PIDiff: Physics Informed Diffusion Model for Protein Pocket-Specific 3D Molecular Generation⋆

Seungyeon Choi[a,1], Samgmin Seo[a], Byung Ju Kim[b], Chihyun Park[c] and Sanghyun Park[a,*]

[a] *Department of Computer Science, Yonsei University, Seoul, 03722, Republic of Korea*
[b] *UBLBio Corporation, Suwon, 16679, Republic of Korea*
[c] *Department of Computer Science and Engineering, Kangwon National University, Chuncheon, 24341, Republic of Korea*

ARTICLE INFO

Keywords:
3D Molecule Generation
Denoising Diffusion Probabilistic Model
Physics-informed Neural Network
Geometric Deep Learning

ABSTRACT

Designing drugs capable of binding to the structure of target proteins for treating diseases is essential in drug development. Recent remarkable advancements in geometric deep learning have led to unprecedented progress in three-dimensional (3D) generation of ligands that can bind to the protein pocket. However, most existing methods primarily focus on modeling the geometric information of ligands in 3D space. Consequently, these methods fail to consider that the binding of proteins and ligands is a phenomenon driven by intrinsic physicochemical principles. Motivated by this understanding, we propose PIDiff, a model for generating molecules by accounting in the physicochemical principles of protein–ligand binding. Our model learns not only the structural information of proteins and ligands but also to minimize the binding free energy between them. To evaluate the proposed model, we introduce an experimental framework that surpasses traditional assessment methods by encompassing various essential aspects for the practical application of generative models to actual drug development. The results confirm that our model outperforms baseline models on the CrossDocked2020 benchmark dataset, demonstrating its superiority. Through diverse experiments, we have illustrated the promising potential of the proposed model in practical drug development.

1. Introduction

Structure-based drug design (SBDD), an approach wherein drugs are developed leveraging protein structures, holds great promise, considering that drugs are substances capable of binding to protein structures to either inhibit or activate specific functions (Anderson, 2003). Accordingly, numerous extensive studies have been conducted to identify ligands that can bind to target proteins (Blundell, 1996; Lyne, 2002; Shoichet, 2004; Pagadala, Syed and Tuszynski, 2017; Hansson, Oostenbrink and van Gunsteren, 2002). However, traditional methods encounter substantial hurdles in drug development, primarily owing to the vast chemical space and computationally-demanding tasks, resulting in high costs and prolonged timelines. To address these challenges, the field of generative models has extensively explored molecule generation, traditionally representing them in one-dimensional (strings) or two-dimensional (topology) formats (Cheng, Gong, Liu, Song and Zou, 2021). Nevertheless, these approaches have limitations in accurately depicting molecules in three-dimensional (3D) space. They fail to capture the intricate atomic interactions within the protein–ligand binding pocket (Xie, Wang, Li, Lai and Pei, 2022). Consequently, this shortfall hinders the accurate prediction of molecules that effectively bind to specific target proteins.

Recent breakthroughs in geometric deep learning (Atz, Grisoni and Schneider, 2021; Isert, Atz and Schneider, 2023) for molecular structures have spurred a shift in molecular generation research toward the utilization of geometric 3D representations, surpassing the earlier string- and topology-based models. Generating the desired molecules based on a 3D representation typically entails modeling both continuous variables (such as the position of atoms) and discrete variables (like the types of atoms). This transition to 3D methodologies facilitates a direct examination of how molecules interact within protein pockets, offering a more realistic and logical strategy for specific scenarios. Particularly in 3D object modeling, the adoption of SE(3)-equivariant neural networks, which incorporate an inductive bias to adeptly navigate the vast search space created by object translation and rotation, has unlocked unprecedented potential in 3D molecular generation Satorras, Hoogeboom and Welling (2021); Xu, Yu, Song, Shi, Ermon and Tang (2022). Additionally, the application of denoising diffusion probabilistic model (DDPM) (Ho, Jain and Abbeel, 2020; Nichol and Dhariwal, 2021), which has shown impressive results in image generation, is emerging as a successful approach in the challenging task of 3D molecular generation (Hoogeboom, Satorras, Vignac and Welling, 2022; Guan, Qian, Peng, Su, Peng and Ma, 2023). Building on previous research, these advancements have laid the foundation for a more rational approach to utilizing geometric information of molecules in drug design.

Nevertheless, when the objective is to generate molecules with the capability to bind to a protein, the binding principles between the protein and ligand need to be reconsidered once again. The binding of a small molecule to a protein pocket implies a greater stability compared to its unbound state.

⋆This research was supported by the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. RS-2023-00229822). The funders did not play any role in the design of the study, data collection, analysis, or preparation of the manuscript.

*Corresponding author
sanghyun@yonsei.ac.kr (S. Park)
ORCID(s): 0000-0002-5196-6193 (S. Park)

[1] This is the first author footnote.

This can be elucidated by the physicochemical principle wherein the binding free energy attains its minimum when the protein and small molecule are combined (Zhou and Gilson, 2009; Luque and Barril, 2012). Therefore, recalling these fundamental principles, designing molecular generation models to learn only the geometric shapes of molecules, as done in previous studies, could be a superficial approach that overlooks the essential principles underlying the problem we seek to solve. Such an approach risks overlooking deep and more critical aspects of molecular interactions and binding processes.

Another important aspect to consider in 3D molecular generation is the dataset utilized. Specifically, in studies focusing on protein-aware 3D molecular generation, the pdb format (Berman, Westbrook, Feng, Gilliland, Bhat, Weissig, Shindyalov and Bourne, 2000) is commonly used. This data type represents proteins and ligands, which vibrate according to thermodynamic principles Fischer, Coleman, Fraser and Shoichet (2014), as fixed relative atomic positions in three-dimensional coordinates. This can be explained by the fact that, whether obtained through crystallization methods such as X-ray crystallography MacArthur, Laskowski and Thornton (1994), predicted by AlphaFold Jumper, Evans, Pritzel, Green, Figurnov, Ronneberger, Tunyasuvunakool, Bates, Žídek, Potapenko et al. (2021), or derived from protein-ligand complexes obtained via Cross-docking Francoeur, Masuda, Sunseri, Jia, Iovanisci, Snyder and Koes (2020), these structures are all subject to constant vibration in real environments, yet are represented as fixed three-dimensional coordinates. Consequently, such data have limitations in capturing the dynamic and constantly changing real-world environment. Crucially, existing generative models have been trained relying on the empirical distribution of these data, posing considerable challenges in synthesizing high-quality data that accurately reflects the true nature of molecular interactions.

To address the two primary limitations in 3D molecular generation, we propose incorporating intrinsic physicochemical principles into the generative model, drawing inspiration from the physics-informed neural network (PINN) (Karniadakis, Kevrekidis, Lu, Perdikaris, Wang and Yang, 2021; Hao, Liu, Zhang, Ying, Feng, Su and Zhu, 2022; Moon, Zhung, Yang, Lim and Kim, 2022). PINN is a comprehensive framework that enhances a model's generalization performance by integrating relevant physical properties as inductive biases within the neural network, rooted in extensive fluid dynamics research. By leveraging this concept, we develop PIDiff, a generative model that defines the inductive bias based on the fundamental physicochemical principle that stable binding between a protein and ligand occurs at positions where the binding free energy is minimized along their reaction coordinates. `PIDiff` is designed not only to learn the geometric information of molecules bound to target proteins but also to be trained to ensure that the generated molecules achieve minimal binding free energy with the target proteins. Binding free energy can be approximated as the sum of intermolecular interaction energies, such as Van der Waals (Bronowska, 2011; Bitencourt-Ferreira, Veit-Acosta and de Azevedo, 2019) interactions. By differentiating this value with respect to the positions of the generated atoms, we can condition the minimization of the binding free energy. To the best of our knowledge, this approach is the first attempt in molecular generation research to integrate the physicochemical principle of binding free energy minimization as an inductive bias to include the intrinsic principles of protein-ligand binding. The design of this 3D molecular generation model allows for the synthesis of higher-quality data that is closer to real-world scenarios. This is achieved by not only learning the empirical distribution of the existing training datasets but also by ensuring compliance with the underlying physical laws implicitly embedded within the datasets. Additionally, the physicochemical principles we aim to incorporate into our generative model are invariant truths that apply equally even in the dynamic and constantly vibrating thermodynamic environment. By doing so, we can more accurately reflect real-world environments compared to previous studies that focus solely on geometric information.

Commencing with the performance evaluation of the proposed model on the CrossDocked2020 (Francoeur et al., 2020) benchmark dataset, we conducted extensive experiments with a primary focus on the potential utility of molecular generative models in the drug development process. As a result, when compared to several contemporary models on the CrossDocked2020 benchmark, our model achieved state-of-the-art performance and demonstrated the stability of the molecules generated by our model in terms of molecular conformation. Furthermore, beyond just benchmark performance, we validated the capability of our model to generate molecules with high binding affinity in pocket structures obtained from various sources for key target proteins responsible for several diseases. We conducted an additional assessment of the proposed model by verifying its selectivity against off-target effects (Xie, Xie and Bourne, 2011; Whitebread, Hamon, Bojanic and Urban, 2005), which are a major concern in SBDD research. Lastly, we conclude our proposed comprehensive evaluation framework by integrating molecular dynamics simulations (MD) (De Vivo, Masetti, Bottegoni and Cavalli, 2016) to more realistically assess the actual behavior of molecules generated by the model within the protein pocket. This multifaceted evaluation provides a thorough understanding of our model's performance, real-world applicability, and selectivity in drug development processes.

Our main contributions can be summarized as follows:

- We introduce `PIDiff`, the first model in 3D molecular generation research to incorporate the physicochemical principle that the binding of proteins and ligands occurs at the state where the binding free-energy landscape is minimized. This approach allows for more feasible molecular generation by satisfying the physical laws that can be applied to the real world.

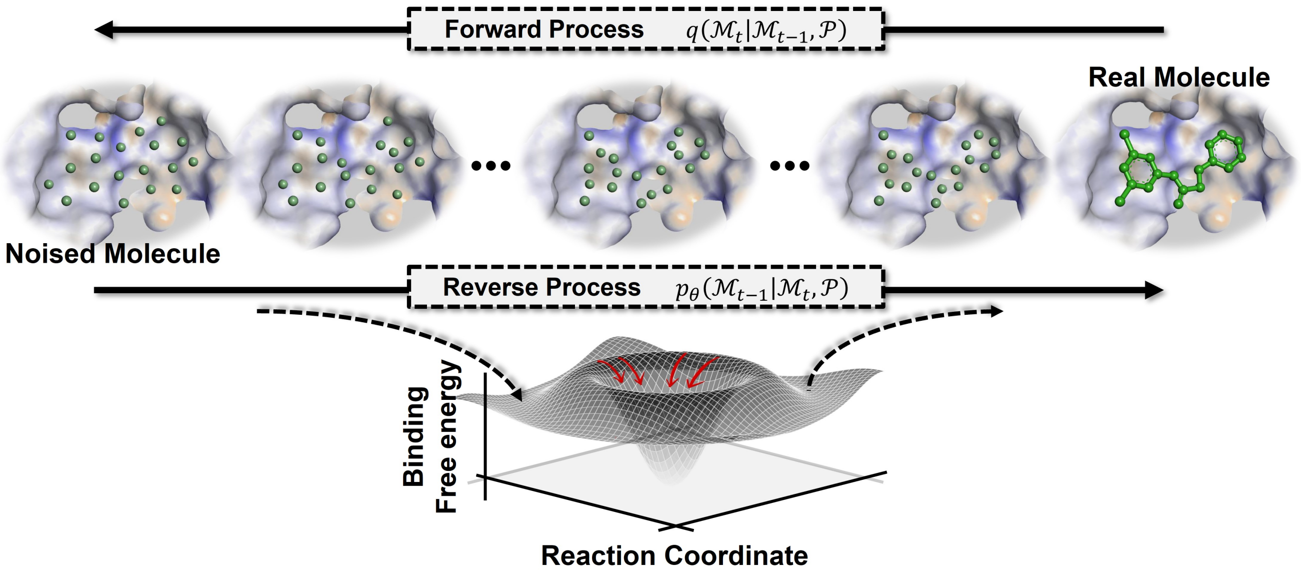

Forward Process $q(\mathcal{M}_t|\mathcal{M}_{t-1},\mathcal{P})$
Real Molecule
Noised Molecule
Reverse Process $p_\theta(\mathcal{M}_{t-1}|\mathcal{M}_t,\mathcal{P})$
Binding Free energy
Reaction Coordinate

Figure 1: Overview of PIDiff. The forward process incrementally injects noise into a Real Molecule($\mathcal{M}_0$) until it transforms into a Noised Molecule($\mathcal{M}_T$) across t timesteps. In contrast, the reverse process progressively removes noise to convert a Noised Molecule back into a Real Molecule through a function parameterized by θ. A critical aspect of this process is that domain knowledge is simultaneously infused during the reverse process to entail that the binding free energy at the reaction coordinate between the inferred Real Molecule and a fixed protein pocket is minimized.

- Our experimental results demonstrate that our model, `PIDiff`, has set a state-of-the-art performance record by surpassing other baseline models in terms of the binding affinity of the molecules it generated for the protein pockets in the CrossDocked2020 benchmark dataset.
- We propose a diverse and rigorous experimental framework to evaluate the performance of generative models in SBDD from the perspective of their utility in drug development. The results offer various perspectives on the applicability of our model to drug development.

2. Methods

2.1. Problem definition

To represent molecules that bind to specific protein pockets, we construct protein-ligand pair $\mathcal{L} = \left\{\left(X^{P,i}, V^{P,i}\right), \left(X^{M,i}, V^{M,i}\right)\right\}_{i=1}^{N_{\text{pair}}}$, where N_{pair} is the number of protein-ligand pairs. Here , X^P and X^M represent the 3D coordinates of atoms within the protein pocket and ligand respectively, and V^P and V^M denote features such as atom types associated with the protein pocket atoms and ligand atoms. For brevity, we represent proteinn pocket as $\mathcal{P} = \left\{\left(X^{P,i}, V^{P,i}\right)\right\}_{i=1}^{N_{\text{pair}}}$, and the molecule binding to the pocket as $\mathcal{M} = \left\{\left(X^{M,i}, V^{M,i}\right)\right\}_{i=1}^{N_{\text{pair}}}$. Our goal is to generate $\mathcal{M}$ that binds to a given protein pocket $\mathcal{P}$.

