Guided Diffusion for molecular generation with interaction prompt

Peng Wu[1], Huabin Du[4], Yingchao Yan[4], Chen Bai[2,4,*], and Song Wu[1,3,*]

1. School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518055, China.
2. Warshel Institute for Computational Biology, School of Life and Health Sciences, School of Medicine, The Chinese University of Hong Kong, Shenzhen, Shenzhen, 518172, Guangdong, China.
3. South China Hospital, Health Science Center, Shenzhen University, Shenzhen 518116, China.
4. MoMed Biotechnology Co., Ltd., Hangzhou 310005, China.

* Correspondence: wusong@szu.edu.cn (S.W.); baichen@cuhk.edu.cn (C.B.)

Abstract

Structured based drug design is a critical strategy for modern drug development. Recently, molecular generative models have demonstrated their potential in designing molecules from scratch with high binding affinities in a pre-determined protein pocket. However, the generative processes are random in current generative models and the atomic interaction information between ligand and protein are ignored. Besides that, the binding mode of ligands in proteins is crucial for drug design and the ligand has high propensity to bind with residues called hotspots. Hot spot residues contribute to the majority of the binding free energy and have been recognized as appealing targets for designing molecules. To this end, we develop an interaction prompt guided diffusion model, InterDiff to deal with the above challenges. Four kinds of atomic interactions are involved in our model and represented as learnable vector embeddings. These embeddings serve as a condition for each residue to guide the molecular generative process. Comprehensive in silico experiments evince that our model could generate molecules with desired interactions. Furthermore, we validate InterDiff on two realistic protein-based therapeutic agents and experiments show that InterDiff could design molecules with similar binding mode with known targeted drugs.

Introduction

Structure based drug discovery (SBDD) strives for design molecules that can bind to a target protein with high binding affinity and specificity, which acts as a critical approach in contemporary biopharmaceutical research[1]. However, SBDD remains a challenge owing to the massive chemical space. It is estimated that the number of “drug-like” molecules range from 10^{20}-10^{60} considering the oral bioavailability and Lipinski's rule-of-five[2]. Traditional *in silico* SBDD methods, such as virtual screening are computationally costly due to the large feasible chemical space and could not find novel drugs. In recent years, molecular generative models have emerged as a promising technique in drug discovery and enabled *de novo* molecular generation. Earlier methods relied on 1D (SMILES strings and SELFIES strings) or 2D (graphs) molecular representations[3-6], and are able to generate diverse and novel molecules. Nevertheless, these models ignore 3D spatial information of molecules and the protein pocket environment, which is essential for molecular properties and protein binding affinity. Consequently, 3D structure-based generative methods have gained lots of attention due to their capability of designing molecules that bind to a specific protein pocket.

Recently, various models have been proposed for 3D structure-based molecular generation, including variational autoencoders (VAEs), flow-based models, autoregressive models and diffusion models[7-14]. In [7], Matthew et al. represent molecules as density grids and a conditional VAE is used to generate new atomic density grids, then atom fitting and bond inference are applied to obtain novel molecules. Although they achieve remarkable results in generating diverse molecules, as pointed in [9, 10, 13], inferencing molecules from density grids is a nontrivial task and irregularities are presented in the generated molecules. Besides, the model is not equivariant and hard to scale to large protein systems. Peng et al. utilize autoregressive model to generate molecules atom by atom in protein pocket[14], but the generation process is inefficient and the deviations are accumulated as a result of the sequential generation. For instance, if the first several atoms are placed at improper positions, this will incur bias in subsequent generation process. On the contrary, diffusion-based models sample atom types and coordinates simultaneously in the light of protein context. Concretely, diffusion model defines a noise schedule and add noise to the molecular geometry in forward process. In the backward (generation) process, the model learns to reverse the noise process to recover the true molecular geometry. There is no mismatch between the training and generating process in diffusion models. Further, geometric symmetries in molecular system like rotation, translation and reflection are respected in diffusion model to improve the generalization ability.

Nonetheless, diffusion models still face one limitation in real scenarios. Essentially, diffusion models pertain to a model class named score-based generative models[15]. Another member in score-based generative models is score matching, which estimates the score of data at different noise scales and samples by gradually decreasing noise levels[16, 17]. As Song et al. pointed, when the number of noise scales go to infinity, score-based generative models can be regarded as a stochastic differential equation (SDE)[15]. While sampling from the SDE, there exists a corresponding ordinary differential equation (ODE) sharing the same marginal probability densities. On this account, the diversity of generated molecules would be decreased[12]. Additionally, the binding mode of proteins with ligand are vital for understanding the biological processes. It has been found that only a fraction of residues in the pocket, called hot spots contribute most to the binding affinity[18, 19]. A mutation in hot spots can cause significant drop in binding affinity and even drive drug resistance in patients[20, 21]. In modern development of drugs, hot spots are crucial for rational drug design and one usually desire that the drug can form interactions with hot spots. However, current diffusion models ignore the protein-ligand interaction information and cannot customize the generated molecules.

Inspired by the fruitful progress of prompt-based learning in nature language processing, we develop a prompt-based diffusion model called InterDiff to tailor the binding mode of generated molecules in the protein pocket. Specifically, we introduce four kinds of learnable prompt embedding to indicate the interaction type of protein residues, including π-π interaction, cation-π interaction, hydrogen bond interaction and halogen bond interaction. We perform an empirical study on CrossDocked2020 dataset[22] and shows that InterDiff is able to generate molecules under prescribed interaction prompts with high probability. In addition, we validate our model on two well-known targets in neural systems and cancers respectively. Experiments implicate that InterDiff could generate molecules having similar binding mode with known targeted drugs. To the best of our knowledge, this is the first work that introducing interaction prompt in structured based drug design.

Results

InterDiff leverages interaction prompts to guide diffusion model and design molecules, resembling the prefix-tuning which optimizes a small continuous task-specific vector (Figure 1 and methods)[23]. To evaluate InterDiff, we first conduct the experiment on a benchmark dataset compared with three recently published methods on five general metrics. Furthermore, we evaluate the model performance in generating molecules with predefined interactions, which is the key characteristic of InterDiff. We also illustrate the potential of InterDiff in real scenarios. Two protein targets with targeted drugs are selected and InterDiff is assessed to design molecules with identical interactions.

Data: The crossDocked2020 was used to train and evaluate InterDiff[22]. We follow the same splitting and filtering criterion described in [10, 14], obtaining 100000 samples for training and 100 samples for testing. The protein-ligand interactions are detected by BINANA2[24], four kinds of interactions are adopted including cation-π, π-π, hydrogen and halogen interaction. We ignore the complexes that have residue detected with more than one interaction. Phenylalanine and Tyrosine were found to have four interactions (Figure S1) and hydrogen bond account for the vast majority of interactions (Figure S1, S2).

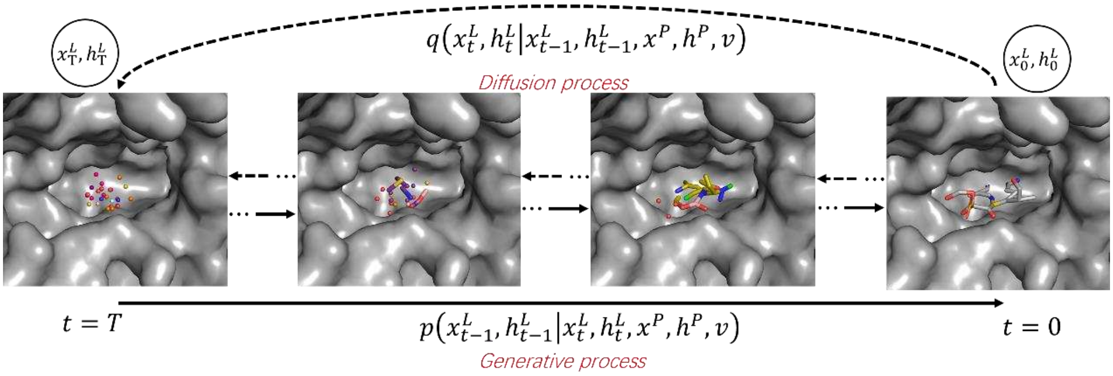

x_T^L, h_T^L
$q\left(x_t^L, h_t^L \middle| x_{t-1}^L, h_{t-1}^L, x^P, h^P, v\right)$
Diffusion process
x_0^L, h_0^L
$t = T$
$p\left(x_{t-1}^L, h_{t-1}^L \middle| x_t^L, h_t^L, x^P, h^P, v\right)$
$t = 0$
Generative process

Figure 1: Overview of InterDiff in a protein conditional generation. In diffusion process q, we simulate a progressively noised ligand point cloud (coordinates and atom types: x^L, h^L) under protein environment (x^P, h^P) over T timesteps. The interaction prompts v are tunable in the diffusion process. In the generative process p, a neural network learns to recover data from a noise

distribution conditioned on protein and prompts.

