Apo2Mol: 3D Molecule Generation via Dynamic Pocket-Aware Diffusion Models

Xinzhe Zheng[1], **Shiyu Jiang**[1], **Gustavo Seabra**[1], **Chenglong Li**[1], **Yanjun Li**[1,2,*]

[1]Department of Medicinal Chemistry, Center for Natural Products, Drug Discovery and Development, University of Florida, Gainesville, FL, USA
[2]Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, USA

Abstract

Deep generative models are rapidly advancing structure-based drug design, offering substantial promise for generating small molecule ligands that bind to specific protein targets. However, most current approaches assume a rigid protein binding pocket, neglecting the intrinsic flexibility of proteins and the conformational rearrangements induced by ligand binding, limiting their applicability in practical drug discovery. Here, we propose Apo2Mol, a diffusion-based generative framework for 3D molecule design that explicitly accounts for conformational flexibility in protein binding pockets. To support this, we curate a dataset of over 24,000 experimentally resolved apo-holo structure pairs from the Protein Data Bank, enabling the characterization of protein structure changes associated with ligand binding. Apo2Mol employs a full-atom hierarchical graph-based diffusion model that simultaneously generates 3D ligand molecules and their corresponding holo pocket conformations from input apo states. Empirical studies demonstrate that Apo2Mol can achieve state-of-the-art performance in generating high-affinity ligands and accurately capture realistic protein pocket conformational changes.

Code — https://github.com/AIDD-LiLab/Apo2Mol
Extended version — https://arxiv.org/abs/2511.14559

1 Introduction

Structure-based drug design (SBDD) leverages the three-dimensional structures of protein targets to guide the rational therapeutic design (Anderson 2003). By capturing the geometry and physicochemical features of binding sites, SBDD enables the generation of ligands with high affinity and specificity. The advances of deep learning have further catalyzed the SBDD field, with deep generative models showing remarkable success in generating ligands tailored to specific pockets (Zhang et al. 2025). Early studies primarily operated on 1D string-based or 2D graph-based molecular representations, while subsequent approaches advanced to directly generating 3D ligand conformations (Bilodeau et al. 2022). More recently, diffusion models have emerged as a powerful framework, iteratively denoising atoms within the pocket to yield novel ligands together with their 3D poses (Guan et al. 2023b; Schneuing et al. 2024; Li et al. 2025; Lin et al. 2025).

*Corresponding author: yanjun.li@ufl.edu

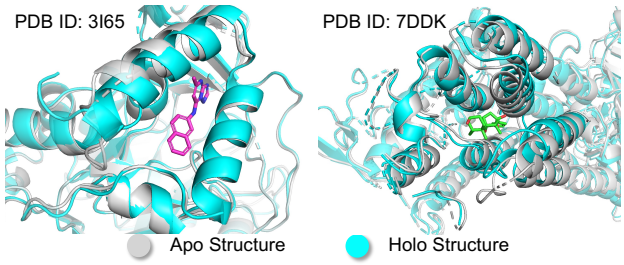

Figure 1: Conformational changes between the apo (unbound) and holo (ligand-bound) states, illustrating ligand-induced structural dynamics.

Despite these advances, a fundamental limitation persists in most current approaches to SBDD: the assumption of a static, rigid protein binding pocket (Guan et al. 2023b; Schneuing et al. 2024; Li et al. 2025). Modeling the receptor pocket as rigid neglects the dynamic nature of molecular recognition (Fig. 1), biasing design toward a single holo (ligand-bound) conformation and potentially missing ligands that would induce alternative, more favorable pocket geometries (Wei and McCammon 2024). This limitation becomes especially acute for emerging targets or those for which only apo (unbound) conformations are available, as current SBDD methods often struggle to generate high-quality ligands in the absence of an appropriate bound-state template. A recent work DynamicFlow (Zhou et al. 2025) relaxes the rigidity assumption by introducing a full-atom stochastic flow-matching model trained on molecular dynamics (MD) simulation trajectories from the MISATO dataset (Siebenmorgen et al. 2024), enabling joint generation of the holo pocket conformation and the bound ligand from an initial apo structure. While promising, MD simulations are computationally expensive and sensitive to force field parameters and timescales (Hollingsworth and Dror 2018). Moreover, simulated dynamics may fail to capture experimental conformational transitions, potentially introducing simulation-specific artifacts into generative models (Hollingsworth and Dror 2018).

To overcome the limitations of rigid-pocket assumptions and the dependency on potentially biased simulation data, we introduce Apo2Mol, a diffusion-based full-atom framework for SBDD that explicitly considers the dynamics of

the protein binding pocket (see Fig. 2). We address two central challenges in tandem: (1) grounding training on a large-scale dataset derived exclusively from experimentally resolved structures, and (2) modeling 3D ligand generation with protein structural dynamics. Instead of relying on MD simulations, we curate a high-quality dataset of experimentally resolved apo-holo protein structure pairs stored in the Protein Data Bank (PDB) by leveraging the PLINDER resource (Durairaj et al. 2024). Rigorous filtering criteria (Apx. A) is applied to retain pharmacologically relevant ligands, high-resolution structures (≤ 2.5Å), and apo-holo pairs with 100% sequence identity. This yields a final dataset comprising $24,601$ apo-holo pairs. Compared with ApoBind (Aggarwal, Gupta, and Priyakumar 2021), which contains $\sim 10K$ samples with $\geq 80\%$ apo-holo sequence identity, our dataset is substantially larger and better aligned with the requirements of this study. Building upon our curated dataset, Apo2Mol employs a diffusion-based framework in which the forward diffusion gradually perturbs the ligand coordinates and atom types, and simultaneously interpolates protein pocket coordinates from holo towards apo conformations. During reverse diffusion, the apo pocket serves as the initial structural condition, the model learns to denoise a randomly initialized ligand while concurrently transforming the pocket conformation from apo to its holo state. To better capture ligand–pocket interactions, Apo2Mol integrates SE(3)-equivariant attention layers within a hierarchical graph-based message-passing framework that aggregates atom-level information into residue-level representations, enabling more expressive and spatially aware modeling. Our experimental results demonstrate that Apo2Mol generates high-affinity ligands and produces reasonable pocket conformational changes, advancing the capabilities of dynamic SBDD beyond static structure assumptions.

