

Target-aware 3D molecular generation based on guided equivariant diffusion

Received: 17 October 2024

Accepted: 5 August 2025

Published online: 25 August 2025

Qiaoyu Hu^{1,4}✉, Changzhi Sun^{2,4}, Huan He^{1,4}, Jiazheng Xu^{3,4}, Danlin Liu^{1,3,4}, Wenqing Zhang¹, Sumeng Shi¹, Kai Zhang^{1,3} & Honglin Li^{1,2}✉

Recent molecular generation models for structure-based drug design (SBDD) often produce unrealistic 3D molecules due to the neglect of structural feasibility and drug-like properties. In this paper, we introduce DiffGui, a target-conditioned E(3)-equivariant diffusion model that integrates bond diffusion and property guidance, to address the above challenges. The combination of atom diffusion and bond diffusion guarantees the concurrent generation of both atoms and bonds by explicitly modeling their interdependencies. Property guidance incorporates the binding affinity and drug-like properties of molecules into the training and sampling processes. Extensive experiments prove that DiffGui outperforms existing methods in generating molecules with high binding affinity, rational chemical structure, and desirable properties. Ablation studies confirm the importance of bond diffusion and property guidance modules. DiffGui demonstrates effectiveness in both de novo drug design and lead optimization, with validation through wet-lab experiments.

Drug discovery not only greatly impacts individuals' physical health and the quality of their lives, but also serves as a vital catalyst for social progress and national economic prosperity. However, the development of innovative drugs is fraught with challenges and uncertainties. This process, involving retrieval of lead compound, lead optimization, preclinical evaluations, and clinical trials, typically spans 10 years and costs billions of dollars on average^{1–3}. The discovery of lead compounds is the most important stage in the entire process because it exerts huge influences on subsequent development steps and determines the fate of the project to a large extent^{4,5}. Traditionally, the identification of potential drug candidates mainly relied on incidental occurrences^{6,7} that are inherently difficult to replicate with consistency. However, with the advancements of techniques in molecular biology, structural biology, combinatorial chemistry, and artificial intelligence (AI), the paradigm of drug discovery has been transferred from random methods to rational drug design, which can significantly increase the success rate and efficiency of drug development.

Rational drug design is composed of two approaches: ligand-based drug design (LBDD) and structure-based drug design (SBDD). LBDD designs new molecules by modifying the existing active ligands

to enhance their binding affinity, selectivity, and pharmacokinetic/pharmacodynamic properties⁸. It is particularly valuable when the three-dimensional (3D) structure of the biological target is unknown. However, with the accurate prediction of biomolecular structures now widely available through AI-based methods such as AlphaFold^{9–11}, LBDD faces limitations for not incorporating the structural information of target proteins. In addition, it is also unsuitable for proteins with few or no known ligands, which is a common situation when developing drugs for novel targets. By contrast, SBDD is believed to be more effective to deliver the ligand molecules inside the binding pockets by considering the drug-target interactions at the molecular level^{12–14}. It includes two main protocols: virtual screening and molecular generation. Virtual screening employs physics-based or data-based scoring functions to estimate the binding affinities between targets and ligands, thereby selecting top-ranked molecules from chemical compound libraries for subsequent wet-lab validation and further optimization^{15,16}. Yet, it is computationally expensive to search the physical libraries that involve 10^6 – 10^7 molecules or the virtual on-demand libraries that contain 10^{10} – 10^{15} molecules, let alone the massive chemical space (10^{60} – 10^{100}) of potential pharmacologically active

¹Innovation Center for AI and Drug Discovery, School of Pharmacy, East China Normal University, Shanghai, China. ²Lingang Laboratory, Shanghai, China.

³School of Computer Science and Technology, East China Normal University, Shanghai, China. ⁴These authors contributed equally: Qiaoyu Hu, Changzhi Sun, Huan He, Jiazheng Xu, Danlin Liu. ✉e-mail: qyuhu@hsc.ecnu.edu.cn; hlli@hsc.ecnu.edu.cn

molecules^{17,18}. Besides, the gigascale screening must be extremely accurate to guard against the false-positive hits that can cheat the scoring function by exploiting its imperfections and approximations¹⁹. Even a minimal false-positive rate of one in a million in a 10^{10} library would result in 10,000 false hits, which may flood out valid hit candidate selection²⁰.

Recent breakthroughs in geometric deep learning techniques^{21–23} have facilitated the emergence of deep generative models^{24–28}, which are capable of directly producing pocket-aware ligands with appropriate 3D conformations. Early pioneers have attempted to represent the molecules as atomic density maps and the 3D space as voxelized grids^{27,29}. They harness 3D convolutional neural networks (3D CNNs) to model the protein-ligand complex and utilize conditional variational autoencoders (VAEs) to generate new molecules. Nonetheless, these models are not equivariant on molecular geometry and suffer from serious scalability problems owing to the exponential growth of the voxels' number as the pocket size increases. To address these issues, the following approaches^{24–26,30,31} represent the molecules as 3D graphs and achieve SE(3)-equivariance through various techniques. For instance, GraphBP²⁶ incorporates the embeddings of atomic distance and bond angle into the training and sampling processes. Pocket2Mol²⁴ employs an E(3)-equivariant graph neural network (GNN) to ensure the rotational and translational equivariance of the system. Despite their improved performance, these models adopt an autoregressive strategy to generate the ligand atoms sequentially, which may suffer from several inherent shortcomings. Firstly, the sequential sampling models impose an unnatural generation order of atoms, thereby neglecting the global context information of the ligand. Secondly, errors introduced during the initial stages of the sampling process may gradually accumulate to promote the formation of invalid structures. Lastly, the autoregressive models frequently encounter the problem of premature termination, thus resulting in the generation of small fragments instead of complete ligands.