Molecular structure and properties

Routinely, we assess the generated molecules from test set in five commonly used metrics: 1) **Vina Score**. A binding affinity indicator calculated by physical-based empirical scoring function. 2) **QED**. QED is a measurement of drug-likeness of molecules. 3) **SA**. SA (synthetic accessibility) measures the feasibility of synthesize molecule based on fragmental analysis in compound database. 4) **Lipinski**. We calculate the Lipinski score by quantifying how many rules are fulfilled in Lipinski's rule of five. 5) **Diversity**. Diversity is computed by averaging the pairwise dissimilarity (one minus *Tanimoto* similarity) of generated molecules in each pocket.

	Vina Score (↓)	QED (↑)	SA (↑)	Lipinski (↑)	Diversity (↑)
Test set	-6.871±2.32	0.476±0.20	0.728±0.14	4.340±1.14	–
Pocket2Mol	-6.561±2.67	**0.573**±0.16	**0.756**±0.13	**4.879**±0.42	0.731±0.12
AR(3D-SBDD)	-6.592±2.08	0.507±0.19	0.634±0.14	4.723±0.65	0.698±0.10
TargetDiff	**-7.163**±1.72	0.472±0.20	0.585±0.12	4.519±0.84	0.717±0.09
InterDiff	-6.584±1.29	0.413±0.18	0.598±0.11	4.405±1.01	**0.820**±0.04

Tab 1: Evaluation results from test set of CrossDocked 2020 dataset. The performance is re-evaluated for baseline methods Pocket2Mol, AR and TargetDiff.

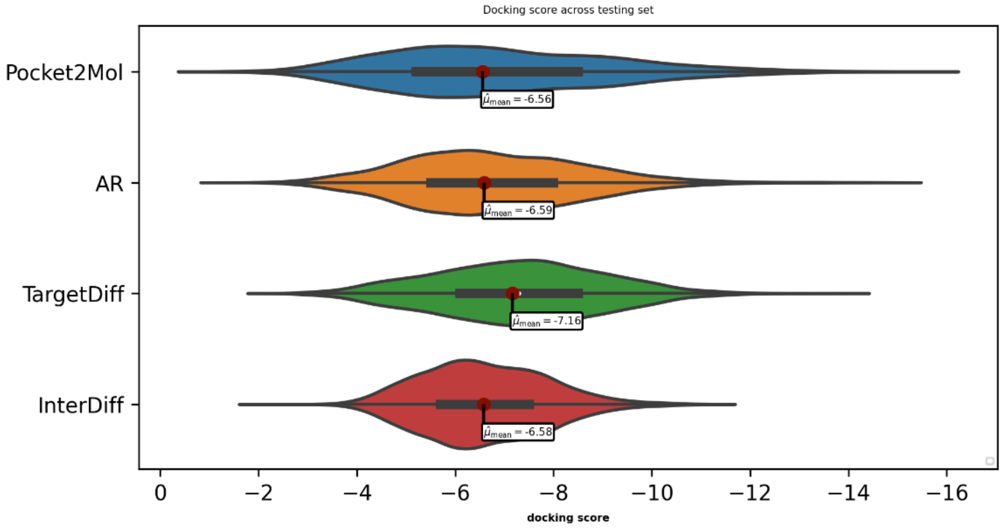


Figure 2: violin plot of distributions of docking score for InterDiff and baseline methods. InterDiff shows a limited range of scores compared to other methods.

Tab 1 displays the results of InterDiff and baseline methods. Overall, InterDiff outperforms baseline methods in the diversity and the other indicators are less ideal. We notice that the average number of atoms in generated molecules sampled by InterDiff is smaller than other methods, ranging from 2-4 atoms. Since a larger molecule tends to have a better docking score, this may account for the lower docking score for our method. TargetDiff achieves best results in vina score and Pocket2Mol performs best in QED, SA and Lipinski. However, this also indicates that InterDiff and TargetDiff could generate novel molecules since QED and SA are calculated based on existing drug database. To evaluate the substructure of generated molecules, we count the percentage of different ring size. Results show that InterDiff and TargetDiff tend to produce a larger proportion of 7-membered ring while AR has more 3-membered rings. (Tab S1). This could partially explain the lower score in QED and SA of TargetDiff and InterDiff, since 7-membered ring rarely appears in common drugs[25]. Besides, we spot that the variance of vina score in InterDiff are much smaller than the other methods, as shown in Figure 2. In the generative process, we keep the interaction prompt of the protein residues the same (see methods part) as reference molecule in test set and the number of atoms is sampled according to the distribution of pocket size and number of atoms in ligand (Figure S3). The interaction prompt for the residues restricts the fluctuating extent of binding pose of the generated molecules, resulting in a binding mode akin to the reference molecules. We will analyze the accuracy of InterDiff in generated molecules with defined interaction prompt. Considering that the vina score is calculated based on the binding pose of ligand, this could shed light on lower variance of vina score in InterDiff.

We further evaluate the molecular chemical space distribution of generated molecules by 2D and 3D molecular fingerprints. The projection of Morgan Fingerprint and USRCAT (Ultrafast Shape Recognition with CREDO Atom types) features by UMAP (Uniform Manifold Approximation and Projection) are displayed in Figure S4 and S5. Morgan Fingerprint derives from 2D molecular representation and takes atom types and connectivity into account. USRCAT is a method that measures 3D shapes of molecules meanwhile considering the pharmacophoric features. Generally, InterDiff has similar distributions in space with TargetDiff and AR shows a dispersed distribution regarding USRCAT projection. Pocket2Mol behaves differently against others, which locates at the upper corner and may correspond to a region with more drug-like molecules. For the 2D Morgan Fingerprint, InterDiff has one dense region while other models have two or more than three dense regions. The interaction prompt may confine the diversities of structures presented in generated molecules, making them similar to the reference molecules. This is in line with the distribution of vina score, which has a narrow range comparing to other methods.