2 Related Work

2.1 Structure-based Drug Design

Structure-based drug design leverages the three-dimensional structures of protein targets to rationally design small molecules with desired binding affinities. Early approaches generated 1D strings, e.g. SMILES (Weininger 1988), or 2D molecular graphs conditioned on protein contexts (Skalic et al. 2019; Atz et al. 2024; Tan, Gao, and Li 2023; Wang et al. 2025). Subsequent methods employed variational auto encoders (VAEs) and autoregressive models to directly produce 3D ligand conformations aligned with receptor binding sites (Liu et al. 2022; Peng et al. 2022; Ragoza, Masuda, and Koes 2022; Zhang et al. 2023b). Recent diffusion-based approaches further advanced SBDD by iteratively denoising atom types and positions within the protein pocket, followed by post-processing bond formation (Guan et al. 2023a; Lin et al. 2025; Schneuing et al. 2024). DecompDiff (Guan et al. 2023b) introduces the decomposition of biochemical priors into scaffold and arm structures. IPDiff (Huang et al. 2024b) further improves upon this by explicitly incorporating binding-aware biochemical priors. Hierarchical consistency diffusion is adopted to integrate atom- and motif-level structural priors for consistent generation (Li et al. 2025). DynamicFlow (Zhou et al. 2025) proposes integrating protein dynamics into SBDD through training on MD trajectories, using a full-atom stochastic flow model. While existing *de novo* design methods are trained exclusively on static holo structures or simulated dynamics, our Apo2Mol explicitly leverages experimentally validated apo-holo structure pairs to jointly denoise protein pockets and ligands, effectively capturing ligand-binding dynamics.

2.2 Modeling Protein-ligand Interaction Dynamics

Proteins are inherently dynamic, and their functional states often involve conformational changes upon ligand binding. Understanding the dynamic process is crucial for applications like drug discovery and design (Wei and McCammon 2024; Li et al. 2022). Molecular dynamics simulations have traditionally served as the cornerstone for exploring protein conformational landscapes and transition pathways at atomic detail (Alonso, Bliznyuk, and Gready 2006; Dror et al. 2012; Hollingsworth and Dror 2018). Recently, machine learning–especially deep generative models–has shown great promise in modeling dynamics and generating conformational ensembles for protein and small molecule (Janson et al. 2023; Kuznetsov et al. 2024; Lu et al. 2025; Wang et al. 2024), as well as protein-ligand docking (Corso et al. 2025; Lu et al. 2024; Morehead and Cheng 2025; Morehead et al. 2024). While the above modeling approaches have incorporated conformational dynamics, such considerations are largely absent from most *de novo* small molecule generation methods (Grisoni 2023; Tang et al. 2024; Xie et al. 2022). In this work, Apo2Mol jointly generates both the ligand molecules and the corresponding holo pocket conformations from apo states, capturing the inherent dynamics of molecular interaction within a unified generative framework.

3 Preliminary

3.1 Problem Statement

In this work, we address the challenge of jointly generating a holo protein pocket and its corresponding rational ligand from an apo protein conformation. The holo protein pocket is represented as $\mathcal{P}_{\mathrm{H}} = \{\boldsymbol{x}^i_{\mathcal{P}_{\mathrm{H}}}, \boldsymbol{v}^i_{\mathcal{P}_{\mathrm{H}}}\}_{i=1}^{N_{\mathcal{P}}}$, and its corresponding apo conformation and the generated small molecule are denoted as $\mathcal{P}_{\mathrm{A}} = \{\boldsymbol{x}^i_{\mathcal{P}_{\mathrm{A}}}, \boldsymbol{v}^i_{\mathcal{P}_{\mathrm{A}}}\}_{i=1}^{N_{\mathcal{P}}}$ and $\mathcal{M} = \{\boldsymbol{x}^i_{\mathcal{M}}, \boldsymbol{v}^i_{\mathcal{M}}\}$, respectively. Here, $N_{\mathcal{P}}$ and $N_{\mathcal{M}}$ represent the number of atoms in the protein pocket and the ligand. $\boldsymbol{x}^i \in \mathbb{R}^3$ refers to the 3D atom coordinates, while $\boldsymbol{v}^i \in \mathbb{R}^K$ represents the atom features. The learning objective can be formulated as modeling the conditional distribution of $p(\mathcal{P}_{\mathrm{H}}, \mathcal{M}|\mathcal{P}_{\mathrm{A}})$.

For the ligand generation, the model directly predicts the ligand atom types and associated 3D coordinates. In contrast, for holo pocket prediction, we refrain from estimating atomic coordinates to preserve the structural integrity of the protein pocket. Following prior works (Lu et al. 2024; Zhang et al. 2023a), we model pocket conformational changes at the residue level by predicting translations $\boldsymbol{tr} \in \mathbb{R}^3$, rotations $\boldsymbol{q} \in \mathbb{H}$, and chi angle updates $\boldsymbol{\mathcal{X}} = \{\mathcal{X}_i \in \mathrm{SO}(2)\}_{i=0}^{4}$, which correspond to side-chain torsion angles. To represent 3D rotations, we adopt quaternions $\boldsymbol{q}$, which offer superior numerical stability over rotation vectors, eliminate singularities such as

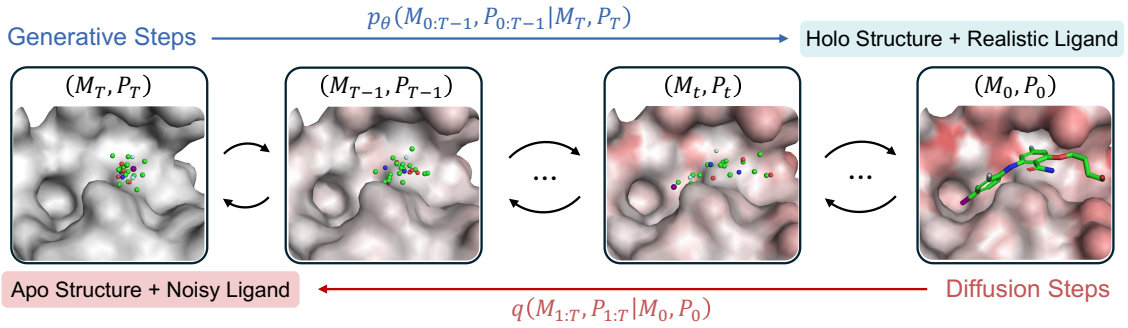

Figure 2: Schematic overview of Apo2Mol. The diffusion process gradually corrupts an experimental holo pocket–ligand pair by injecting noise into the ligand and linearly interpolating the pocket from its holo conformation toward the apo state. The generative process learns to denoise the corrupted inputs and recover the joint distribution of the holo pocket conformations and their corresponding ligand types and poses. Structural deviations between the apo and holo conformations are illustrated in red.

gimbal lock, and enable smooth interpolation (Diebel et al. 2006). This formulation enables physically plausible transformations between apo and holo states.