Diffusion-based methods^{32–35} alleviate the aforementioned problems via the implementation of non-autoregressive generation scheme. By integrating diffusion probabilistic models³⁶ and equivariant neural networks^{37–39}, these methods can accomplish the task of pocket-conditioned molecular generation within continuous 3D space. Generally, each atom in the protein-ligand complex is characterized by continuous atom coordinates and discrete atom types, with noise being incrementally introduced during the forward diffusion process. The equivariant GNN is utilized to not only update the atom embeddings by message passing mechanisms, but also preserve the rotation, translation, and permutation symmetries. In reverse diffusion, atom types and positions are predicted by denoising from categorical and Gaussian distributions, respectively. However, the diffusion-based models are often inclined to produce unrealistic molecules with distorted structures, such as three- or four-membered rings, extra-large rings, and fused rings, which are energetically unstable and synthetically difficult. This may stem from the manner that the complete molecules are constructed. After acquiring the atom positions, current models typically predict the bond types based on canonical bond lengths and assemble them into intact molecules using the OpenBabel toolkit⁴⁰. As a consequence, minor deviations in atom coordinates can give rise to incorrect identification of bond types, subsequently affecting the overall structure of the generated ligand. Although DecompDiff⁴¹ incorporates molecular inductive bias into the training process by pre-decomposing ligands into arms and scaffolds, and leverages the validity guidance to instruct the sampling procedure, it cannot fully resolve the issue of ill-conformations because of the complexity and overwhelming diversity of inductive biases. On the other hand, most diffusion models aim to yield high-affinity binders without explicitly considering the essential drug-like properties such as drug-likeness^{42,43}, synthetic accessibility⁴⁴, and the octanol-water partition coefficient⁴⁵, which serve as crucial criteria for choosing

favorable compounds. They aspire to implicitly extract the relevant information from existing protein-ligand datasets, albeit acknowledging that the molecules contained in these datasets may not uniformly exhibit optimal or satisfactory properties.

In this work, inspired by the work that solves the atom-bond inconsistency problem⁴⁶ and classifier-free diffusion guidance⁴⁷, we propose DiffGui, a novel guided diffusion model to tackle the above issues. It can not only mitigate the ill-conformational problem by introducing bond diffusion as a guidance to generate atom coordinates, but also address the attribute issue by employing property guidance during training and sampling processes. Extensive experiments presented in this study have demonstrated that DiffGui can effectively generate novel 3D molecules with high estimated binding affinities, plausible chemical structures and desired molecular properties inside the given protein pockets. It achieves the state-of-the-art (SOTA) performance on various evaluation metrics for the PDBbind dataset and exhibits competitive outcomes for the CrossDocked dataset. Case studies further confirm the superiority of DiffGui in the realms of de novo drug design and lead optimization. The generation experiments for mutated targets suggest that DiffGui is sensitive to minor changes within the protein pocket environment, underscoring its capacity to capture the complicated topological and geometrical information.

Results

This section is organized as follows: First, we describe the overall framework of the DiffGui model. Second, we compared the quality, molecular metrics and properties of the ligands generated by our method with those produced by other existing SOTA methods. Subsequently, we conducted ablation studies to determine the respective roles of bond diffusion and property guidance modules. Finally, we demonstrated the practical value of DiffGui by applying it to structure-based drug design for protein targets, lead optimization based on fragments, and molecule generation for mutated targets. Specifically, the quality of generated molecules is primarily evaluated by the Jensen-Shannon (JS) divergence between the distributions of bonds, angles, and dihedrals for the reference and generated ligands. The RMSD (root mean square deviation) values between the generated geometries and optimized/predicted conformations are also utilized as an evaluation metric for quality. The basic molecular metrics include atom stability, molecular stability, PoseBusters validity (PB-validity), RDKit validity, novelty, uniqueness, similarity with reference ligands, and similarity of protein-ligand interaction fingerprints. The molecular properties encompass estimated binding affinity (Vina Score), quantitative estimate of drug-likeness (QED), synthetic accessibility (SA), octanol-water partition coefficient (LogP), and topological polar surface area (TPSA).

Overview of DiffGui framework

DiffGui is a bond- and property-guided, non-autoregressive generative model for target-aware molecule generation based on the equivariant diffusion framework³⁶. It integrates the mechanism of atom diffusion and bond diffusion into the forward process, while leveraging an array of molecular properties such as affinity, QED, SA, LogP, and TPSA to guide the reverse generative process (Fig. 1a). Essentially, during the forward process $q(\mathbf{x}^t|\mathbf{x}^{t-1}, \mathbf{p}, \mathbf{c})$ (\mathbf{p} and \mathbf{c} represent the protein pocket and the condition of molecular properties, respectively), noise is gradually injected into the atoms and bonds of the ligand based on different noise schedules (Fig. 1b). This divides the forward process into two distinct phases. In the first phase, the bond types are gradually diffused towards the prior distribution (none-bond type), while the atom types and their positions undergo marginal disruption. Injecting a small amount of noise into the atoms, rather than rigidly fixing their states in this phase, is important for enhancing the model's robustness. This approach provides more flexibility in predicting the bond types,

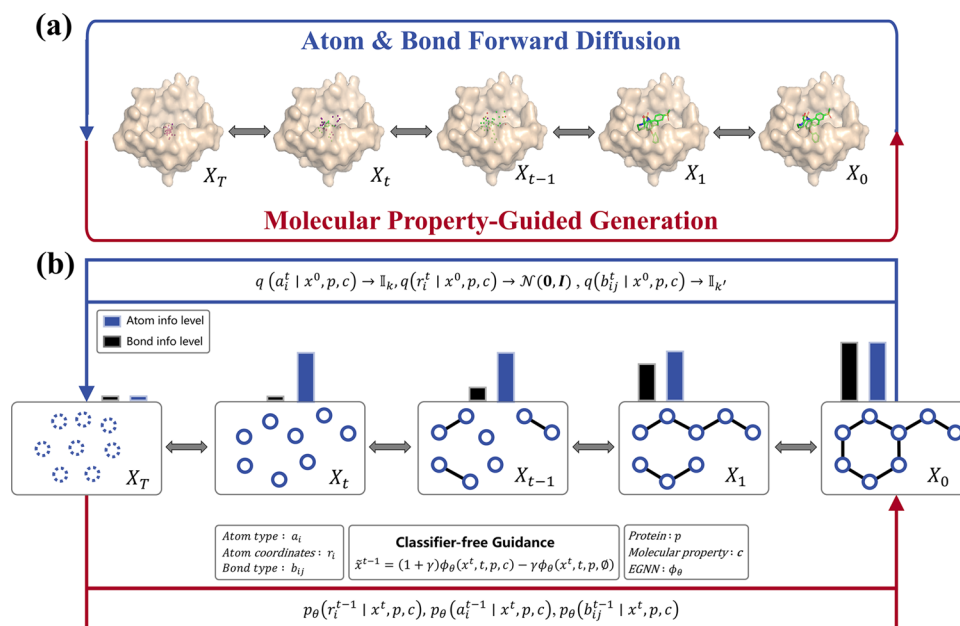


Fig. 1 | Overview of DiffGui framework. a The diagram of diffusion and generative processes in DiffGui. **b** The details of the DiffGui structure.

as they can now be inferred from the dynamic atom distances within a specified range, rather than relying solely on the static values. In the second phase, the atom types and positions are both perturbed to their prior distributions. By this means, the model circumvents learning bond types with bond lengths that significantly deviate from the ground truth during the diffusion process. The E(3)-equivariant GNN is also modified to update the representations of both atom and bond within the message passing framework. Since atom coordinates are continuous while atom/bond types are discrete, we utilize a Gaussian distribution to model the former and categorical distributions to represent the latter. Thus, the joint molecular distribution can be formulated as a product of atom coordinate distribution and atom/bond type distributions.

