Binding-Adaptive Diffusion Models for Structure-Based Drug Design

Zhilin Huang[1,2*], Ling Yang[3*], Zaixi Zhang[4], Xiangxin Zhou[5], Yu Bao[6], Xiawu Zheng[2], Yuwei Yang[6], Yu Wang[2†], Wenming Yang[1,2†]

[1]Shenzhen International Graduate School, Tsinghua University
[2]Peng Cheng Laboratory
[3]Peking University
[4]University of Science and Technology of China
[5]University of Chinese Academy of Sciences
[6]ByteDance
zerinhwang03@pku.edu.cn, zaixi@mail.ustc.edu.cn, {zhouxiangxin1998, nlp.baoy}@gmail.com, yuwei.yang@bytedance.com, {zhengxw01, wangy20}@pcl.ac.cn, {yangling0818, yangelwm}@163.com

Abstract

Structure-based drug design (SBDD) aims to generate 3D ligand molecules that bind to specific protein targets. Existing 3D deep generative models including diffusion models have shown great promise for SBDD. However, it is complex to capture the essential protein-ligand interactions exactly in 3D space for molecular generation. To address this problem, we propose a novel framework, namely **Bind**ing-Adaptive **D**iffusion **M**odels (BINDDM). In BINDDM, we adaptively extract *subcomplex*, the essential part of binding sites responsible for protein-ligand interactions. Then the selected protein-ligand subcomplex is processed with SE(3)-equivariant neural networks, and transmitted back to each atom of the complex for augmenting the target-aware 3D molecule diffusion generation with binding interaction information. We iterate this hierarchical complex-subcomplex process with *cross-hierarchy interaction node* for adequately fusing global binding context between the complex and its corresponding subcomplex. Empirical studies on the CrossDocked2020 dataset show BINDDM can generate molecules with more realistic 3D structures and higher binding affinities towards the protein targets, with up to **-5.92** Avg. Vina Score, while maintaining proper molecular properties. Our code is available at https://github.com/YangLing0818/BindDM

Introduction

Designing ligand molecules that can bind to specific protein targets and modulate their function, also known as *structure-based drug design* (SBDD) (Anderson 2003; Batool, Ahmad, and Choi 2019), is a fundamental problem in drug discovery and can lead to significant therapeutic benefits. SBDD requires models to synthesize drug-like molecules with stable 3D structures and high binding affinities to the target. Nevertheless, it is challenging and involves massive computational efforts because of the enormous space of synthetically feasible chemicals (Ragoza, Masuda, and Koes 2022a) and freedom degree of both compound and protein structures (Hawkins 2017).

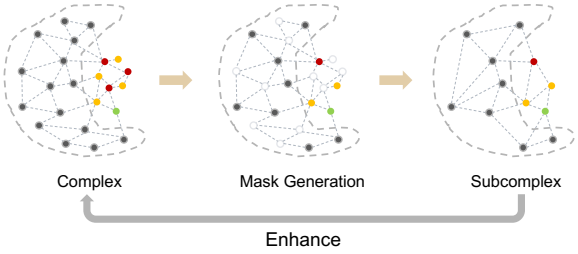

Figure 1: BINDDM extracts subcomplex from protein-ligand complex, and utilizes it to enhance the binding-adaptive 3D molecule generation in complex.

Recent advances in modeling geometric structures of biomolecules (Bronstein et al. 2021; Atz, Grisoni, and Schneider 2021) motivate a promising direction for SBDD (Gaudelet et al. 2021; Zhang et al. 2023b). Several new generative methods have been proposed for the SBDD task (Li, Pei, and Lai 2021; Luo et al. 2021; Peng et al. 2022; Powers et al. 2022; Ragoza, Masuda, and Koes 2022b; Zhang et al. 2023a), which learn to generate ligand molecules by modeling the complex spatial and chemical interaction features of the binding site. For instance, some methods adopt autoregressive models (ARMs) (Luo and Ji 2021; Liu et al. 2022a; Peng et al. 2022) and show promising results in SBDD tasks, which generate 3D molecules by iteratively adding atoms or bonds based on the target binding site. However, ARMs tend to suffer from error accumulation, and it also is difficult to find an optimal generation order.

An alternative to address these limitations of ARMs is to sample the atomic coordinates and types of all the atoms at once (Du et al. 2022). Recent diffusion-based SBDD methods (Guan et al. 2023a; Schneuing et al. 2023; Lin et al. 2022; Guan et al. 2023b) adopt diffusion models (Ho, Jain, and Abbeel 2020; Song et al. 2020) to model the distribution of atom types and positions from a standard Gaussian prior with post-processing to assign bonds.

*These authors contributed equally.
†Corresponding author.

These diffusion-based methods learn the joint generative process with a SE(3)-equivariant diffusion models (Hoogeboom et al. 2022) to capture both spatial and chemical interactions between atoms, and have achieved comparable performance with previous autoregressive models.

Despite the state-of-the-art performance, existing methods pay little attention to the binding-specific substructure of protein-ligand complex, *i.e.*, the essential part of binding sites responsible for protein-ligand interactions, which plays a crucial role in generating molecules with high binding affinities towards the protein targets (Bajusz et al. 2021; Kozakov et al. 2015). Although recent FLAG (Zhang et al. 2023a) and DrugGPS (Zhang and Liu 2023) learn to generate pocket-aware 3D molecules fragment-by-fragment, these massive pre-defined fragments (*e.g.*, motifs or subpockets) are still complex for the model to exactly discover essential protein-ligand interactions from highly diverse protein pockets in nature (Spitzer, Cleves, and Jain 2011; Basanta et al. 2020). Consequently, these limit their practical use in designing high-affinity molecules for new protein targets.

To address these issues, we propose BINDDM, a new binding-adaptive diffusion model for SBDD. Instead of using pre-defined fragments of pockets or molecules (*e.g.*, subpockets or motifs), at each time step of denoising process, we directly extract essential binding *subcomplex* from protein-ligand complex with a learnable structural pooling. Then we process the selected subcomplex with SE(3)-equivariant GNNs, and transmit them back to the complex as enhanced binding context to improve the atomic target-aware 3D molecule generation. To facilitate the exchange between the complex and its subcomplex, we iterate the above process via our designed *cross-hierarchy interaction nodes*. Extensive experiments demonstrate that BINDDM can generate molecules with more realistic 3D structures and higher binding affinities towards the protein targets, while maintaining proper molecular properties. We highlight our main contributions as follows:

- We propose a hierarchical complex-subcomplex diffusion model for structure-based drug design, which incorporates essential binding-adaptive subcomplex for 3D molecule diffusion generation.
- We design and incorporate cross-hierarchy interaction nodes into our iterative denoising networks in the generation process for sufficiently fusing context information.
- Empirical results on CrossDocked2020 dataset demonstrate that our BINDDM achieves better performance compared with previous methods, higher affinity with target protein and other drug properties.