3.2 Diffusion Models

Diffusion models (Ho, Jain, and Abbeel 2020; Song et al. 2021) are a class of generative models that have achieved remarkable success in multiple domains (Zhu et al. 2023; Epstein et al. 2023; Vignac et al. 2023; Hoogeboom et al. 2022). These models are inspired by stochastic differential equations (SDEs) that transform a complex data distribution p_{data} into a simple prior distribution, typically the Gaussian distribution, through the gradual noise injection:

$$\mathrm{d}\mathbf{x}_t = \boldsymbol{\mu}(\mathbf{x}_t, t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}\mathbf{w}_t, \tag{1}$$

where $t \in [0, T]$, $T > 0$ is a fixed constant, $\mathbf{w}_t$ denotes Brownian motion, and $\boldsymbol{\mu}$, σ are drift and diffusion coefficients, respectively. A notable property of this process is the existence of a deterministic reverse-time ordinary differential equation (ODE), called the probability flow ODE, whose solution follows the same marginal p_t:

$$\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = \boldsymbol{\mu}(\mathbf{x}_t, t) - \frac{1}{2}\sigma(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x}_t), \tag{2}$$

where $\nabla \log p_t(\mathbf{x})$ is the score function. When $\boldsymbol{\mu} = 0$ and $\sigma(t) = \sqrt{2t}$, this becomes a pure Gaussian diffusion, and p_t remains analytically tractable.

To estimate the score term, diffusion models employ a neural network $\boldsymbol{s}_\theta(\mathbf{x}, t) \approx \nabla \log p_t(\mathbf{x})$, trained using score matching to obtain an empirical estimate of the probability flow ODE:

$$\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = -t\boldsymbol{s}_\theta(\mathbf{x}, t). \tag{3}$$

Once trained, samples can be generated by solving Eq. 2 backward in time, typically starting from a Gaussian prior and integrating to $t = \epsilon > 0$ for numerical stability. This yields an approximate sample from the data distribution p_{data}.

4 Method

In this work, we propose Apo2Mol, a framework that captures the dynamic nature of protein pockets for SBDD. As illustrated in Fig. 2, Apo2Mol interpolates the protein pocket between its apo and holo conformations, generating interaction pairs by combining these intermediate pocket states with noisy ligands at each time step. The red regions on the protein pocket highlight conformational deviations from the apo structure. The following section presents a detailed overview of the Apo2Mol framework.

4.1 Data Preparation

Pocket Alignment We first apply root-mean-square deviation (RMSD) alignment to superimpose the apo and holo pocket structures in 3D space. Then, the Kabsch algorithm (Kabsch 1976) is utilized for calculating the residue-level translations $\boldsymbol{tr}$ and rotations $\boldsymbol{q}$, centered on the C_α atoms, that align the holo backbone atoms N-C_α-C to the apo structure: $\boldsymbol{tr}, \boldsymbol{q} = \text{Kabsch}(\boldsymbol{x}^{\text{holo}}_{N\text{-}C_\alpha\text{-}C} - \boldsymbol{x}^{\text{apo}}_{N\text{-}C_\alpha\text{-}C})$. Finally, using the utilities provided in DynamicBind (Lu et al. 2024), we extract the chi angles for both holo and apo structures, and calculate the chi angle update as: $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{X}}^{\text{apo}} - \boldsymbol{\mathcal{X}}^{\text{holo}}$.

Pseudo-interaction Pair Given a paired protein pocket and its corresponding ligand, we construct the pseudo-interaction pair as the model input at each diffusion step t.

For the ligand, we adopt the noise injection scheme from prior works (Huang et al. 2024b,a), where both position and atom-type representations are corrupted progressively. At each time step t, the forward process is defined as:

$$\begin{aligned} q(\boldsymbol{x}_t^{\mathcal{M}} \mid \boldsymbol{x}_0^{\mathcal{M}}) &= \mathcal{N}(\boldsymbol{x}_t^{\mathcal{M}}; \sqrt{\overline{\alpha}_t}\boldsymbol{x}_0^{\mathcal{M}}, (1-\overline{\alpha}_t)\mathbf{I}), \\ q(\boldsymbol{v}_t^{\mathcal{M}} \mid \boldsymbol{v}_0^{\mathcal{M}}) &= \mathcal{C}\left(\boldsymbol{v}_t^{\mathcal{M}} \mid \overline{\alpha}_t \boldsymbol{v}_0^{\mathcal{M}} + (1-\overline{\alpha}_t)/K\right), \end{aligned} \tag{4}$$

where $\mathcal{N}$ and $\mathcal{C}$ denote the Gaussian and categorical distributions respectively. The scalar α_t is determined by a predefined noise schedule.

Unlike the ligand, the conformation of the protein pocket at step T is not sampled from a fixed prior but rather corresponds to the apo state. To simulate intermediate pocket conformations along the apo–holo trajectory, we adopt a residue-level interpolation strategy that ensures structural coherence. Specifically, we interpolate residue-level translations, rotations, and chi angles updates. For translations and

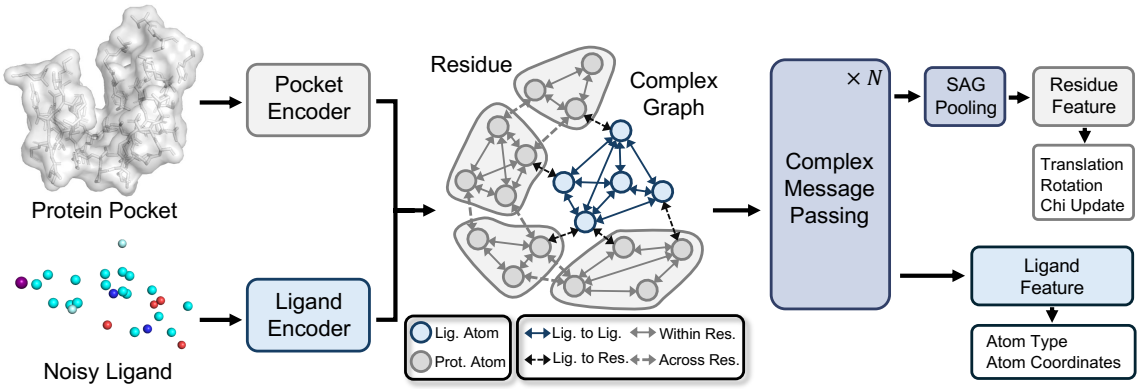

Figure 3: Illustration of the Apo2Mol framework, which jointly models ligand generation and protein pocket refinement through hierarchical graph message passing.

chi angles, the interpolated states at time t are modeled as:

$$\begin{aligned} \boldsymbol{tr}_t &\sim \mathcal{N}((t/T)\boldsymbol{tr}_0, \lambda_{tr}\beta_t \mathbf{I}), \\ \boldsymbol{\mathcal{X}}_t &\sim \mathcal{N}\left((t/T)\boldsymbol{\mathcal{X}}_0, \lambda_{\mathcal{X}}\beta_t \mathbf{I}\right), \end{aligned} \tag{5}$$

where λ_{tr} and $\lambda_{\mathcal{X}}$ control the noise scale applied to translations and chi angles, respectively. β_t is equal to $1 - \alpha_t$.