2.2. Preliminaries

When aiming to generate a molecule that binds to a given protein pocket, the application of a DDPM (Ho et al., 2020) requires tractable distributions capable of modeling the molecule's atom coordinates and types. Previous research (Ho et al., 2020; Hoogeboom, Nielsen, Jaini, Forré and Welling, 2021; Guan et al., 2023) has developed a diffusion framework that utilizes Gaussian distribution $\mathcal{N}$ for modeling continuous variables (like atom coordinates) and categorical distribution $\mathcal{C}$ for discrete variables (such as atom types). The atom coordinates and types of the ligand are independently perturbed by introducing a small Gaussian noise and uniform noise at each timestep t, respectively. This process follows a Markov chain characterized by predetermined variance schedules β_t, as shown in Eq.(1). The arbitrary step for both distributions can be expressed using $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \Pi_{s=1}^{t} \alpha_s$ in Eq.(2), and the posterior of atom coordinates and atom types can be defined in closed-form as in Eq.(3) using the Bayes' rule. Detailed information is described in Supplementary Material (1).

$$\begin{aligned} q\left(\mathcal{M}_t \mid \mathcal{M}_{t-1}, \mathcal{P}\right) = \mathcal{N}\left(X_t^M; \sqrt{1-\beta_t} X_{t-1}^M, \beta_t \mathbf{I}\right) \cdot \\ \mathcal{C}\left(V_t^M \mid \left(1-\beta_t\right) V_{t-1}^M + \beta_t / K\right) \end{aligned} \quad (1)$$

$$\begin{aligned} q\left(X_t^M \mid X_0^M\right) &= \mathcal{N}\left(X_t^M; \sqrt{\bar{\alpha}_t} X_0^M, \left(1-\bar{\alpha}_t\right) \mathbf{I}\right) \\ q\left(V_t^M \mid V_0^M\right) &= \mathcal{C}\left(V_t^M \mid \bar{\alpha}_t V_0^M + \left(1-\bar{\alpha}_t\right) / K\right) \end{aligned} \quad (2)$$

$$q\left(X_{t-1}^M \mid X_t^M, X_0^M\right) = \mathcal{N}\left(X_{t-1}^M; \tilde{\boldsymbol{\mu}}_t\left(X_t^M, X_0^M\right), \tilde{\beta}_t \mathbf{I}\right)$$

$$q\left(V_{t-1}^M \mid V_t^M, V_0^M\right) = \mathcal{C}\left(V_{t-1}^M \mid \tilde{\boldsymbol{c}}_t\left(V_t^M, V_0^M\right)\right) \tag{3}$$

The critical aspect in generating molecules in 3D space, particularly during learning of the reverse transition kernel $p_\theta\left(\mathcal{M}_0 \mid \mathcal{P}\right)$, is to design a model that approximates a likelihood invariant to the translation and rotation of the protein-ligand complex. This approach is an essential inductive bias enables efficient exploration of the extensive search space associated with SE(3) transformations. Previous studies (Guan et al., 2023; Satorras et al., 2021; Xu et al., 2022), aiming to generate various 3D objects, have demonstrated that if the Center of Mass (CoM) of a protein pocket is shifted to zero and the reverse transition kernel $p\left(X_{t-1}^M \mid X_t^M, X^P\right)$ satisfies SE(3)-equivariance, then the likelihood $p_\theta\left(\mathcal{M}_0 \mid \mathcal{P}\right)$ for inferring molecules relative to a given protein pocket remains invariant to translation and rotation.

2.3. Overview of proposed model

As illustrated in Fig.1, we introduce `PIDiff`, a diffusion-based generative model infused with the chemical principles of stable states between proteins and ligands. Our model operates by fixing the position of protein atoms and injecting noise into the position and type of atoms in the *Real Molecule*, transforming it into a *Noised Molecule* through a forward process. It then undergoes a reverse process, removing the noise to revert back to the *Real Molecule*. A critical aspect of our model is that during the reverse process, it is not merely reconstructing the position and type of *Real Molecule* atoms but is designed to ensure that the ligand satisfies the physicochemical principles for stable binding to the target protein.

2.4. Equivariant reverse process

The reverse process, which is the opposite of the forward process that injects noise into the coordinates and types of atoms constituting a molecule as expressed in Eq.(3), can be represented as shown in Eq.(4). Our objective is to approximate these reverse processes through a neural network parameterized by θ.

$$\begin{aligned}
&p_\theta\left(\mathcal{M}_{t-1} \mid \mathcal{M}_t, \mathcal{P}\right)\\
&= \mathcal{N}\left(X_{t-1}^M; \boldsymbol{\mu}_\theta\left(\mathcal{M}_t, t, \mathcal{P}\right), \sigma_t^2 I\right) \cdot \mathcal{C}\left(V_{t-1}^M \mid \boldsymbol{c}_\theta\left(\mathcal{M}_t, t, \mathcal{P}\right)\right)\\
&= \mathcal{N}\left(X_{t-1}^M; \boldsymbol{\mu}_\theta\left(\left[X_t^M, V_t^M\right], t, \left[X^P, V^P\right]\right), \sigma_t^2 I\right) \cdot\\
&\quad \mathcal{C}\left(V_{t-1}^M \mid \boldsymbol{c}_\theta\left(\left[X_t^M, V_t^M\right], t, \left[X^P, V^P\right]\right)\right)
\end{aligned} \tag{4}$$

Here, $[\cdot, \cdot]$ is the concatenation operator, X^M and V^M represent the 3D coordinates and atom types of the atoms constituting the ligand, respectively, while X^P and V^P similarly denote the 3D coordinates and atom types of the atoms constituting the protein. To ensure the generative process for molecule inference remains unaffected by the translation and rotation of complex structures, we have employed an approximation function defined by an SE(3)-equivariant graph transformer architecture (Guan et al., 2023; Hoogeboom et al., 2022; Guan, Qian, Ma, Ma and Peng, 2021). The model is designed to predict X_0^M and V_0^M, as in Eq.(5). From these predicted values, it derives $\boldsymbol{\mu}_\theta$ and $\boldsymbol{c}_\theta$ of the reverse transition kernel, enabling the sampling process. Additionally, the methods for calculating $\hat{X}_0^M$ and $\hat{V}_0^M$ slightly differ from each other. A detailed description of the layers that constitute ϕ_θ can be found in the Supplementary Material (2).

$$\left[\hat{X}_0^M, \hat{V}_0^M\right] = \phi_\theta\left(\left[X_t^M, V_t^M\right], t, \left[X^P, V^P\right]\right) \tag{5}$$

2.5. Physics informed optimization

It crucial to conduct optimization both from a data perspective (to align the geometric distribution of molecules inferred by the generative model with the distribution of real data) and from a physics-informed perspective (to minimize the binding free energy between the inferred molecules and the target protein). Training from the data perspective follows previous research(Ho et al., 2020; Guan et al., 2023; Hoogeboom et al., 2021) by optimizing the variational bound of the negative log likelihood. For continuous variables, such as atom coordinates, training involves minimizing the KL-divergence between the forward Gaussian process $q\left(X_{t-1}^M \mid X_t^M, X_0^M\right)$ and the reverse Gaussian process $p_\theta\left(X_{t-1}^M \mid X_t^M\right)$, which can be precisely defined in a closed form (indicated in Eq.(6)). For discrete variables, representing atom types, since the distribution is not Gaussian but categorical, training progresses by minimizing the KL divergence of the posterior for both forward and reverse processes at each time step, utilizing the predicted value$\hat{V}_0^M$, as outlined in Eq (7). Additional explanations for the two loss functions are summarized in Supplementary Material (4).

$$\begin{aligned}
L_{t-1}^{(X)} &= \mathrm{KL}\left(\mathcal{N}\left(X_{t-1}^M; \tilde{\mu}_t\left(X_t^M, X_0^M\right)\right) \mid \mathcal{N}\left(X_{t-1}^M; \mu_\theta\left(\mathcal{M}_t, t, \mathcal{P}\right)\right)\right)\\
&= \frac{1}{2\sigma_t^2}\left\|\tilde{\boldsymbol{\mu}}_t\left(X_t^M, X_0^M\right) - \boldsymbol{\mu}_\theta\left(\left[X_t^M, V_t^M\right], t, \left[X^P, V^P\right]\right)\right\|^2\\
&= \gamma_t\left\|X_0^M - \hat{X}_0^M\right\|^2
\end{aligned} \tag{6}$$

$$\begin{aligned}
L_{t-1}^{(V)} &= \mathrm{KL}\left(\tilde{\boldsymbol{c}}_t\left(V_t^M, V_0^M\right) \mid \boldsymbol{c}_\theta\left(\mathcal{M}_t, t, \mathcal{P}\right)\right)\\
&= \sum_k \tilde{\boldsymbol{c}}\left(V_t^M, v_0^M\right)_k \log \frac{\tilde{\boldsymbol{c}}\left(V_t^M, V_0^M\right)_k}{\boldsymbol{c}_\theta\left(\mathcal{M}_t, t, \mathcal{P}\right)_k}
\end{aligned} \tag{7}$$

where $\gamma_t = \frac{1}{2\sigma_q^2(t)} \frac{\bar{\alpha}_{t-1}(1-\alpha_t)^2}{(1-\bar{\alpha}_t)^2}$ and

$$\boldsymbol{c}_\theta\left(\mathcal{M}_t, t, \mathcal{P}\right)_k = \frac{\left[\alpha_t V_t^M + (1-\alpha_t)/K\right] \odot \left[\bar{\alpha}_{t-1}\hat{V}_0^M + (1-\bar{\alpha}_{t-1})/K\right]}{\sum_{V_t^M \in S}\left[\alpha_t V_t^M + (1-\alpha_t)/K\right] \odot \left[\bar{\alpha}_{t-1}\hat{V}_0^M + (1-\bar{\alpha}_{t-1})/K\right]}.$$

To perform optimization from physics-informed perspective, the binding free energy must be calculated based on the positions of atoms inferred by the generative model. This calculation begins with evaluating the energy associated with Van der Waals interactions, which is a critical step in determining the interaction dynamics between the protein and ligand (Pacholczyk and Kimmel, 2011). The minimization of binding free energy is of paramount importance to

Algorithm 1 Inference procedure

```
Input: Protein pocket atoms $\mathcal{P}$ and trained PIDiff model $\phi_\theta$
Output: Generated molecule $\mathcal{M}$
Sample initial molecular atom coordinates $X_T^M$ and atom types $V_T^M$
    $X_T^M \in \mathcal{N}(0, \boldsymbol{I})$
    $V_T^M = \text{one\_hot}\left(\arg\max g_i\right),$ where $g \sim \text{Gumbel}(0, 1)$
for $t$ in $T, T-1, ..., 1$ do
    Predict $\left[\hat{X}_0^M, \hat{V}_0^M\right]$ from $\left[X_t^M, V_t^M\right]$ with $\phi_\theta$ : $\left[\hat{X}_0^M, \hat{V}_0^M\right] = \phi_\theta\left(\left[X_t^M, V_t^M\right], t, \left[X_t^P, V_t^P\right]\right)$
    $X_{t-1}^M = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})X_t + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)\hat{X}_0}{1-\bar{\alpha}_t} + \sqrt{\frac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}}\mathbf{I}$
    $V_{t-1}^M = \frac{\left[\alpha_t V_t^M + (1-\alpha_t)/K\right] \odot \left[\bar{\alpha}_{t-1}\hat{V}_0^M + (1-\bar{\alpha}_{t-1})/K\right]}{\sum_{V_t^M \in S}\left[\alpha_t V_t^M + (1-\alpha_t)/K\right] \odot \left[\bar{\alpha}_{t-1}\hat{V}_0^M + (1-\bar{\alpha}_{t-1})/K\right]}$
end for
return $X_0^M, V_0^M$
```

achieve a stable and favorable bond between the protein and ligand. Therefore, we prioritize the minimization of free energy as our primary target within the optimization process. To this end, we introduce constraints to ensure that derivative of the binding free energy with respect to the inferred atomic positions is zero. This methodology guides the generated ligand and target protein toward a state of minimized binding free energy, effectively promoting the formation of stable interactions. The total Lennard-Jones potential energy is represented as the sum of the energies for all possible pairs of protein atoms and ligand atoms, as expressed in Eq (8).

$$\mathrm{E}_{\mathrm{LJ}} = \sum_{j \in P}\sum_{i \in M} \epsilon \left[\left(\frac{d'_{ij}}{d_{ij}}\right)^{12} - 2\left(\frac{d'_{ij}}{d_{ij}}\right)^{6}\right] \tag{8}$$

$$d_{ij} = \left\| \left[\hat{X}^M_{0,[i,0]}, \hat{X}^M_{0,[i,1]}, \hat{X}^M_{0,[i,2]}\right] - \left[X^P{}_{[j,0]}, X^P{}_{[j,1]}, X^P{}_{[j,2]}\right] \right\|$$

where M and P denote the index sets of ligand and protein atoms, respectively, while d_{ij} represents the distance between a ligand atom and a protein atom. r_i and r_j refer to the 3D coordinates of atoms constituting the ligand and protein, defined as $\mathbf{r}_i = \left[\hat{X}^M_{0,[i,0]}, \hat{X}^M_{0,[i,1]}, \hat{X}^M_{0,[i,2]}\right] \in \mathbb{R}^{M \times 3}$ and $\mathbf{r}_j = \left[X^P{}_{[j,0]}, X^P{}_{[j,1]}, X^P{}_{[j,2]}\right] \in \mathbb{R}^{P \times 3}$, respectively. ϵ and d'_{ij} are constants since their values are not influenced by the positions of the generated atoms.