Related Work

Structure-Based Drug Design

As the increasing availability of 3D-structure protein-ligand data (Kinnings et al. 2011), structure-based drug design (SBDD) becomes a hot research area and it aims to generate diverse molecules with high binding affinity to specific protein targets (Luo et al. 2021; Yang et al. 2022; Schneuing et al. 2023; Tan, Gao, and Li 2023). Early attempts learn to generate SMILES strings or 2D molecular graphs given protein contexts (Skalic et al. 2019; Xu, Ran, and Chen 2021). However, it is uncertain whether the resulting compounds with generated strings or graphs could really fit the geometric landscape of the 3D structural pockets. More works start to involve 3D structures of both proteins and molecules (Li, Pei, and Lai 2021; Ragoza, Masuda, and Koes 2022b; Zhang et al. 2023a). Luo et al. (2021), Liu et al. (2022b), and Peng et al. (2022) adopt autoregressive models to generate 3D molecules in an atom-wise manner. Recently, powerful diffusion models (Sohl-Dickstein et al. 2015; Song and Ermon 2019; Ho, Jain, and Abbeel 2020) begin to play a role in SBDD, and have achieved promising generation results with non-autoregressive sampling (Lin et al. 2022; Schneuing et al. 2023; Guan et al. 2023a). TargetDiff (Guan et al. 2023a), DiffBP (Lin et al. 2022), and DiffSBDD (Schneider et al. 1999) utilize E(n)-equivariant GNNs (Satorras, Hoogeboom, and Welling 2021) to parameterize conditional diffusion models for protein-aware 3D molecular generation. Despite progress, existing methods pay little attention to binding-specific protein-ligand substructures. In contrast, We propose BINDDM to automatically extracts essential binding-adaptive *subcomplex*, and design a hierarchical equivariant molecular diffusion model for SBDD.

Diffusion Models for SBDD

As a new family of deep generative models, diffusion models (Sohl-Dickstein et al. 2015; Ho, Jain, and Abbeel 2020; Song et al. 2020) have been recently applied in SBDD tasks. They usually represent the protein-ligand complex by treating protein binding pockets and ligand molecules as atom point sets in the 3D space, and define a diffusion process for both continuous atom coordinates and discrete atom types for reverse diffusion generation. TargetDiff (Guan et al. 2023a) and DiffBP (Lin et al. 2022) both propose a target-aware molecular diffusion process with a SE(3)-equivariant GNN denoiser. DecompDiff (Guan et al. 2023b) proposes a two-stage diffusion model, which uses an open-source software to obtain molecule-agnostic binding priors as templates for the generation process. In contrast, our BINDDM is a single-stage approach that generates molecules from scratch without relying on external knowledge. It adaptively mines binding-related subcomplexes from the original complex to enhance the generation process, fully considering the interaction between protein pockets and ligands.

Preliminary

The SBDD task from the perspective of generative models can be defined as generating molecules which can bind to a given protein pocket. The target protein and molecule can be represented as $\mathcal{P} = \{(\boldsymbol{x}_P^{(i)}, \boldsymbol{v}_P^{(i)})\}_{i=1}^{N_P}$ and $\mathcal{M} = \{(\boldsymbol{x}_M^{(i)}, \boldsymbol{v}_M^{(i)})\}_{i=1}^{N_M}$, respectively. Here N_P (resp. N_M) refers to the number of atoms of the protein $\mathcal{P}$ (resp. the molecule $\mathcal{M}$). $\boldsymbol{x} \in \mathbb{R}^3$ and $\boldsymbol{v} \in \mathbb{R}^K$ denote the position and type of the atom, respectively. And K denotes the number of atom types. In the sequel, matrices are denoted by uppercase boldface. For a matrix $\mathbf{X}$, $\mathbf{x}_i$ denotes the vector on its i-th row, and $\mathbf{X}_{1:N}$ denotes the submatrix comprising its 1-st

to N-th rows. For brevity, the molecule is denoted as $\mathbf{M} = [\mathbf{X}_M, \mathbf{V}_M]$ where $\mathbf{X}_M \in \mathbb{R}^{N_M \times 3}$ and $\mathbf{V}_M \in \mathbb{R}^{N_M \times K}$, and the protein is denoted as $\mathbf{P} = [\mathbf{X}_P, \mathbf{V}_P]$ where $\mathbf{X}_P \in \mathbb{R}^{N_P \times 3}$ and $\mathbf{V}_P \in \mathbb{R}^{N_P \times K}$. The task can be formulated as modeling the conditional distribution $p(\mathbf{M}|\mathbf{P})$.

Denoising Diffusion Probabilistic Models (DDPMs) equipped with SE(3)-invariant prior and SE(3)-equivariant transition kernel have been applied on the SBDD task (Guan et al. 2023a; Schneuing et al. 2023; Lin et al. 2022). Specifically, types and positions of the ligand molecule are modeled by DDPM, while the number of atoms N_M is usually sampled from an empirical distribution (Hoogeboom et al. 2022; Guan et al. 2023a) or predicted by a neural network (Lin et al. 2022), and bonds are determined as post-processing.

In the forward diffusion process, a small Gaussian noise is gradually injected into data as a Markov chain. Because noises are only added on ligand molecules but not proteins in the diffusion process, we denote the atom positions and types of the ligand molecule at time step t as $\mathbf{X}_t$ and $\mathbf{V}_t$ and omit the subscript M without ambiguity. The diffusion transition kernel can be defined as follows:

$$q(\mathbf{M}_t|\mathbf{M}_{t-1}, \mathbf{P}) = \prod_{i=1}^{N_M} \mathcal{N}(\mathbf{x}_{i,t}; \sqrt{1-\beta_t}\mathbf{x}_{i,t-1}, \beta_t \boldsymbol{I}) \cdot \mathcal{C}(\mathbf{v}_{i,t}|(1-\beta_t)\mathbf{v}_{i,t-1} + \beta_t/K), \tag{1}$$

where $\mathcal{N}$ and $\mathcal{C}$ stand for the Gaussian and categorical distribution respectively, β_t is defined by fixed variance schedules. The corresponding posterior can be derived as follows:

$$q(\mathbf{M}_{t-1}|\mathbf{M}_t, \mathbf{M}_0, \mathbf{P}) = \prod_{i=1}^{N_M} \mathcal{N}(\mathbf{x}_{i,t-1}; \tilde{\boldsymbol{\mu}}(\mathbf{x}_{i,t}, \mathbf{x}_{i,0}), \tilde{\beta}_t \boldsymbol{I}) \cdot \mathcal{C}(\mathbf{v}_{i,t-1}|\tilde{\boldsymbol{c}}(\mathbf{v}_{i,t}, \mathbf{v}_{i,0})), \tag{2}$$

where $\tilde{\boldsymbol{\mu}}(\mathbf{x}_{i,t}, \mathbf{x}_{i,0}) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\mathbf{x}_{i,0} + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_{i,t}$, $\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$, $\alpha_t = 1-\beta_t$, $\bar{\alpha}_t = \prod_{s=1}^{t}\alpha_s$, $\tilde{\boldsymbol{c}}(\mathbf{v}_{i,t}, \mathbf{v}_{i,0}) = \frac{\boldsymbol{c}^*}{\sum_{k=1}^{K} c_k^*}$, and $\boldsymbol{c}^*(\mathbf{v}_{i,t}, \mathbf{v}_{i,0}) = [\alpha_t \mathbf{v}_{i,t} + (1-\alpha_t)/K] \odot [\bar{\alpha}_{t-1}\mathbf{v}_{i,0} + (1-\bar{\alpha}_{t-1})/K]$.