For rotations, which are represented as quaternions, we use spherical linear interpolation (Slerp) (Shoemake 1985) to smoothly transition from the holo-to-apo rotation $\boldsymbol{q}$ to the unit quaternion $\boldsymbol{q}_{\text{unit}}$ (no rotation):

$$\boldsymbol{q}_t = \text{Slerp}\left(\boldsymbol{q}_{\text{unit}}, \boldsymbol{q}_0, t/T\right) \otimes \boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \sim \mathcal{N}_{\mathbb{H}}\left(\mathbf{1}, \lambda_q \beta_t\right), \tag{6}$$

where $\otimes$ represents quaternion multiplication, and $\mathcal{N}_{\mathbb{H}}(\mathbf{1}, \lambda_q\beta_t)$ is a perturbation distribution centered at the identity quaternion with variance scaled by $\lambda_q\beta_t$. Noise is added to improve robustness, helping the model escape local minima and generalize across diverse conformational pathways.

The pocket coordinates at time t are then computed by applying the interpolated transformations to the holo structure:

$$\boldsymbol{x}_t^{\mathcal{P}} = (\boldsymbol{\mathcal{X}}_t)(\boldsymbol{q}_t \boldsymbol{x}_0^{\mathcal{P}} \boldsymbol{q}_t^* + \boldsymbol{tr}_t), \tag{7}$$

where $\boldsymbol{q}_t^*$ is the conjugate version of the quaternion.

Finally, by assembling the components described above, we obtain the pseudo interaction pair as $(\boldsymbol{x}_t^{\mathcal{M}}, \boldsymbol{v}_t^{\mathcal{M}}; \boldsymbol{x}_t^{\mathcal{P}}, \boldsymbol{v}_t^{\mathcal{P}})$.

4.2 Apo2Mol Framework

Fig. 3 illustrates the network architecture of Apo2Mol. The following content provides details about each component.

Feature Extraction The protein pocket and the noisy ligand at time step t are encoded separately. Ligand atoms are one-hot encoded by atom type, while protein atoms are represented by three one-hot vectors that specify atom type, amino acid identity, and a backbone indicator. These features are projected into a shared d-dimensional embedding space using two-layer multilayer perceptrons (MLPs). We denote the resulting embeddings as $\boldsymbol{h}^{\mathcal{M}} \in \mathbb{R}^{N_{\mathcal{M}} \times d}$ for the ligand and $\boldsymbol{h}^{\mathcal{P}} \in \mathbb{R}^{N_{\mathcal{P}} \times d}$ for the pocket.

Complex Graph Construction After obtaining the hidden features, Apo2Mol constructs a protein-ligand complex graph to enable graph neural network (GNN) training. A k-nearest neighbor strategy is employed to establish edges based on spatial proximity, connecting nodes from both the protein pocket and the ligand. To facilitate learning of meaningful protein-ligand interaction, while capturing residue-level conformational dynamics of the pocket, four distinct types of edges are defined: 1) intra-ligand edges; 2) ligand–residue edges; 3) intra-residue edges, and 4) inter-residue edges across different residues. This hierarchical design provides a multi-scale representation that supports accurate modeling of ligand binding and pocket refinement.

Complex Message Passing To capture intricate molecular interactions during the generative process, Apo2Mol employs a SE(3)-equivariant (Satorras, Hoogeboom, and Welling 2021) GNN model with attention mechanism over the constructed protein–ligand complex graph. Each node in the graph is associated with both: a 3D equivariant feature $\boldsymbol{x} \in \mathbb{R}^3$, representing the spatial position, and a 3D invariant feature $\boldsymbol{h} \in \mathbb{R}^d$ of an atom, representing its latent chemical context. During message passing, the model processes local neighborhoods using attention-based aggregation. The update rules for node i at l-th layer are defined as:

$$\begin{aligned} \boldsymbol{h}_i^{l+1} &= \boldsymbol{h}_i^l + \sum_{j \in N_i} f_{\boldsymbol{h}}(r_{ij}^l, \boldsymbol{h}_i^l, \boldsymbol{h}_j^l, \boldsymbol{e}_{ij}), \\ \boldsymbol{x}_i^{l+1} &= \boldsymbol{x}_i^l + \sum_{j \in N_i} \Delta \boldsymbol{x}_{ij}^l \, f_{\boldsymbol{x}}(r_{ij}^l, \boldsymbol{h}_i^{l+1}, \boldsymbol{h}_j^{l+1}, \boldsymbol{e}_{ij}). \end{aligned} \tag{8}$$

where N_i denotes the set of neighboring nodes of node i, $\boldsymbol{e}_{ij}$ denotes the edge features between nodes i and j, $r_{ij}^l = \|\boldsymbol{x}_j^l - \boldsymbol{x}_i^l\|$, and $\Delta \boldsymbol{x}_{ij}^l = \boldsymbol{x}_i^l - \boldsymbol{x}_j^l$. The functions $f_{\boldsymbol{h}}$ and $f_{\boldsymbol{x}}$ are implemented as attention-based neural networks that model how feature similarity and geometric distance contribute to updates of invariant and equivariant features, respectively.

The iterative updates of both positional and chemical context features simulate the dynamic interactions between the ligand and the pocket, facilitating the model to capture rich information for the final predictions.

Final Prediction Following the complex message passing stage, the learned invariant features $\boldsymbol{h}$ and equivariant coordinates $\boldsymbol{x}$ are utilized to generate the final predictions.

For the ligand generation, the coordinates produced at the final GNN layer are taken directly as the predicted atom positions $\hat{\boldsymbol{x}}_0^{\mathcal{M}}$, while the corresponding hidden features $\boldsymbol{h}^{L,\mathcal{M}}$ are passed through an additional MLP head to output the predicted atom types $\hat{\boldsymbol{v}}_0^{\mathcal{M}}$.

For the protein pocket refinement, we employ a Self-Attention Graph Pooling (SAGPooling) layer (Lee, Lee, and Kang 2019) to aggregate atom-level representations into residue-level features. Specifically, the concatenated features $[\boldsymbol{x}^{L,\mathcal{P}}, \boldsymbol{h}^{L,\mathcal{P}}]$ are fed into the SAGPooling module, which adaptively selects structurally informative nodes based on learned attention scores. The resulting residue-level representations are then passed through separate MLP heads to predict the translation vectors $\hat{\boldsymbol{tr}}_0$, rotation quaternions $\hat{\boldsymbol{q}}_0$, and chi angle updates $\hat{\boldsymbol{\mathcal{X}}}_0$.