A crucial point to note here is that the energy of Van der Waals interaction we aim to model must satisfy the condition of being invariant to rotation and translation regarding the positions of atoms inferred by the generative model and the positions of protein atoms. As the likelihood of the reverse process in the diffusion model is mentioned to be invariant to translation and rotation in Section 2.2, we demonstrate the SE(3)-invariance of the Van der Waals interaction as in Proof:

Proof. Let $T_g(\mathbf{x})$ can be written explicitly as $T_g(\mathbf{x}) = \boldsymbol{R}\mathbf{x} + \mathbf{b}$, where $\boldsymbol{R} \in \mathbb{R}^{3\times3}$ is the rotation matrix and $\boldsymbol{b} \in \mathbb{R}^3$ is the translation vector.

$$\begin{aligned}
\mathrm{E}_{\mathrm{LJ}} &= \sum_{j \in P}\sum_{i \in M} \epsilon \left[\left(\frac{d'_{ij}}{\left\|\mathbf{r}_i - \mathbf{r}_j\right\|^2}\right)^{12} - 2\left(\frac{d'_{ij}}{\left\|\mathbf{r}_i - \mathbf{r}_j\right\|^2}\right)^{6}\right] \\
&= \sum_{j \in P}\sum_{i \in M} \epsilon \left[\left(\frac{d'_{ij}}{T_g\left(\left\|\mathbf{r}_i - \mathbf{r}_j\right\|^2\right)}\right)^{12} - 2\left(\frac{d'_{ij}}{T_g\left(\left\|\mathbf{r}_i - \mathbf{r}_j\right\|^2\right)}\right)^{6}\right] \\
&= \sum_{j \in P}\sum_{i \in M} \epsilon \left[\left(\frac{d'_{ij}}{\left\|\left(\boldsymbol{R}\mathbf{r}_i + \boldsymbol{b}\right) - \left(\boldsymbol{R}\mathbf{r}_j + \boldsymbol{b}\right)\right\|^2}\right)^{12} - 2\left(\frac{d'_{ij}}{\left\|\left(\boldsymbol{R}\mathbf{r}_i + \boldsymbol{b}\right) - \left(\boldsymbol{R}\mathbf{r}_j + \boldsymbol{b}\right)\right\|^2}\right)^{6}\right] \\
&= \sum_{j \in P}\sum_{i \in M} \epsilon \left[\left(\frac{d'_{ij}}{\left(\mathbf{r}_i - \mathbf{r}_j\right)^{\top}\boldsymbol{R}^{\top}\boldsymbol{R}\left(\mathbf{r}_i - \mathbf{r}_j\right)}\right)^{12} - 2\left(\frac{d'_{ij}}{\left(\mathbf{r}_i - \mathbf{r}_j\right)^{\top}\boldsymbol{R}^{\top}\boldsymbol{R}\left(\mathbf{r}_i - \mathbf{r}_j\right)}\right)^{6}\right] \\
&= \sum_{j \in P}\sum_{i \in M} \epsilon \left[\left(\frac{d'_{ij}}{\left(\mathbf{r}_i - \mathbf{r}_j\right)^{\top}\mathbf{I}\left(\mathbf{r}_i - \mathbf{r}_j\right)}\right)^{12} - 2\left(\frac{d'_{ij}}{\left(\mathbf{r}_i - \mathbf{r}_j\right)^{\top}\mathbf{I}\left(\mathbf{r}_i - \mathbf{r}_j\right)}\right)^{6}\right] \\
&= \sum_{j \in P}\sum_{i \in M} \left[\left(\frac{d'_{ij}}{\left\|\mathbf{r}_i - \mathbf{r}_j\right\|^2}\right)^{12} - 2\left(\frac{d'_{ij}}{\left\|\mathbf{r}_i - \mathbf{r}_j\right\|^2}\right)^{6}\right]
\end{aligned}$$

□

We formalize the derivative of the distance between the predicted positions of the protein and ligand atoms with respect to the potential energy derived from the positions of atoms inferred by the generative model, as expressed in equation (9).

$$L^{(D)} = \sum_{j \in P}\sum_{i \in M} \left(\frac{\partial\left(\epsilon\left[\left(\frac{d'_{ij}}{d_{ij}}\right)^{12} - 2\left(\frac{d'_{ij}}{d_{ij}}\right)^{6}\right]\right)}{\partial d_{ij}} \right)^2 \tag{9}$$

Therefore, our ultimate objective is to concurrently minimize the generative loss related to Eqs (6) and (7), as well as the derivative loss related to Eq (9), as expressed in Eq (10). The generative process of the trained PIDiff is described in Alg.1.

$$L_{\text{total}} = L^{(X)}_{t-1} + L^{(V)}_{t-1} + L^{(D)} \tag{10}$$

3. Experiments

We remind ourselves of the purpose that the proposed model should be utilized to enhance efficiency and accelerate the drug development process, and pose the following research questions(RQs) to determine whether the proposed model can actually be applied in drug development, especially in SBDD.

- **RQ1.** Can molecules generated by the model effectively bind to target proteins? **-** Section 3.1
- **RQ2.** Can molecules generated by the model maintain a stable state when bound to target proteins? **-** Section 3.2
- **RQ3.** Can molecules generated by the model effectively bind to proteins targeting actual diseases according to structures obtained from various sources? **-** Section 3.3
- **RQ4.** Can molecules generated by the model for a specific protein structure exhibit selectivity against proteins with different structures? **-** Section 3.4
- **RQ5.** Can molecules generated by the model maintain stable binding within actual biological systems when compared with real drugs? **-** Section 3.5

To address these five RQs, we implemented various evaluation frameworks and conducted the corresponding experiments. The significance of each RQ and the details of the corresponding experiments are provided in the relevant subsections. The source code of the proposed model and the experiments conducted in this study is available at https://github.com/hello-maker/PIDiff.

All experiments were conducted on Ubuntu 18.04.6 LTS with 64 GB of memory and a GeForce RTX 3090. Our model generally converges within 12 hours and 100k steps. When the batch size is 4, the GPU memory usage is typically around 18GB. The generation of 1000 molecules using the trained model typically takes 2.5 hours. Additionally, PIDiff was implemented using Python 3.8 and PyTorch 1.21.1. For data preprocessing and final molecule generation, widely used cheminformatics tools RDKit 23.03 and OpenBabel 3.1 were utilized.

3.1. Analysis of binding affinity of generated molecules to target proteins

To verify whether the trained generative model can infer molecules that can bind to a protein pocket, we employed the CrossDocked2020 (Francoeur et al., 2020) dataset. Furthermore, to clearly understand the model's generalization ability, we applied filtering and splitting to the dataset like in previous studies (Luo, Guan, Ma and Peng, 2021; Peng, Luo, Guan, Xie, Peng and Ma, 2022). The test and training sets were divided under the condition that the protein sequence identity between the two subsets was below 40%, and the sequence identity was computed using the MMseq2 (Steinegger and Söding, 2017). From this procedure, we obtained a high-quality training set containing 100,000 protein–ligand pairs and a test set composed of 100 proteins. We evaluated the superiority of the proposed `PIDiff` model by comparing the mean and median binding affinity of 100 molecules generated per protein pocket in the test set against various baseline models. The baseline models used for comparison with `PIDiff` were retrieved from recent studies, and descriptions of each model are provided in Supplementary Materials (3). The binding affinity was estimated by calculating the binding energy via the widely-used AutoDock Vina (Trott and Olson, 2010; Eberhardt, Santos-Martins, Tillack and Forli, 2021) docking tool.

The experimental results listed in Tab.1 verify the ability of `PIDiff` to generate molecules with high binding affinity to target proteins, as confirmed by various evaluation metrics. Using the AutoDock Vina tool, we obtained three versions of binding energy: *Vina Dock, Vina Min, and Vina Score*. The widely-used ***Vina Dock*** (Zhang, Zhang, Jin, Zhang, Hu, Shen, Cao, Du, Kang, Deng et al., 2023; Zhang and Liu, 2023; Schneuing, Du, Harris, Jamasb, Igashov, Du, Blundell, Lió, Gomes, Welling et al., 2022; Peng et al., 2022) metric yields the minimum binding energy ($\simeq$ maximum binding affinity) following docking, which fine-tunes the structure of the synthesized molecules and explores various binding poses. Furthermore, ***Vina Min*** assesses the binding affinity post-local energy minimization. Nonetheless, these methods(*Vina Dock*, *Vina Min*) may change the coordinates of the created atoms, possibly making them inappropriate for an unbiased evaluation of the generative model. Thus, we also utilized ***Vina Score***, which evaluates the binding affinity while preserving the atomic coordinates as predicted by the model. ***High Affinity*** denotes the proportion of generated molecules that exhibit superior binding to each test protein compared with a reference molecule that is bound to a protein pocket on the test set. ***SR (Success Rate)*** is defined as the success rate, representing the proportion of the 100 protein pockets in the test set where at least one generated molecule exhibits stronger binding affinity than the ligand bound in the test set pocket. When we reiterate that these generative models have the responsibility to create structures of molecules that could potentially become new drugs, it is crucial not only to measure the binding affinity of all generated molecules but also to analyze the binding affinity of molecules within the synthesizable range. To this end, we adopted a threshold for the generally synthesizable SA (synthetic accessibility) values mentioned in (Popova, Isayev and Tropsha, 2018). We measured the binding affinity of molecules generated by each model that surpassed the SA threshold, defined as ***Vina Score^{SA}***. Finally, we adopted the PoseBusters Buttenschoen, Morris and Deane (2024) test suite to evaluate the physical and chemical validity of the generated molecules. This benchmark tool performs a total of 12 tests from three perspectives: chemical validity and consistency, intramolecular validity, and intermolecular validity. A generated molecule that passes all 12 tests is defined as a valid molecule. Additionally, the performance

Metric Method	VinaScore (↓)		VinaMin (↓)		VinaDock (↓)		HighAffinity (↑)		VinaScore^{SA}(↓)		SR(↑)	Valid^{PB}(↑)
	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.		
TestSet	-6.36	-6.46	-6.71	-6.49	-7.45	-7.26	-	-	-6.28	-6.34	-	93.0%
AR	-5.75	-5.64	-6.18	-5.88	-6.75	-6.62	0.379	0.310	-5.59	-5.48	74.7%	68.2%
liGAN	N/A	N/A	N/A	N/A	-6.33	-6.2	0.211	0.111	N/A	N/A	68.4%	N/A
GraphBP	N/A	N/A	N/A	N/A	-4.80	-4.70	0.142	0.067	N/A	N/A	57.1%	N/A
Pocket2Mol	-5.14	-4.70	-6.42	5.82	-7.15	-6.79	0.483	0.510	-5.12	-5.48	88.7%	**69.5%**
DiffSBDD	52.78	44.57	16.45	0.49	-6.65	-7.15	0.452	0.454	51.53	43.46	83.0%	1.5%
TargetDiff	-5.47	-6.30	-6.64	-6.83	-7.80	-7.91	0.579	0.625	-5.31	-6.13	91.9%	52.5%
ResGen	13.79	5.75	-1.53	-3.36	-4.90	-5.26	0.232	0.000	13.73	5.75	40.7%	14.6%
PIDiff [Ours]	**-6.58**	**-7.37**	**-7.52**	**-7.57**	**-8.10**	**-8.35**	**0.641**	**0.666**	**-6.03**	**-6.89**	**100%**	58.1%

Table 1
Evaluation of generated molecules in terms of binding affinity by our model(PIDiff) and other baselines for proteins from CrossDocked testset.

for various metrics that characterize the molecular properties is provided in the supplementary materials (8).

As listed in Table 1, the proposed PIDiff, substantially outperforms the comparison models across the six evaluation metrics. Compared with the previous state-of-the-art Targetdiff model, PIDiff shows a notable improvement with a 21% increase in *Vina Score*, 12% in *High Affinity*, 13% in *Vina Score*^{SA}, and 9% in *SR*. Notably, as seen in the *SR* metric, our model can generate molecules that exhibit superior binding to all protein pockets in the test set compared with the reference molecules. These results indicate that prior chemical knowledge (that is, the complex structures of a protein and ligand should be located at a minimum in the binding free-energy landscape) is crucial for rational molecule design. This is particularly evident when comparing our model to Targetdiff, which similarly uses a diffusion generative model to generate molecules. The ablation case of the PIDiff model, which does not consider physicochemical principles, corresponds to the performance of TargetDiff. Additionally, an interesting observation is that for some comparison models, the binding energy (*Vina Score*) between the generated molecules and the target protein can result in positive(+) values. This indicates unrealistic binding poses, and docking of these binding poses yields negative(-) values for the binding energy (*Vina Dock*). Consequently, despite the generative model producing molecules with unrealistic binding poses, the *Vina Dock* metric adjusts them to realistic binding poses for calculating the binding energy. This observation highlights that the evaluation based on the *Vina Score*, which calculates the binding energy with fixed molecular positions inferred by the generative model, offers a fair assessment of the model performance.

Regarding the Valid^{PB} metric, PIDiff's performance ranks third, following Pocket2Mol and AR. According to Supplementary Material (9), the performance gap between our model and the Pocket2Mol (or AR) models is primarily due to many molecules failing the bond angle test, one of the 12 tests conducted. For models like Pocket2Mol and AR, the molecule generation process adopts an autoregressive sampling approach, where atoms are sequentially predicted to form a complete molecule. This process ensures that the geometric rationality among previously predicted atoms is maintained when predicting the next atom. Consequently, many molecules pass the bond angle test in the PoseBusters suite. However, our model employs a non-autoregressive sampling method, inferring an entire molecule at once rather than predicting atoms sequentially. As a result, it shows comparatively lower performance in the bond angle test than autoregressive sampling models. However, autoregressive sampling models are prone to cumulative errors and exposure bias due to discrepancies between the training and sampling processes Xiao, Wu, Guo, Li, Zhang, Qin and Liu (2023). Consequently, they perform significantly worse than PIDiff in all evaluation metrics except for the Valid^{PB} metric. This indicates that these models do not generate molecules with strong binding affinities to the target protein, making them unsuitable as drug candidates. Therefore, autoregressive models are not ideal for generative tasks in structure-based drug design.