In the approximated reverse process, also known as the generative process, a neural network parameterized by θ learns to recover data by iteratively denoising. The reverse transition kernel can be approximated with predicted atom types $\hat{\mathbf{v}}_{i,0}$ and atom positions $\hat{\mathbf{x}}_{i,0}$ as follows:

$$p_\theta(\mathbf{M}_{t-1}|\mathbf{M}_t, \mathbf{P}) = \prod_{i=1}^{N_M} \mathcal{N}(\mathbf{x}_{i,t-1}; \tilde{\boldsymbol{\mu}}(\mathbf{x}_{i,t}, \hat{\mathbf{x}}_{i,0}), \tilde{\beta}_t \boldsymbol{I}) \cdot \mathcal{C}(\mathbf{v}_{i,t-1}|\tilde{\boldsymbol{c}}(\mathbf{v}_{i,t}, \hat{\mathbf{v}}_{i,0})). \tag{3}$$

The Proposed BINDDM

As discussed in previous sections, we aim to develop a hierarchical binding-specific diffusion model for SBDD. We here present our proposed BINDDM, as illustrated in Figure 2. In this subsection, we will describe how to introduce the selected protein-ligand binding *subcomplex* into the design of the neural network ϕ_θ which predicts (*i.e.*, reconstructs) $\mathbf{M}_0 = [\mathbf{X}_0, \mathbf{V}_0]$ in the reverse generation process:

$$[\hat{\mathbf{X}}_0, \hat{\mathbf{V}}_0] = \phi_\theta([\mathbf{X}_t, \mathbf{V}_t], t, \mathbf{P}). \tag{4}$$

To extract essential interaction binding-adaptive protein-ligand subcomplex, we design a learnable structural pooling to filter the nodes in the original complex graph. To sufficiently utilize both the complex and the subcomplex, we apply SE(3)-equivariant neural networks on them. Finally, we design cross-hierarchy interaction nodes to iteratively exchange information between the complex and the subcomplex, and facilitate the target-aware 3D molecule generation.

Binding-Adaptive Subcomplex Extraction

We first elaborate on how we adaptively extract essential binding subcomplex at each time step t. Different from the denoising networks in previous SBDD methods (Guan et al. 2023a,b; Schneuing et al. 2023) that only process the full-atom complex graph, our BINDDM produces a binding-adaptive subcomplex graph from the complex graph with a learnable structural pooling. Formally, we have a k-nearest neighbors graph $\mathcal{G}_C^l$ based on the protein-ligand complex $\mathbf{C} = [\mathbf{M}_t, \mathbf{P}]$ at each denoising time step, where the superscripts l denotes the l-th graph layer of the denoising network, and $[\cdot]$ denotes the concatenation along the first dimension. We aim to extract a binding subcomplex graph $\tilde{\mathcal{G}}_S^l = (\tilde{\mathbf{H}}_S^l, \tilde{\mathbf{X}}_S^l)$ from $\mathcal{G}_C^l = (\mathbf{H}_C^l, \mathbf{X}_C^l)$ (and $\mathcal{G}_S^l = (\mathbf{H}_S^l, \mathbf{X}_S^l)$, for $l > 0$), where $\mathbf{H} \in \mathbb{R}^{N \times d}$ is the node hidden state matrix (initialized with $\mathbf{V}_t$ in first layer), $\mathbf{X} \in \mathbb{R}^{N \times 3}$ is the node position matrix. We calculate the confidence scores $\mathbf{Z}^l \in \mathbb{R}^{N \times 1}$ of all nodes in the complex graph $\mathcal{G}_C^l$ contributing to the molecule generation with provided binding sites:

$$\hat{\mathbf{H}}_C^l = \begin{cases} f_\theta(\mathbf{H}_C^0), l = 0 \\ f_\theta([\mathbf{H}_C^l, \text{pad}(\mathbf{H}_S^l, \text{idx}^l)]), l > 0 \end{cases} \tag{5}$$

$$\mathbf{Z}^l = \sigma(\mathbf{D}_C^{l\,-\frac{1}{2}} \mathbf{A}_C^l \mathbf{D}_C^{l\,-\frac{1}{2}} \hat{\mathbf{H}}_C^l \Phi_{att}) \tag{6}$$

where $\mathbf{A}_C^l \in \mathbb{R}^{N \times N}$ is the adjacency matrix with pair-wise node connections defined on k-nn graph according to $\mathbf{X}_C^l$, $\mathbf{D}_C^l \in \mathbb{R}^{N \times N}$ is the degree matrix of $\mathbf{A}_C^l$, $f_\theta(\cdot)$ is an MLP, Φ_{att} is the learnable parameter, $\text{pad}(\cdot)$ is the operation of filling empty nodes into the position of filtered nodes according to the indices of selected nodes. In this way, the padded subcomplex graph has the same number of nodes as the complex graph. The idx^l is indices of the top $\lceil rN \rceil$ nodes which are selected based on confidence scores $\mathbf{Z}^l$, and $r \in (0, 1]$ is the selection ratio that determines the number of nodes to keep:

$$\text{idx}^l = \text{top-rank}(\mathbf{Z}^l, \lceil rN \rceil) \tag{7}$$

where the top-rank is the function that returns indices of top $\lceil rN \rceil$ values. In practice, we set $r = 0.5$. Then, the hidden state matrix $\tilde{\mathbf{H}}_S^l$ and the position matrix $\tilde{\mathbf{X}}_S^l$ of subcomplex are obtained:

$$\tilde{\mathbf{H}}_S^l = \mathbf{H}_{C,\text{idx}^l,:}^l \odot \mathbf{Z}_{\text{idx}^l}^l, \tag{8}$$

$$\tilde{\mathbf{X}}_S^l = \mathbf{X}_{C,\text{idx}^l,:}^l \tag{9}$$

where $\cdot_{\text{idx}^l}$ is an indexing operation, $\odot$ is the broadcasted elementwise product, $\mathbf{H}_{C,\text{idx}^l,:}^l$ and $\mathbf{X}_{C,\text{idx}^l,:}^l$ are the row-wise (i.e. node-wise) indexed matrix $\mathbf{H}_C^l$ and $\mathbf{X}_C^l$, respectively. Next, we process the selected subcomplex to better leverage binding context for molecule generation.