To summarize, at each time step t, the overall mapping learned by Apo2Mol can be expressed as:

$$(\hat{\boldsymbol{x}}_0^{\mathcal{M}}, \hat{\boldsymbol{v}}_0^{\mathcal{M}}), (\hat{\boldsymbol{tr}}_0, \hat{\boldsymbol{q}}_0, \hat{\boldsymbol{\mathcal{X}}}_0) = \boldsymbol{s}_\theta(\boldsymbol{x}_t^{\mathcal{M}}, \boldsymbol{v}_t^{\mathcal{M}}; \boldsymbol{x}_t^{\mathcal{P}}, \boldsymbol{v}_t^{\mathcal{P}}). \quad (9)$$

4.3 Training Objective

To train Apo2Mol, we approximate the diffusion model to the score function described in Eq. 2.

The total training objective is composed of multiple loss components, each designed to improve the prediction accuracy of specific outputs for the ligand and the protein pocket.

Ligand Loss For the ligand, we incorporate the same objective function as prior work (Guan et al. 2023a). The ligand atom position loss is defined as:

$$\mathcal{L}_x^{\mathcal{M}} = \|\hat{\boldsymbol{x}}_0^{\mathcal{M}} - \boldsymbol{x}_0^{\mathcal{M}}\|^2. \quad (10)$$

To supervise ligand atom types, we compute the KL-divergence of categorical distributions as follows:

$$\mathcal{L}_v^{\mathcal{M}} = \sum_k \boldsymbol{c}(\boldsymbol{v}_t^{\mathcal{M}}, \boldsymbol{v}_0^{\mathcal{M}})_k \log \frac{\boldsymbol{c}(\boldsymbol{v}_t^{\mathcal{M}}, \boldsymbol{v}_0^{\mathcal{M}})_k}{\boldsymbol{c}(\boldsymbol{v}_t^{\mathcal{M}}, \hat{\boldsymbol{v}}_0^{\mathcal{M}})_k}. \quad (11)$$

Pocket Loss For the protein pocket, at each time step t, the model is trained to reverse the noisy transformations and recover the original holo state.

The translation loss is defined using mean squared error, where the target is the inverse translation vector obtained via quaternion rotation:

$$\mathcal{L}_{tr}^{\mathcal{P}} = \|\hat{\boldsymbol{tr}}_0 - (-\boldsymbol{q}_0^* \boldsymbol{tr}_0 \boldsymbol{q}_0)\|^2. \quad (12)$$

For rotation prediction, we apply an L1 loss, along with a unit norm regularization:

$$\mathcal{L}_q^{\mathcal{P}} = |\hat{\boldsymbol{q}}_0 - \boldsymbol{q}_0^*| + (1 - \hat{\boldsymbol{q}}_0 \hat{\boldsymbol{q}}_0^*). \quad (13)$$

For chi angle updates, we use a cosine-based loss to capture angular differences:

$$\mathcal{L}_{\mathcal{X}}^{\mathcal{P}} = 1 - \cos(\hat{\boldsymbol{\mathcal{X}}}_0 - (-\boldsymbol{\mathcal{X}}_0)). \quad (14)$$

The overall loss function is formulated as a weighted sum of the above five individual loss terms: $\mathcal{L} = \lambda_1 \mathcal{L}_x^{\mathcal{M}} + \lambda_2 \mathcal{L}_v^{\mathcal{M}} + \lambda_3 \mathcal{L}_{tr}^{\mathcal{P}} + \lambda_4 \mathcal{L}_q^{\mathcal{P}} + \lambda_5 \mathcal{L}_{\mathcal{X}}^{\mathcal{P}}$.

The training and sampling details are summarized in Apx. B, where we also provide pseudocode for each procedure.

5 Experiment

5.1 Experimental Settings

Dataset Curation To evaluate the effectiveness of Apo2Mol, we build a new benchmark dataset of experimental structural pairs of protein–ligand complexes and their corresponding apo conformations. The PLINDER dataset (Durairaj et al. 2024), which contains more than $400K$ protein–ligand complexes with comprehensive structural annotations, serves as the initial source of structures. To ensure pharmacologically relevant and high-quality data, we retain only triplets comprising experimentally resolved apo and holo protein structures (with 100% sequence identity and resolution ≤ 2.5Å) paired with their drug-like small molecule ligands, excluding ions, cofactors, artifacts, molecular fragments, and other undesired ligands. Additional stringent filtering yields a final benchmark of $24,601$ apo-holo-ligand triplets. For each complex, the holo pocket is defined as all residues within 10Å of the ligand atoms, and the corresponding apo pocket is obtained by extracting the same residues from the unbound structure. The dataset is chronologically split into $23,052$ training, $1,071$ validation, and 478 test samples to reflect real-world settings. Full details on the data processing pipeline and statistics are provided in Apx. A.

Baseline Methods We compare Apo2Mol against four representative baseline methods in SBDD: 1) **Pocket2Mol** (Peng et al. 2022): a generative model that constructs 3D molecules by sequentially placing atoms around a given protein pocket. 2) **TargetDiff** (Guan et al. 2023a): a diffusion-based method that generates ligand atom coordinates and types, with bond information determined via a post-processing step. 3) **DecompDiff** (Guan et al. 2023b): a diffusion model that decomposes ligands into scaffold and arm substructures and diffuses them separately using bond-level guidance and structural priors. 4) **IPDiff** (Huang et al. 2024b): it integrates protein–ligand interaction priors into both the forward and reverse diffusion processes to improve binding-aware generation.

While FlexSBDD (Zhang, Wang, and Liu 2024) and DynamicFlow (Zhou et al. 2025) provide alternative frameworks for incorporating pocket dynamics into molecule generation, we exclude them from benchmarking since their implementations are not publicly available, preventing reproducibility.

Evaluation Metrics To evaluate the quality of the generated ligands, we consider six widely used metrics. 1) **Binding Affinity**, measured by the Vina min score computed using AutoDock Vina (Eberhardt et al. 2021) under standard settings following prior work (Zhou et al. 2025). 2) **QED** (Bickerton et al. 2012), a quantitative estimate of drug-likeness. 3) **SA Score** (Ertl and Schuffenhauer 2009), reflecting synthetic accessibility. 4) **logP** (Ghose, Viswanadhan, and Wendoloski 1999), the octanol–water partition coefficient; values between -0.4 and 5.6 indicate favorable pharmacokinetic properties. 5) **Lipinski** (Lipinski et al. 2012), denoting the number of molecular properties that satisfy Lipinski's Rule of Five. 6) **High Affinity** measures the percentage of generated ligands with binding affinity higher than the reference molecules. All evaluation metrics, except High Affinity, are reported using both their mean and median values.

