina Min			Vina Dock			Drug-Likeness	
	Evina	IMP%	MPBG%	Evina	IMP%	MPBG%	Evina	IMP%	MPBG%	QED	SA
GRAPHBP (2022)	–	0.00	–	–	1.67	–	-4.57	10.86	-30.03	0.44	0.64
POCKET2MOL (2022)	-5.23	31.06	-15.03	-6.03	38.04	-4.95	-7.05	48.07	-0.17	0.39	0.65
TARGETDIFF (2023)	-5.71	38.21	-22.80	-6.43	47.09	-1.60	-7.41	51.99	5.38	0.49	0.60
FLAG (2023)	–	0.04	–	–	3.44	–	-3.65	11.78	-47.64	0.41	0.58
D3FG (2023)	–	3.70	–	-2.59	11.13	-67.37	-6.78	28.90	-8.85	0.49	0.66
DECOMPDIFF (2023)	-5.18	19.66	-17.17	-6.04	34.84	-6.78	-7.10	48.31	-1.59	0.49	0.66
DIFFSBDD (2024)	–	12.67	–	-2.15	22.24	–	-5.53	29.76	-23.51	0.49	0.34
MOLCRAFT (2024)	-6.59	54.86	-1.86	-7.17	**61.02**	10.42	-7.83	58.05	8.17	0.50	**0.67**
DIFFBP (2025)	–	8.60	–	–	19.68	–	-7.34	49.24	6.23	0.47	0.59
Ours	**-6.94**	**56.50**	**7.95**	**-7.30**	60.70	**12.94**	**-8.06**	**60.78**	**10.78**	**0.54**	**0.67**

Note: Evina > 0 are considered invalid and are represented by "–". Baseline ligand results are from CBGbench (Lin et al. 2025b).

Table 1: Comparative performance of baseline models and our method on binding affinity and drug-likeness metrics.

Experiments

Experimental Setup

Dataset. We use the CrossDocked (Francoeur et al. 2020) dataset, comprising 22.5 million docked protein–ligand complexes. Following standard protocols (Luo et al. 2021; Lin et al. 2025b; Qu et al. 2024), we select samples with Root-Mean-Square Deviation (RMSD) < 1 Å and sequence identity $< 30\%$. This yields 100,000 complexes for training and 100 for testing. For each test complex, 100 ligands are generated to ensure robust evaluation.

Metrics. We evaluate model performance across four categories. ↑ indicates higher is better, and ↓ indicates lower is better.

Binding Affinity. Following prior works, we use AUTODOCK VINA (Trott and Olson 2010) to evaluate binding energy from three perspectives: (a) Vina Score, which measures the energy of the original docked pose; (b) Vina Min, referring to the minimum energy after local minimization; and (c) Vina Dock, which represents the energy obtained through global redocking. Additionally, we compute **Evina** (↓), the mean binding energy across all generated ligands; **IMP%** (↑), the percentage of generated ligands outperforming the reference; and **MPBG%** (↑), the average binding energy improvement over the reference.

Drug-likeness. We assess drug-likeness using two standard metrics: Quantitative Estimate of Drug-likeness **(QED)** (↑), which reflects oral bioavailability, and Synthetic Accessibility **(SA)** (↑), which estimates synthetic feasibility.

Structural Plausibility. Following Guan et al. (Guan et al. 2023a), we compare the distributions of bond lengths, C–C distances within 2 Å, and all-atom distances within 12 Å. Bond length categories include C–C, C=C, C:C, C–N, C=N, C:N, C–O, and C=O, where the symbols "–", "=", and ":" respectively denote single, double, and aromatic bonds. These distributions are quantified using Jensen–Shannon Divergence, reported as **JSD_BL** (↓), **JSD_CC_2Å** (↓), and **JSD_ALL_12Å** (↓), where lower values indicate better alignment with empirical statistics.

Conformational Stability. Following Harris et al. (Harris et al. 2023), we assess the conformational stability of generated ligands using two metrics: **Strain Energy (SE)** (↓), which quantifies the internal energetic stability of the ligand and is reported at the 25th, 50th, and 75th percentiles, and **Steric Clashes (Clash)** (↓), which measures the number of atomic overlaps between ligand and protein, reflecting spatial feasibility within the binding pocket.

Baselines. We compare our method against a diverse set of baseline models, including both auto-regressive and non-auto-regressive approaches. Detailed descriptions of these baselines are provided in the Appendix.