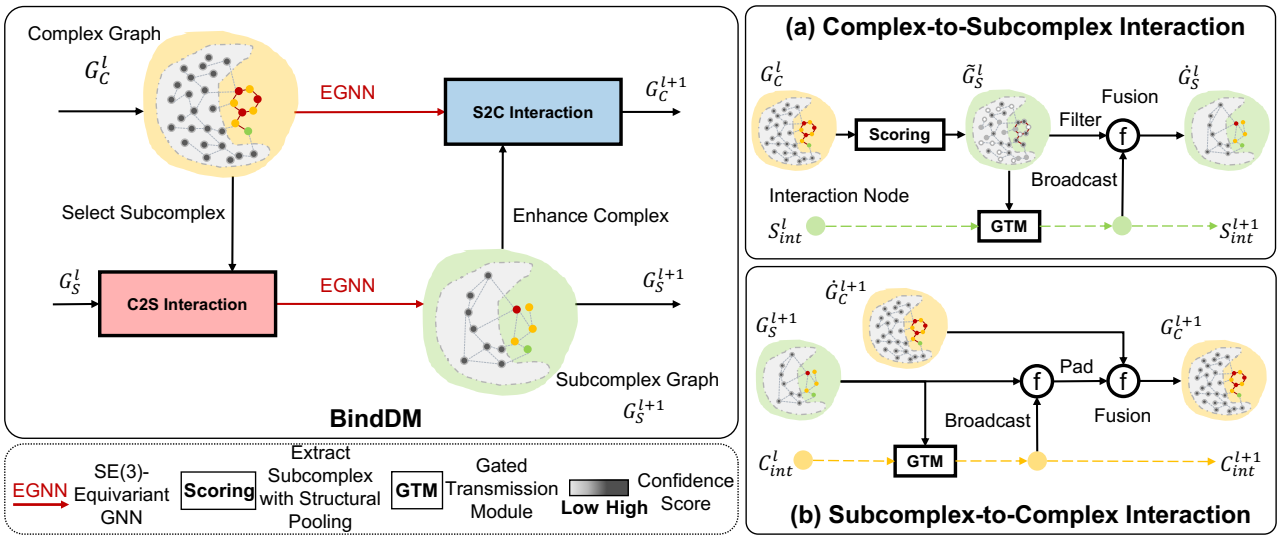

Figure 2: The overview of BINDDM.

3D Equivariant Complex-Subcomplex Processing

Our goal is to generate 3D molecules based on target protein binding sites, the model needs to generate both continuous atom coordinates and discrete atom types, while being SE(3)-equivariant to global translation and rotation during the entire generative process. This property is a critical inductive bias for generating 3D molecules (Hoogeboom et al. 2022; Schneuing et al. 2023; Guan et al. 2023a), and an invariant distribution composed with an equivariant transition function will result in an invariant distribution. Thus, for our hierarchical complex-subcomplex denoising network, we have the following proposition in the setting of protein-aware molecule generation.

Proposition 1. *Denoting SE(3)-transformation as T, we can achieve invariant likelihood w.r.t T on both the protein-ligand complex and its subcomplex: $p_\theta(T\mathbf{M}_0|T\mathbf{P}) = p_\theta(\mathbf{M}_0|\mathbf{P})$ if we shift the Center of Mass (CoM) of protein atoms to zero and parameterize the Markov transition $p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_P)$ with a SE(3)-equivariant network.*

We apply two SE(3)-equivariant neural networks on the k-nn graphs ($\mathcal{G}_C^l$ and $\dot{\mathcal{G}}_S^l$) of the protein-ligand complex and its corresponding subcomplex in the denoising process, respectively. For the subcomplex graph $\dot{\mathcal{G}}_S^l$ updated through the complex-to-subcomplex (C2S) interaction, the SE(3)-invariant hidden states $\dot{\mathbf{H}}_S^l$ and SE(3)-equivariant positions $\dot{\mathbf{X}}_S^l$ are updated as follows to obtain the $\mathcal{G}_S^{l+1}$:

$$\begin{aligned}
\mathbf{h}_{S,i}^{l+1} =& \dot{\mathbf{h}}_{S,i}^l + \sum_{j\in\mathcal{N}_i} f_{S,h}^l\left(\left\|\dot{\mathbf{x}}_{S,i}^l - \dot{\mathbf{x}}_{S,j}^l\right\|, \dot{\mathbf{h}}_{S,i}^l, \dot{\mathbf{h}}_{S,j}^l, \mathbf{e}_{S,ij}\right) \\
\mathbf{x}_{S,i}^{l+1} =& \dot{\mathbf{x}}_{S,i}^l + \sum_{j\in\mathcal{N}_i} \left(\dot{\mathbf{x}}_{S,i}^l - \dot{\mathbf{x}}_{S,j}^l\right) \cdot \\
& f_{S,x}^l\left(\left\|\dot{\mathbf{x}}_{S,i}^l - \dot{\mathbf{x}}_{S,j}^l\right\|, \mathbf{h}_{S,i}^{l+1}, \mathbf{h}_{S,j}^{l+1}, \mathbf{e}_{S,ij}\right) \cdot \mathbb{1}_{\text{mol}}
\end{aligned} \tag{10}$$

where $\mathcal{N}_i$ is the set of k-nearest neighbors of atom i on the subcomplex graph, $\mathbf{e}_{S,ij}$ indicates the atom i and atom j are both protein atoms or both ligand atoms or one protein atom and one ligand atom, and $\mathbb{1}_{\text{mol}}$ is the ligand atom mask since the protein atom coordinates are known and thus supposed to remain unchanged during this update. The similar process are applied on the complex graph $\mathcal{G}_C^l = (\mathbf{H}_C^l, \mathbf{X}_C^l)$ to obtained the $\dot{\mathcal{G}}_C^{l+1} = (\dot{\mathbf{H}}_C^{l+1}, \dot{\mathbf{X}}_C^{l+1})$.

Iterative Cross-Hierarchy Interaction

In BINDDM, we introduce two *cross-hierarchy interaction nodes* to facilitate the information exchange between the binding contexts of two hierarchies, the complex graph and its subcomplex graph. Specifically, we initialize the interaction node of subcomplex graph via sum pooling:

$$\mathbf{c}_{int}^l = \text{Pool}_{\text{sum}}(\mathcal{G}_S^l). \tag{11}$$

Then the binding context of the extracted subcomplex graph is transmitted back to the complex graph $\mathcal{G}_C^l$ for cross-hierarchy information fusion via the interaction node $\mathbf{c}_{int}^l$ and the gated transmission module as shown in Figure 3:

$$\hat{\mathbf{c}}_{int}^l = \text{Attn}(\text{Query}(\mathbf{c}_{int}^l), \text{Key}(\mathbf{H}_S^l)) \cdot \text{Value}(\mathbf{H}_S^l) \tag{12}$$

$$\alpha_{\mathbf{c}}^l = \sigma(f_1(\mathbf{c}_{int}^l) + f_2(\hat{\mathbf{c}}_{int}^l)) \tag{13}$$