Method	Vina min (↓)		QED (↑)		SA (↑)		Logp		Linpiski (↑)		High Affinity (↑)
	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	
Reference	-7.41	-7.38	0.61	0.64	0.71	0.71	2.08	2.56	4.67	5.00	\
IPDiff	-6.40	-6.56	0.51	0.52	0.60	0.60	1.90	2.27	4.65	5.00	29.6%
TargetDiff	-5.19	-5.20	0.37	0.38	0.57	0.57	1.69	1.52	4.68	5.00	33.8%
Pocket2Mol	-3.30	-3.29	0.46	0.46	**0.85**	**0.86**	0.83	1.02	**4.98**	5.00	\
DecompDiff	-6.37	-6.40	0.56	0.58	0.66	0.66	1.53	1.76	4.73	5.00	34.3%
Apo2Mol - Gen	**-6.79**	**-7.09**	**0.59**	**0.63**	0.61	0.61	3.25	3.04	4.78	5.00	**42.7%**

Table 1: Benchmark on ligand generation from **apo structures**. Baseline models are evaluated on apo structures, while Apo2Mol generates both refined pockets and novel ligands from the apo inputs, with ligand accessed against the generated pockets. The best and second results are highlighted in **bold** and underlined. This convention is used throughout subsequent tables.

Method	Vina min (↓)		QED (↑)		SA (↑)		Logp		Linpiski (↑)		High Affinity (↑)
	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	
Reference	-7.41	-7.38	0.61	0.64	0.71	0.71	2.08	2.56	4.67	5.00	\
IPDiff	-7.09	-7.08	0.52	0.54	0.61	0.61	2.17	2.42	4.67	5.00	44.9%
TargetDiff	-5.50	-5.56	0.39	0.39	0.57	0.57	1.79	1.70	4.71	5.00	35.1%
Pocket2Mol	-3.32	-3.35	0.46	0.46	**0.87**	**0.88**	0.92	1.12	**4.99**	5.00	\
DecompDiff	-6.39	-6.41	0.56	0.58	0.65	0.65	1.53	1.76	4.74	5.00	31.9%
Apo2Mol - holo	**-7.86**	**-8.03**	**0.61**	**0.65**	0.59	0.62	3.49	3.35	4.76	5.00	**52.9%**

Table 2: Benchmark on ligand generation from **holo structures**. Baseline models are evaluated on native holo structures, while Apo2Mol generates novel ligands directly from the apo inputs, with ligand assessed against the native holo pockets.

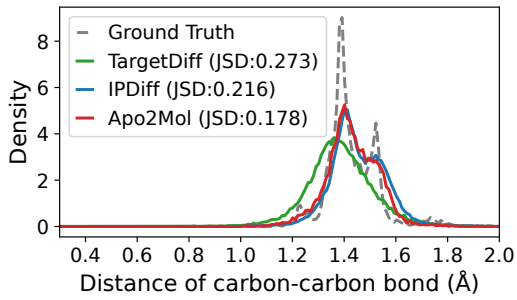

Figure 4: Distribution comparison for distances of carbon-carbon pairs for ground truth molecules in the test set (gray) and model-generated molecules (color). JSD between two distributions is reported.

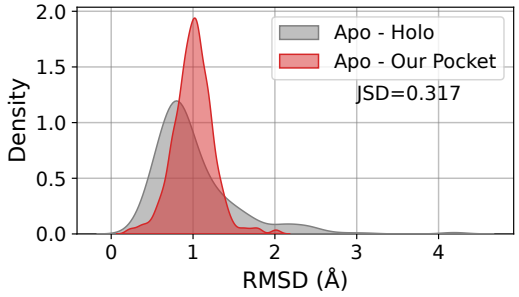

Figure 5: RMSD distribution between apo and holo, and between apo and generated pockets. JSD denotes the Jensen–Shannon divergence between the two distributions.

To assess the plausibility of the generated pockets, we compute the RMSD to their apo counterparts and calculate the Jensen–Shannon Divergence (JSD) between RMSD distributions of generated and experimental holo pockets, quantifying how well the model captures apo-to-holo transformations.

5.2 Main Results

Binding Affinity and Molecule Properties To evaluate the effectiveness of Apo2Mol in ligand generation with dynamic pocket refinement, we conduct two sets of comparative experiments. In the first setting, all baseline models are trained on holo pockets following their original protocols. During evaluation, only apo conformations are provided to both the baselines and Apo2Mol, simulating a realistic scenario in which the target protein lacks an experimentally determined ligand-bound state. Since Apo2Mol jointly generates both a refined pocket and a ligand, its ligand evaluation is performed on the generated pocket conformation. As shown in Tab. 1, Apo2Mol achieves the best performance across key metrics: the average Vina min score improves from -6.40 (best baseline, IPDiff) to -6.79, and the median score increases from -6.56 to -7.09, indicating a strong shift toward higher-

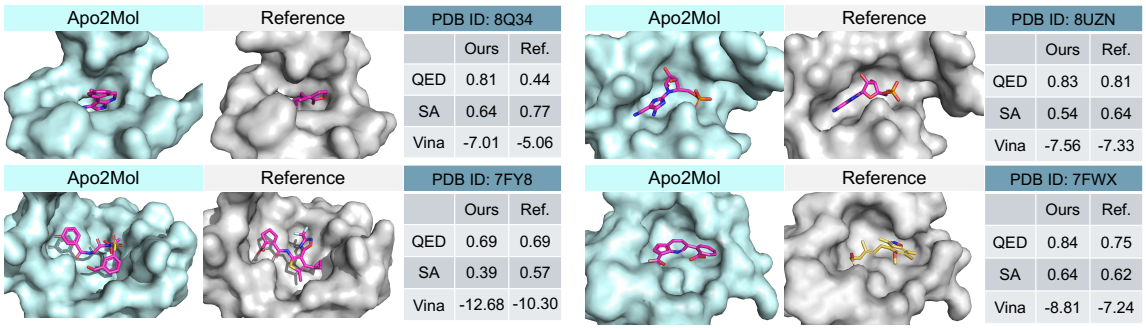

Figure 6: Visualization of ligands and pockets generated by Apo2Mol, compared with the corresponding reference ligands and holo structures. QED, SA, and Vina min scores are reported.

Method	Vina min (↓)		QED (↑)	
	Avg.	Med.	Avg.	Med.
Apo2Mol	**-6.79**	**-7.09**	**0.587**	**0.629**
w/o complex graph	-6.18	-6.22	0.524	0.529
w/o quaternion	-6.51	-6.44	0.523	0.534

Table 3: Ablation study on key network components.

affinity binding. Apo2Mol also outperforms baselines on molecular property metrics like QED, suggesting that refining pockets from the apo to holo state enables the generation of ligands with more drug-like profiles.

In the second setting, all baselines are trained and evaluated directly on holo pockets, following the previous *de novo* design setup disregarding protein conformational dynamics. By contrast, Apo2Mol continues to generate the ligands from the apo inputs, with ligand quality assessed against the corresponding native holo pocket. As shown in Tab. 2, although the baselines perform better under this idealized setting than the previous one, Apo2Mol remains competitive, achieving a substantial gain in binding affinity with an average Vina min score of -7.86 versus -7.09 for the second best model IPDiff, and a median of -8.03 versus -7.08. In addition, Apo2Mol also attains a higher High Affinity rate of 52.9%, compared with 44.9% for IPDiff.