Results

The performance of our method is evaluated across multiple metrics to comprehensively assess its effectiveness in generating high-quality ligands for SBDD. Results are shown in Tables 1, 2, and 3, with the best scores in bold and the second-best underlined. Our method consistently outperforms baseline models across all dimensions, including binding affinity, structural plausibility, and conformational stability.

Binding Affinity and Drug-Likeness. As shown in Table 1, our method achieves the best Vina Score of -6.94, with 56.50% of generated ligands outperforming reference ligands. In terms of MPBG%, our method is the only one yielding a positive score 7.95%, indicating overall improved binding affinity without any post-processing. For Vina Min and Vina Dock, SculptDrug also achieves optimal values of -7.30 and -8.06, respectively, demonstrating superior local energy minima and docking robustness. In terms of drug-likeness, SculptDrug yields competitive QED (0.54) and SA (0.67) scores, reflecting its potential to generate compounds with acceptable pharmacological and synthetic properties.

Structural Plausibility. Table 2 summarizes the Jensen–Shannon Divergence (JSD) scores for ligand structural distributions. SculptDrug consistently achieves the lowest divergence across all metrics: 0.1522 for average bond lengths (JSD_BL), 0.1163 for short-range C–C bonds

	JSD_BL	JSD_CC_2Å	JSD_All_12Å
GRAPHBP	0.5316	0.5229	0.3936
POCKET2MOL	0.5602	0.5236	0.2364
TARGETDIFF	0.2409	0.2285	0.0600
FLAG	0.4003	0.3611	0.0676
D3FG	0.3894	0.3416	0.1119
DECOMPDIFF	0.2277	0.1895	0.0486
DIFFSBDD	0.4610	0.4553	0.1614
MOLCRAFT	0.2251	0.1737	0.0417
DIFFBP	0.5904	0.5926	0.3364
Ours	**0.1522**	**0.1163**	**0.0331**

Table 2: Comparative analysis of ligand bond length distribution across generative models.

	SE_25	SE_50	SE_75	Clash
GRAPHBP	-	-	-	193.67
POCKET2MOL	-	-	-	8.13
TARGETDIFF	252.02	878.72	-	9.33
FLAG	129.11	353.60	885.80	42.59
D3FG	460.07	1332.99	-	29.01
DECOMPDIFF	132.37	473.16	1904.26	8.51
DIFFSBDD	895.07	-	-	71.18
MOLCRAFT	83.92	197.10	536.22	7.03
DIFFBP	-	-	-	78.17
Ours	**72.90**	**170.17**	**523.66**	**6.41**

"-" indicates the value exceeds 10,000.

Table 3: Evaluation of strain energy (SE) and steric clashes (Clash) in ligand-protein complexes.

within 2 Å (JSD_CC_2Å), and 0.0331 for all-atom distances within 12 Å (JSD_All_12Å). These results demonstrate that SculptDrug generates chemically valid structures with high fidelity to empirical spatial patterns, capturing local bonding accuracy.

Conformational Stability. As shown in Table 3, SculptDrug generates ligand conformations with reduced strain energy and fewer steric clashes, indicating improved spatial compatibility with the binding pocket and enhanced structural plausibility.

Ablation Studies

To evaluate the contribution of each key component in SculptDrug , we conduct an ablation study by systematically removing individual modules: (1) **w/o Multi_edge**, replacing the multi-type edge construction with a K-nearest neighbor (KNN) scheme; (2) **w/o Global**, which removes the global attention layer responsible for integrating global structural information through virtual atoms; and (3) **w/o Surface**, which omits the Boundary Awareness Block that incorporates protein surface geometry into ligand generation. The complete **SculptDrug** model serves as the baseline. Figure 4 summarizes the performance metrics for each

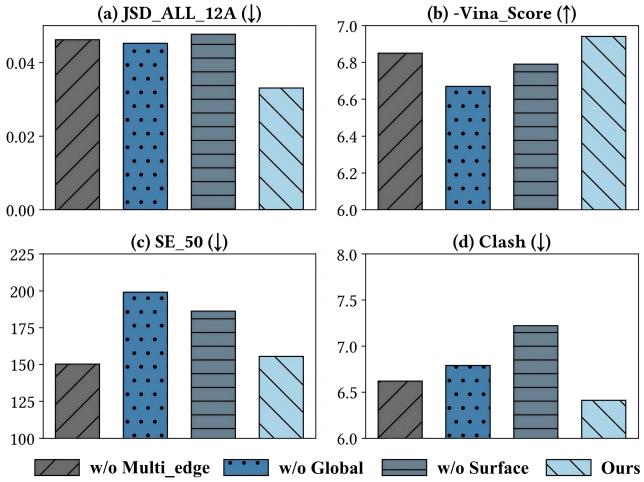

Figure 4: Impact of variants on SculptDrug 's performance.

variant alongside the complete model. We can observe that removing any single component leads to a consistent decrease in performance across all metrics.