3.2. Stability of generated molecules in terms of binding energy

Docking tools widely used in SBDD typically return the optimal pose and structure of an input ligand through various optimization processes(Meng, Zhang, Mezei and Cui, 2011). Inspired by molecular docking simulation, we conducted experiments to verify the stability of molecules generated by our model from an energy perspective. This assessment involved comparisons of the molecular structures inferred by the generative model with those optimized through docking simulations. If the molecular conformations before and after docking have a large difference, it implies that the molecular structure and binding pose generated by the model was chemically unfavorable for binding with the target protein, leading to substantial adjustments to the input molecular structure during docking. In such cases, the generative model may be insufficient for producing

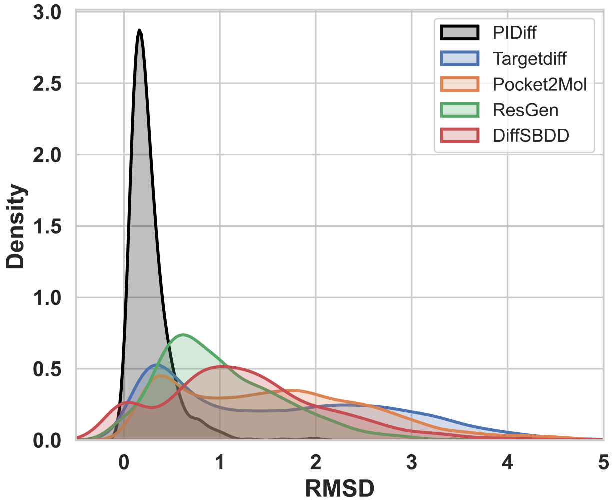


Figure 2: Distribution of the change in structure difference between pre-docking and post-docking for the generated molecule

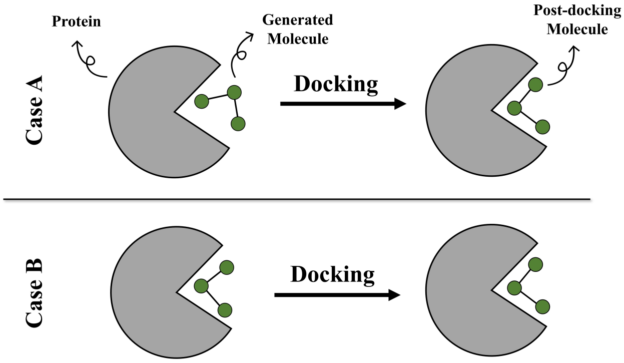


Figure 3: Case.A indicates that there is a relatively large difference in conformation between the generated molecule and the post-docking molecule. This means that significant adjustments to the generated molecule were made during docking to achieve an ideal protein-ligand binding. Consequently, it implies that the generated molecule has a structure and pose disadvantageous for binding to the protein. In contrast, Case.B shows that the difference in conformation between the generated molecule and the post-docking molecule is relatively small. This implies that not many adjustments were needed for the generated molecule during docking to achieve ideal protein-ligand binding. Consequently, it suggests that the generated molecule has a structure and pose advantageous for binding to the protein.

molecular conformations that are reasonable from a free-energy perspective. A brief example illustrating this aspect is described in Fig.3. Considering the interpretability of this evaluation method, the experimental results shown in Fig.2 allow us to assess the stability of molecules generated by the proposed model, including comparison models.

The experimental result showcass the distribution of conformational differences in molecules generated by each model before and after docking. The conformational differences were obtained by calculating the root mean square deviation (RMSD) of molecule positions before and after docking. The molecules generated by the proposed `PIDiff` model exhibit substantially less structural change after docking than those generated by five comparison models. This result clearly indicates that our model can infer the most favorable molecules for binding. Moreover, it highlights the importance of not only learning the distribution of molecular positions and structures during model training but also incorporating prior physicochemical knowledge into the generative model.

For a deeper analysis, we conducted several case studies. We visualized molecules generated for specific protein pockets both in their original form and after docking. For concise notation, we will henceforth refer to the undocked generated molecules (original form) as pre-docking molecules, and those subjected to docking as post-docking molecules. The protein–ligand complexes were categorized into three groups as molecules: 1) obtainable from the CrossDocked2020 dataset (reference), 2) generated by the proposed `PIDiff` model, and 3) generated by the latest model, ResGen (Zhang et al., 2023). In Fig.4, the molecules in cyan color spectrum indicate pre-docking molecules, while those in red color spectrum signify post-docking molecules. The molecules generated by the proposed `PIDiff` model exhibit negligible conformational differences between pre- and post-docking. In contrast, molecules generated by ResGen show significant structural changes after docking, implying the need for considerable fine-tuning for favorable binding with the protein. Furthermore, this case study provides a visual rationale for the distribution of low RMSD value observed for `PIDiff` in the previous experiment (Fig.2), as well as the distribution of high RMSD value noted for other comparative models. Additionally, the visualization results from this case study underscore the need for a cautious approach when evaluating the performance of target-aware 3D molecule generation models based on the outcomes (*Vina Dock*) of docking simulations. Additional explanations regarding the molecules presented in the Fig.4 are available in Supplementary Material (7).

3.3. Applicability of generated molecules generated to target proteins of various diseases

While achieving high performance on benchmark sets such as CrossDocked2020 is undoubtedly important, it is essential to remember that the target protein-aware 3D molecule generative models, including the model proposed in this study, are intended to accelerate the SBDD task. Accordingly, to evaluate the practical utility of 3D molecular generation, it is crucial to verify whether these models can design reasonable molecules for proteins that are actual drug targets, not just those proteins included in the benchmark dataset. To bridge the gap between artificial intelligence (AI)

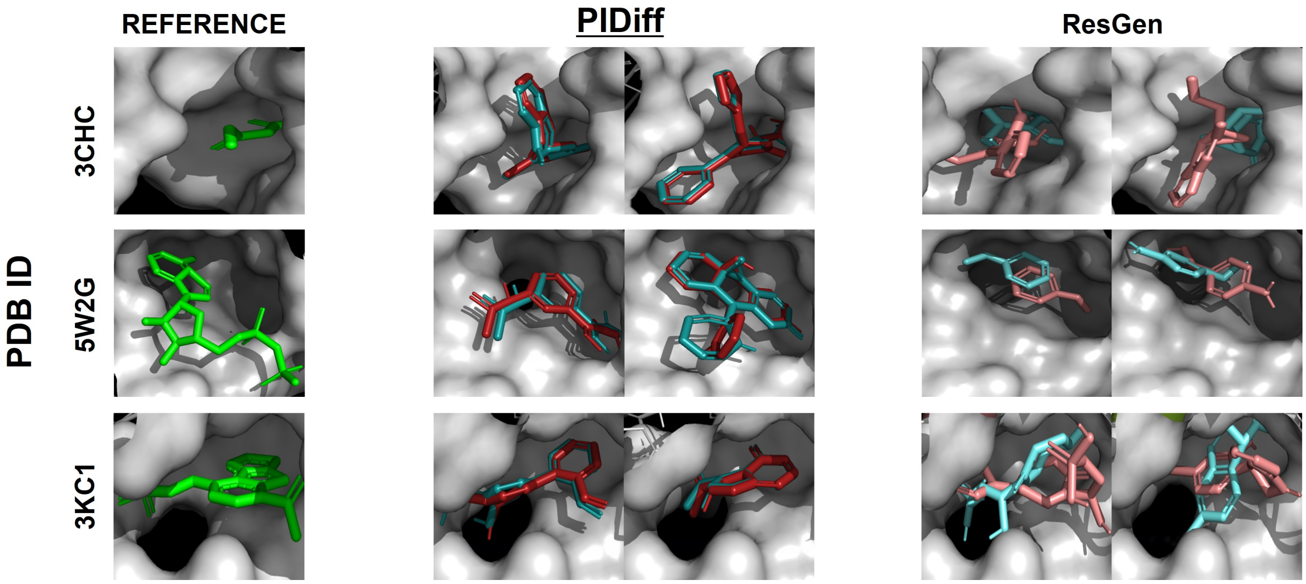


Figure 4: Visualization of binding pose of reference molecules and molecules generated by `PIDiff`, and ResGen on protein 3CHC, 5W2G, 3KC1(PDB ID). Among the generated molecules, those in the cyan color spectrum represent pre-docking molecules, while those in the red color spectrum signify post-docking molecules.

generative models and practical SBDD, we conducted experiments to explore the feasibility of generating candidate compounds for inhibitors of kinases, which play a major role in the oncogenesis and metastasis of various cancers (Bhullar, Lagarón, McGowan, Parmar, Jha, Hubbard and Rupasinghe, 2018).

For a proper evaluation, we carried out an extensive literature review related to kinase inhibitors, selecting several representative and experimentally active target proteins and drugs (Knowles, Murray-Rust, Kjær, Scott, Hanrahan, Santoro, Ibáñez and McDonald, 2006; Wilson, Agafonov, Hoemberger, Kutter, Zorba, Halpin, Buosi, Otten, Waterman, Theobald et al., 2015; Davis, Hunt, Herrgard, Ciceri, Wodicka, Pallares, Hocker, Treiber and Zarrinkar, 2011; Shaw, Friboulet, Leshchiner, Gainor, Bergqvist, Brooun, Burke, Deng, Liu, Dardaei et al., 2016; Qiu, Tarrant, Choi, Sathyamurthy, Bose, Banjade, Pal, Bornmann, Lemmon, Cole et al., 2008; Hammam, Saez-Ayala, Rebuffet, Gros, Lopez, Hajem, Humbert, Baudelet, Audebert, Betzi et al., 2017). The protein crystal structures were obtained from PDB (Berman et al., 2000) database. The proteins selected for verification were RET, ERBB2, ABL1, ALK, EGFR, and KIT. The PDB IDs for the structures of each protein are described in the Supplementary Material (5). We measured the binding affinity of 1000 molecules generated by the trained generative model per protein pocket using the *Vina score*, as listed in Tab. 1. We verified the real-world competitiveness of the molecules generated by `PIDiff` by comparing the distribution of binding affinities of 1000 molecules generated for each protein by our proposed model, with the binding affinities of compounds known to bind to those proteins as illustrated in Fig. 5.

Observing the Vina score patterns of molecules generated by `PIDiff` across all six cases, considering the reference drugs are compounds actually binding to the proteins, it can be inferred that the generated molecules possess the capacity to produce numerous molecular structures capable of functioning as kinase inhibitors. A quantitative analysis reveals that the proportions of molecules with superior binding affinity to the reference drugs for RET, ERBB2, ABL1, EGFR, ALK, and KIT are 36%, 13%, 38%, 84%, 35%, and 52%, respectively. Notably, the most of molecules generated for the EGFR protein demonstrate superior binding affinity to the binding drug, which is a remarkable result. Based on this case study, our proposed model, `PIDiff`, suggests that it not only achieves high performance on benchmark datasets but also possesses the capability to generate molecules with competitive binding affinities for target proteins aimed at treating various diseases in practice.

However, when broadening the perspective on experiments within real-world settings, the possibility of the absence of the target protein structure cannot be overlooked. Considering that, out of the 250 million proteins with known sequences available through the UniProt (Universal Protein Resource) (uni, 2023), only approximately 210,000 protein structure entries are available for 67,000 unique proteins via the PDB (Berman et al., 2000) database, the mention of the above possibility appears to be reasonable. To address this bottleneck, we conducted additional experiments incorporating the AlphaFold (Jumper et al., 2021) system, which offers unprecedented accuracy in protein structure prediction.

For this purpose, we selected FLT3 and VEGFR2 proteins, which are experimentally known to be active with the

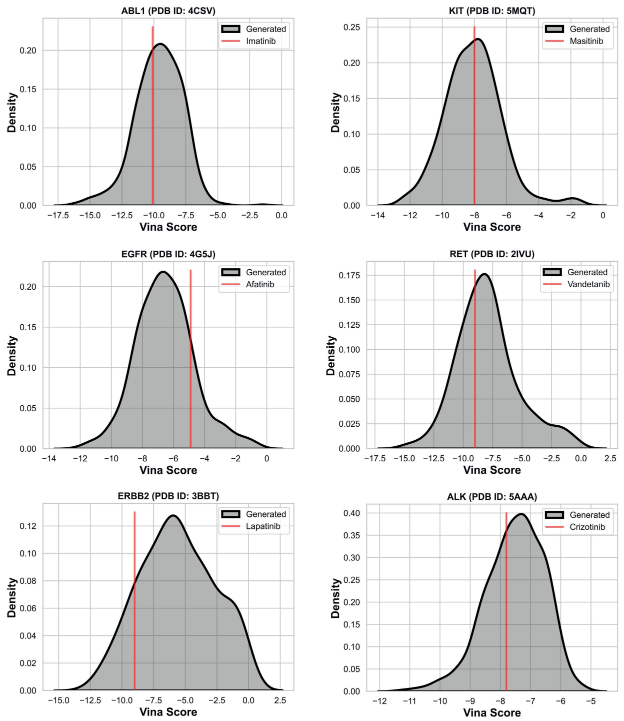


Figure 5: The distribution of Vina Score for molecules generated by `PIDiff` for six target proteins of kinase inhibitors, for which structures are obtainable from the PDB database. The red bars represent the Vina Score of drugs known to actually bind to each target protein.

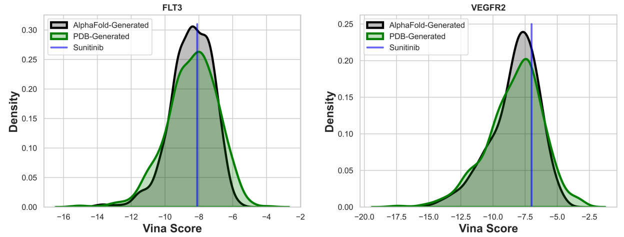


Figure 6: The distribution of Vina Score for molecules generated by `PIDiff` for two target proteins of kinase inhibitors. The Vina scores for molecules generated using both the AlphaFold version and the PDB version of the protein structures were measured against the PDB version of the protein structure.

target drug sunitinib (Roskoski Jr, 2007). These proteins are well-studied as target proteins for kinase inhibitors (Shibuya, 2010). We extracted the sequence information for the ATP binding sites of these two proteins and used AlphaFold (Varadi, Anyango, Deshpande, Nair, Natassia, Yordanova, Yuan, Stroe, Wood, Laydon et al., 2022; Varadi, Bertoni, Magana, Paramval, Pidruchna, Radhakrishnan, Tsenkov, Nair, Mirdita, Yeo et al., 2024) to obtain the structural information for these sequences. Subsequently, we generated 1000 molecules for each protein using the same manner as the experiment for Figure 5. The Vina scores of these molecules were then compared to the Vina score of the reference drug sunitinib. Additionally, to verify the reliability of the molecules generated using the AlphaFold-predicted protein structures, we compared their Vina scores with those of molecules generated based on experimentally determined structures. In other words, we obtained PDB files (FLT3: 4RT7, VEGFR2: 1YWN) containing the actual structural information for the sequences used to obtain the protein structures via AlphaFold, and then generated 1000 molecules for these experimentally determined structures as well. The distribution of Vina scores for the 1000 molecules generated for the AlphaFold version of the protein is shown in gray, while the distribution for the 1000 molecules generated for the PDB version of the protein is shown in green. The Vina score of the reference drug, sunitinib, is indicated by a blue bar. Among the molecules generated for the AlphaFold version of the protein, the proportions of molecules with stronger binding affinities than sunitinib were 58% for FLT3 and 74% for VEGFR2 (In this comparison, the Vina scores for both the generated molecules and sunitinib were calculated using the PDB version of the protein.). As shown in the figure 6, the distribution of Vina score for the molecules generated using the structure obtained from the PDB and those generated using the structure obtained from AlphaFold are not significantly different. This indicates that generating new molecules using structures predicted by AlphaFold is a valid approach even when the exact structure of a protein is unknown. In summary, we can conclude that our model, PIDiff, has the capability to generate competitive molecules for target drugs using structures obtained from AlphaFold.