Performance of InterDiff in design specific Interactions

To evaluate the capability of InterDiff in generating molecules with designated interactions, we re-generate molecules in the test set for 100 times and the number of atoms is identical to the reference molecule for each sample. Additionally, we excluded test samples that did not detect any interaction and 99 samples are left for testing. After generation, the interactions were detected by BINANA2 using the conformers generated by InterDiff in protein pocket. We compute the accuracy of accomplishing designated interactions for generated molecules. As an illustration, if *['GLU', 14, 'Hydrogen']* (fourteenth residue GLU with hydrogen bond) and *['TYR', 39, 'caption']* are given as the condition for generating, and only *('TYR', 39, 'caption')* is obtained in the generated molecular conformer, the accuracy would be 50%. The results are exhibited in Figure 3:

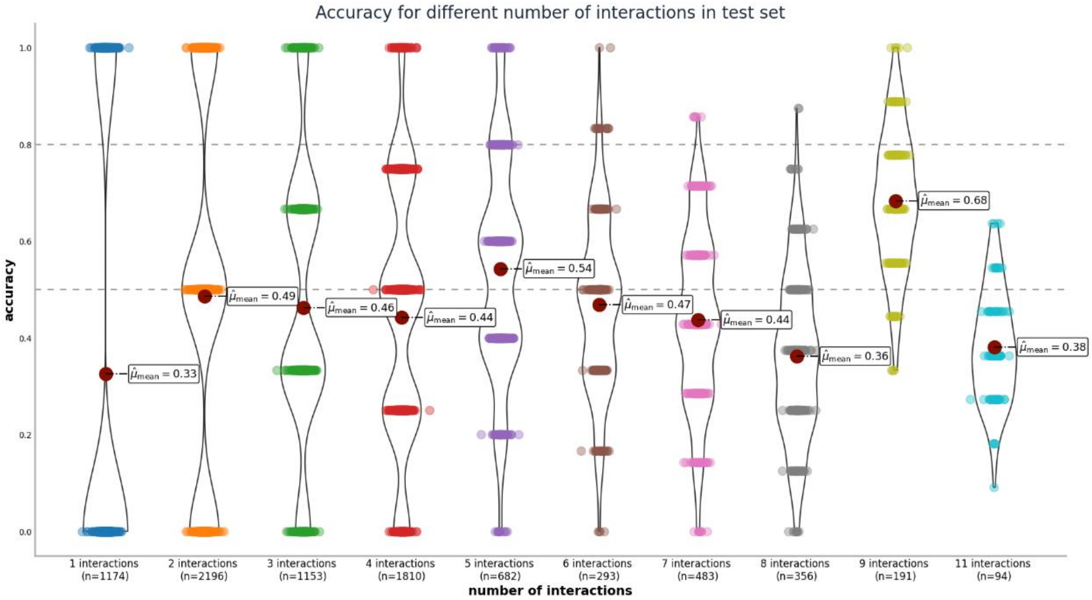


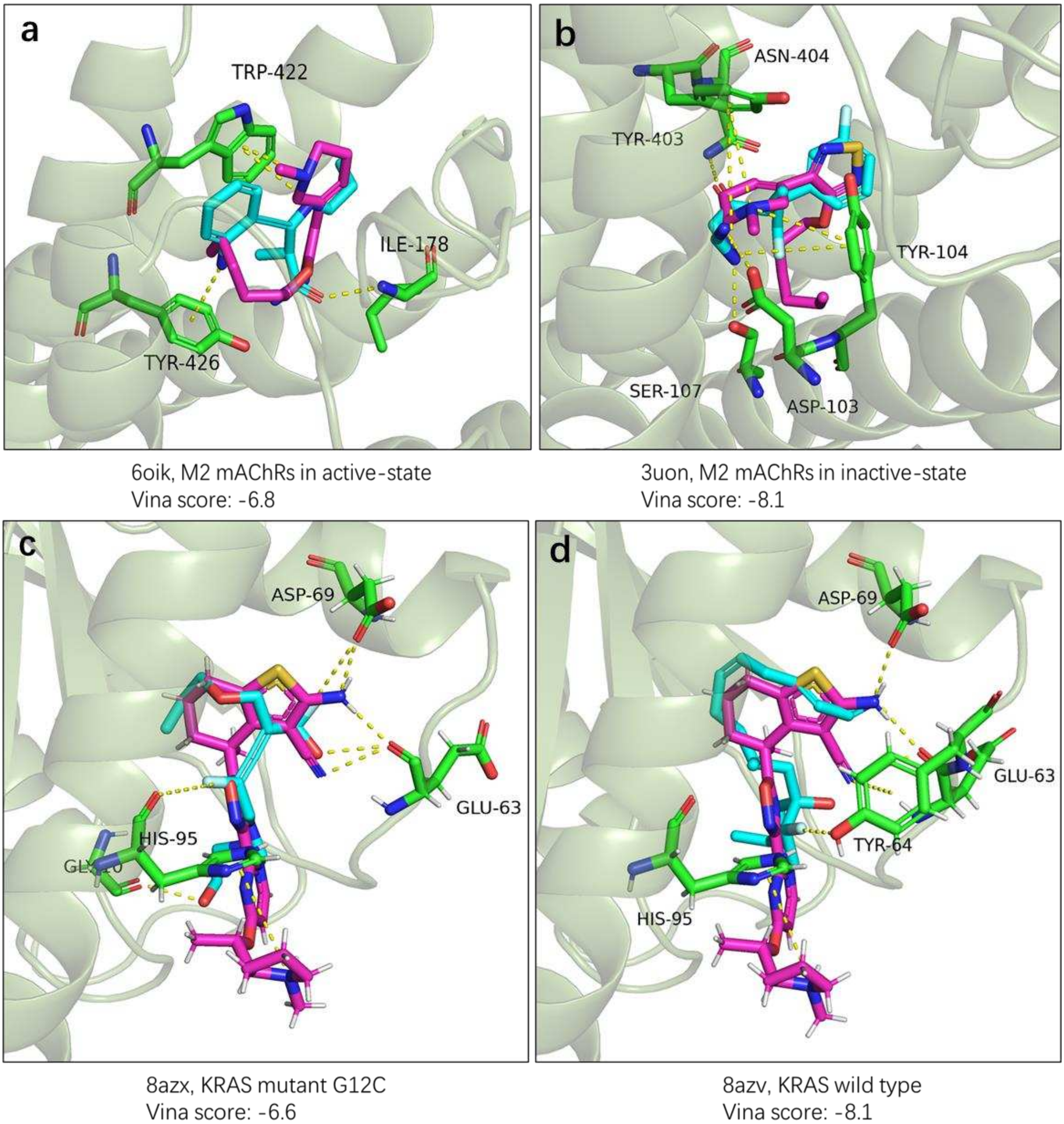
Figure 3: Accuracy of achieving designated interactions in generated molecules in test set. The samples are sorted by the number of interactions.

The number of interactions in the test set ranges from 1 to 11 except for the 10 interactions and we generate 100 samples for each protein pocket. Overall, InterDiff illustrates excellent performance in designing molecules under specific interaction prompts. We achieve the highest accuracy under 9 interactions (mean = 68%, n = 191) and the lowest accuracy under 1 interaction (mean = 33%, n = 1174). In the most difficult setting 11 interactions, InterDiff still reach an average accuracy of 38% in 94 generated molecules (6 molecules fail to reconstruct). Among the 94 generated molecules, 4 samples realize the highest accuracy with 7 interactions agreed with the reference. What' s more, in 9 interaction cases, 5 molecules attain the same interaction types as the reference and 23 molecules are in accord with 8 interactions. To verify if the generated conformers agree with the pose after docking, we compare the raw conformers from InterDiff with the docking poses generated by QuickVina. We plot the resulting RMSD density distribution of 9 conformers (ordered by the docking score) generated by QuickVina (Figure S7). For the besting scoring pose, QuickVina agrees with 9% of generated conformers (RMSD below 2 angstrom), which is similar to the work that Schneuing et al. reported[9]. Additionally, we also observe a significant drop of the proportion of agreed molecules in the less confident poses (4% in the 9th pose). This indicates InterDiff can generate molecular conformers approaching the steady binding pose. Furthermore, to estimate the ability of InterDiff in achieving disparate interactions, we assess the accuracy of four interactions in each sample and the results are shown in Figure S6.

Overall, InterDiff behaves distinctly in designing disparate interaction. The probability of hydrogen bonds being designed is the highest, followed by halogen bond. We note that InterDiff does not perform well in π-π interactions, which is understandable since the requirements of π-π interaction are more complicated than the others. Accordant with BINANA2's criterion, three standardized aromatic residues (phenylalanine, tyrosine and histidine) are involved. The aromatic ring center distance between ligand and protein must be less than 4.4 angstroms and the ring atoms in could not deviate from planarity by more than 15 degrees. Last but not least, the angle of two normal vectors in the ring planes needs to within 30 degrees. Compared to other interactions, π-π interactions demand aromatic ring structure on the ligand and the ring has to be positioned in a proper manner. In addition, we noticed that the π-π interactions only comprise around 8 percent of the total interaction samples. On account of this, we speculate that the imbalanced interaction distribution may also impact the model performance in π-π interactions.