In addition to the protein pocket, molecular properties are also considered as a distinct condition that is incorporated into the atomic features. Instead of sampling along the gradient direction of a label-specific classifier⁴⁸, we embrace classifier-free guidance⁴⁷ that jointly trains the unconditional and conditional models by randomly setting the property label to a null taken \emptyset with a probability. This simplifies the training pipeline, as it eliminates the need for training an additional label classifier. During the reverse process $p_\theta(x^{t-1} | x^t, p, c)$, the sampling can be performed using a linear combination of the conditional and unconditional score estimates, where γ is a parameter that controls the strength of property guidance (Fig. 1b). Besides, due to the strong relationship between the bond type and bond length, the generation of atom positions is further guided by the confidence of a bond predictor, which takes the atom types and coordinates as input to predict the bond types. In this manner, the atoms can be placed in the correct positions, thus facilitating the generation of accurate 3D conformations of molecules. For an in-depth exploration of DiffGui and its underlying methodologies, please refer to the Methods section for more details.

Quality of generated molecules

As previously stated, we mainly rely on two key metrics, JS divergence and RMSD values, to compare the sub-structural and global geometry of molecules generated by diverse methods. The JS divergence⁴⁹ is a method of measuring the similarity between two probability distributions. It is a symmetrized and smoothed version of the Kullback-Leibler (KL) divergence⁵⁰, and lower JS values indicate greater

similarity. Hence, JS divergence is employed here to assess the extent to which the sub-structures of generated molecules can effectively capture the true geometric distributions presented within the reference ligands. As displayed in Figs. 2 and 3, comparing with other generative models (ResGen³⁰, PocketFlow³¹, GCDM³², TargetDiff³³, DiffSBDD³⁴, and PMDM³⁵, briefly explained in the Methods section), DiffGui achieves the lowest JS divergences of 0.1815 and 0.0486 in the distributions of C-C bond distance and all-atom pairs distance, respectively, when evaluated on the PDBbind dataset. The true C-C bond distance (Fig. 2) predominantly spans the range of 1.3 to 1.6 Å, featuring two distinct peaks located approximately at 1.4 and 1.5 Å. Despite presenting two peaks, DiffSBDD and PMDM have peak densities that do not closely align with the ground truth, whereas ResGen, PocketFlow, and TargetDiff exhibit a single peak at around 1.4 Å. The JS divergence of GCDM approaches that of DiffGui, yet remains slightly higher. Regarding the distance between all-atom pairs (Fig. 3), three prominent peaks emerge approximately at 1.5, 2.5, and 3.5 Å in true distribution, accompanied by shoulders extending between 4 and 6.5 Å. In matching the true distribution, ResGen, GCDM, and PMDM perform poorly, while PocketFlow, TargetDiff, and DiffSBDD demonstrate suboptimal performance. This signifies their inability to adequately model both the short- and long-range molecular interactions. On the CrossDocked dataset, DiffGui attains the second-lowest JS divergences of 0.1923 and 0.0704 in the distributions of C-C bond distance and all-atom pairs distance, respectively, as depicted in Supplementary Figs. 1 and 2.

Furthermore, to facilitate a more comprehensive evaluation, we have calculated the average JS divergences of bonds, angles, and dihedrals in generated molecules across different methods. The values are listed in Table 1. In comparison to other baselines, DiffGui exhibits the lowest JS divergence among all metrics on the PDBbind dataset, while achieving either the lowest or highly comparable values on the CrossDocked dataset. The better performance of DiffGui on PDBbind over CrossDocked is primarily because our model was trained specifically on PDBbind, which allows it to capture the underlying patterns and features within this dataset more effectively. As a result, the model is able to leverage the knowledge it has acquired to make more accurate predictions when evaluated on PDBbind. The CrossDocked dataset differs from PDBbind in various aspects, such as data distribution, complex structure, and binding