3.4. Selectivity of generated molecules

Kinases that regulate various signaling pathways in human diseases, including cancer, vascular diseases, diabetes, inflammation, and degenerative diseases, have become attractive targets for new drug development. Most kinase inhibitors bind to the ATP(Adenosine TriPhosphate) binding site within the kinase catalytic domain. However, the high structural conservation across the ATP binding sites of various kinases can lead to off-target effects (Attwood, Fabbro, Sokolov, Knapp and Schiöth, 2021), with kinase inhibitors binding to kinases other than the target protein. This can cause undesirable side effects, ultimately undermining the clinical effectiveness of a drug (Grünwald, Heinzer and Fiedler, 2007; Wolter, Stefan, Decallonne, Dumez, Bex, Carmeliet and Schöffski, 2008) and remaining a challenging issue.

To address this challenge, we considered the hypothesis that molecules generated based on the structure of a specific protein have a relatively weak binding to proteins with different structures and aimed to validate this hypothesis using our model. For the experiment, we selected sorafenib, a kinase inhibitor targeting the B-Raf protein approved by the FDA(Food and Drug Administration) for treating liver and kidney cancer, and identified its off-target proteins (PKB, ERK1, PIM1, MEK1, CDK1, PKC, IGFR1, EGFR, c-MET,

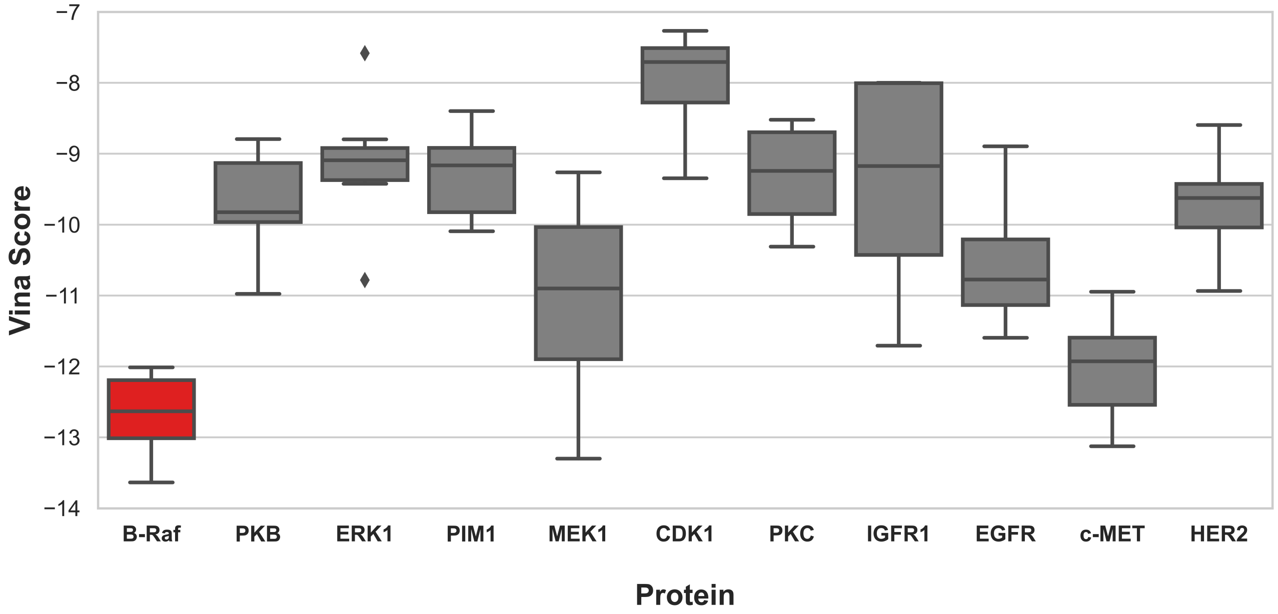


Figure 7: Box plot of Vina scores for molecules generated for the B-Raf protein, shown for on-target (red box) and off-targets (gray box). Superior Vina scores for on-target and less favorable scores for off-targets imply excellent selectivity.

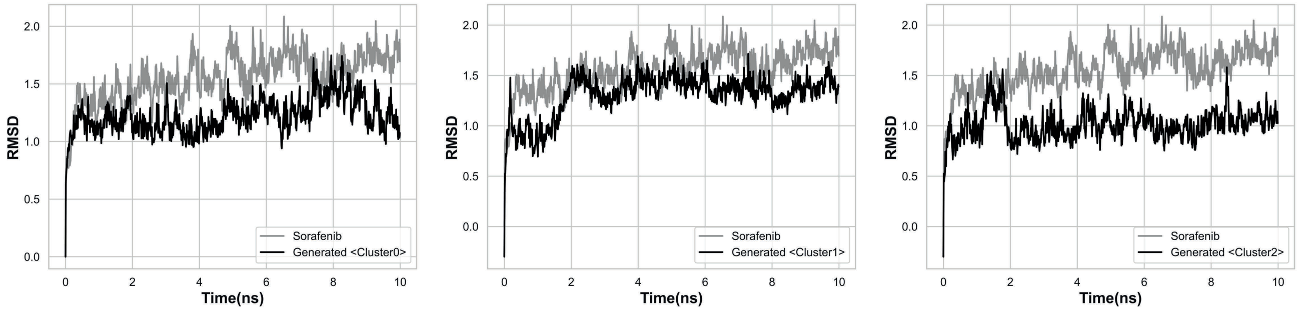


Figure 8: Changes in the distance between the protein and ligand over time during MD simulations for the B-Raf protein with two molecules (Sorafenib, generated molecule)

HER2) through our curation, drawing on previous studies (Karaman, Herrgard, Treiber, Gallant, Atteridge, Campbell, Chan, Ciceri, Davis, Edeen et al., 2008). The PDB IDs for the structures of each protein are described in Supplementary Material (6).

After calculating the Vina score for the B-Raf protein, we selected molecules with binding affinities superior to sorafenib and that surpassed the SA value threshold used during the calculation of *Vina Score*^{SA} in Section 3.1. These selected molecules were then calculated for their vina scores against 10 off-targets. The distribution of Vina scores for the selected molecules, both on-target (B-Raf) and off-targets, are shown in Fig.7. The molecules generated for B-Raf tend to bind less strongly to off-target proteins with structures different from the on-target protein. This observation aligns with our hypothesis to some extent. Overall, molecules generated by `PIDiff` for a target protein pocket can achieve selective binding by not adhering to other proteins. This demonstrates the potential applicability of our model for SBDD, suggesting its utility in achieving targeted binding with high selectivity.

3.5. Stability of generated molecules from thermodynamics perspective

A molecular docking simulation has exceptional efficiency in measuring the binding affinity of various molecules with proteins and rapidly optimizing those molecules. For fast processing, it superficially handles various aspects such as the solvation energy without properly constructing the environment of real biological systems, consequently sacrificing accuracy (Śledź and Caflisch, 2018; Kitchen, Decornez, Furr and Bajorath, 2004; Carlson, 2002). To overcome these limitations, *in-silico* based drug design traditionally employs MD simulation, which models the thermodynamic behavior of individual particles over time, allowing for a detailed understanding of protein-ligand interactions within a realistic biological system environment (e.g., aqueous state) on a temporal basis. However, to date, attempts to evaluate the performance of generative models using MD simulation have not been made in molecular generation research. This gap represents one of the reasons why the divide between actual drug design and molecular generation studies has not been narrowed.

Therefore, we propose a novel experimental framework in our molecular generation research for SBDD that involves validating molecules generated by the generative model through MD simulations. In this framework, after the broad and general performance of the generative model has been assessed through a docking simulation, a more detailed evaluation is conducted by comparing the MD simulation results of molecules generated by the model with the behavior of actual drugs. This comparison enables a rigorous assessment of the performance of individual molecules.

We performed clustering based on structural similarity among the generated molecules and selected one molecule from each cluster for MD simulation. This approach aims to assess the results of the MD simulations by selecting representative molecules from each cluster, thereby allowing for an indirect evaluation of the entire set of generated molecules. We used ECFP (Rogers and Hahn, 2010) fingerprints to represent the molecular structures and employed k-means clustering with $k = 3$. To visualize the results of the k-means clustering, we plotted the ECFP fingerprints using PCA (Principal Component Analysis) and displayed the clusters as shown supplementary materials (10). From each cluster, we selected one molecule using the same criteria applied in Section 3.4 and performed MD simulations for these representative molecules. The target protein was selected as the B-Raf protein, which was utilized in section 3.4, and MD simulations were performed on Sorafenib, known to bind to the B-Raf protein. The preprocessing of the protein and ligand for simulation was supported by the CHARMM-GUI (Jo, Kim, Iyer and Im, 2008) tool, and MD simulations were conducted over 10 ns using the ABMER (Salomon-Ferrer, Case and Walker, 2013; Arantes, Polêto, Pedebos and Ligabue-Braun, 2021) tool. The detailed procedure is outlined in our source code. From the simulation results, we measured the distance (i.e., RMSD) between the protein pocket and ligand across all trajectory files over time for both for sorafenib and for the three generated molecules. As shown in Fig.8, compared with the RMSD over time for the reference sorafenib drug, the three molecular structures generated by `PIDiff` exhibits a stable fluctuation range over time, displaying an overall lower RMSD profile. An increasing RMSD indicates either an unstable ligand structure or weaker binding affinity with the protein (De Vivo et al., 2016). Thus, we can conclude that the molecular structure generated by our model can stably bind to the target protein. These experimental results are a critical indication that molecules generated by generative models, such as the proposed `PIDiff`, possess competitive binding capabilities to target proteins in real physiological environments, even when compared with approved drugs.

4. Conclusion

In this study, we propose the `PIDiff` model that can generate rational and realistic drugs capable of binding to the structure of a target protein pocket. This research stems from the necessity of considering not only geometric information of complex structures but also the intrinsic principles of binding between proteins and ligands. To assess our model, we first evaluated the binding affinity of molecules generated for target proteins on a benchmark dataset. Then, we investigated the structures of target proteins that caused real diseases to determine the practical applicability of the model. Additionally, we verified whether the molecules generated by our model could exhibit selectivity, a critical challenge in SBDD. Finally, we validated whether the molecules produced by our model demonstrated stable binding patterns within biological systems when compared with real drugs. Overall, experiments from various perspectives have confirmed the potential of `PIDiff` as a valuable tool for accelerating drug development.

While generating molecules that stably bind to target proteins is a crucial step, it is not the endpoint in drug development. To be viable drug candidates, the generated molecules must also meet various criteria, such as ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) Ferreira and Andricopulo (2019). Although addressing these issues is essential, our study primarily focused on generating molecules that bind to target proteins, rather than optimizing for efficacy or other drug-like properties. Consequently, we did not explore these aspects in depth. To address these challenges, we believe that latent space optimization Rombach, Blattmann, Lorenz, Esser and Ommer (2022); Wallace, Gokul, Ermon and Naik (2023) or manipulation Park, Kwon, Choi, Jo and Uh (2023) and guidance sampling techniques Song, Sohl-Dickstein, Kingma, Kumar, Ermon and Poole (2020); Dhariwal and Nichol (2021) will be key solutions in the future. In particular, applying guidance sampling methods, which are gaining attention in various condition generation tasks such as the inverse problem, to our drug development tasks could be highly promising. This approach could enable the generation of molecules with desired properties for structure-based drug design (SBDD) tasks, making it an exciting area of future research.

References

, 2023. Uniprot: the universal protein knowledgebase in 2023. Nucleic Acids Research 51, D523–D531.

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

Arantes, P.R., Polêto, M.D., Pedebos, C., Ligabue-Braun, R., 2021. Making it rain: cloud-based molecular simulations for everyone. Journal of Chemical Information and Modeling 61, 4852–4856.

Attwood, M.M., Fabbro, D., Sokolov, A.V., Knapp, S., Schiöth, H.B., 2021. Trends in kinase drug discovery: targets, indications and inhibitor design. Nature Reviews Drug Discovery 20, 839–861.

Atz, K., Grisoni, F., Schneider, G., 2021. Geometric deep learning on molecular representations. Nature Machine Intelligence 3, 1023–1032.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein data bank. Nucleic acids research 28, 235–242.

Bhullar, K.S., Lagarón, N.O., McGowan, E.M., Parmar, I., Jha, A., Hubbard, B.P., Rupasinghe, H.V., 2018. Kinase-targeted cancer therapies: progress, challenges and future directions. Molecular cancer 17, 1–20.

Bitencourt-Ferreira, G., Veit-Acosta, M., de Azevedo, W.F., 2019. Van der waals potential in protein complexes. Docking Screens for Drug Discovery , 79–91.

Blundell, T.L., 1996. Structure-based drug design. Nature 384, 23.

Bronowska, A.K., 2011. Thermodynamics of ligand-protein interactions: implications for molecular design, in: Thermodynamics-Interaction Studies-Solids, Liquids and Gases. IntechOpen.

Buttenschoen, M., Morris, G.M., Deane, C.M., 2024. Posebusters: Ai-based docking methods fail to generate physically valid poses or generalise to novel sequences. Chemical Science .

Carlson, H.A., 2002. Protein flexibility and drug design: how to hit a moving target. Current opinion in chemical biology 6, 447–452.

Cheng, Y., Gong, Y., Liu, Y., Song, B., Zou, Q., 2021. Molecular design in drug discovery: a comprehensive review of deep generative models. Briefings in bioinformatics 22, bbab344.

Davis, M.I., Hunt, J.P., Herrgard, S., Ciceri, P., Wodicka, L.M., Pallares, G., Hocker, M., Treiber, D.K., Zarrinkar, P.P., 2011. Comprehensive analysis of kinase inhibitor selectivity. Nature biotechnology 29, 1046–1051.

De Vivo, M., Masetti, M., Bottegoni, G., Cavalli, A., 2016. Role of molecular dynamics and related methods in drug discovery. Journal of medicinal chemistry 59, 4035–4061.

Dhariwal, P., Nichol, A., 2021. Diffusion models beat gans on image synthesis. Advances in neural information processing systems 34, 8780–8794.

Eberhardt, J., Santos-Martins, D., Tillack, A.F., Forli, S., 2021. Autodock vina 1.2. 0: New docking methods, expanded force field, and python bindings. Journal of chemical information and modeling 61, 3891–3898.

Ferreira, L.L., Andricopulo, A.D., 2019. Admet modeling approaches in drug discovery. Drug discovery today 24, 1157–1165.