Application of InterDiff in real scenarios

In this section, we investigate the potential of InterDiff in designing drugs when the binding mode of a reference molecule is available. We select two protein targets with different subtypes and use InterDiff to design molecules with similar binding mode as existing drugs. The first target is muscarinic acetylcholine receptor (mAChR), acting an important target in central nervous system diseases, for instance, Alzheimers's disease and schizophrenia[26]. Xanomeline was developed as an agonist to mAChRs and studies have found that it has almost identical binding affinity to all mAChR subtypes (M1-M5), but stimulates them to appreciably different extent[27]. A recent study termed this phenomenon as "efficacy-driven selectivity" and the authors found that Xanomeline's binding mode differs between inactive states and active states of mAChRs[28]. We use InterDiff to design molecules for M2 type mAChR in both inactive state and active state, conditioning on the binding mode of Xanomeline in two states. The second target is KRAS, commonly mutated in cancers and serving as a therapeutic targeting in various cancers, such as lung cancer, colorectal cancer and pancreatic cancer. Current inhibitors only target KRAS G12C mutants but the non-G12C mutants constitute the most in KRAS driven cancers. Recently, Kim et al. reported a non-covalent inhibitor BI-2865, which can bind to a wide range of KRAS altercations[29]. In like manner, we design molecules with InterDiff and take the binding mode of BI-2865 in KRAS G12C and another mutant G13D as references. We sample 300 molecules for each state or mutant of two targets and check the interactions after docking by QuickVina. The original interactions for two drugs are list in Table S3. Among the generated molecules, we successfully obtain molecules that have identical interactions as existing drugs, and we randomly select four molecules for illustration.



6oik, M2 mAChRs in active-state
Vina score: -6.8

3uon, M2 mAChRs in inactive-state
Vina score: -8.1

8azx, KRAS mutant G12C
Vina score: -6.6

8azv, KRAS wild type
Vina score: -8.1

Figure 4: Pose of generated molecules and native drug in protein target. **a, b**: The illustration of Xanomeline and generated molecules in M2 mAChRs active state and inactive state. **c, d**: Generated molecules and BI-2865 in KRAS wild type and G12C mutant. The binding poses of generated molecules and Xanomeline are obtained by QuickVina while the pose of BI-2865 are obtained from cocrystal structure. Residues that have interactions with molecules are colored with green. BI-2865 and Xanomeline are colored with purple and generated molecules are colored with blue. The structures of all molecules are available in Figure S8.

As demonstrated in Figure 4, we present the poses of generated molecules after docking together with the targeted drugs. The poses of Xanomeline are acquired by docking for mAChRs and the poses of BI-2865 are acquired from cocrystal structure in PDB database for KRAS. We can see that the docking pose of designed molecules overlap well with the reference drug. What's more, InterDiff successfully generates similar interactions as the reference drug in three of the four protein targets. For the last target (Figure 4d, KRAS wild type), three of the four interactions are consistent.

The docking pose of generated molecules and Xanomeline for M2 mAChRs in active and inactive state are illustrated in Figure 4a and 4b. For the active state (PDB:6oik), the original interaction is TRP-422 with interaction cation-π. InterDiff successfully realizes the same interaction and additionally introduces two new interactions, TYR-426 with cation-π and ILE-178 with hydrogen bond. For the inactive state (PDB:3uon), the primary interactions are cation-π in TYR-403 and TYR-104. InterDiff also reproduces the same interactions and three hydrogen bonds are formed in ASN-404, SER-107and ASP-103. In the second case, three interactions are discovered by BINANA2 in KRAS mutant G12C cocrystal structure (PDB:8azx), ASP-69 with hydrogen bond, GLU-63 with hydrogen bond and HIS-95 with cation-π. While in KRAS wild type (PDB:8azv), an additional interaction, TYR-64 with cation-π is found. InterDiff could discover molecules with the similar binding mode in KRAS mutant G12C (form an extra interaction GLY-10 with hydrogen, Figure 4c) and KRAS wild type except for HIS-95 with cation-π. Besides that, we notice that the BI-2865 has 5 ring structures, and there are no rings in molecules generated by InterDiff. Currently, InterDiff could not control the sub-structures in generating process and this could be a future direction.

Fragment growing by inpainting

In this part, we investigate the potential of InterDiff in fragment-based drug design (FBDD). FBDD enables designing molecules conditioned on a potent substructure. It is very common that one may desire to optimize certain parts of a molecule while fix the molecular scaffold. To this end, we additionally train an unconditional diffusion model which learns the joint distribution of ligand atoms and protein atoms. The model structure and training process is identical to the conditional InterDiff except for the training objective (protein atoms are included in the loss function). To generate molecules under a given scaffold, we modify the sampling process by injecting the fixed context in the denoising step and replacing the corresponding parts from the model. This technique is named inpainting and initially introduced in image imputation[30, 31]. Formally, in each denoising step, we do following operations:

$$x_{t-1}^{known} \sim \mathcal{N}\left(x_t \middle| \sqrt{1-\beta_t} x_0, \beta_t \boldsymbol{I}\right),$$
$$x_{t-1}^{unknown} \sim \mathcal{N}\left(\tilde{\mu}_\theta(x_t), \tilde{\beta}_t \boldsymbol{I}\right),$$
$$x_{t-1} = m \odot x_{t-1}^{known} + (1-m) \odot x_{t-1}^{unknown},$$

where x_{t-1}^{known} indicates the reference samples, $x_{t-1}^{unknown}$ indicates the samples from the model and m is a binary mask which signifies the fixed context. In the experiments, the pocket atoms and the native ligand are the x_{t-1}^{known} and the denoising samples from the model are $x_{t-1}^{unknown}$. It is evident that by iterating this step during the sampling process the molecular scaffold can be preserved in the final generated molecules.