Molecule Structure We assess the structural plausibility of Apo2Mol-generated molecules by comparing carbon–carbon bond distance distributions. As shown in Fig. 4, Apo2Mol achieves the lowest JSD (0.178) to ground truth, outperforming IPDiff (0.216) and TargetDiff (0.273), demonstrating its ability to generate more realistic molecular structures.

Pocket Structure We compare the RMSD distributions of experimental apo–holo pocket pairs with those of apo–Apo2Mol-generated pockets to evaluate how effectively Apo2Mol reproduces realistic ligand-binding pocket conformations. As shown in Fig. 5, the generated pockets generally follow the apo–holo transition distribution (Apx. A) but exhibit a mild shift (JSD = 0.317), suggesting a potential limitation of Apo2Mol in fully modeling the conformational variability of individual apo pockets.

Taken together, these results demonstrate that by jointly modeling ligand generation, pocket dynamics, and ligand–pocket interactions within a unified diffusion framework, Apo2Mol can produce high-quality drug-like ligands and corresponding ligand-pocket conformations even in the absence of holo structures, offering a robust solution to real-world SBDD scenarios where only apo structures are available.

5.3 Ablation Study

We conduct ablations to evaluate the contribution of key components in Apo2Mol. As shown in Tab. 3, removing the complex graph—by using only a single edge type within pocket atoms, leads to notable drops in both binding affinity and drug-likeness (Vina min: -6.18; QED Avg: 0.524). Replacing quaternions with rotation vectors for conformational modeling also degrades performance (Vina min: -6.51; QED Avg: 0.523). These results highlight the effectiveness of both the hierarchical graph design and quaternion-based transformations in generating high-quality ligand–pocket structures.

5.4 Visualization

Fig. 6 illustrates four representative examples of ligands and pockets generated by Apo2Mol, shown alongside the corresponding reference ligands and holo structures. Across diverse targets, Apo2Mol generates ligands with favorable binding affinity and drug-likeness, while also recovering plausible pocket geometries. These visualizations further highlight the ability of Apo2Mol to produce chemically and structurally realistic complexes directly from apo inputs.

6 Conclusion and Limitation

We present Apo2Mol, a full-atom SE(3)-equivariant diffusion framework that jointly generates small molecule ligands and refined protein pockets directly from apo states. Through extensive evaluations, Apo2Mol demonstrates strong performance compared to existing baselines and effectively models the protein dynamics during the ligand *de novo* design process. Despite these advances, a modest distribution gap remains between generated and native holo pockets, suggesting that additional refinement is needed to fully capture fine-grained conformational transitions for specific protein targets. Future work may benefit from pretraining on diverse protein structures to further close this gap.

Acknowledgements

This work was supported in part by the University of Florida (UF Startup Fund, UF Health Cancer Institute Pilot Grant #UFS-2023-08, and UF Research AI Award to Y. L.), National Institutes of Health (R21EB037868 to Y. L.), and the Bodor Professorship Fund (to C. L.). We acknowledge UFIT Research Computing for providing computational resources.

References

Aggarwal, R.; Gupta, A.; and Priyakumar, U. 2021. Apobind: a dataset of ligand unbound protein conformations for machine learning applications in de novo drug design. *arXiv preprint arXiv:2108.09926*.

Alonso, H.; Bliznyuk, A. A.; and Gready, J. E. 2006. Combining docking and molecular dynamic simulations in drug design. *Medicinal research reviews*, 26(5): 531–568.

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

Atz, K.; Cotos, L.; Isert, C.; Håkansson, M.; Focht, D.; Hilleke, M.; Nippa, D. F.; Iff, M.; Ledergerber, J.; Schiebroek, C. C.; et al. 2024. Prospective de novo drug design with deep interactome learning. *Nature Communications*, 15(1): 3408.

Bickerton, G. R.; Paolini, G. V.; Besnard, J.; Muresan, S.; and Hopkins, A. L. 2012. Quantifying the chemical beauty of drugs. *Nature chemistry*, 4(2): 90–98.

Bilodeau, C.; Jin, W.; Jaakkola, T.; Barzilay, R.; and Jensen, K. F. 2022. Generative models for molecular discovery: Recent advances and challenges. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 12(5): e1608.

Corso, G.; Somnath, V. R.; Getz, N.; Barzilay, R.; Jaakkola, T.; and Krause, A. 2025. Composing Unbalanced Flows for Flexible Docking and Relaxation. In *The Thirteenth International Conference on Learning Representations*.

Diebel, J.; et al. 2006. Representing attitude: Euler angles, unit quaternions, and rotation vectors. *Matrix*, 58.

Dror, R. O.; Dirks, R. M.; Grossman, J.; Xu, H.; and Shaw, D. E. 2012. Biomolecular simulation: a computational microscope for molecular biology. *Annual review of biophysics*, 41(1): 429–452.

Durairaj, J.; Adeshina, Y.; Cao, Z.; Zhang, X.; et al. 2024. PLINDER: The protein-ligand interactions dataset and evaluation resource. In *ICML'24 Workshop ML for Life and Material Science: From Theory to Industry Applications*.

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.

Epstein, D.; Jabri, A.; Poole, B.; Efros, A.; and Holynski, A. 2023. Diffusion self-guidance for controllable image generation. *Advances in Neural Information Processing Systems*, 36: 16222–16239.

Ertl, P.; and Schuffenhauer, A. 2009. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of cheminformatics*, 1(1): 8.

Ghose, A. K.; Viswanadhan, V. N.; and Wendoloski, J. J. 1999. 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.

Grisoni, F. 2023. Chemical language models for de novo drug design: Challenges and opportunities. *Current Opinion in Structural Biology*, 79: 102527.

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 *The Eleventh 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 *Proceedings of the 40th International Conference on Machine Learning*, 11827–11846.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.

Hollingsworth, S. A.; and Dror, R. O. 2018. Molecular dynamics simulation for all. *Neuron*, 99(6): 1129–1143.

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

Huang, Z.; Yang, L.; Zhou, X.; Qin, C.; Yu, Y.; Zheng, X.; Zhou, Z.; Zhang, W.; Wang, Y.; and Yang, W. 2024a. Interaction-based retrieval-augmented diffusion models for protein-specific 3d molecule generation. In *Forty-first International Conference on Machine Learning*.

Huang, Z.; Yang, L.; Zhou, X.; Zhang, Z.; Zhang, W.; Zheng, X.; Chen, J.; Wang, Y.; Cui, B.; and Yang, W. 2024b. Protein-ligand interaction prior for binding-aware 3d molecule diffusion models. In *The Twelfth International Conference on Learning Representations*.