Fischer, M., Coleman, R.G., Fraser, J.S., Shoichet, B.K., 2014. Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. Nature chemistry 6, 575–583.

Francoeur, P.G., Masuda, T., Sunseri, J., Jia, A., Iovanisci, R.B., Snyder, I., 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, 4200–4215.

Grünwald, V., Heinzer, H., Fiedler, W., 2007. Managing side effects of angiogenesis inhibitors in renal cell carcinoma. Oncology Research and Treatment 30, 519–524.

Guan, J., Qian, W.W., Ma, W.Y., Ma, J., Peng, J., 2021. Energy-inspired molecular conformation optimization, in: international conference on learning representations.

Guan, J., Qian, W.W., Peng, X., Su, Y., Peng, J., Ma, J., 2023. 3d equivariant diffusion for target-aware molecule generation and affinity prediction, in: International Conference on Learning Representations.

Hammam, K., Saez-Ayala, M., Rebuffet, E., Gros, L., Lopez, S., Hajem, B., Humbert, M., Baudelet, E., Audebert, S., Betzi, S., et al., 2017. Dual protein kinase and nucleoside kinase modulators for rationally designed polypharmacology. Nature Communications 8, 1420.

Hansson, T., Oostenbrink, C., van Gunsteren, W., 2002. Molecular dynamics simulations. Current opinion in structural biology 12, 190–196.

Hao, Z., Liu, S., Zhang, Y., Ying, C., Feng, Y., Su, H., Zhu, J., 2022. Physics-informed machine learning: A survey on problems, methods and applications. arXiv preprint arXiv:2211.08064 .

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

Hoogeboom, E., Nielsen, D., Jaini, P., Forré, P., Welling, M., 2021. Argmax flows and multinomial diffusion: Learning categorical distributions. Advances in Neural Information Processing Systems 34, 12454–12465.

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

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

Jo, S., Kim, T., Iyer, V.G., Im, W., 2008. Charmm-gui: a web-based graphical user interface for charmm. Journal of computational chemistry 29, 1859–1865.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al., 2021. Highly accurate protein structure prediction with alphafold. Nature 596, 583–589.

Karaman, M.W., Herrgard, S., Treiber, D.K., Gallant, P., Atteridge, C.E., Campbell, B.T., Chan, K.W., Ciceri, P., Davis, M.I., Edeen, P.T., et al., 2008. A quantitative analysis of kinase inhibitor selectivity. Nature biotechnology 26, 127–132.

Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S., Yang, L., 2021. Physics-informed machine learning. Nature Reviews Physics 3, 422–440.

Kitchen, D.B., Decornez, H., Furr, J.R., Bajorath, J., 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. Nature reviews Drug discovery 3, 935–949.

Knowles, P.P., Murray-Rust, J., Kjær, S., Scott, R.P., Hanrahan, S., Santoro, M., Ibáñez, C.F., McDonald, N.Q., 2006. Structure and chemical inhibition of the ret tyrosine kinase domain. Journal of biological chemistry 281, 33577–33587.

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

Luque, J., Barril, X., 2012. Physico-chemical and computational approaches to drug discovery. 23, Royal society of chemistry.

Lyne, P.D., 2002. Structure-based virtual screening: an overview. Drug discovery today 7, 1047–1055.

MacArthur, M.W., Laskowski, R.A., Thornton, J.M., 1994. Knowledge-based validation of protein structure coordinates derived by x-ray crystallography and nmr spectroscopy. Current Opinion in Structural Biology 4, 731–737.

Meng, X.Y., Zhang, H.X., Mezei, M., Cui, M., 2011. Molecular docking: a powerful approach for structure-based drug discovery. Current computer-aided drug design 7, 146–157.

Moon, S., Zhung, W., Yang, S., Lim, J., Kim, W.Y., 2022. Pignet: a physics-informed deep learning model toward generalized drug–target interaction predictions. Chemical Science 13, 3661–3673.

Nichol, A.Q., Dhariwal, P., 2021. Improved denoising diffusion probabilistic models, in: International Conference on Machine Learning, PMLR. pp. 8162–8171.

Pacholczyk, M., Kimmel, M., 2011. Exploring the landscape of protein-ligand interaction energy using probabilistic approach. Journal of Computational Biology 18, 843–850.

Pagadala, N.S., Syed, K., Tuszynski, J., 2017. Software for molecular docking: a review. Biophysical reviews 9, 91–102.

Park, Y.H., Kwon, M., Choi, J., Jo, J., Uh, Y., 2023. Understanding the latent space of diffusion models through the lens of riemannian geometry. Advances in Neural Information Processing Systems 36, 24129–24142.

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

Popova, M., Isayev, O., Tropsha, A., 2018. Deep reinforcement learning for de novo drug design. Science advances 4, eaap7885.

Qiu, C., Tarrant, M.K., Choi, S.H., Sathyamurthy, A., Bose, R., Banjade, S., Pal, A., Bornmann, W.G., Lemmon, M.A., Cole, P.A., et al., 2008. Mechanism of activation and inhibition of the her4/erbb4 kinase. Structure 16, 460–467.

Rogers, D., Hahn, M., 2010. Extended-connectivity fingerprints. Journal of chemical information and modeling 50, 742–754.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022. High-resolution image synthesis with latent diffusion models, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 10684–10695.

Roskoski Jr, R., 2007. Sunitinib: a vegf and pdgf receptor protein kinase and angiogenesis inhibitor. Biochemical and biophysical research communications 356, 323–328.

Salomon-Ferrer, R., Case, D.A., Walker, R.C., 2013. An overview of the amber biomolecular simulation package. Wiley Interdisciplinary Reviews: Computational Molecular Science 3, 198–210.

Satorras, V.G., Hoogeboom, E., Welling, M., 2021. E (n) equivariant graph neural networks, in: International conference on machine learning, PMLR. pp. 9323–9332.

Schneuing, A., Du, Y., Harris, C., Jamasb, A., Igashov, I., Du, W., Blundell, T., Lió, P., Gomes, C., Welling, M., et al., 2022. Structure-based drug design with equivariant diffusion models. arXiv preprint arXiv:2210.13695 .

Shaw, A.T., Friboulet, L., Leshchiner, I., Gainor, J.F., Bergqvist, S., Brooun, A., Burke, B.J., Deng, Y.L., Liu, W., Dardaei, L., et al., 2016. Resensitization to crizotinib by the lorlatinib alk resistance mutation l1198f. New England Journal of Medicine 374, 54–61.

Shibuya, M., 2010. Tyrosine kinase receptor flt/vegfr family: its characterization related to angiogenesis and cancer. Genes & cancer 1, 1119–1123.

Shoichet, B.K., 2004. Virtual screening of chemical libraries. Nature 432, 862–865.

Śledź, P., Caflisch, A., 2018. Protein structure-based drug design: from docking to molecular dynamics. Current opinion in structural biology 48, 93–102.

Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B., 2020. Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:2011.13456 .

Steinegger, M., Söding, J., 2017. Mmseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nature biotechnology 35, 1026–1028.

Trott, O., Olson, A.J., 2010. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of computational chemistry 31, 455–461.

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al., 2022. Alphafold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic acids research 50, D439–D444.

Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., Tsenkov, M., Nair, S., Mirdita, M., Yeo, J., et al., 2024. Alphafold protein structure database in 2024: providing structure coverage for over 214 million protein sequences. Nucleic Acids Research 52, D368–D375.

Wallace, B., Gokul, A., Ermon, S., Naik, N., 2023. End-to-end diffusion latent optimization improves classifier guidance, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 7280–7290.

Whitebread, S., Hamon, J., Bojanic, D., Urban, L., 2005. Keynote review: in vitro safety pharmacology profiling: an essential tool for successful drug development. Drug discovery today 10, 1421–1433.

Wilson, C., Agafonov, R., Hoemberger, M., Kutter, S., Zorba, A., Halpin, J., Buosi, V., Otten, R., Waterman, D., Theobald, D., et al., 2015. Using ancient protein kinases to unravel a modern cancer drug's mechanism. Science 347, 882–886.

Wolter, P., Stefan, C., Decallonne, B., Dumez, H., Bex, M., Carmeliet, P., Schöffski, P., 2008. The clinical implications of sunitinib-induced hypothyroidism: a prospective evaluation. British journal of cancer 99, 448–454.

Xiao, Y., Wu, L., Guo, J., Li, J., Zhang, M., Qin, T., Liu, T.y., 2023. A survey on non-autoregressive generation for neural machine translation and beyond. IEEE Transactions on Pattern Analysis and Machine Intelligence .

Xie, L., Xie, L., Bourne, P.E., 2011. Structure-based systems biology for analyzing off-target binding. Current opinion in structural biology 21, 189–199.

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

Xu, M., Yu, L., Song, Y., Shi, C., Ermon, S., Tang, J., 2022. Geodiff: A geometric diffusion model for molecular conformation generation, in: International Conference on Learning Representations. URL: https://openreview.net/forum?id=PzcvxEMzvQC.

Zhang, O., Zhang, J., Jin, J., Zhang, X., Hu, R., Shen, C., Cao, H., Du, H., Kang, Y., Deng, Y., et al., 2023. Resgen is a pocket-aware 3d molecular generation model based on parallel multiscale modelling. Nature Machine Intelligence 5, 1020–1030.

Zhang, Z., Liu, Q., 2023. Learning subpocket prototypes for generalizable structure-based drug design. ICML .

Zhou, H.X., Gilson, M.K., 2009. Theory of free energy and entropy in noncovalent binding. Chemical reviews 109, 4092–4107.

Supplementary Material

1. Supplementary Material (1)

As mentioned in the Manuscript, to utilize diffusion for the 3D molecule generation task, it is necessary to model both the coordinates and types of atoms simultaneously. Coordinates are composed of continuous variables, while types are composed of discrete variables. To model these, Gaussian distribution (N) for coordinates and Categorical distribution (C) for types must be used respectively. Utilizing this, we have defined a forward process that injects noise into the coordinates and types of atoms, as shown in *Suppl.Eq.(1)*. Following the derivation process below, Eq.(2) can be derived.

$$\begin{aligned}
X_t &= \sqrt{\alpha_t} X_{t-1} + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}_{t-1} \\
&= \sqrt{\alpha_t}\left(\sqrt{\alpha_{t-1}} X_{t-2} + \sqrt{1-\alpha_{t-1}}\epsilon_{t-2}\right) + \sqrt{1-\alpha_t}\epsilon_{t-1} \\
&= \sqrt{\alpha_t\alpha_{t-1}} X_{t-2} + \sqrt{\alpha_t - \alpha_t\alpha_{t-1}}\epsilon_{t-2} + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}_{t-1} \\
&= \sqrt{\alpha_t\alpha_{t-1}} X_{t-2} + \sqrt{\sqrt{\alpha_t - \alpha_t\alpha_{t-1}}^2 + \sqrt{1-\alpha_t}^2}\boldsymbol{\epsilon}_{t-2} \\
&= \sqrt{\alpha_t\alpha_{t-1}} X_{t-2} + \sqrt{\alpha_t - \alpha_t\alpha_{t-1} + 1 - \alpha_t}\boldsymbol{\epsilon}_{t-2} \\
&= \sqrt{\alpha_t\alpha_{t-1}} X_{t-2} + \sqrt{1 - \alpha_t\alpha_{t-1}}\epsilon_{t-2} \\
&= \ldots \\
&= \sqrt{\prod_{i=1}^{t}\alpha_i} X_0 + \sqrt{1 - \prod_{i=1}^{t}\alpha_i}\boldsymbol{\epsilon}_0 \\
&= \sqrt{\bar{\alpha}_t} X_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_0 \\
&\sim \mathcal{N}\left(X_t;\ \sqrt{\bar{\alpha}_t} X_0, (1-\bar{\alpha}_t)\mathbf{I}\right)
\end{aligned} \tag{1}$$

$$\begin{aligned}
V_t &= \alpha_t V_{t-1} + (1-\alpha_t)/K \\
&= \alpha_t\left(\alpha_{t-1} V_{t-2} + (1-\alpha_{t-1})/K\right) + (1-\alpha_t)/K \\
&= \alpha_t\alpha_{t-1} V_{t-2} + \alpha_t(1-\alpha_{t-1})/K + (1-\alpha_t)/K \\
&= \alpha_t\alpha_{t-1} V_{t-2} + (1-\alpha_t\alpha_{t-1})/K \\
&= \ldots \\
&= \prod_{i=1}^{t}\alpha_i V_0 + (1 - \prod_{i=1}^{t}\alpha_i)/K \\
&= \bar{\alpha}_t V_0 + (1-\bar{\alpha}_t)/K \\
&\sim \mathcal{C}\left(V_t \mid \bar{\alpha}_t V_0 + (1-\bar{\alpha}_t)/K\right)
\end{aligned} \tag{2}$$

Additionally, starting with the Bayes rule expansion, the derivation of the $q\left(X_{t-1} \mid X_t, X_0\right)$, $q\left(V_{t-1} \mid V_t, V_0\right)$ in the manuscript's Eq.(3) can be performed as *Suppl.Eq.(3)* and *Suppl.Eq.(4)*, in accordance with previous studies [1, 2, 3, 4].