Notably, the **w/o Global** variant leads to the most pronounced degradation in **Vina Score**, reflecting impaired ligand–protein compatibility due to missing global structural context. The **w/o Surface** variant results in elevated steric clashes and a slight increase in all-atom JSD, suggesting that surface-aware conditioning plays an important role in guiding spatially plausible ligand placement. The **w/o Multi_edge** variant experiences a moderate decline, demonstrating that multiple edge types are essential for capturing complex local interactions and enhancing spatial modeling fidelity. In contrast, the complete **SculptDrug** model achieves the best performance overall, validating the synergistic effect of integrating hierarchical structure, surface information, and fine-grained interaction modeling.

Conclusion and Future Works

In this study, we propose SculptDrug, a novel Bayesian Flow Network-based SBDD model that effectively addresses three key challenges in ligand generation: boundary condition constraints, hierarchical structural integration, and spatial modeling fidelity. By employing a progressive denoising strategy and two tailored components—the Boundary Awareness Block and the Hierarchical Encoder—our method integrates protein surface geometry and multi-level structural information. A comprehensive evaluation on CrossDocked2020 demonstrates that SculptDrug surpasses state-of-the-art models in binding affinity, drug-likeness, structural plausibility, and conformational stability. Nonetheless, we observe that a small portion of generated ligands still exhibit high strain energy, indicating room for improvement. In addition, future work may consider incorporating dynamic protein modeling to further enhance biological relevance.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (62472174), the Shanghai Frontiers Science Center of Molecule Intelligent Syntheses, and the ECNU Multifunctional Platform for Innovation (001) for high-performance computing resources.

References

Anderson, A. C. 2003. The process of structure-based drug design. *Chemistry & biology*, 10(9): 787–797.

Arthur, D.; and Vassilvitskii, S. 2007. k-means++: the advantages of careful seeding. In *SODA*, 1027–1035.

Bjerrum, E. J.; and Threlfall, R. 2017. Molecular Generation with Recurrent Neural Networks (RNNs). *CoRR*, abs/1705.04612.

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

Francoeur, P. G.; Masuda, T.; Sunseri, J.; Jia, A.; Iovanisci, R. B.; Snyder, I.; and Koes, D. R. 2020. 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.

Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A. C.; and Bengio, Y. 2020. Generative adversarial networks. *Commun. ACM*, 63(11): 139–144.

Graves, A.; Srivastava, R. K.; Atkinson, T.; and Gomez, F. J. 2023. Bayesian Flow Networks. *CoRR*, abs/2308.07037.

Guan, J.; Qian, W. W.; Peng, X.; Su, Y.; Peng, J.; and Ma, J. 2023a. 3D Equivariant Diffusion for Target-Aware Molecule Generation and Affinity Prediction. In *ICLR*.

Guan, J.; Zhou, X.; Yang, Y.; Bao, Y.; Peng, J.; Ma, J.; Liu, Q.; Wang, L.; and Gu, Q. 2023b. DecompDiff: Diffusion Models with Decomposed Priors for Structure-Based Drug Design. In *ICML*, volume 202, 11827–11846.

Harris, C.; Didi, K.; Jamasb, A. R.; Joshi, C. K.; Mathis, S. V.; Lio, P.; and Blundell, T. 2023. Benchmarking Generated Poses: How Rational is Structure-based Drug Design with Generative Models? *arXiv preprint arXiv:2308.07413*.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising Diffusion Probabilistic Models. In *NeurIPS*, volume 33, 6840–6851.

Isert, C.; Atz, K.; and Schneider, G. 2023. Structure-based drug design with geometric deep learning. *Current Opinion in Structural Biology*, 79: 102548.

Jin, W.; Barzilay, R.; and Jaakkola, T. S. 2018. Junction Tree Variational Autoencoder for Molecular Graph Generation. In *ICML*, volume 80, 2328–2337.

Jorgensen, W. L. 1991. Rusting of the lock and key model for protein-ligand binding. *Science*, 254(5034): 954–955.

Kingma, D. P.; and Welling, M. 2014. Auto-Encoding Variational Bayes. In *ICLR*.

Lin, H.; Huang, Y.; Zhang, O.; Liu, Y.; Wu, L.; Li, S.; Chen, Z.; and Li, S. Z. 2023. Functional-Group-Based Diffusion for Pocket-Specific Molecule Generation and Elaboration. In *NeurIPS*, volume 36, 34603–34626.