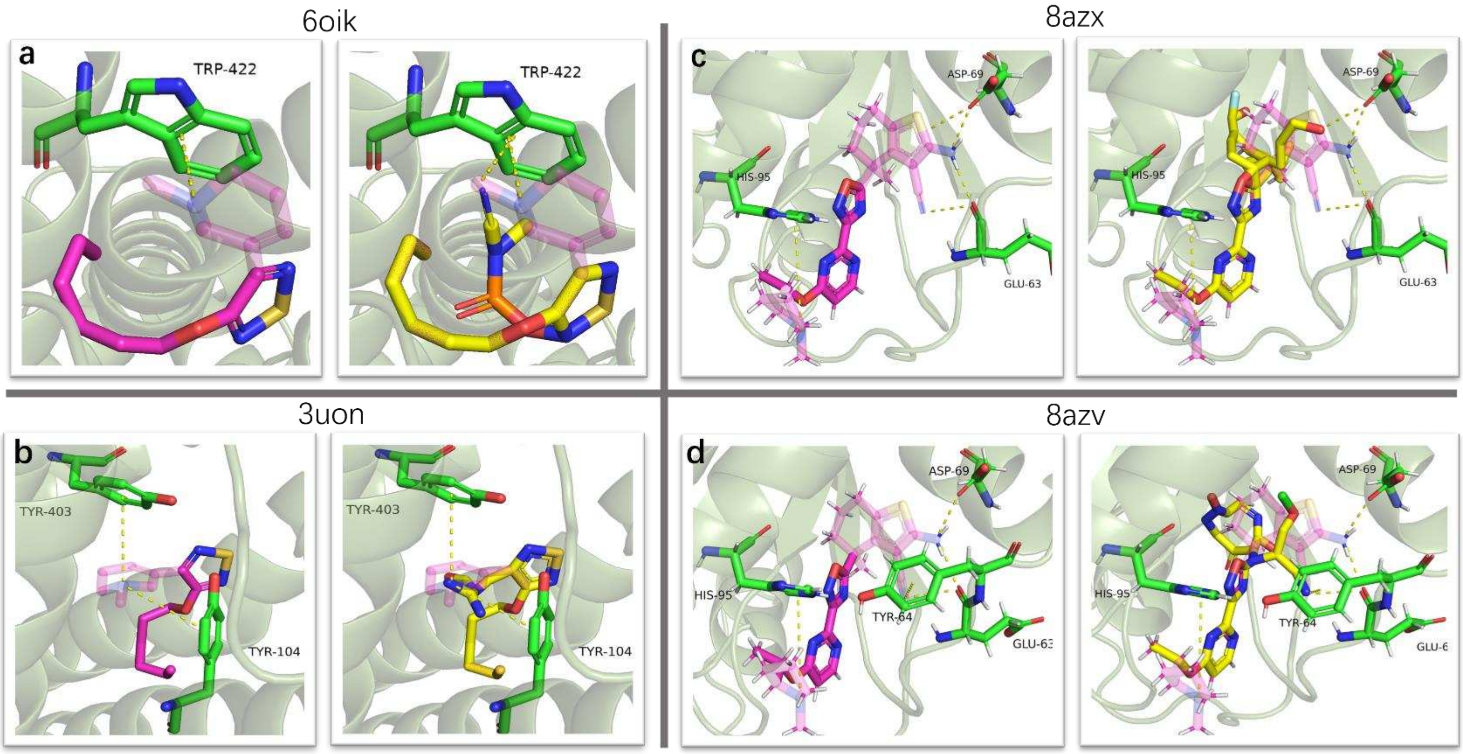


Figure 5: Pose of designed fragments with scaffold and native drug in protein target. The transparent parts of native drug are fragments with interactions. Residues that have interactions with molecules are colored with green. BI-2865 and Xanomeline are colored with purple and generated molecules are colored with yellow.

The same protein targets are used in experiment as the previous section and 100 molecules are sampled for each state of the two targets. We evaluate InterDiff in FBDD by removing the fragments (Figure 5, transparent parts) that have interactions with protein residues and keep the rest of the molecules as the fixed scaffolds. Four illustrative examples are shown in Figure 5, and we successfully design fragments with the same interactions as the native drug in M2 mAChRs based on the scaffold. For KRAS mutant G12C (PDB:8azx) and KRAS wild type (PDB:8azv), one (ASP-69 with hydrogen) of three and two (ASP-69 with hydrogen, TYR-64 with cation-π) of four interactions are achieved. The docking poses are generated by the model and our method can inpaint new fragments with desired interactions around the fixed scaffolds. However, we also noticed that the accuracy of InterDiff in inpainting mode is lower than the pocket-conditional mode. The model has to estimate the positions of both protein and ligand atoms in the denoising steps and on the contrary, only ligand atoms are estimated in the pocket-conditional mode. The errors in protein atoms could hamper the model to design the correct interactions in ligand atoms. In addition, in case PDB:6oik, we found that the new fragments are anchored in an alternative position on the 5-membered ring. It would be interesting to add the information of anchor point in diffusion model and generate diverse molecules.

Related work

Diffusion model for molecular design

Diffusion models are a new kind of generative model inspired by diffusion process. Impressive progresses have been made in distinct generating task such as images, audios and even videos[32-34]. In molecular science, Hoogeboom et al. first proposed E(3) Equivariant Diffusion Model (EDM) for molecular generation which notably outperforms previous 3D generative methods[35]. Shortly after their work, Schneuing developed a diffusion model for structured based drug design named DiffSBDD, which is the first of its kind[9]. Two strategies are introduced under their framework, protein-conditional and ligand-inpainting generation. Specifically, ligand-inpainting method learns the joint distribution of protein-ligand complex, and new ligands are completed in inference stage. Experiments exhibit that both strategies can produce novel and drug-like ligands. In silico docking assessment also verify the potential in generating ligands with high binding affinity. Similar work were done in [13] and the difference lies in a dual diffusion was used to capture the local and global protein environment. In addition, Guan et al. presented a target-aware diffusion model. Unlike previous work that need to evaluate generated molecules through docking method like AutoDock, their model can estimate and rank the binding affinity of molecules. The authors raise a problem that the bond inference is implemented in a post-processing manner and irrational structures may appear in generated molecules, which is also pointed out in Schneuing's work[9]. Huang et al. tackle this problem by setting distance threshold for covalent bond[12], but the bond distance can vary depending on the particular chemical structure. Alternatively, Wu et al. developed a diffusion model guided by a prior diffusion bridge[36], which can guarantee a desirable output. Specifically, AMBER inspired physical energy and statistical energy were incorporated as priors to guide training process. Another solution to this issue is integrating the chemical bonds into the diffusion process to enhance the quality of generated molecules[37].

Prompt learning for molecular design

Prompt-based learning is initially a strategy to train large language models (LLMs), serving as an alternative to the fine-tuning paradigm so the LLMs can adapt to different tasks without re-training[38]. Afterward this technique was introduced to vision-language model and greatly improved the performance over all evaluation tasks[39]. Very recently, several attempts have been made to incorporate prompt-based learning to molecular design[40-43]. These works combine SMILES representation of molecules with other modalities including chemical structure texts[40, 41], pharmacological properties[41-43], medical description texts[42], and protein pocket[43]. In [43], Gao et al. propose a unified model called PrefixMol considering both chemical properties and binding pocket via generative pre-trained transformer (GPT). The pocket information is transformed into an embedding by geometric vector transformer (GVF) and used as a prefix condition together with other conditional embeddings. PrefixMol demonstrates excellent performance in single and multiple conditional molecular generation. But still, PrefixMol is an autoregressive model, and the global context of ligands are lost during the generation process. Moreover, their model treats pocket residues equally and the outputs are 1D SMILES representation, which makes it hard to apply in real scenarios.

Discussion

In this work, we propose a novel diffusion model named InterDiff to guide the molecular generation by residue interaction prompt. Our model could generate molecules under certain interaction conditions with high probability, which is critical for structure-based drug design. Besides, we demonstrate that InterDiff could be easily modified into fragment-based generative model and improve molecules by introducing interactions with certain hot spots. This characteristic is of benefit for modern drug design and could help optimizing lead compound. Nevertheless, InterDiff still faces several problems such as the infeasible structures presented in the generated molecules, which are also commonly seen in other methods. Although this may be the flaws in molecular reconstruction algorithms, efforts are needed to increase the molecular structural rationality. Luckily, potential solutions have been proposed as we discussed in the related work part. We will attempt to optimize the sub-structures of generated molecules to ameliorate the drug-likeness and synthetic accessibility. Currently, the choice the interaction prompts for residues depends on the reference molecules or one's experience, and existing tools like FTMap may be helpful to this problem[44]. In addition, the accuracy in accomplishing π-π interactions is still room for improvement and could be an interesting future direction.