$$\mathbf{c}_{int}^l = \text{GRU}(\mathbf{c}_{int}^{l-1}, \alpha_{\mathbf{c}}^l \cdot \mathbf{c}_{int}^l + (1-\alpha_{\mathbf{c}}^l) \cdot \hat{\mathbf{c}}_{int}^l) \tag{14}$$

$$\mathbf{h}_{C,i}^l = f_3(\dot{\mathbf{h}}_{C,i}^l, \mathbf{c}_{int}^l) \tag{15}$$

where σ is sigmoid, f_1, f_2 and f_3 are MLPs, $\dot{\mathbf{h}}_{C,i}^l$ is the SE(3)-invariant hidden state of i-th node in $\dot{\mathcal{G}}_C^l$. Equations (12) to (14) mix the messages between the subcomplex graph and the interaction node through gated recurrent unit (GRU) (Chung et al. 2014) for updating $\mathbf{c}_{int}^l$. Equation (15) perform node-wise fusion with $\mathbf{c}_{int}^l$ for subcomplex-to-complex (S2C) interaction. Similarly, we also have the interaction node $\mathbf{s}_{int}^l$ for complex-to-subcomplex (C2S) interaction, and we iterate these cross-hierarchy processes for sufficiently incorporating binding-adaptive subcomplex into the 3D molecule generation process as shown in Figure 2.

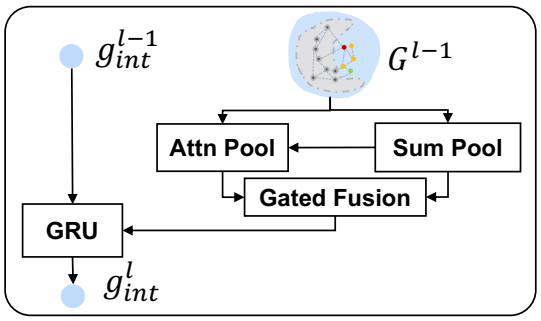

Figure 3: The overview of the gated transmission module.

Training and Sampling

To train BINDDM (*i.e.*, optimize the evidence lower bound induced by BINDDM), we use the same objective function as Guan et al. (2023a). The atom position loss and atom type loss at time step $t-1$ are defined as follows respectively:

$$\begin{aligned}\mathbf{L}_{t-1}^{(x)} &= \frac{1}{2\tilde{\beta}_t^2}\sum_{i=1}^{N_M}\|\tilde{\boldsymbol{\mu}}(\mathbf{x}_{i,t},\mathbf{x}_{i,0})-\tilde{\boldsymbol{\mu}}(\mathbf{x}_{i,t},\hat{\mathbf{x}}_{i,0})\|^2 \\ &= \gamma_t\sum_{i=1}^{N_M}\|\mathbf{x}_{i,0}-\hat{\mathbf{x}}_{i,0}\|\end{aligned} \tag{16}$$

$$\mathbf{L}_{t-1}^{(v)} = \sum_{i=1}^{N_M}\sum_{k=1}^{K}\tilde{\boldsymbol{c}}(\mathbf{v}_{i,t},\mathbf{v}_{i,0})_k\log\frac{\tilde{\boldsymbol{c}}(\mathbf{v}_{i,t},\mathbf{v}_{i,0})_k}{\tilde{\boldsymbol{c}}(\mathbf{v}_{i,t},\hat{\mathbf{v}}_{i,0})_k} \tag{17}$$

where $\hat{\mathbf{X}}_0$ and $\hat{\mathbf{V}}_0$ are predicted from $\mathbf{X}_t$ and $\mathbf{V}_t$, and $\gamma_t = \frac{\bar{\alpha}_{t-1}\beta_t^2}{2\tilde{\beta}_t^2(1-\bar{\alpha}_t)^2}$. Kindly recall that $\mathbf{x}_{i,t}$, $\mathbf{v}_{i,t}$, $\hat{\mathbf{x}}_{i,0}$, and $\hat{\mathbf{v}}_{i,0}$ correspond to the i-th row of $\mathbf{X}_t$, $\mathbf{V}_t$, $\hat{\mathbf{X}}_0$, and $\hat{\mathbf{V}}_0$, respectively. The final loss is a weighted sum of atom coordinate loss and atom type loss with a hyperparameter λ as: $\mathbf{L} = \mathbf{L}_{t-1}^{(x)} + \lambda\mathbf{L}_{t-1}^{(v)}$.

Experiments

Experimental Settings

Datasets and Baseline Methods As for molecular generation, following the previous work (Luo et al. 2021; Peng et al. 2022; Guan et al. 2023a), we train and evaluate BINDDM on the CrossDocked2020 dataset (Francoeur et al. 2020). We follow the same data preparation and splitting as Luo et al. (2021), where the 22.5 million docked binding complexes are refined to high-quality docking poses (RMSD between the docked pose and the ground truth $<$ 1Å) and diverse proteins (sequence identity $<$ 30%). This produces 100,000 protein-ligand pairs for training and 100 proteins for testing. We compare our model with four recent representative methods for SBDD. **LiGAN** (Ragoza, Masuda, and Koes 2022a) is a conditional VAE model trained on an atomic density grid representation of protein-ligand structures. **AR** (Luo et al. 2021) and **Pocket2Mol** (Peng et al. 2022) are autoregressive schemes that generate 3D molecules atoms conditioned on the protein pocket and previous generated atoms. **TargetDiff** (Guan et al. 2023a) and **DecompDiff** (Guan et al. 2023b) are recent state-of-the-art diffusion methods which generate atom coordinates and atom types in a non-autoregressive way.

Evaluation We comprehensively evaluate the generated molecules from three perspectives: **molecular structures**, **target binding affinity**, and **molecular properties**. In terms of **molecular structures**, we calculate the Jensen-Shannon divergences (JSD) in empirical distributions of atom/bond distances between the generated molecules and the reference ones. To estimate the **target binding affinity**, following previous work (Luo et al. 2021; Ragoza, Masuda, and Koes 2022b; Guan et al. 2023a), we adopt AutoDock Vina (Eberhardt et al. 2021) to compute and report the mean and median of binding-related metrics, including *Vina Score*, *Vina Min*, *Vina Dock* and *High Affinity*. Vina Score directly estimates the binding affinity based on the generated 3D molecules; Vina Min performs a local structure minimization before estimation; Vina Dock involves an additional re-docking process and reflects the best possible binding affinity; High affinity measures the ratio of how many generated molecules binds better than the reference molecule per test protein. To evaluate **molecular properties**, we utilize *QED*, *SA*, *Diversity* as metrics following Luo et al. (2021); Ragoza, Masuda, and Koes (2022a). QED is a quantitative estimation of drug-likeness combining several desirable molecular properties; SA (synthesize accessibility) is a measure estimation of the difficulty of synthesizing the ligands; Diversity is computed as average pairwise dissimilarity between all generated ligands. All sampling and evaluation procedures follow Guan et al. (2023a) for fair comparison.