$$
\begin{aligned}
q\left(X_{t-1} \mid X_t, X_0\right) &= \frac{q\left(X_t \mid X_{t-1}, X_0\right) q\left(X_{t-1} \mid X_0\right)}{q\left(X_t \mid X_0\right)} \\
&= \frac{\mathcal{N}\left(X_t;\ \sqrt{\alpha_t} X_{t-1}, \left(1-\alpha_t\right) \mathbf{I}\right) \mathcal{N}\left(X_{t-1};\ \sqrt{\bar{\alpha}_{t-1}} X_0, \left(1-\bar{\alpha}_{t-1}\right) \mathbf{I}\right)}{\mathcal{N}\left(X_t;\ \sqrt{\bar{\alpha}_t} X_0, \left(1-\bar{\alpha}_t\right) \mathbf{I}\right)} \\
&\propto \exp\left\{-\left[\frac{\left(X_t-\ \sqrt{\alpha_t} X_{t-1}\right)^2}{2\left(1-\alpha_t\right)}+\frac{\left(X_{t-1}-\ \sqrt{\bar{\alpha}_{t-1}} X_0\right)^2}{2\left(1-\bar{\alpha}_{t-1}\right)}-\frac{\left(X_t-\ \sqrt{\bar{\alpha}_t} X_0\right)^2}{2\left(1-\bar{\alpha}_t\right)}\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\frac{\left(-2 \sqrt{\alpha_t} X_t X_{t-1}+\alpha_t X_{t-1}^2\right)}{1-\alpha_t}+\frac{\left(X_{t-1}^2-2 \sqrt{\bar{\alpha}_{t-1}} X_{t-1} X_0\right)}{1-\bar{\alpha}_{t-1}}+C\left(X_t, X_0\right)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[-\frac{2 \sqrt{\alpha_t} X_t X_{t-1}}{1-\alpha_t}+\frac{\alpha_t X_{t-1}^2}{1-\alpha_t}+\frac{X_{t-1}^2}{1-\bar{\alpha}_{t-1}}-\frac{2 \sqrt{\bar{\alpha}_{t-1}} X_{t-1} X_0}{1-\bar{\alpha}_{t-1}}\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\left(\frac{\alpha_t}{1-\alpha_t}+\frac{1}{1-\bar{\alpha}_{t-1}}\right) X_{t-1}^2-2\left(\frac{\sqrt{\alpha_t} X_t}{1-\alpha_t}+\frac{\sqrt{\bar{\alpha}_{t-1}} X_0}{1-\bar{\alpha}_{t-1}}\right) X_{t-1}\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\frac{1-\bar{\alpha}_t}{\left(1-\alpha_t\right)\left(1-\bar{\alpha}_{t-1}\right)} X_{t-1}^2-2\left(\frac{\sqrt{\alpha_t} x_t}{1-\alpha_t}+\frac{\sqrt{\bar{\alpha}_{t-1}} x_0}{1-\bar{\alpha}_{t-1}}\right) X_{t-1}\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left(\frac{1-\bar{\alpha}_t}{\left(1-\alpha_t\right)\left(1-\bar{\alpha}_{t-1}\right)}\right)\left[X_{t-1}^2-2 \frac{\left(\frac{\sqrt{\alpha_t} X_t}{1-\alpha_t}+\frac{\sqrt{\bar{\alpha}_{t-1}} x_0}{1-\bar{\alpha}_{t-1}}\right)\left(1-\alpha_t\right)\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t} X_{t-1}\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left(\frac{1}{\frac{\left(1-\alpha_t\right)\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t}}\right)\left[X_{t-1}^2-2 \frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right) X_t+\ \sqrt{\bar{\alpha}_{t-1}}\left(1-\alpha_t\right) X_0}{1-\bar{\alpha}_t} X_{t-1}\right]\right\} \\
&\propto \mathcal{N}(X_{t-1}; \underbrace{\frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right) X_t+\ \sqrt{\bar{\alpha}_{t-1}}\left(1-\alpha_t\right) X_0}{1-\bar{\alpha}_t}}_{\tilde{\boldsymbol{\mu}}\left(X_t, X_0\right)}, \frac{\left(1-\alpha_t\right)\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t} \mathbf{I}\Bigg)
\end{aligned}
\tag{3}
$$

$$
\begin{aligned}
q\left(V_{t-1} \mid V_t, V_0\right) &= \frac{q\left(V_t \mid V_{t-1}, V_0\right) q\left(V_{t-1} \mid V_0\right)}{q\left(V_t \mid V_0\right)} \\
&= \frac{q\left(V_t \mid V_{t-1}, V_0\right) q\left(V_{t-1} \mid V_0\right)}{\sum_{V_{t-1} \in \mathcal{S}} q\left(V_t \mid V_{t-1}, V_0\right) q\left(V_{t-1} \mid V_0\right)} \\
&= \frac{q\left(V_t \mid V_{t-1}\right) q\left(V_{t-1} \mid V_0\right)}{\sum_{V_{t-1} \in \mathcal{S}} q\left(V_t \mid V_{t-1}\right) q\left(V_{t-1} \mid V_0\right)} \\
&= \frac{\left[\alpha_t V_t+\left(1-\alpha_t\right) / K\right] \odot\left[\bar{\alpha}_{t-1} V_0+\left(1-\bar{\alpha}_{t-1}\right) / K\right]}{\sum_{V_t \in \mathcal{S}}\left[\alpha_t V_t+\left(1-\alpha_t\right) / K\right] \odot\left[\bar{\alpha}_{t-1} V_0+\left(1-\bar{\alpha}_{t-1}\right) / K\right]} \\
&= \tilde{\boldsymbol{c}}_t\left(V_t, V_0\right)
\end{aligned}
\tag{4}
$$

2. Supplementary Material (2)

Ultimately, the training of a diffusion-based generative model involves approximating the mean (μ_θ) of the reverse process at each timestep to the mean ($\tilde{\mu}$) of the forward process. At that time, learning the μ_θ of the reverse process can be broadly classified into three methods through the use of a reparameterization trick suitable for each objective: deriving X_0 from X_t at an arbitrary timestep and then learning based on X_0, deriving noise from X_t at an arbitrary timestep and then learning based on the noise, and lastly, deriving the score from X_t and learning based on that score. Among these three methods, we adopt the approach of TargetDiff [4], which has recently been successfully experimented with, to train our diffusion model. Additionally, the training of the diffusion process for the categorical distribution follows the approach of Argmax Flow [3], which has been successful in generation of discrete data.

To generate molecules in 3D form, we need to model both the three-dimensional coordinates of each atom and the type of that atom, respectively. Therefore, the backbone architecture for performing the diffusion process can be represented as $\left[\hat{X}_0^M, \hat{V}_0^M\right] = \phi_\theta\left(\left[X_t^M, V_t^M\right], t, \left[X^P, V^P\right]\right)$, as mentioned in the manuscript. Each function used for predicting X_0^M and V_0^M is defined as follows:

$$\boldsymbol{h}^{i,l+1} = \boldsymbol{h}^{i,l} + \sum_{j\in\mathcal{V},i\neq j} f_{\boldsymbol{h}}\left(d_{ij}^{l}, \boldsymbol{h}^{i,l}, \boldsymbol{h}^{j,l}, e_{ij}; \theta_{\boldsymbol{h}}\right) \tag{5}$$

$$X_t^{i,l+1} = X_t^{i,l} + \sum_{j\in\mathcal{V},i\neq j} \left(X_t^{i,l} - X_t^{j,l}\right) f_X\left(d_{ij}^{l}, \boldsymbol{h}_i^{l}, \boldsymbol{h}_j^{l}, e_{ij}; \theta_X\right) \cdot \mathbf{1}_{X^M} \tag{6}$$

where index i represents the atom index to be modeled and $\mathcal{V}$ denotes the set of atoms in proximity to the atom with index i. Further, $\boldsymbol{h}$ is obtained from the embedding matrix encoding the atomic type, and $\boldsymbol{h}^{i,l}$ indicates the embedding value of the atom corresponding to index i in the l-th layer. d_{ij} represents the Euclidean distance between two atoms i and j, and e_{ij} is an indicator for the connection between the two atoms. X represents the 3D coordinates of the protein and ligand atoms, and it can be rewritten as $[X^P$,$X^M]$. The mask $\mathbf{1}_{X^M}$ is designed to fix the coordinates of atoms related to protein while allowing updates only for the ligand atoms. The layers that make up $f_{\boldsymbol{h}}$ and f_X employ the graph transformer architecture [5], and the parameters of these two types of layers($f_{\boldsymbol{h}}$,f_X) are partially shared. The last layer embedding $\boldsymbol{h}_L$ is processed through a multi-layer perceptron and a softmax function, resulting in the derivation of V_0^M, and X_0^M is defined as the value corresponding to the index of the ligand atom in the last layer X_t^L.

3. Supplementary Material (3)

(a) **AR:** This 3D generative model estimates the probability density of atoms in 3D space, given a designated protein binding site as context. It employs an auto-regressive sampling scheme to sequentially generate 3D molecules, stopping when no space remains for new atoms.

(b) **liGAN:** This deep learning system generates 3D molecular structures conditioned on a receptor binding site, approach in structure-based drug discovery. It utilizes a conditional variational autoencoder trained on atomic density grids of cross-docked protein-ligand structures. Atom fitting and bond inference are applied to construct valid molecular conformations from generated atomic densities. The system evaluates the properties of generated molecules, showing significant changes when conditioned on mutated receptors. The latent space of the generative model is explored through sampling and interpolation.

(c) **GraphBP:** This framework generates 3D molecules binding to specific proteins using a 3D graph neural network. It obtains geometry-aware and chemically informative representations from contextual information, including the binding site and previously placed atoms. The generation process sequentially generates atom types and their relative locations using a local spherical coordinate system.

(d) **Pocket2Mol:** An E(3)-equivariant generative network for efficient molecular sampling based on 3D protein pockets. It comprises a graph neural network for spatial and bonding relationships in binding pockets and an algorithm for sampling drug candidates conditioned on pocket representations. The network considers 3D coordinates, bond types, and functional groups, sampling drug candidates from a tractable distribution without MCMC methods.

(e) **DiffSBDD:** Introduces protein-conditioned and ligand-inpainting generation strategies. Protein-conditioned generation treats the protein as a fixed context, while ligand-inpainting models the joint distribution of the protein-ligand complex, inpainting new ligands during inference.

(f) **TargetDiff:** This study proposes a 3D equivariant diffusion model for target-aware molecule generation and affinity prediction. It learns a joint generative process for atom coordinates and types using a SE(3)-equivariant network. The model overcomes challenges of voxelized atom densities and autoregressive sampling, being rotation-equivariant and respecting geometric constraints. It serves as an unsupervised feature extractor for binding affinity estimation and improves binding affinity ranking and prediction without retraining.

(g) **ResGen:** This method predicts new atom types and positions based on neighboring atoms' structural and geometric information. It uses scalar and vector features, along with a Gaussian mixture distribution, to determine the position distribution of newly generated atoms.

4. Supplementary Material (4)

$$
\begin{aligned}
L_{t-1}^{(X)} &= \mathrm{KL}\left(\mathcal{N}\left(X_{t-1}^{M}; \tilde{\mu}_t\left(X_t^M, X_0^M\right)\right) \mid \mathcal{N}\left(X_{t-1}^{M}; \mu_\theta\left(\mathcal{M}_t, t, \mathcal{P}\right)\right)\right) \\
&= \frac{1}{2\sigma_t^2}\left\|\tilde{\boldsymbol{\mu}}_t\left(X_t^M, X_0^M\right) - \boldsymbol{\mu}_\theta\left(\left[X_t^M, V_t^M\right], t, \left[X^P, V^P\right]\right)\right\|^2
\end{aligned} \tag{7}
$$

We match μ_θ to the derived result in Eq.(S3) to approximate μ_θ.

$$
\begin{aligned}
\tilde{\boldsymbol{\mu}}_t\left(X_t^M, X_0^M\right) &= \frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)X_t + \sqrt{\bar{\alpha}_{t-1}}\left(1-\alpha_t\right)X_0}{1-\bar{\alpha}_t} \\
\boldsymbol{\mu}_\theta\left(\left[X_t^M, V_t^M\right], t, \left[X^P, V^P\right]\right) &= \frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)X_t + \sqrt{\bar{\alpha}_{t-1}}\left(1-\alpha_t\right)\hat{X}_0}{1-\bar{\alpha}_t}
\end{aligned} \tag{8}
$$

$$
\begin{aligned}
&\frac{1}{2\sigma_t^2}\left\|\tilde{\boldsymbol{\mu}}_t\left(X_t^M, X_0^M\right) - \boldsymbol{\mu}_\theta\left(\left[X_t^M, V_t^M\right], t, \left[X^P, V^P\right]\right)\right\|^2 \\
&= \frac{1}{2\sigma_t^2}\left\|\frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)X_t + \sqrt{\bar{\alpha}_{t-1}}\left(1-\alpha_t\right)X_0^M}{1-\bar{\alpha}_t} - \frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)X_t + \sqrt{\bar{\alpha}_{t-1}}\left(1-\alpha_t\right)\hat{X}_0^M}{1-\bar{\alpha}_t}\right\|^2 \\
&= \frac{1}{2\sigma_t^2}\left\|\frac{\sqrt{\bar{\alpha}_{t-1}}\left(1-\alpha_t\right)X_0^M}{1-\bar{\alpha}_t} - \frac{\sqrt{\bar{\alpha}_{t-1}}\left(1-\alpha_t\right)\hat{X}_0^M}{1-\bar{\alpha}_t}\right\|^2 \\
&= \frac{1}{2\sigma_t^2}\left\|\frac{\sqrt{\bar{\alpha}_{t-1}}\left(1-\alpha_t\right)}{1-\bar{\alpha}_t}\left(X_0^M - \hat{X}_0^M\right)\right\|^2 \\
&= \frac{1}{2\sigma_t^2}\frac{\sqrt{\bar{\alpha}_{t-1}}\left(1-\alpha_t\right)}{1-\bar{\alpha}_t}\left\|X_0^M - \hat{X}_0^M\right\|^2
\end{aligned} \tag{9}
$$

Similarly, to define $L_{t-1}^{(V)}$, we take the derived result from Eq.(S4) and match it with $\boldsymbol{c}_\theta$.

$$
\tilde{\boldsymbol{c}}\left(V_t^M, V_0^M\right) = \frac{\left[\alpha_t V_t^M + \left(1-\alpha_t\right)/K\right] \odot \left[\bar{\alpha}_{t-1} V_0^M + \left(1-\bar{\alpha}_{t-1}\right)/K\right]}{\sum_{V_t^M \in \mathcal{S}}\left[\alpha_t V_t^M + \left(1-\alpha_t\right)/K\right] \odot \left[\bar{\alpha}_{t-1} V_0^M + \left(1-\bar{\alpha}_{t-1}\right)/K\right]}
$$

$$
\boldsymbol{c}_\theta\left(\mathcal{M}_t, t, \mathcal{P}\right) = \frac{\left[\alpha_t V_t^M + \left(1-\alpha_t\right)/K\right] \odot \left[\bar{\alpha}_{t-1} \hat{V}_0^M + \left(1-\bar{\alpha}_{t-1}\right)/K\right]}{\sum_{V_t^M \in \mathcal{S}}\left[\alpha_t V_t^M + \left(1-\alpha_t\right)/K\right] \odot \left[\bar{\alpha}_{t-1} \hat{V}_0^M + \left(1-\bar{\alpha}_{t-1}\right)/K\right]}
$$

5. Supplementary Material (5)

Protein name : PDB ID (link)

(a) **RET :** 2IVU (https://www.rcsb.org/structure/2IVU)

(b) **ERBB2 :** 3BBT (https://www.rcsb.org/structure/3BBT)

(c) **ABL1 :** 4CSV (https://www.rcsb.org/structure/4CSV)