Janson, G.; Valdes-Garcia, G.; Heo, L.; and Feig, M. 2023. Direct generation of protein conformational ensembles via machine learning. *Nature Communications*, 14(1): 774.

Kabsch, W. 1976. A solution for the best rotation to relate two sets of vectors. *Foundations of Crystallography*, 32(5): 922–923.

Kuznetsov, M.; Ryabov, F.; Schutski, R.; Shayakhmetov, R.; Lin, Y.-C.; Aliper, A.; and Polykovskiy, D. 2024. COSMIC: Molecular Conformation Space Modeling in Internal Coordinates with an Adversarial Framework. *Journal of Chemical Information and Modeling*, 64(9): 3610–3620.

Lee, J.; Lee, I.; and Kang, J. 2019. Self-attention graph pooling. In *International conference on machine learning*, 3734–3743. pmlr.

Li, G.; Jiang, C.; Gao, Z.; Liu, Y.; Liu, C.; Chen, J.; Huang, Y.; and Li, J. 2025. Molecule generation for target protein binding with hierarchical consistency diffusion model. *Chemical Science*, 16(39): 18263–18277.

Li, Y.; Zhou, D.; Zheng, G.; Li, X.; Wu, D.; and Yuan, Y. 2022. DyScore: a boosting scoring method with dynamic properties for identifying true binders and nonbinders in structure-based

drug discovery. *Journal of chemical information and modeling*, 62(22): 5550–5567.

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

Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; and Feeney, P. J. 2012. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, 64: 4–17.

Liu, M.; Luo, Y.; Uchino, K.; Maruhashi, K.; and Ji, S. 2022. Generating 3D Molecules for Target Protein Binding. In *International Conference on Machine Learning*, 13912–13924. PMLR.

Lu, J.; Chen, X.; Lu, S. Z.; Shi, C.; Guo, H.; Bengio, Y.; and Tang, J. 2025. Structure Language Models for Protein Conformation Generation. In *The Thirteenth International Conference on Learning Representations*.

Lu, W.; Zhang, J.; Huang, W.; Zhang, Z.; Jia, X.; Wang, Z.; Shi, L.; Li, C.; Wolynes, P. G.; and Zheng, S. 2024. DynamicBind: predicting ligand-specific protein-ligand complex structure with a deep equivariant generative model. *Nature Communications*, 15(1): 1071.

Morehead, A.; and Cheng, J. 2025. FlowDock: Geometric flow matching for generative protein–ligand docking and affinity prediction. *Bioinformatics*, 41.

Morehead, A.; Giri, N.; Liu, J.; and Cheng, J. 2024. Deep Learning for Protein-Ligand Docking: Are We There Yet? In *ICML 2024 AI for Science Workshop*.

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.

Ragoza, M.; Masuda, T.; and Koes, D. R. 2022. 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.

Schneuing, A.; Harris, C.; Du, Y.; Didi, K.; Jamasb, A.; Igashov, I.; Du, W.; Gomes, C.; Blundell, T. L.; Lio, P.; et al. 2024. Structure-based drug design with equivariant diffusion models. *Nature Computational Science*, 4(12): 899–909.

Shoemake, K. 1985. Animating rotation with quaternion curves. In *Proceedings of the 12th annual conference on Computer graphics and interactive techniques*, 245–254.

Siebenmorgen, T.; Menezes, F.; Benassou, S.; Merdivan, E.; Didi, K.; Mourão, A. S. D.; Kitel, R.; Liò, P.; Kesselheim, S.; Piraud, M.; et al. 2024. MISATO: machine learning dataset of protein–ligand complexes for structure-based drug discovery. *Nature computational science*, 4(5): 367–378.

Skalic, M.; Sabbadin, D.; Sattarov, B.; Sciabola, S.; and De Fabritiis, G. 2019. From target to drug: generative modeling for the multimodal structure-based ligand design. *Molecular pharmaceutics*, 16(10): 4282–4291.

Song, Y.; Sohl-Dickstein, J.; Kingma, D. P.; Kumar, A.; Ermon, S.; and Poole, B. 2021. Score-Based Generative Modeling through Stochastic Differential Equations. In *International Conference on Learning Representations*.

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.

Tang, X.; Dai, H.; Knight, E.; Wu, F.; Li, Y.; Li, T.; and Gerstein, M. 2024. A survey of generative AI for de novo drug design: new frontiers in molecule and protein generation. *Briefings in Bioinformatics*, 25(4): bbae338.

Vignac, C.; Krawczuk, I.; Siraudin, A.; Wang, B.; Cevher, V.; and Frossard, P. 2023. DiGress: Discrete Denoising diffusion for graph generation. In *The Eleventh International Conference on Learning Representations*.

Wang, L.; Shen, Y.; Wang, Y.; Yuan, H.; Wu, Y.; Gu, Q.; et al. 2024. Protein Conformation Generation via Force-Guided SE (3) Diffusion Models. In *Forty-first International Conference on Machine Learning*.

Wang, Y.; Wang, Z.; Li, Y.; Yan, P.; and Li, X. 2025. Improving Covalent and Noncovalent Molecule Generation via Reinforcement Learning with Functional Fragments. *Journal of Chemical Information and Modeling*.

Wei, H.; and McCammon, J. A. 2024. Structure and dynamics in drug discovery. *npj Drug Discovery*, 1(1): 1.

Weininger, D. 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1): 31–36.

Xie, W.; Wang, F.; Li, Y.; Lai, L.; and Pei, J. 2022. Advances and challenges in de novo drug design using three-dimensional deep generative models. *Journal of Chemical Information and Modeling*, 62(10): 2269–2279.

Zhang, Y.; Zhang, Z.; Zhong, B.; Misra, S.; and Tang, J. 2023a. Diffpack: A torsional diffusion model for autoregressive protein side-chain packing. *Advances in Neural Information Processing Systems*, 36: 48150–48172.

Zhang, Z.; Min, Y.; Zheng, S.; and Liu, Q. 2023b. Molecule generation for target protein binding with structural motifs. In *The eleventh international conference on learning representations*.

Zhang, Z.; Wang, M.; and Liu, Q. 2024. Flexsbdd: Structure-based drug design with flexible protein modeling. *Advances in Neural Information Processing Systems*, 37: 53918–53944.

Zhang, Z.; Yan, J.; Huang, Y.; Liu, Q.; Chen, E.; Wang, M.; and Zitnik, M. 2025. Structure-Based Drug Design with Geometric Deep Learning: A Comprehensive Survey. *ACM Computing Surveys*.

Zhou, X.; Xiao, Y.; Lin, H.; et al. 2025. Integrating Protein Dynamics into Structure-Based Drug Design via Full-Atom Stochastic Flows. In *The Thirteenth International Conference on Learning Representations*.

Zhu, Y.; Li, Z.; Wang, T.; He, M.; and Yao, C. 2023. Conditional text image generation with diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14235–14245.