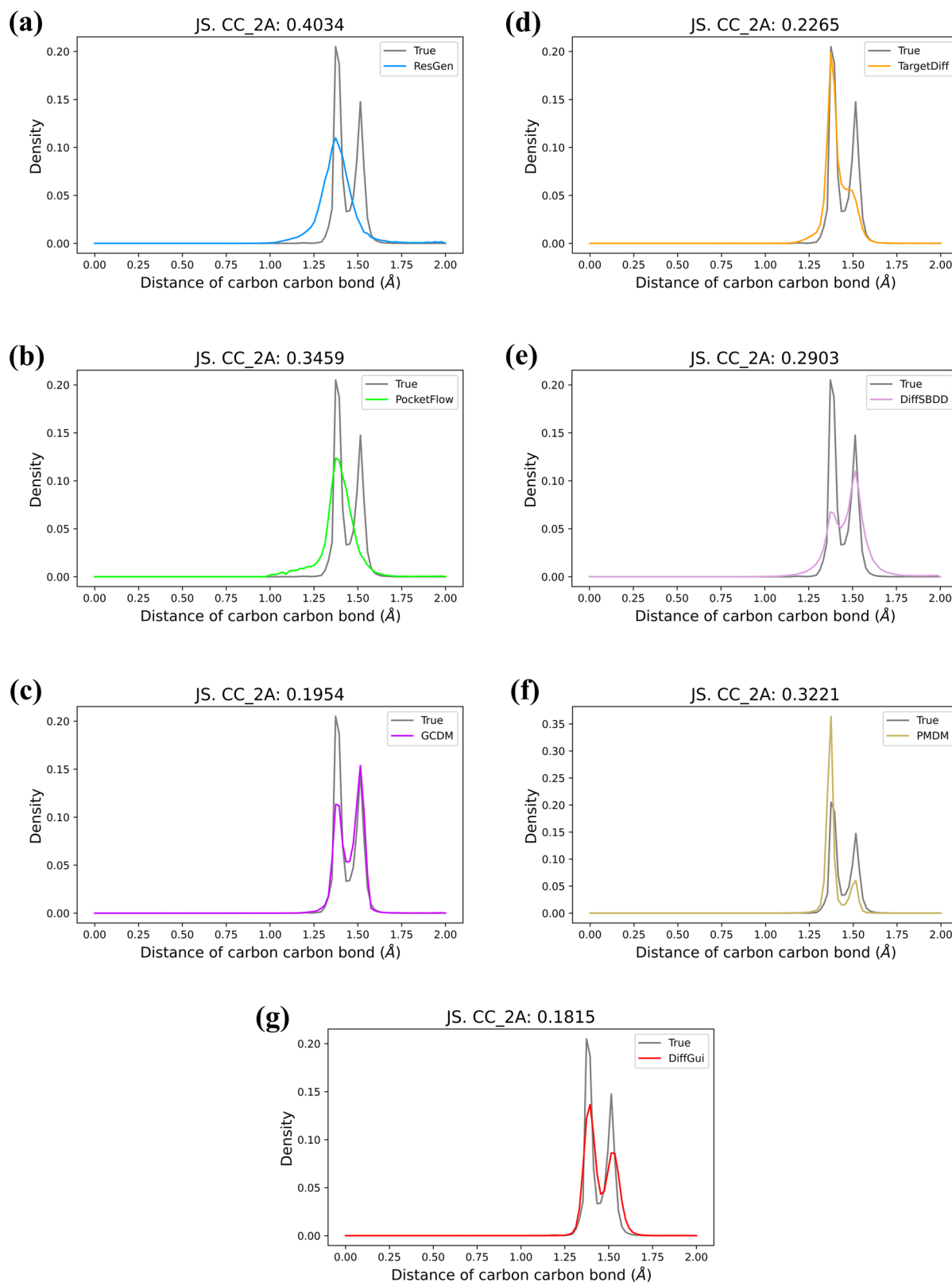


Fig. 2 | Comparison of the distributions of C-C bond distance for the PDBbind dataset. a ResGen, **(b)** PocketFlow, **(c)** GCDM, **(d)** TargetDiff, **(e)** DiffSBDD, **(f)** PMDM, **(g)** DiffGui. The reference molecules and generated molecules are depicted

by gray and colored lines, respectively. Jensen-Shannon divergence of C-C bond distance within 2.0 Å (JS. CC_2A) is listed on the top of each sub-figure. Source data are provided as a Source Data file.

affinity range. Due to the absence of a true affinity label, we cannot train DiffGui on this dataset. However, the fact that our model is able to perform reasonably well on this different dataset indicates that it has good generalization capabilities and it learns the generalizable

features that are not limited to the PDBbind dataset. Overall, the results underscore the ability of our method to more accurately capture the entanglement of 3D molecular information conditioned on the protein pockets, thereby generating molecules that exhibit a

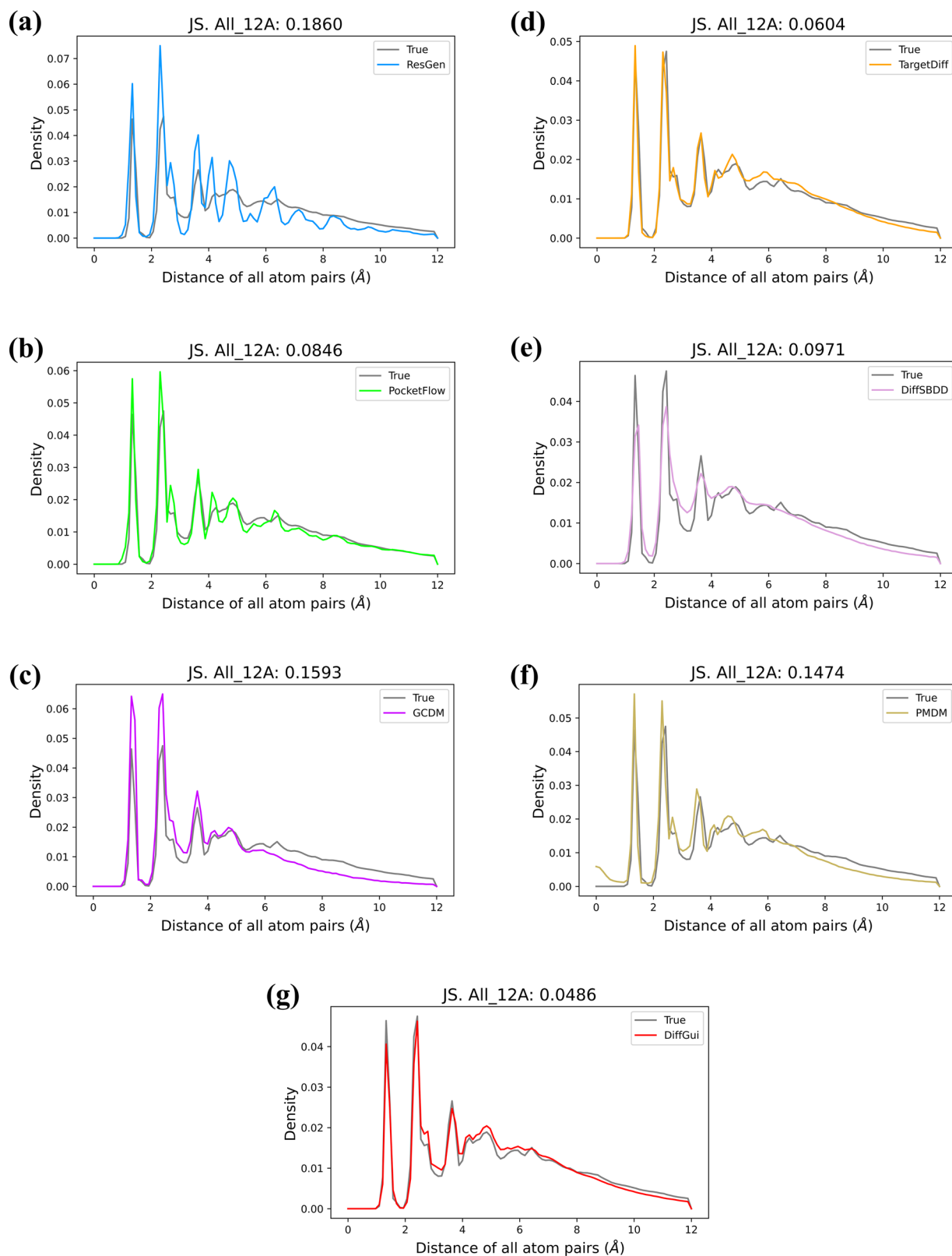


Fig. 3 | Comparison of the distributions of all-atom pairs distance for the PDBbind dataset. **a** ResGen, **(b)** PocketFlow, **(c)** GCDM, **(d)** TargetDiff, **(e)** DiffSBDD, **(f)** PMDM, **(g)** DiffGui. The reference molecules and generated molecules