(d) **ALK :** 5AAA (https://www.rcsb.org/structure/5AAA)

(e) **EGFR :** 4G5J (https://www.rcsb.org/structure/4G5J)

(f) **KIT :** 5MQT (https://www.rcsb.org/structure/5MQT)

6. Supplementary Material (6)

Protein name : PDB ID (link)

(a) **B-Raf :** 1UWH (https://www.rcsb.org/structure/1UWH)

(b) **PKB :** 6HH1 (https://www.rcsb.org/structure/6HH1)

(c) **ERK1 :** 4QTB (https://www.rcsb.org/structure/4QTB)

(d) **PIM1 :** 1YI3 (https://www.rcsb.org/structure/1YI3)

(e) **MEK1 :** 1CDK (https://www.rcsb.org/structure/1CDK)

(f) **CDK1 :** 6GU2 (https://www.rcsb.org/structure/6GU2)

(g) **PKC :** 2IW4 (https://www.rcsb.org/structure/2IW4)

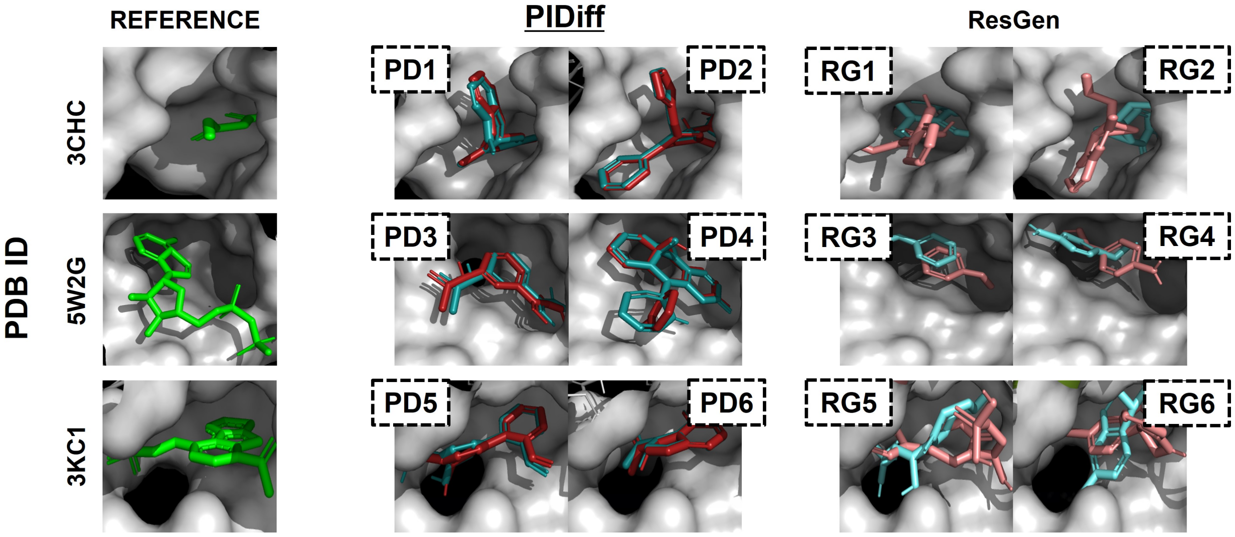
(h) **IGFR1 :** 2OJ9 (https://www.rcsb.org/structure/2OJ9)

(i) **EGFR :** 1M17 (https://www.rcsb.org/structure/1M17)

(j) **c-MET :** 3NW7 (https://www.rcsb.org/structure/3NW7)

(k) **HER2 :** 3PPO (https://www.rcsb.org/structure/3PPO)

7. Supplementary Material (7)



	3CHC				5W2G				3KC1			
	PD1	PD2	RG1	RG2	PD3	PD4	RG3	RG4	PD5	PD6	RG5	RG6
Vina Dock (↓)	-7.95	-9.85	-5.66	-5.27	-7.90	-7.13	-3.99	-3.88	-8.65	-8.05	-6.61	-7.24
RMSD (↓)	0.20	0.05	2.77	2.64	0.37	0.28	1.95	1.64	0.28	0.13	1.80	3.59

Table 1: The binding affinity of molecules generated by each model and the RMSD difference of the changed molecular conformation before and after docking.

The protein-ligand complex used as the 'REFERENCE' is a sample obtained from CrossDocked2020 [6], and the data can be accessed on the provided GitHub.

(a) `example/3chc_B_rec.pdb`
`example/3chc_B_rec_ligand.sdf`

(b) `example/5w2g_A_rec.pdb`
`example/5w2g_A_rec_ligand.sdf`

(c) `example/3kc1_A_rec.pdb`
`example/3kc1_A_rec_ligand.sdf`

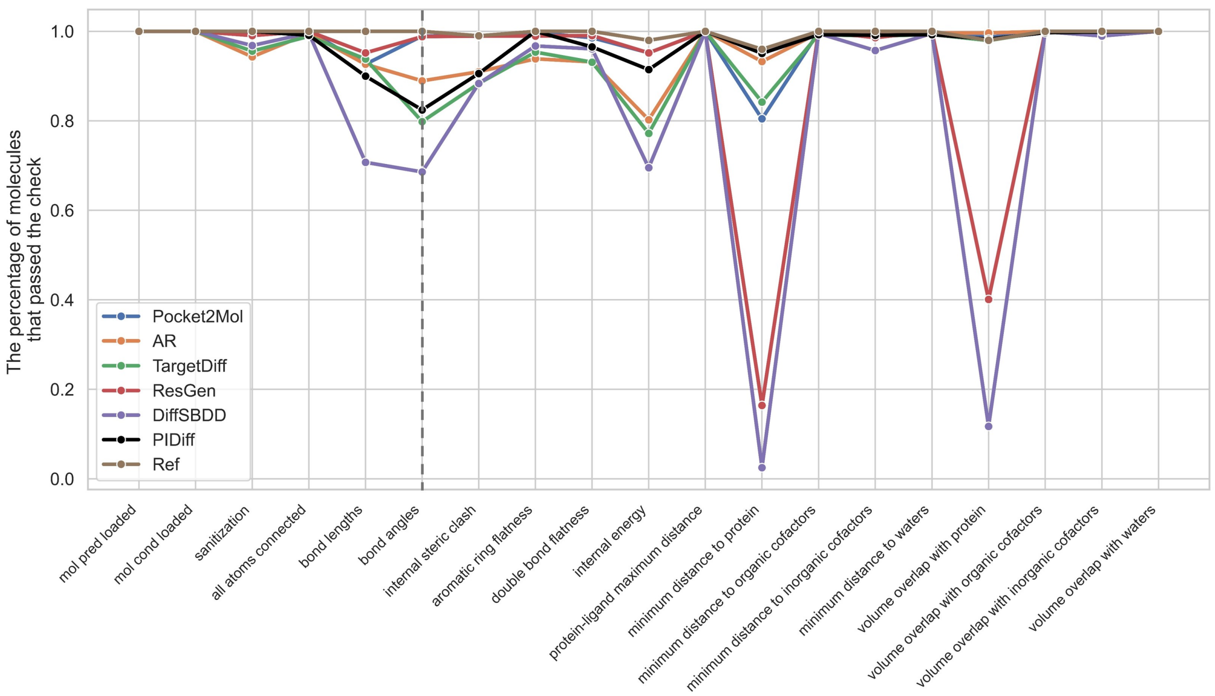
Similarly, molecules generated by each model can also be downloaded from the GitHub.

8. Supplementary Material (8)

	Ref	AR	liGAN	GraphBP	P2M	DiffSBDD	ResGen	TargetDiff	**PIDiff**
QED (Avg.)	0.475	0.508	0.502	0.501	0.572	0.480	0.548	0.479	0.483
QED (Med.)	0.467	0.499	0.496	0.499	0.577	0.493	0.532	0.480	0.491
Lipinski (Avg.)	4.270	4.750	4.787	4.883	4.880	4.487	4.951	4.512	4.729
Lipinski (Med.)	5.000	5.000	5.000	5.000	5.000	5.000	5.000	5.000	5.000
SA (Avg.)	0.73	0.63	0.59	0.49	0.74	0.61	0.68	0.58	0.58
SA (Med.)	0.74	0.63	0.57	0.48	0.75	0.60	0.69	0.58	0.57
Vina Dock (Top1)	-7.45	*N/A*	*N/A*	-9.332	-9.247	-10.099	-9.622	-11.784	-12.293
Vina Dock (Top3)	-7.45	*N/A*	*N/A*	-8.809	-9.042	-9.505	-9.453	-11.161	-11.989
Vina Dock (Top5)	-7.45	*N/A*	*N/A*	-8.515	-8.924	-9.123	-9.343	-10.841	-11.434
Vina Dock (Top10)	-7.45	*N/A*	*N/A*	-8.060	-8.730	-8.220	-9.077	-10.364	-10.921

Table 2: Molecular Properties

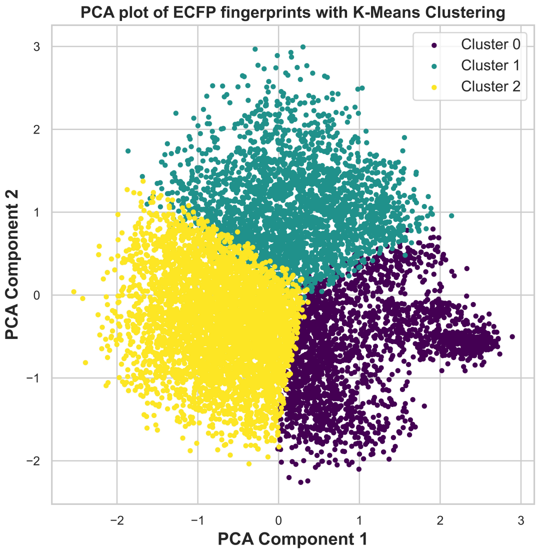
9. Supplementary Material (9)



Compared to the high-performing Pocket2Mol and AR models, which showed superior results on the ValidPB metric, our model, `PIDiff`, exhibited a relatively lower pass rate for the ”bond angles” criterion among the 12 evaluation cases. This outcome can be attributed to the differences in the operational mechanisms of each model used for molecule generation. The PoseBusters experimental results indicate that the top three models with the highest pass rates for the bond angle criterion are Pocket2Mol, AR, and ResGen. A commonality among these three models is their use of autoregressive sampling during molecule generation. Autoregressive sampling involves sequentially predicting each atom to form a complete molecule by iteratively predicting the position of the next atom based on the previously predicted atoms. At this point, when predicting the position of the next atom, the models ensure that the geometric rationality among the previously predicted atoms is maintained, resulting in superior pass rates in the bond angle criterion tests. However, the autoregressive sampling method introduces exposure bias due to the discrepancy between the model's behavior during training and sampling. Additionally, this approach may generate unrealistic fragments because it cannot grasp the overall context of the molecular structure during the initial stages of sampling. Furthermore, in this approach, the output of each step is used as the input for the next step. Errors that occur at each stage can accumulate over iterations, adversely affecting the quality of the final output. Consequently, this can lead to the generation of unrealistic and functionally inadequate molecules. To overcome these issues, models like ours, including TargetDiff and DiffSBDD, generate molecules using a non-autoregressive approach. This non-autoregressive method creates an entire molecule at once rather than adding atoms step by step. By generating the molecule in a single step, we can address the limitations associated with the autoregressive sampling method.

To summarize, the lower pass rates in the bond angle criterion for models like ours, including TargetDiff and DiffSBDD, can be explained by the fact that these models do not adopt a step-by-step approach to add atoms based on geometric rationality. However, as confirmed in the revised Table 1, the non-autoregressive sampling method is more effective for the primary objective of creating molecules that strongly bind to the target protein. Furthermore, both Pocket2Mol and AR models, which use autoregressive sampling, generate molecules with weaker binding affinities compared to the molecules in the test set. Therefore, the molecules generated by these models cannot be considered strong candidates for drug discovery, as they do not exhibit strong binding to the target protein. This suggests that while autoregressive sampling ensures the geometric validity of bond angles, it does not necessarily achieve strong binding affinities to the target protein. This implies that if the geometric rationality of bond angles between atoms in molecules generated using non-autoregressive sampling can be further improved, the PB-valid metric could become competitive with those of autoregressive models. To this end, future research focusing on enhancing the validity of bond angles in generated molecules—such as incorporating the torsion angles of atoms constituting the ligand during model training—holds great promise. This approach could significantly improve the geometric rationality of generated molecules, making it an exciting and promising area for further investigation.

10. Supplementary Material (10)



11. Supplementary Material (11)

The backbone architecture of PIDiff follows a graph transformer framework and is composed of 16 attention heads. The attention block consists of 10 equivariant layers, each with a hidden dimension of 128. Key/value embeddings are generated through a 2-layer MLP. Layer normalization is applied to all layers, and the Swish activation function is used. In the diffusion process, a cosine schedule is adopted to inject noise into the original data, consisting of a total of 1000 diffusion steps. Our model utilizes the AdamW gradient descent method, with an initial learning rate of 0.001, betas set to (0.95, 0.999), and a weight decay parameter of 0.0001. The learning rate is scheduled to decay exponentially by a factor of 0.6, with a minimum learning rate of 1e-6. The learning rate is decayed if there is no improvement in the validation loss for 10 consecutive evaluations. Evaluations are performed every 1000 training steps. To balance the scales of the atom coordinate loss, atom type loss, and derivative loss, we multiply the atom type loss by a factor of α=100 and the derivative loss by a factor of β=0.001.

References

[1] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, Advances in neural information processing systems 33 (2020) 6840–6851.

[2] C. Luo, Understanding diffusion models: A unified perspective, arXiv preprint arXiv:2208.11970 (2022).

[3] E. Hoogeboom, D. Nielsen, P. Jaini, P. Forré, M. Welling, Argmax flows and multinomial diffusion: Learning categorical distributions, Advances in Neural Information Processing Systems 34 (2021) 12454–12465.

[4] J. Guan, W. W. Qian, X. Peng, Y. Su, J. Peng, J. Ma, 3d equivariant diffusion for target-aware molecule generation and affinity prediction, in: International Conference on Learning Representations, 2023.

[5] J. Guan, W. W. Qian, W.-Y. Ma, J. Ma, J. Peng, Energy-inspired molecular conformation optimization, in: international conference on learning representations, 2021.

[6] P. G. Francoeur, T. Masuda, J. Sunseri, A. Jia, R. B. Iovanisci, I. Snyder, D. R. Koes, Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design, Journal of chemical information and modeling 60 (9) (2020) 4200–4215.