Methods

Molecular diffusion model

We build our model upon the framework develop by Guan et al.[10]. Let $\mathcal{M} = (x, h)$ denote the molecular 3D point cloud data with $x = \left[x^{(L)}, x^{(P)}\right] \in \mathcal{R}^{N\times 3}$ and $h = \left[h^{(L)}, h^{(P)}\right] \in \mathcal{R}^{N\times M}$. In our setting, $\left[x^{(L)}, x^{(P)}\right]$ indicate atom coordinates of ligand and protein, and $\left[h^{(L)}, h^{(P)}\right]$ represent the atom categorial features, where N is the number of atoms. We use diffusion model to learn the distributions of protein-ligand complexes. Diffusion model learns two Markov processes, a diffusion process q and a denoising process p. Diffusion process adds Gaussian noise to data $\mathcal{M}_t$ in time step t, where $t = 0, \cdots, T-1$ is the predefined time steps ($T = 1000$ in the implementation):

$$q(\mathcal{M}_t|\mathcal{M}_{t-1}) = \mathcal{N}(\mathcal{M}_t|\alpha_t\mathcal{M}_{t-1}, \sigma_t^2\boldsymbol{I}),$$

where α_t is the schedule that controls how much signals are preserved in the diffusion process and σ_t is the noise schedule that controls how much noises are added. For the 3D molecular point cloud data, the atom types are categorical data while the atom coordinates are continuous data. At time step t, We add Gaussian noise and uniform noise to the atom coordinate feature and atom type feature respectively[45]. Following the convention in [10, 45], the joint distribution states as:

$$q(\mathcal{M}_t|\mathcal{M}_{t-1}) = \mathcal{N}\left(x_t\middle|\sqrt{1-\beta_t}x_{t-1}, \beta_t \boldsymbol{I}\right) \cdot \mathcal{C}(h_t|(1-\beta_t)h_{t-1} + \beta_t/K),$$

where $\mathcal{C}$ indicates a categorical distribution with parameters after $|$, β_t is the variance schedules and K is the number of atom types with $k = 1, \cdots, K$. In the implementation, the variance is reduced as the steps grow. For the generative denoising process, the posterior distribution $p(\cdot)$ can be computed in a close form by the Bayesian formula:

$$p(x_{t-1}|x_t, x_0) = p(x_t|x_{t-1}, x_0)\frac{p(x_{t-1}|x_0)}{p(x_t|x_0)} = \mathcal{N}\left(x_{t-1}\middle|\tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t \boldsymbol{I}\right),$$

$$p(h_{t-1}|h_t, h_0) = p(h_t|h_{t-1}, h_0)\frac{p(h_{t-1}|h_0)}{p(h_t|h_0)} = \mathcal{C}\left(h_{t-1}\middle|\boldsymbol{\theta}_{post}(h_t, h_0)\right),$$

where $\boldsymbol{\theta}_{post}(h_t, h_0) = \tilde{\boldsymbol{\theta}}/\sum_{k=1}^{K}\tilde{\boldsymbol{\theta}}_k$, $\tilde{\boldsymbol{\theta}} = [\alpha_t h_t + (1-\alpha_t)/K] \odot [\bar{\alpha}_{t-1}h_0 + (1-\bar{\alpha}_{t-1})/K]$, $\tilde{\mu}_t(x_t, x_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t$, $\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$ and $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{s=1}^{t}\alpha_s$. In the denoising process, x_0 and h_0 are approximated by neural network, and we denote the approximation of x_0 and h_0 as $\hat{x}_0, \hat{h}_0 = \Phi_\Omega(x_t, h_t, t)$, where Φ is a neural network parameterized by Ω. The training objective is the summation for atom coordinates and atom types. The atom coordinate loss states as:

$$L_{t-1}^{x} = \gamma_t||x_0 - \hat{x}_0||^2 + C,$$

where γ_t is the weight for MSE loss and C is a constant. In the implementation, we set $\gamma_t = 1$ for all time steps. The atom type loss is computed by KL-divergence of two categorical distributions:

$$L_{t-1}^{h} = \sum\nolimits_k \boldsymbol{\theta}_{post}(h_t, h_0)_k \cdot log\frac{\boldsymbol{\theta}_{post}(h_t, h_0)_k}{\boldsymbol{\theta}_{post}(h_t, \hat{h}_0)_k}.$$

The final loss is calculated by the weighted summation of MSE loss, KL-divergence and a classification loss:

$$L = \lambda_x L_{t-1}^{x} + \lambda_h L_{t-1}^{h} + \lambda_c Cls(h^p),$$

where λ_x, λ_h and λ_c are the weight for MSE loss, KL-divergence and classification loss respectively and h^p indicates the protein atom features. The classification loss classified the protein atom features according to their atomic interaction types and the cross entropy loss are used in the experiment.

Equivariant diffusion under prompt guidance

In this section, we elaborate our proposed InterDiff model. InterDiff is a graph neural network in which the atom denotes the nodes and the Euclidean distance between atoms denotes the edges. We define an edge among two nodes when the Euclidean distance is below 7 angstroms. Let $v = \left(v_I^{(d)}, v_I^{(c)}\right)$ denotes the interaction prompts, where $v_I^d \in \mathcal{R}^4$ are one-hot representation prompts, $I \in (cation - pi, halogen, hydrogen, pi - pi)$ and v_I^c are learnable continuous embeddings. The atom node features are also encoded by one-hot vectors and transformed by a single linear layer: $h^{0,(p)} = Linear\left(\left[h^p, v_I^{(d)}\right]\right), h^{0,(L)} = Linear(h^L)$.

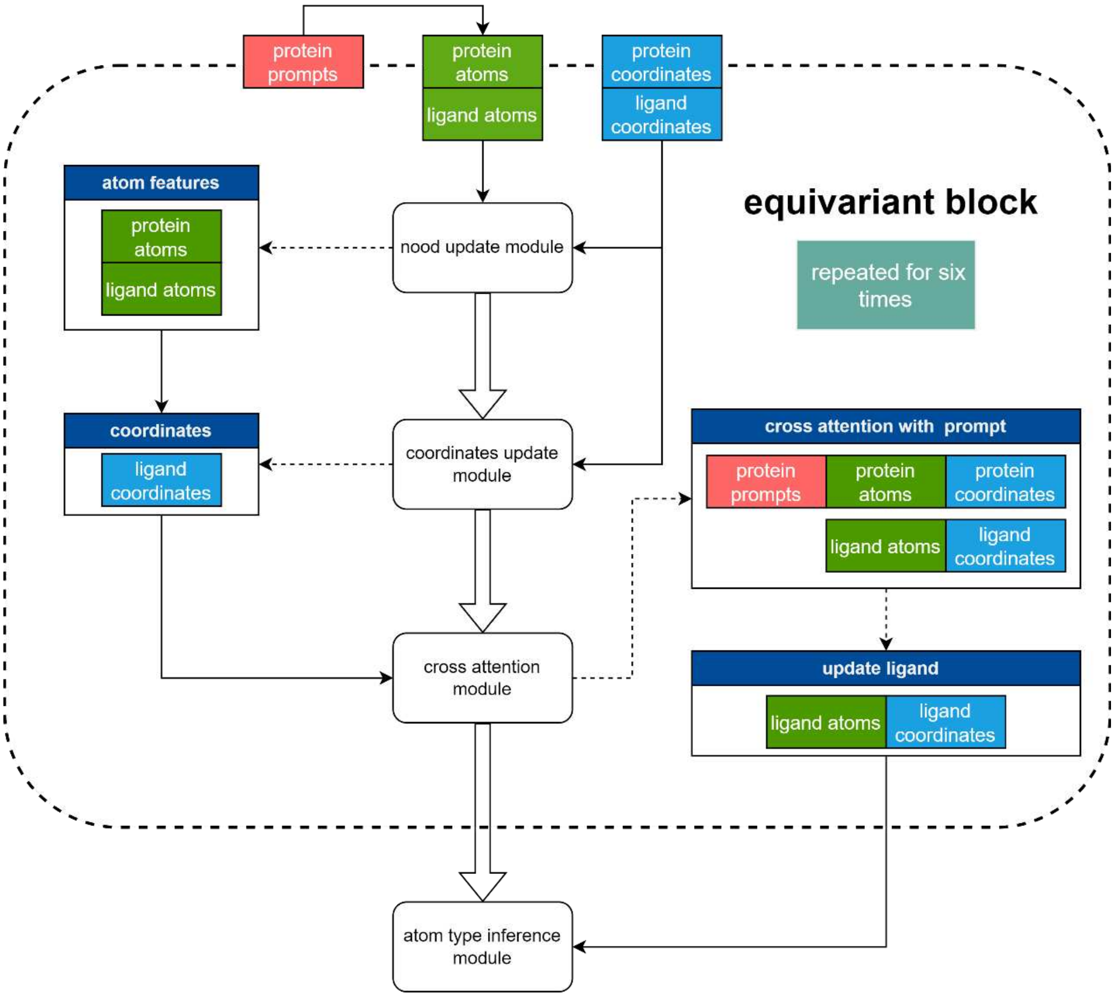


Figure 6: Structure of equivariant block used in InterDiff. Modules are represented by rounded rectangle with white context while data are shown by rectangle with distinct colors. Input and output flows are shown with arrowhead and dashed arrowhead respectively.