are depicted by gray and colored lines, respectively. Jensen-Shannon divergence of all-atom pairs distance within 12.0 Å (JS. All_12A) is listed at the top of each subfigure. Source data are provided as a Source Data file.

higher degree of rationality on chemical structures. The detailed JS information on 20 covalent bond types, 13 bond angle types, and 15 dihedral angle types is presented in Supplementary Tables 1–6. Although DiffGui does not obtain minimum JS scores on certain

individual items, it is superior to other methods in light of its lowest mean values and overall performance.

The global geometry of the generated 3D conformation is assessed by computing the RMSD values between the original conformations

Table 1 | Average JS divergences between reference ligands and generated molecules

Datasets	JS div.	ResGen	PocketFlow	GCDM	TargetDiff	DiffSBDD	PMDM	DiffGui
PDB bind	bonds (↓)	0.576 ± 0.14	0.515 ± 0.12	0.461 ± 0.16	0.473 ± 0.17	0.469 ± 0.14	0.474 ± 0.17	0.363 ± 0.15
	angles (↓)	0.312 ± 0.11	0.257 ± 0.06	0.323 ± 0.07	0.267 ± 0.11	0.334 ± 0.08	0.239 ± 0.09	0.211 ± 0.08
	dihedrals (↓)	0.450 ± 0.09	0.404 ± 0.08	0.483 ± 0.14	0.372 ± 0.08	0.402 ± 0.07	0.454 ± 0.10	0.361 ± 0.10
Cross Docked	bonds (↓)	0.553 ± 0.15	0.483 ± 0.09	0.451 ± 0.17	0.467 ± 0.16	0.488 ± 0.14	0.402 ± 0.17	0.392 ± 0.18
	angles (↓)	0.238 ± 0.06	0.277 ± 0.07	0.361 ± 0.11	0.258 ± 0.06	0.327 ± 0.05	0.194 ± 0.05	0.198 ± 0.05
	dihedrals (↓)	0.404 ± 0.06	0.422 ± 0.08	0.490 ± 0.19	0.441 ± 0.11	0.481 ± 0.11	0.544 ± 0.13	0.423 ± 0.13

JS divergence is computed on the distributions of bond, angle, and dihedral. Bold fonts indicate the best results.

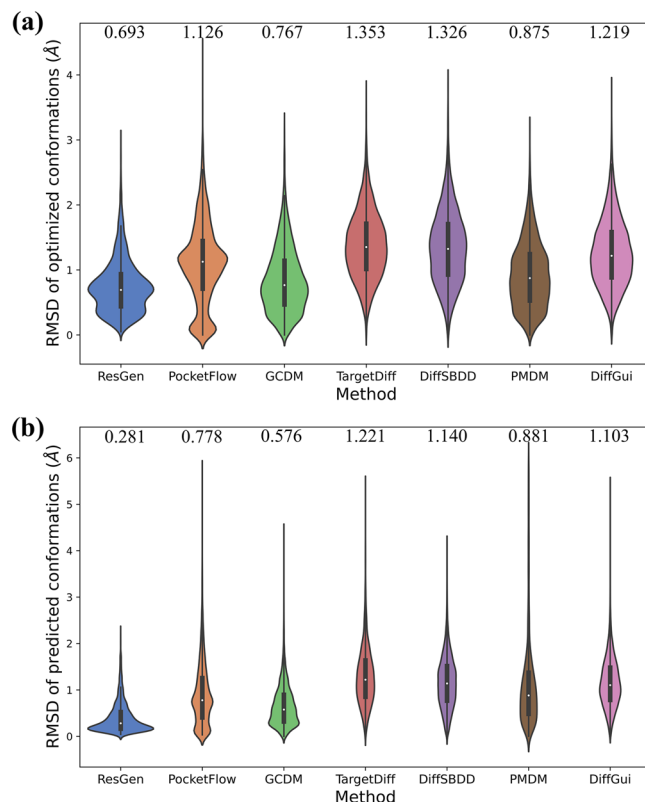


Fig. 4 | Comparison of conformational similarity between generated and optimized/predicted structures for the PDBbind dataset. (a) RMSD distributions between generated and optimized conformations, **(b)** RMSD distributions between generated and predicted conformations. Median RMSD value is listed on the top of each sub-figure. Each method samples 10,000 molecules, collecting one optimized conformation and twenty predicted conformations per molecule. The statistical descriptors for each box plot (minimum, maximum, median, and 25th/75th percentiles) are provided in Supplementary Table 17. Source data are provided as a Source Data file.

extracted from the generative models and the optimized/predicted conformations produced by the RDKit software (<https://www.rdkit.org/>). The generated conformation is optimized by the Merck Molecular Force Field (MMFF)⁵¹, while 20 conformations are predicted for each molecule using the ETKDG conformation generation algorithm⁵², followed by relaxation using the UFF force field⁵³. In Fig. 4 and Supplementary Fig. 3, we visually depicted the RMSD distributions by violin plots, with the median RMSD values displayed at the top of each sub-figure. Among the entire spectrum of models, ResGen and GCDM stand out as they reveal the two lowest median RMSD values on both PDBbind and CrossDocked datasets. Meanwhile, the remaining models showcase comparable levels of RMSD, with the highest median value being below 1.6 Å. The lower RMSD values of ResGen and GCDM can be attributed to distinct reasons. ResGen, as an autoregressive model, is prone to premature termination during generation, often producing small molecular fragments instead of complete molecules. GCDM adopts a fully-connected 3D graph that imposes significant computational costs for large molecules. This forces the model to simplify the generation process and prioritize the fragmentary outputs. In short, DiffGui consistently achieves ~1 Å RMSD for all scenarios, indicating its efficacy to generate molecules with appropriate sizes and plausible 3D conformations that optimally fit protein binding pockets.