Lin, H.; Huang, Y.; Zhang, O.; Ma, S.; Liu, M.; Li, X.; Wu, L.; Wang, J.; Hou, T.; and Li, S. Z. 2025a. Diffbp: Generative diffusion of 3d molecules for target protein binding. *Chemical Science*, 16: 1417–1431.

Lin, H.; Zhao, G.; Zhang, O.; Huang, Y.; Wu, L.; Tan, C.; Liu, Z.; Gao, Z.; and Li, S. Z. 2025b. CBGBench: Fill in the Blank of Protein-Molecule Complex Binding Graph. In *ICLR*.

Lionta, E.; Spyrou, G.; K Vassilatis, D.; and Cournia, Z. 2014. Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Current topics in medicinal chemistry*, 14(16): 1923–1938.

Liu, M.; Luo, Y.; Uchino, K.; Maruhashi, K.; and Ji, S. 2022. Generating 3D Molecules for Target Protein Binding. In *ICML*, volume 162, 13912–13924.

Liu, Q.; Allamanis, M.; Brockschmidt, M.; and Gaunt, A. L. 2018. Constrained Graph Variational Autoencoders for Molecule Design. In *NeurIPS*, volume 31, 7806–7815.

Luo, S.; Guan, J.; Ma, J.; and Peng, J. 2021. A 3D Generative Model for Structure-Based Drug Design. In *NeurIPS*, volume 34, 6229–6239.

Masuda, T.; Ragoza, M.; and Koes, D. R. 2020. Generating 3D Molecular Structures Conditional on a Receptor Binding Site with Deep Generative Models. *CoRR*, abs/2010.14442.

Peng, X.; Luo, S.; Guan, J.; Xie, Q.; Peng, J.; and Ma, J. 2022. Pocket2Mol: Efficient Molecular Sampling Based on 3D Protein Pockets. In *ICML*, volume 162, 17644–17655.

Pinheiro, P. O.; Jamasb, A. R.; Mahmood, O.; Sresht, V.; and Saremi, S. 2024. Structure-based drug design by denoising voxel grids. In *ICML*, volume 235, 40795–40812.

Qu, Y.; Qiu, K.; Song, Y.; Gong, J.; Han, J.; Zheng, M.; Zhou, H.; and Ma, W. 2024. MolCRAFT: Structure-Based Drug Design in Continuous Parameter Space. In *ICML*, volume 235, 41749–41768.

Reymond, J.-L.; Ruddigkeit, L.; Blum, L.; and Van Deursen, R. 2012. The enumeration of chemical space. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(5): 717–733.

Sanner, M. F.; Olson, A. J.; and Spehner, J.-C. 1996. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers*, 38(3): 305–320.

Schneuing, A.; Harris, C.; Du, Y.; Didi, K.; Jamasb, A. R.; Igashov, I.; Du, W.; Gomes, C. P.; Blundell, T. L.; Lio, P.; Welling, M.; Bronstein, M. M.; and Correia, B. E. 2024. Structure-based drug design with equivariant diffusion models. *Nat. Comput. Sci.*, 4(12): 899–909.

Segler, M. H.; Kogej, T.; Tyrchan, C.; and Waller, M. P. 2018. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS central science*, 4(1): 120–131.

Song, Y.; Gong, J.; Zhou, H.; Zheng, M.; Liu, J.; and Ma, W.-Y. 2024a. Unified Generative Modeling of 3D Molecules with Bayesian Flow Networks. In *ICLR*.

Song, Z.; Huang, T.; Li, L.; and Jin, W. 2024b. SurfPro: Functional Protein Design Based on Continuous Surface. In *ICML*, volume 235, 46074–46088.

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

Zhang, O.; Wang, T.; Weng, G.; Jiang, D.; Wang, N.; Wang, X.; Zhao, H.; Wu, J.; Wang, E.; Chen, G.; et al. 2023a. Learning on topological surface and geometric structure for 3D molecular generation. *Nature Computational Science*, 3(10): 849–859.

Zhang, O.; Zhang, J.; Jin, J.; Zhang, X.; Hu, R.; Shen, C.; Cao, H.; Du, H.; Kang, Y.; Deng, Y.; et al. 2023b. ResGen is a pocket-aware 3D molecular generation model based on parallel multiscale modelling. *Nature Machine Intelligence*, 5(9): 1020–1030.

Zhang, Z.; Min, Y.; Zheng, S.; and Liu, Q. 2023c. Molecule Generation For Target Protein Binding with Structural Motifs. In *ICLR*.