InterDiff is composed of six equivariant block layers (Figure 6), and each block consists of three modules. Formally, the first module updates the node features:

$$h^{l-1,(P)} = h^{l-1,(P)} + v_I^{(c)},$$

$$\begin{aligned}
&m_{i,j} = cat\left(d_{ij,}h_i^{l-1}, h_j^{l-1}\right),\\
&h_k = MLP\left(m_{i,j}\right), h_v = MLP\left(m_{i,j}\right), h_q = MLP\left(h^{l-1}\right),\\
&h^l = h^{l-1} + attention\left(h_k, h_q, h_v\right),
\end{aligned}$$

where $h^{l,(P)}$ indicates protein atom features in lth layer, d_{ij} is the Euclidean distance between atom i and j, and $cat(\cdot)$ indicates the concatenation operation. The second module updates the ligand coordinates:

$$\begin{aligned}
&m_{i,j} = cat\left(d_{ij,}h_i^{l-1}, h_j^{l-1}\right),\\
&h_k = MLP\left(m_{i,j}\right), h_v = MLP\left(m_{i,j}\right), h_q = MLP\left(h^{l-1}\right),\\
&h_v = h_v \cdot [x_i^{l-1} - x_{j,j\in Ne(i)}^{l-1}]_{i=1}^{N}\\
&x^{l,(L)} = x^{l-1,(L)} + attention\left(h_k, h_q, h_v\right),
\end{aligned}$$

where $x^{l,(L)}$ indicates ligand atom coordinates in lth layer, $x_{j,j\in Ne(i)}^{l}$ represents that jth atom coordinates and j is the neighbors of atom i. The third module updates ligand atom features and coordinates simultaneously with cross attention:

$$\begin{aligned}
&\tilde{d}_{ij} = dis\_encoding\left(d_{ij}\right)\\
&cxt = cat\left(h^{l_h,(P)}, v_I^{(c)}\right)\\
&h^{l,(L)}, x^{l,(L)} = crossatte\left(cxt, \tilde{d}_{ij}, d_{ij}, h^{l_h,(L)}, x^{l_x,(L)}, x^{l_x,(p)}\right),
\end{aligned}$$

where $dis\_encoding(\cdot)$ is the encoding of distance matrix between ligand and protein, $h^{l_h,(L)}$, $x^{l_x,(L)}$ and $x^{l_x,(P)}$ indicate the node features and coordinates of ligand and coordinates of protein from the first and second module. For the distance encoding, we use multilayer perceptron in the implementation. The $crossatte(\cdot)$ is computed as follows:

$$\begin{aligned}
&h_{qk}^{l,(L)} = Linear\left(h^{l_h,(L)}\right), cxt_{qk} = Linear(cxt),\\
&sim = MLP\left(cat\left(h_{qk}^{l,(L)} \cdot cxt_{qk}, \tilde{d}_{ij}\right)\right),\\
&\tilde{h}^{l,(L)} = softmax(sim)\left[:\frac{1}{2}nheads\right] \cdot cxt_v,\\
&\tilde{x}^{l,(L)} = mean\left(softmax(sim)\left[\frac{1}{2}nheads:\right]\right),\\
&h^{l,(L)} = h^{l_h,(L)} + MLP\left(\tilde{h}^{l,(L)}\right),\\
&x^{l,(L)} = x^{l_x,(L)} + [x_i^{l_x,(L)} - x_j^{l_x,(p)}]_{ij} \cdot MLP\left(\tilde{x}^{l,(L)}\right),
\end{aligned}$$

where $nheads$ is the number of heads in cross attention mechanism, sim is the attention map between ligand and protein. Noted that we ignore the operations of dimension rearrangement in above formulas and simplify Einstein summation to dot product here. Identical to previous work [35, 46, 47], we use 'subspace-trick' by limiting the center of mass (CoM) of training samples to zero to ensure the model can achieve translation invariance in the generative process. For the SE(3)-equivariance of Markov transition, the proof of the first two modules is similar to [10] and we prove the equivariance for the cross attention module in the supplementary.

Training and sampling details

InterDiff consists of 6 equivariant blocks and each block has three modules with transformer like structure. The diffusion steps are set to 1000 in training and sampling. We utilize a sigmoid β scheduler for atom coordinates and a cosine β scheduler for atom types. The number of heads is 16 for the first two modules and 32 for the cross attention module. The dimension is 128 for the atom features and interaction prompt $v_I^{(c)}$. We use Adam[48] method to optimize the model with an initial learning rate 0.001, betas=(0.95,0.999) and the batch size is set to 8. A 'plateau' scheduler was applied to decay the learning rate with a factor 0.8 when the validation performance is stuck for 4 evaluation steps. The minimum learning rate is 1e-6. The loss weight is 100 for atom type loss and 1 for MSE loss. We train InterDiff on one NVIDIA V100S GPU and the model converges within 32 hours. In addition, we empirically found that the model could be further improved on validation set (randomly selected from training data for validation) when fix the prompt embedding and fine tune after convergence.

In the sampling process, the interaction prompts for each sample are provided in keeping with the molecule in test set. The center of mass is subtracted from the coordinates of protein atoms and the number of atoms is sample according to the pocket size (Figure S3). The initial coordinates of ligand atoms are sampled from a normal distribution and atom types are sampled from a Gumbel distribution and then transformed into one-hot vectors.

Featurization of atoms and distance

Atoms in ligand and protein are represented by one-hot vector initially and then transformed by a linear layer. We use a mixed representation for protein atoms and ligand atoms as described in [10]. Specifically, the protein atom features encode the information about amino acid types, atom types and whether the atom is backbone atoms. The ligand atom features encode the atom types and aromatic information. The distance between atoms and bond types are used to construct graph edges. Four types of bonds are considered by one-hot vector, which indicates the connection between ligand atoms, protein atoms, ligand-protein atoms and protein-ligand atoms. The edge feature are then encoded by gaussian radial basis functions with learnable parameters of mean and variance, for the details please refer to [49].

Characterizations and parameters of interactions

In this paper, we consider four types of interactions, and the characterizations of interactions are consistent with BINANA2. Cation-π interactions comprise of a charged functional group and an aromatic ring. The coordinate of charged functional groups is projected to the plane of the aromatic ring and cation-π interaction is accepted if the distance of two center points between pairs is less than a threshold. π-π interactions have two types of forms, pi-pi stacking (face to face) and T-stacking (edge to face). To detect π-π interactions, distance of the projection of center points on two aromatic rings and the angle of two vectors normal to planes for each ring are calculated. If the distance and angle satisfy certain thresholds, a π-π interaction is identified. Hydrogen bond is composed of a hydrogen bond donor and a hydrogen bond acceptor. In BINANA2, thiol, amine, and hydroxyl groups are allowed as donors and nitrogen, sulfur and oxygen atoms can act as receptors. Likewise the distance between donor and receptor and the dihedral angle between hydrogen atoms, donor and receptor must locate in a certain range. Halogen bonds also consist of a donor and a receptor. The donors include O-X, N-X, S-X, and C-X, where X is F, Cl, Br, or I. The acceptors could be nitrogen, sulfur and oxygen atoms. The threshold of distance for halogen bonds tends to be longer than hydrogen bonds and the dihedral angle is the same. The details of threshold values are listed in Table S2.