Main Results

Generated 3D Molecular Structures We compare our BINDDM and the representative methods in terms of molecular structures. We compute different bond distributions of the generated molecules and compare them against the corresponding reference empirical distributions in Tab. 1. Our model has a comparable performance with TargetDiff and DecompDiff and substantially outperforms all other baselines across all major bond types, indicating the great potential of BINDDM for generating stable molecular structures.

Bond	liGAN	AR	Pocket2 Mol	Target Diff	Decomp Diff	Ours
C−C	0.601	0.609	0.496	0.369	**0.359**	0.380
C=C	0.665	0.620	0.561	0.505	0.537	**0.229**
C−N	0.634	0.474	0.416	0.363	0.344	**0.265**
C=N	0.749	0.635	0.629	0.550	0.584	**0.245**
C−O	0.656	0.492	0.454	0.421	0.376	**0.329**
C=O	0.661	0.558	0.516	0.461	0.374	**0.249**
C:C	0.497	0.451	0.416	0.263	**0.251**	0.282
C:N	0.638	0.552	0.487	0.235	0.269	**0.130**

Table 1: Jensen-Shannon divergence between bond distance distributions of reference and generated molecules, lower values indicate better performances. "-", "=", and ":" represent single, double, and aromatic bonds, respectively.

Methods	Vina Score (↓)		Vina Min (↓)		Vina Dock (↓)		High Affinity (↑)		QED (↑)		SA (↑)		Diversity (↑)	
	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.
Reference	-6.36	-6.46	-6.71	-6.49	-7.45	-7.26	-	-	0.48	0.47	0.73	0.74	-	-
LiGAN	-	-	-	-	-6.33	-6.20	21.1%	11.1%	0.39	0.39	0.59	0.57	0.66	0.67
GraphBP	-	-	-	-	-4.80	-4.70	14.2%	6.7%	0.43	0.45	0.49	0.48	**0.79**	**0.78**
AR	<u>-5.75</u>	-5.64	-6.18	-5.88	-6.75	-6.62	37.9%	31.0%	<u>0.51</u>	0.50	<u>0.63</u>	<u>0.63</u>	0.70	0.70
Pocket2Mol	-5.14	-4.70	-6.42	-5.82	-7.15	-6.79	48.4%	51.0%	**0.56**	**0.57**	**0.74**	**0.75**	0.69	0.71
TargetDiff	-5.47	<u>-6.30</u>	-6.64	-6.83	-7.80	-7.91	58.1%	59.1%	0.48	0.48	0.58	0.58	0.72	0.71
DecompDiff	-5.67	-6.04	<u>-7.04</u>	<u>-6.91</u>	<u>-8.39</u>	**-8.43**	<u>64.4%</u>	<u>71.0%</u>	0.45	0.43	0.61	0.60	0.68	0.68
BINDDM	**-5.92**	**-6.81**	**-7.29**	**-7.34**	**-8.41**	<u>-8.37</u>	**64.8%**	**71.6%**	<u>0.51</u>	<u>0.52</u>	0.58	0.58	<u>0.75</u>	<u>0.74</u>

Table 2: Summary of different properties of reference molecules and molecules generated by our model and other baselines. (↑) / (↓) denotes a larger / smaller number is better. Top 2 results are highlighted with bold text and underlined text, respectively.

Target Binding Affinity and Molecule Properties We evaluate the effectiveness of BINDDM in terms of binding affinity. We can see in Tab. 2 that our BINDDM outperforms baselines in binding-related metrics. Specifically, BINDDM surpasses strong autoregressive method Pocket2Mol by a large margin of **15.2%**, **13.6%** and **17.6%** in Avg. Vina Score, Vina Min and Vina Dock, respectively. Compared with the state-of-the-art diffusion-based method DecompDiff, BINDDM not only increased the binding-related metrics Avg. Vina Score and Vina Min by **4.4%** and **3.6%**, respectively, but also significantly increased the property-related metric Avg. QED by **13.3%**. In terms of high-affinity binder, we find that on average **64.8%** of the BINDDM molecules show better binding affinity than the reference molecule, which is significantly better than other baselines. These gains demonstrate that the proposed BINDDM effectively captures significant binding-related subcomplex to enable generating molecules with improved target binding affinity. Moreover, we can see a trade-off between property-related metrics QED and binding-related metrics in previous methods. DecompDiff performs better than AR and Pocket2Mol in binding-related metrics, but falls behind them in QED scores. In contrast, our BINDDM not only achieves state-of-the-art binding-related scores but also maintains proper QED scores, achieving a better trade-off than DecompDiff. Nevertheless, we put less emphasis on QED and SA because they are often applied as rough screening metrics in real drug discovery scenarios, and it would be fine as long as they are within a reasonable range. Figure 4 shows some examples of generated ligand molecules and their properties. Molecules generated by our model have valid structures and reasonable binding poses to the target.

Model Analysis

Effect of the Iterative Cross-Hierarchy Interaction on Target-specific Molecule Generation We conduct a set of ablation experiments to study the effect of iterative cross-hierarchy interaction on the generation ability of diffusion models for the target-specific molecules: (1) **Exp0**: the baseline model without applying the iterative cross-hierarchy interaction, (2) **Exp1**: we replace the binding-adaptive subcomplex extraction (BASE) module with a random selection of atoms from the complex for constructing the subcomplex. The selection ratio is set to 0.5, (3) **Exp2**: we remove subcomplex graphs in iterative cross-hierarchy interaction and keep interaction nodes $\mathbf{c}_{int}$ unchanged for information propagation between cross-layer complex graphs, (4) **Exp3**: we remove interaction nodes $\mathbf{c}_{int}$ and $\mathbf{s}_{int}$ and keep the extracted subcomplex graphs unchanged, (5) **Exp4**: we remove the gated transmission module in the update of interaction nodes $\mathbf{c}_{int}$ and $\mathbf{s}_{int}$. The results are present in Tab. 3.

In the comparison between Exp0 and Exp1, we can find that randomly selected subcomplex can not provide useful information about pocket-ligand binding. And the comparison between Exp1 and BINDDM suggests that BASE is more effective than random selection in exploring binding-related clues from the complex. The effectiveness of BASE is beneficial for BINDDM in generating molecules that are tightly bound to the given protein pocket. In comparing Exp2 with BINDDM, it is evident that solely relying on global interaction nodes for information propagation between cross-layer complex graphs does not provide significant binding-related information for pocket-specific molecular generation. In comparing Exp3 with BINDDM, we observe that the utilization of global interaction nodes for information exchange between complex and subcomplex not only improves the performance of BINDDM in binding-related metrics but also contributes to the molecular property-related ones. And the same conclusion is also observed in the comparison between Exp4 and BINDDM.