For the percentage of ring sizes (Supplementary Tables 7 and 8), five- and six-membered rings constitute the majority (exceeding 95%) of rings in the reference molecules. Conversely, rings composed of three, four, seven, eight, and nine atoms are scarce due to their low

Table 2 | Basic molecular metrics of generated molecules for the PDBbind dataset

Metrics	ResGen	PocketFlow	GCDM	TargetDiff	DiffSBDD	PMDM	DiffGui
Atom Stab. (↑)	0.7849	0.8803	0.9507	0.9107	0.9039	0.7501	0.9509
Mol. Stab. (↑)	0.0738	0.3159	0.3423	0.1666	0.2945	0.0948	0.4069
PB-valid. (↑)	0.8805	0.9100	0.6101	0.8966	0.6792	0.7061	0.9537
RDKit-valid. (↑)	0.7205	0.8488	0.8172	0.7500	0.7895	0.8138	0.8722
Novel. (↑)	1.0000	0.8703	0.9923	0.9984	0.9920	0.9844	0.9942
Uniq. (↑)	0.8861	0.6074	0.8377	0.9969	0.9859	0.9545	0.9812
Sim. Ref. (↓)	0.3749	0.4098	0.4267	0.4913	0.4628	0.4882	0.3261
Inter. Sim. (↑)	0.4160	0.4444	0.3708	0.6063	0.5646	0.4393	0.6353

Bold fonts indicate the best results.

Table 3 | Average molecular properties of reference ligands and generated molecules for the PDBbind dataset

Metrics	Ref.	ResGen	PocketFlow	GCDM	TargetDiff	DiffSBDD	PMDM	DiffGui
Vina Score (↓)	−7.747 ± 3.56	−2.177 ± 1.68	−3.596 ± 1.80	−2.095 ± 1.29	−5.263 ± 2.54	−5.011 ± 2.28	−4.990 ± 2.41	−6.700 ± 2.55
Vina Min (↓)	−8.605 ± 2.44	−4.704 ± 2.16	−4.823 ± 1.85	−3.545 ± 1.49	−6.959 ± 2.58	−5.894 ± 2.22	−6.516 ± 2.61	−7.655 ± 2.51
Vina Dock (↓)	−9.133 ± 2.14	−7.154 ± 2.29	−6.245 ± 2.00	−5.130 ± 1.66	−8.195 ± 2.23	−6.853 ± 2.03	−7.577 ± 2.54	−8.448 ± 2.24
QED (↑)	0.459 ± 0.22	0.577 ± 0.15	0.501 ± 0.17	0.477 ± 0.13	0.460 ± 0.22	0.466 ± 0.19	0.474 ± 0.21	0.631 ± 0.21
SA (↑)	0.720 ± 0.10	0.784 ± 0.11	0.784 ± 0.15	0.706 ± 0.13	0.604 ± 0.13	0.633 ± 0.13	0.607 ± 0.15	0.678 ± 0.15
LogP	1.731 ± 2.74	1.749 ± 1.88	2.855 ± 1.80	0.807 ± 1.67	1.707 ± 2.38	1.384 ± 2.02	1.515 ± 2.60	1.977 ± 3.01
TPSA	115.91 ± 69.43	60.11 ± 29.91	38.46 ± 45.58	53.45 ± 34.28	103.06 ± 60.60	85.66 ± 45.87	117.74 ± 130.83	100.49 ± 62.97

Bold fonts indicate the best results. Statistical analyses are performed using two-sided paired *t* tests without adjustment for multiple comparisons.

chemical stability, limited synthetic accessibility, and notably high toxicity. Among all the methods evaluated, DiffGui emerges as the one that most closely replicates the percentages of five- and six-membered rings found in the reference molecules. ResGen and PocketFlow exhibit a tendency to generate a higher proportion of six-membered rings and a corresponding decrease in five-membered rings. GCDM, TargetDiff, DiffSBDD, and PMDM, on the contrary, yield fewer six-membered rings. It is also worth noting that our method displays an elevated proportion of seven-membered rings, which account for around 10% of all rings. This feature is commonly observed in methods using diffusion models, such as GCDM, TargetDiff, DiffSBDD, and PMDM. It represents a limitation of current diffusion-based approaches and raises an intriguing direction for future improvements. Diffusion models typically treat atoms as nodes and, during the inference stage, the initialization of ligand atom's number is primarily determined by the number of atoms within the protein pocket. Occasionally, this initialization might result in a slight excess of one or two atoms. This slight imbalance may favor the formation of seven-membered rings, which serve as a way to accommodate the extra atoms while maintaining the structural stability and connectivity within the generated molecules. Constructing diffusion models at the scale of a molecular fragment could address this issue. By focusing on pre-defined fragments rather than individual atoms, the model might be more adept at managing variations in atom number and mitigating the propensity to form seven-membered rings. This fragment-based approach could also utilize the statistical patterns and chemical properties of known fragments, guiding the generation process towards more pharmacological relevant and synthetically accessible molecules.

Molecular metrics and properties of generated molecules

We have calculated the basic molecular metrics of generated molecules to assess the generative abilities of various methods. As shown in Table 2 and Supplementary Table 9, our method, DiffGui, demonstrates superior performance over other methods in terms of atom

stability, molecular stability, PB-validity, and RDKit-validity. Regarding novelty and uniqueness, except for PocketFlow, the variations among different methods are relatively minor, with their respective values all approaching 1.0. Besides, DiffGui exhibits the lowest 2D similarity and the highest protein-ligand interaction similarity when compared to other approaches. This indicates that DiffGui is capable of generating more novel molecular scaffolds while maintaining interactions with key binding site residues. In addition, the generated molecules serve as the basis for computing an array of crucial molecular properties, including Vina Score, QED, SA, LogP, and TPSA. The mean values of these properties are summarized in Table 3 and Supplementary Table 10. In terms of performance on the PDBbind dataset (Table 3), DiffGui outperforms other models on nearly all metrics except for SA, suggesting that it is capable of generating more tightly binding drug-like molecules. The AutoDock Vina program⁵⁴ is utilized to estimate the binding affinity, and three types of scores (Vina Score, Vina Min, and Vina Dock) are reported. The Vina Score is computed directly on the generated 3D conformations, while the Vina Min and Vina Dock are calculated after local minimization and redocking of the generated molecules, respectively. DiffGui reveals the lowest values (−6.700, −7.655, and −8.448) in these three Vina scores, demonstrating its superiority to create potential binders with higher affinity for given pockets. Moreover, DiffGui surpasses other methods on the QED metric, which is indicative of its capability to generate more drug-like molecules. The LogP values of all methods range from 1.384 to 2.855, fitting within the universally recognized LogP range (1–3) of drug-like molecules. Although ResGen and PocketFlow show the highest SA scores of 0.784, they possess low TPSA values of 60.11 and 38.46, respectively. This validates the trend of autoregressive models to generate small fragments that may not fully occupy the entire pocket and thus compromise their specificity towards protein targets. Among the diffusion models that tend to generate complete molecules, DiffGui distinguishes itself with the highest SA score of 0.678. In brief, DiffGui excels over other baselines when assessed through the aforementioned molecular properties. This superior performance can be