Data availability

The CrossDocked 2020 can be obtained at https://bits.csb.pitt.edu/files/crossdock2020/; Structured models used in studies are deposited in Protein Data Band with accession codes 6oik, 3uon, 8azx and 8azv. The source code will be available on Github when publised.

Competing interests

The authors declare no conflict of interest.

Reference

1. Anderson, A.C., *The process of structure-based drug design.* Chemistry & biology, 2003. **10**(9): p. 787-797.
2. Lipinski, C.A., et al., *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.* Advanced drug delivery reviews, 2012. **64**: p. 4-17.
3. Gómez-Bombarelli, R., et al., *Automatic chemical design using a data-driven continuous representation of molecules.* ACS central science, 2018. **4**(2): p. 268-276.
4. Grisoni, F., et al., *Bidirectional molecule generation with recurrent neural networks.* Journal of chemical information and modeling, 2020. **60**(3): p. 1175-1183.
5. Jin, W., R. Barzilay, and T. Jaakkola. *Junction tree variational autoencoder for molecular graph generation*. in *International conference on machine learning*. 2018. PMLR.
6. Luo, Y., K. Yan, and S. Ji. *Graphdf: A discrete flow model for molecular graph generation*. in *International Conference on Machine Learning*. 2021. PMLR.
7. Ragoza, M., T. Masuda, and D.R. Koes, *Generating 3D molecules conditional on receptor binding sites with deep generative models.* Chemical science, 2022. **13**(9): p. 2701-2713.
8. Liu, M., et al., *Generating 3d molecules for target protein binding.* arXiv preprint arXiv:2204.09410, 2022.
9. Schneuing, A., et al., *Structure-based drug design with equivariant diffusion models.* arXiv preprint arXiv:2210.13695, 2022.
10. Guan, J., et al., *3d equivariant diffusion for target-aware molecule generation and affinity prediction.* arXiv preprint arXiv:2303.03543, 2023.
11. Lin, H., et al., *Diffbp: Generative diffusion of 3d molecules for target protein binding.* arXiv preprint arXiv:2211.11214, 2022.
12. Huang, L., et al. *Mdm: Molecular diffusion model for 3d molecule generation*. in *Proceedings of the AAAI Conference on Artificial Intelligence*. 2023.
13. Huang, L., *A dual diffusion model enables 3D binding bioactive molecule generation and lead optimization given target pockets.* bioRxiv, 2023: p. 2023.01. 28.526011.

14. Peng, X., et al. *Pocket2mol: Efficient molecular sampling based on 3d protein pockets*. in *International Conference on Machine Learning*. 2022. PMLR.
15. Song, Y., et al., *Score-based generative modeling through stochastic differential equations.* arXiv preprint arXiv:2011.13456, 2020.
16. Song, Y. and S. Ermon, *Generative modeling by estimating gradients of the data distribution.* Advances in neural information processing systems, 2019. **32**.
17. Song, Y., et al. *Sliced score matching: A scalable approach to density and score estimation*. in *Uncertainty in Artificial Intelligence*. 2020. PMLR.
18. Zerbe, B.S., et al., *Relationship between hot spot residues and ligand binding hot spots in protein–protein interfaces.* Journal of chemical information and modeling, 2012. **52**(8): p. 2236-2244.
19. Kozakov, D., et al., *Structural conservation of druggable hot spots in protein–protein interfaces.* Proceedings of the National Academy of Sciences, 2011. **108**(33): p. 13528-13533.
20. Friedman, R., *Computational studies of protein–drug binding affinity changes upon mutations in the drug target.* Wiley Interdisciplinary Reviews: Computational Molecular Science, 2022. **12**(1): p. e1563.
21. Wan, S., et al., *The effect of protein mutations on drug binding suggests ensuing personalised drug selection.* Scientific Reports, 2021. **11**(1): p. 13452.
22. Francoeur, P.G., et al., *Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design.* Journal of chemical information and modeling, 2020. **60**(9): p. 4200-4215.
23. Li, X.L. and P. Liang, *Prefix-tuning: Optimizing continuous prompts for generation.* arXiv preprint arXiv:2101.00190, 2021.
24. Young, J., N. Garikipati, and J.D. Durrant, *BINANA 2: characterizing receptor/ligand interactions in Python and JavaScript.* Journal of chemical information and modeling, 2022. **62**(4): p. 753-760.
25. Taylor, R.D., M. MacCoss, and A.D.J.J.o.m.c. Lawson, *Rings in drugs: Miniperspective.* 2014. **57**(14): p. 5845-5859.
26. Burger, W.A., et al., *Toward an understanding of the structural basis of allostery in muscarinic acetylcholine receptors.* 2018. **150**(10): p. 1360-1372.
27. Bodick, N.C., et al., *Effects of xanomeline, a selective muscarinic receptor agonist, on cognitive function and behavioral symptoms in Alzheimer disease.* 1997. **54**(4): p. 465-473.
28. Powers, A.S., et al., *Structural basis of efficacy-driven ligand selectivity at GPCRs.* 2023: p. 1-10.
29. Kim, D., et al., *Pan-KRAS inhibitor disables oncogenic signalling and tumour growth.* 2023: p. 1-7.
30. Lugmayr, A., et al. *Repaint: Inpainting using denoising diffusion probabilistic models*. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.
31. Song, Y., et al., *Score-based generative modeling through stochastic differential equations.* 2020.
32. Kingma, D., et al., *Variational diffusion models.* Advances in neural information processing

systems, 2021. **34**: p. 21696-21707.

33. Kong, Z., et al., *Diffwave: A versatile diffusion model for audio synthesis.* arXiv preprint arXiv:2009.09761, 2020.

34. Ho, J., et al., *Imagen video: High definition video generation with diffusion models.* arXiv preprint arXiv:2210.02303, 2022.

35. Hoogeboom, E., et al. *Equivariant diffusion for molecule generation in 3d*. in *International conference on machine learning*. 2022. PMLR.

36. Wu, L., et al., *Diffusion-based molecule generation with informative prior bridges.* Advances in Neural Information Processing Systems, 2022. **35**: p. 36533-36545.

37. Peng, X., et al., *MolDiff: Addressing the Atom-Bond Inconsistency Problem in 3D Molecule Diffusion Generation.* arXiv preprint arXiv:2305.07508, 2023.

38. Liu, P., et al., *Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing.* 2023. **55**(9): p. 1-35.

39. Zhou, K., et al., *Learning to prompt for vision-language models.* 2022. **130**(9): p. 2337-2348.

40. Liu, S., et al., *Multi-modal molecule structure-text model for text-based retrieval and editing.* arXiv preprint arXiv:2212.10789, 2022.

41. Dollar, O.W., et al., *MolJET: Multimodal Joint Embedding Transformer for Conditional de novo Molecular Design and Multi-Property Optimization.* 2022.

42. Liu, Z., et al., *MolXPT: Wrapping Molecules with Text for Generative Pre-training.* arXiv preprint arXiv:2305.10688, 2023.

43. Gao, Z., et al., *Prefixmol: Target-and chemistry-aware molecule design via prefix embedding.* arXiv preprint arXiv:2302.07120, 2023.

44. Kozakov, D., et al., *The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins.* Nature protocols, 2015. **10**(5): p. 733-755.

45. Hoogeboom, E., et al., *Argmax flows and multinomial diffusion: Learning categorical distributions.* Advances in Neural Information Processing Systems, 2021. **34**: p. 12454-12465.

46. Köhler, J., L. Klein, and F. Noé. *Equivariant flows: exact likelihood generative learning for symmetric densities*. in *International conference on machine learning*. 2020. PMLR.

47. Xu, M., et al., *Geodiff: A geometric diffusion model for molecular conformation generation.* arXiv preprint arXiv:2203.02923, 2022.

48. Kingma, D.P. and J.J.a.p.a. Ba, *Adam: A method for stochastic optimization.* 2014.

49. Luo, S., et al., *One transformer can understand both 2d & 3d molecular data.* arXiv preprint arXiv:2210.01765, 2022.