Influence of Extracting Binding Clues from Complex, Pocket and Ligand Since the presence of binding clues in both the molecular ligands and protein pockets, we conduct three experiments to explore the effects of extracting binding-related clues from different structures on how tightly the generated molecules bind to the specific protein pockets: the binding-related substructures are extracted from (1) the molecular ligands, (2) the protein pockets, and (3) the complexes (treating the molecule and pocket as a unified entity) to enhance the generation of molecular ligands binding tightly to specific protein pockets, respectively. As present in Tab. 4, BINDDM can achieve the best performance on binding-related metrics when binding-related substructures are extracted from complexes and used to enhance the generation process of protein-specific molecular ligands.

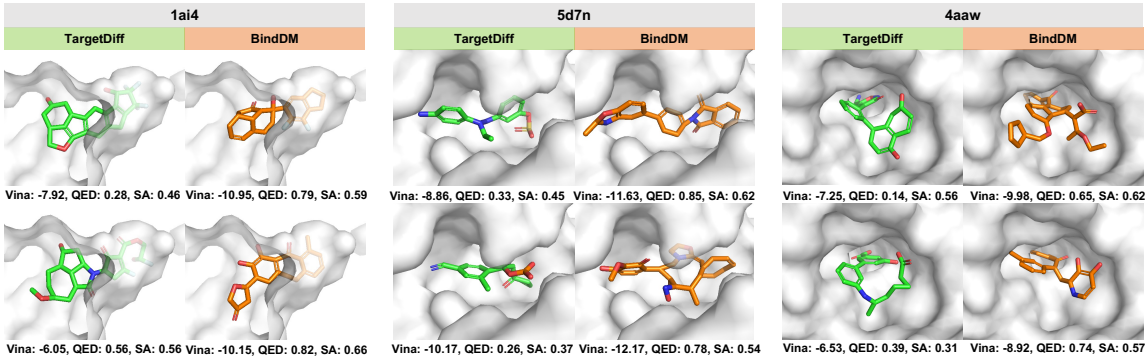

Figure 4: The generated ligand molecules of TargetDiff (Guan et al. 2023a) and BINDDM for the given protein pockets. Carbon atoms in ligands generated by TargetDiff and BINDDM are visualized in green and orange, respectively.

Methods	Vina Score (↓)		Vina Min (↓)		QED (↑)	
	Avg.	Med.	Avg.	Med.	Avg.	Med.
Exp0	-5.04	-5.75	-6.38	-6.52	0.46	0.46
Exp1	-4.79	-5.92	-6.36	-6.66	0.50	0.51
Exp2	-5.65	-6.25	-6.64	-6.65	0.45	0.45
Exp3	-5.62	-6.74	-6.83	-6.92	0.47	0.46
Exp4	-5.60	-6.28	-6.78	-6.83	0.47	0.47
BINDDM	**-5.92**	**-6.81**	**-7.29**	**-7.34**	**0.51**	**0.52**

Table 3: Effect of the iterative cross-hierarchy interaction on target-specific molecule generation.

Methods	Vina Score (↓)		Vina Min (↓)		QED (↑)	
	Avg.	Med.	Avg.	Med.	Avg.	Med.
baseline	-5.04	-5.75	-6.38	-6.52	0.46	0.46
Pocket	-5.37	**-6.84**	-7.03	**-7.38**	**0.51**	**0.52**
Ligand	-5.46	-6.77	-6.98	-7.27	**0.51**	**0.52**
Complex	**-5.92**	-6.81	**-7.29**	-7.34	**0.51**	**0.52**

Table 4: Influence of extracting binding clues from complex, pocket and ligand.

Correlation between Adaptively Extracted Subcomplex and Pocket-Ligand Binding Clues To validate the presence of binding-related clues in the subcomplex extracted by BASE in BINDDM and their suitability for generating molecular ligands tightly bound to specific protein pockets, we initially employed the pre-trained binding affinity prediction model BAPNet (Li, Pei, and Lai 2021). By predicting binding affinity of the complex which consists of the given protein pocket and the molecules generated by BINDDM, the binding-related subcomplex is obtained by ranking the complex atoms according to contributions to the binding affinity prediction. We take the subcomplexes predicted by BAPNet as the reference subcomplexes, and calculate the accuracy of the subcomplexes predicted by BASE as a metric to assess the effectiveness of BASE in extracting binding-related subcomplex. Considering that the sampling process consists of 1000 steps, we calculate the average accuracy

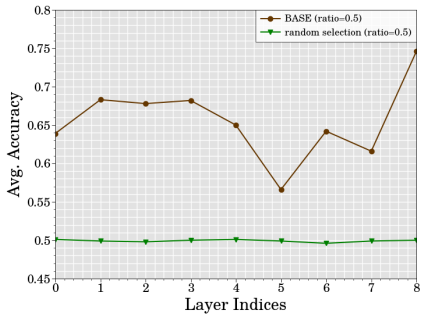

Figure 5: The subcomplex prediction accuracy of BASE in each layer of the de-nosing network.

by comparing the reference subcomplex with all the subcomplexes extracted by each layer of the denoising network in BINDDM (9 layers in total) throughout the entire sampling process. As shown in Figure 5, each layer of the denoising network of BINDDM achieves an accuracy rate of around 0.65 for subcomplex extracted through BASE, which is higher than the accuracy rate of around 0.5 for random selection. This suggests that BASE in BINDDM is capable of extracting binding-related subcomplex to a certain degree. In addition, we replace the process of using BASE to adaptively select binding-related subcomplex with randomly selecting atoms from the complex to construct a substructure. The performance present in Tab. 3 (Exp1) demonstrates subcomplexes extracted from BASE benefit the final performance.

Conclusion

In this paper, we propose an effective diffusion model BINDDM to adaptively extract the essential part of binding sites responsible for protein-ligand interactions, *subcomplex*, for enhancing protein-aware 3D molecule generation. The *cross-hierarchy interaction node* is designed to exchange the hierarchical information between complex and subcomplex. Empirical results demonstrate that BINDDM can generate realistic 3D molecules with high binding affinities and proper molecular properties toward protein targets.

Acknowledgments

The work was partly supported by the National Natural Science Foundation of China (No.62171251), the Major Key Project of GZL under Grant SRPG22-001 and the Major Key Project of PCL under Grant PCL2023A09.

References

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

Atz, K.; Grisoni, F.; and Schneider, G. 2021. Geometric deep learning on molecular representations. *Nature Machine Intelligence*, 3(12): 1023–1032.

Bajusz, D.; Wade, W. S.; Satała, G.; Bojarski, A. J.; Ilaš, J.; Ebner, J.; Grebien, F.; Papp, H.; Jakab, F.; Douangamath, A.; et al. 2021. Exploring protein hotspots by optimized fragment pharmacophores. *Nature Communications*, 12(1): 3201.

Basanta, B.; Bick, M. J.; Bera, A. K.; Norn, C.; Chow, C. M.; Carter, L. P.; Goreshnik, I.; Dimaio, F.; and Baker, D. 2020. An enumerative algorithm for de novo design of proteins with diverse pocket structures. *Proceedings of the National Academy of Sciences*, 117(36): 22135–22145.