ascribed to the guidance of chemical bonds and property labels in our approach, which directs the reverse diffusion process toward the generation of molecules with desired characteristics.

In the evaluation of the CrossDocked dataset (Supplementary Table 10), even though the molecules generated by DiffGui do not attain the highest level of docking score, they showcase competitive results (highlighted in gray) against the best method. Besides, it is noteworthy that the Vina scores achieved by DiffGui are lower than those of the reference ligands for this dataset (especially after minimization and redocking), a phenomenon that is not observed in the PDBbind case. The reason could be that the PDBbind dataset, derived from experimentally determined protein-ligand complex data, involves more challenging ligands. In contrast, the ligands in the CrossDocked dataset are not native binders and may form unrealistic interactions within the binding sites. This hypothesis is further confirmed by the lower Vina scores of reference molecules in the PDBbind dataset. The SA and TPSA scores continue to expose the inherent drawback of autoregressive models (ResGen and PocketFlow), which tend to produce small fragments by sampling the local optimum atom instead of considering the global information of the ligand.

Ablation analyses

To investigate the impact of individual components on model performance, we conducted ablation experiments on the PDBbind dataset and obtained three variants of the full DiffGui model: (1) DiffGui-nobond, a model trained without the bond diffusion process; (2) DiffGui-nolab, a model trained without property label guidance; (3) DiffGui-noboth, a model trained without both above modules. The generation ability of different models and the quality of generated molecules are displayed in Supplementary Tables 11 and 12, respectively. It appears that the two techniques are devoid of any significant detrimental effect on the basic metrics, including validity, connectivity, novelty, uniqueness, and diversity (Supplementary Table 11). Notably, the validity (0.9427) of DiffGui-nolab is higher than those of other models, which is reasonable because apart from the benefits of property guidance, it also inferences with the normal generation process to a certain extent. As shown in Supplementary Table 12, the removal of bond diffusion or property guidance leads to a deterioration in model performance across the JS divergences (detailed information in Supplementary Tables 13–15), Vina scores, and QED metric. Furthermore, the simultaneous exclusion of both modules results in an even more pronounced decline in performance, thus validating their synergistic effects. The values of SA, LogP, and TPSA are all situated within reasonable ranges. Overall, the ablation study demonstrates that the components of bond diffusion and property guidance can contribute to the generation of more realistic molecules with enhanced 3D structural rationality and desired molecular attributes.

De novo drug design on protein targets

Given that GCDM underperformed all comparable methods on three Vina Scores (Table 3 and Supplementary Table 10), we excluded it from the following de novo drug design experiments. We selected 1w51, 3ctj (PDBid) from the PDBbind test set and 7ew4, 8ju6 (PDBid) outside the PDBbind dataset as protein targets. The binding sites of experimentally active compounds are utilized as pockets to enable protein-conditioned molecule generation by various methods. The protein targets of 1w51, 3ctj, 7ew4, and 8ju6 correspond to beta-secretase 1, tyrosine kinase, G protein-coupled receptor, and ion channel, respectively, covering diverse types of proteins. We generated a set of 100 molecules for each target/model and visualized the ligands with the best docking scores among those PB-valid in Figs. 5 and 6. Essentially, the molecules produced by DiffGui possess better-defined chemical structures, higher docking scores and more favorable molecular properties when benchmarked against active molecules and those produced by alternative methods. They closely resemble the binding

poses of positive ligands and fully occupy the designated binding pockets. Furthermore, the binding free energies (ΔG) of these molecules are calculated by the MMGBSA method⁵⁵. As shown in Supplementary Table 16, while the binding free energy does not perfectly correlate with the docking score, the overall trend persists, and the ligands generated by DiffGui exhibit the lowest ΔG values among all produced ligands. Remarkably, with the exception of 3ctj, our model generates molecules with lower ΔG values than the reference ligands, further substantiating its capability to identify high-affinity ligands for a variety of protein targets. The 3D pharmacophore overlap between the generated molecules and reference ligands is computed by Schrödinger's Maestro program. As displayed in Supplementary Fig. 4, the molecules generated by DiffGui possess the highest number of 3D pharmacophore overlaps for 1w51 and 3ctj. However, for transmembrane proteins (7ew4 and 8ju6) in Supplementary Fig. 5, DiffGui does not demonstrate a notable advantage over other methods. This discrepancy arises because, unlike 1w51 and 3ctj, the reference ligands in 7ew4 and 8ju6 disclose limited interactions with the binding pocket residues. Consequently, our method prioritizes the creation of novel pharmacophores to establish new interactions with crucial residues, rather than replicating the sparse binding patterns of the reference ligands.