Batool, M.; Ahmad, B.; and Choi, S. 2019. A structure-based drug discovery paradigm. *International journal of molecular sciences*, 20(11): 2783.

Bronstein, M. M.; Bruna, J.; Cohen, T.; and Veličković, P. 2021. Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges. arXiv:2104.13478.

Chung, J.; Gulcehre, C.; Cho, K.; and Bengio, Y. 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. arXiv:1412.3555.

Du, Y.; Fu, T.; Sun, J.; and Liu, S. 2022. MolGenSurvey: A Systematic Survey in Machine Learning Models for Molecule Design. arXiv:2203.14500.

Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; and Forli, S. 2021. AutoDock Vina 1.2. 0: New docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling*, 61(8): 3891–3898.

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.

Gaudelet, T.; Day, B.; Jamasb, A. R.; Soman, J.; Regep, C.; Liu, G.; Hayter, J. B.; Vickers, R.; Roberts, C.; Tang, J.; et al. 2021. Utilizing graph machine learning within drug discovery and development. *Briefings in bioinformatics*, 22(6): bbab159.

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 *International Conference on Learning Representations*.

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 Krause, A.; Brunskill, E.; Cho, K.; Engelhardt, B.; Sabato, S.; and Scarlett, J., eds., *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, 11827–11846. PMLR.

Hawkins, P. C. 2017. Conformation generation: the state of the art. *Journal of Chemical Information and Modeling*, 57(8): 1747–1756.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33: 6840–6851.

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

Kinnings, S. L.; Liu, N.; Tonge, P. J.; Jackson, R. M.; Xie, L.; and Bourne, P. E. 2011. A machine learning-based method to improve docking scoring functions and its application to drug repurposing. *Journal of chemical information and modeling*, 51(2): 408–419.

Kozakov, D.; Hall, D. R.; Jehle, S.; Luo, L.; Ochiana, S. O.; Jones, E. V.; Pollastri, M.; Allen, K. N.; Whitty, A.; and Vajda, S. 2015. Ligand deconstruction: Why some fragment binding positions are conserved and others are not. *Proceedings of the National Academy of Sciences*, 112(20): E2585–E2594.

Li, Y.; Pei, J.; and Lai, L. 2021. Structure-based de novo drug design using 3D deep generative models. *Chemical science*, 12(41): 13664–13675.

Lin, H.; Huang, Y.; Liu, M.; Li, X.; Ji, S.; and Li, S. Z. 2022. DiffBP: Generative Diffusion of 3D Molecules for Target Protein Binding. arXiv:2211.11214.

Liu, M.; Luo, Y.; Uchino, K.; Maruhashi, K.; and Ji, S. 2022a. Generating 3D Molecules for Target Protein Binding. In *International Conference on Machine Learning*.

Liu, M.; Luo, Y.; Uchino, K.; Maruhashi, K.; and Ji, S. 2022b. Generating 3D Molecules for Target Protein Binding. arXiv:2204.09410.

Luo, S.; Guan, J.; Ma, J.; and Peng, J. 2021. A 3D generative model for structure-based drug design. *Advances in Neural Information Processing Systems*, 34: 6229–6239.

Luo, Y.; and Ji, S. 2021. An autoregressive flow model for 3d molecular geometry generation from scratch. In *International Conference on Learning Representations*.

Peng, X.; Luo, S.; Guan, J.; Xie, Q.; Peng, J.; and Ma, J. 2022. Pocket2mol: Efficient molecular sampling based on 3d protein pockets. In *International Conference on Machine Learning*, 17644–17655. PMLR.

Powers, A. S.; Yu, H. H.; Suriana, P.; and Dror, R. O. 2022. Fragment-based ligand generation guided by geometric deep learning on protein-ligand structure. *bioRxiv*, 2022–03.

Ragoza, M.; Masuda, T.; and Koes, D. R. 2022a. Generating 3D molecules conditional on receptor binding sites with deep generative models. *Chem Sci*, 13: 2701–2713.

Ragoza, M.; Masuda, T.; and Koes, D. R. 2022b. Generating 3D molecules conditional on receptor binding sites with

deep generative models. *Chemical science*, 13(9): 2701–2713.

Satorras, V. G.; Hoogeboom, E.; and Welling, M. 2021. E (n) equivariant graph neural networks. In *International conference on machine learning*, 9323–9332. PMLR.

Schneider, G.; Neidhart, W.; Giller, T.; and Schmid, G. 1999. "Scaffold-hopping" by topological pharmacophore search: a contribution to virtual screening. *Angewandte Chemie International Edition*, 38(19): 2894–2896.

Schneuing, A.; Du, Y.; Harris, C.; Jamasb, A.; Igashov, I.; Du, W.; Blundell, T.; Lió, P.; Gomes, C.; Welling, M.; Bronstein, M.; and Correia, B. 2023. Structure-based Drug Design with Equivariant Diffusion Models. arXiv:2210.13695.

Skalic, M.; Jiménez, J.; Sabbadin, D.; and De Fabritiis, G. 2019. Shape-based generative modeling for de novo drug design. *Journal of chemical information and modeling*, 59(3): 1205–1214.

Sohl-Dickstein, J.; Weiss, E.; Maheswaranathan, N.; and Ganguli, S. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, 2256–2265. PMLR.

Song, Y.; and Ermon, S. 2019. Generative modeling by estimating gradients of the data distribution. *Advances in Neural Information Processing Systems*, 32.

Song, Y.; Sohl-Dickstein, J.; Kingma, D. P.; Kumar, A.; Ermon, S.; and Poole, B. 2020. Score-Based Generative Modeling through Stochastic Differential Equations. In *International Conference on Learning Representations*.

Spitzer, R.; Cleves, A. E.; and Jain, A. N. 2011. Surface-based protein binding pocket similarity. *Proteins: Structure, Function, and Bioinformatics*, 79(9): 2746–2763.

Tan, C.; Gao, Z.; and Li, S. Z. 2023. Target-aware molecular graph generation. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 410–427. Springer.

Xu, M.; Ran, T.; and Chen, H. 2021. De novo molecule design through the molecular generative model conditioned by 3D information of protein binding sites. *Journal of Chemical Information and Modeling*, 61(7): 3240–3254.

Yang, Y.; Ouyang, S.; Dang, M.; Zheng, M.; Li, L.; and Zhou, H. 2022. Knowledge Guided Geometric Editing for Unsupervised Drug Design.

Zhang, Z.; and Liu, Q. 2023. Learning Subpocket Prototypes for Generalizable Structure-based Drug Design. In *International Conference on Machine Learning*.

Zhang, Z.; Min, Y.; Zheng, S.; and Liu, Q. 2023a. Molecule generation for target protein binding with structural motifs. In *The Eleventh International Conference on Learning Representations*.

Zhang, Z.; Yan, J.; Liu, Q.; Chen, E.; and Zitnik, M. 2023b. A Systematic Survey in Geometric Deep Learning for Structure-based Drug Design. arXiv:2306.11768.