In contrast, ResGen prefers to create small molecular building blocks that are confined in sub-pockets and may cause off-target effects. Despite displaying high SA scores, their Vina scores are relatively poor, with even a positive value (2.003) in the case of 1w51 (beta-secretase 1, Fig. 5a). PocketFlow favors the generation of linear molecules with alternating single and double bonds. These molecules are fairly flexible and may possess low drug-likeness. In 3ctj (tyrosine kinase, Fig. 5b), the alkene group even protrudes outside the pocket, hindering the protein-ligand interactions and ultimately leading to the reduced docking score. Molecules generated by TargetDiff typically have superior docking scores in comparison to the molecules produced by other methods. However, there exist several limitations in their chemical structures that impair the structural rationality and synthetic accessibility. First, seven-membered rings or even larger ones, that are uncommon in drug-like molecules, frequently occur in the generated structures. Second, the fused rings are not aromatic, and these non-aromatic structures have low chemical stability and high synthesis difficulty. Last but not the least, the molecules in 7ew4 (GPCR, Fig. 6a) and 8ju6 (ion channel, Fig. 6b) incorporate six-membered rings accompanied with only two double bonds. Nonetheless, these rings are expected to be aromatic when judging from their planar conformations and overall molecular structures. Hence, we conclude that the post-operation of adding chemical bonds via the OpenBabel toolkit⁴⁰ cannot guarantee the accuracy of bond types. Moreover, the 3D conformation of the molecule in 1w51 undergoes distortion to accommodate the pocket shape, probably giving rise to its high strain energy. The other two diffusion-based methods, DiffSBDD and PMDM, generate molecules that bind more loosely than our method. Their chemical structures are also unreasonable. For instance, in the 1w51 case, the molecule of DiffSBDD has three-membered and seven-membered rings, while the molecule of PMDM forms a macrocyclic ring with two consecutive peptide bonds. In the cases of 7ew4 and 8ju6, DiffSBDD yields macrocyclic compounds with multiple hydroxy groups, which contribute to their low LogP values and reflect their high hydrophilicity. The molecules of PMDM in the pockets of 3ctj and 7ew4 both feature three seven-membered rings, whereas the ligands entirely lack aromatic groups. This structural characteristic is consistent with the high proportion of seven-membered rings (21.2% in Supplementary Table 7 and 17.6% in Supplementary Table 8) observed in PMDM-generated molecules.

In addition, to examine the diversity of molecules generated by DiffGui, we visualized the top-ranked ligands for protein targets of 4b5d, 5ni7, and 5ywy (PDBid) in Supplementary Fig. 6. These targets

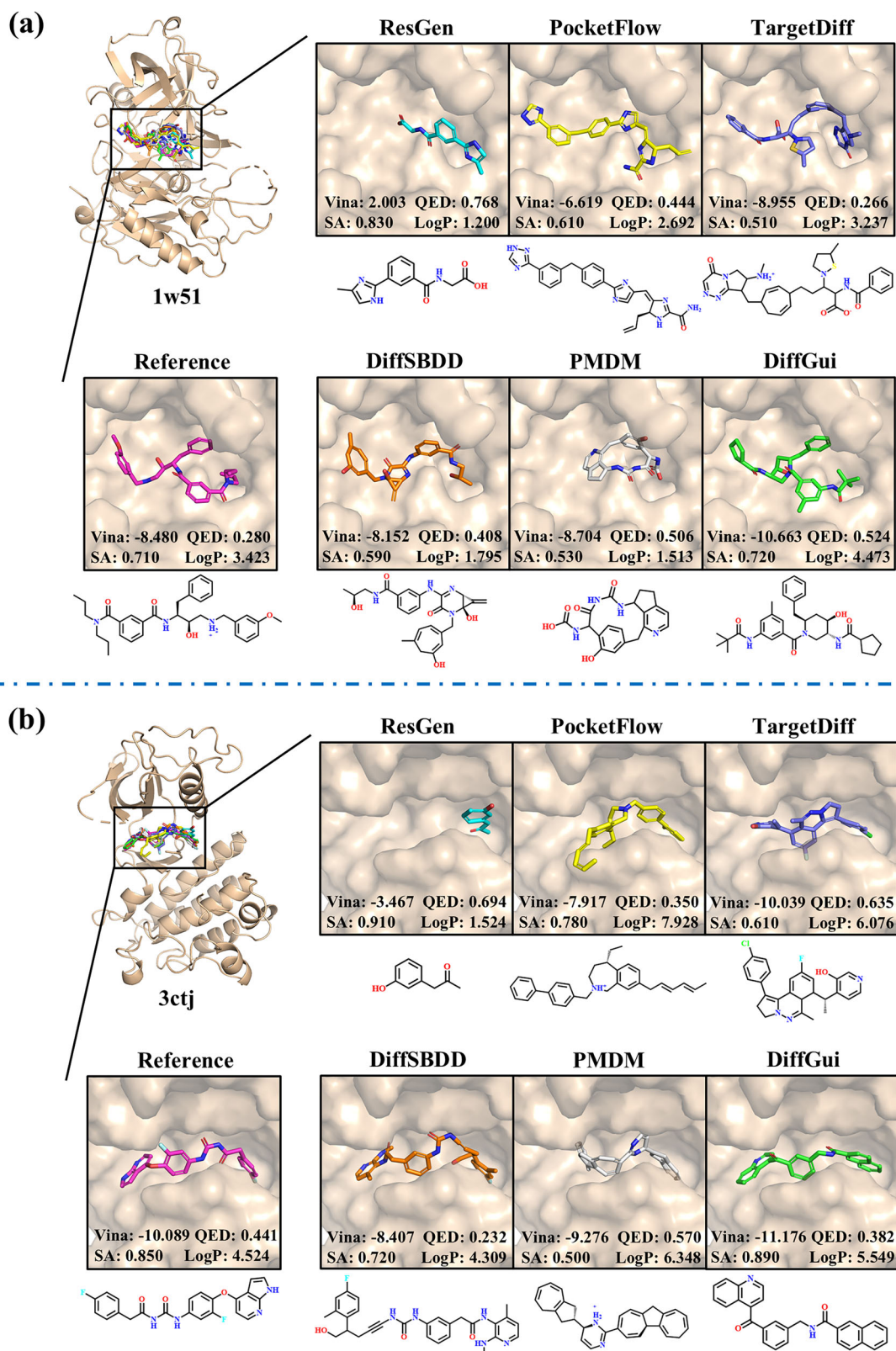


Fig. 5 | Molecules generated for targets from the PDBbind test set. The protein targets of 1w51 (a) and 3ctj (b) are beta-secretase 1 and tyrosine kinase, respectively. The visualized molecules have the best docking scores among those PB-valid.

are Capitella teleta AChBP, nuclear receptor ROR-gamma, and prostaglandin E2 receptor, respectively. All generated molecules fit perfectly with the 3D geometry of the pockets, whether they are shallow or deep. The generated molecules are diverse and exhibit better docking scores and properties than the reference ligands. Besides, the

diversity of molecules produced for the PDBbind test set is computed to be 0.7256 (Supplementary Table 11), thus proving that our method can provide various promising candidates that can be employed for further drug development. Moreover, we conducted wet-lab experimental validation on molecules generated for RSK4 (ribosomal S6

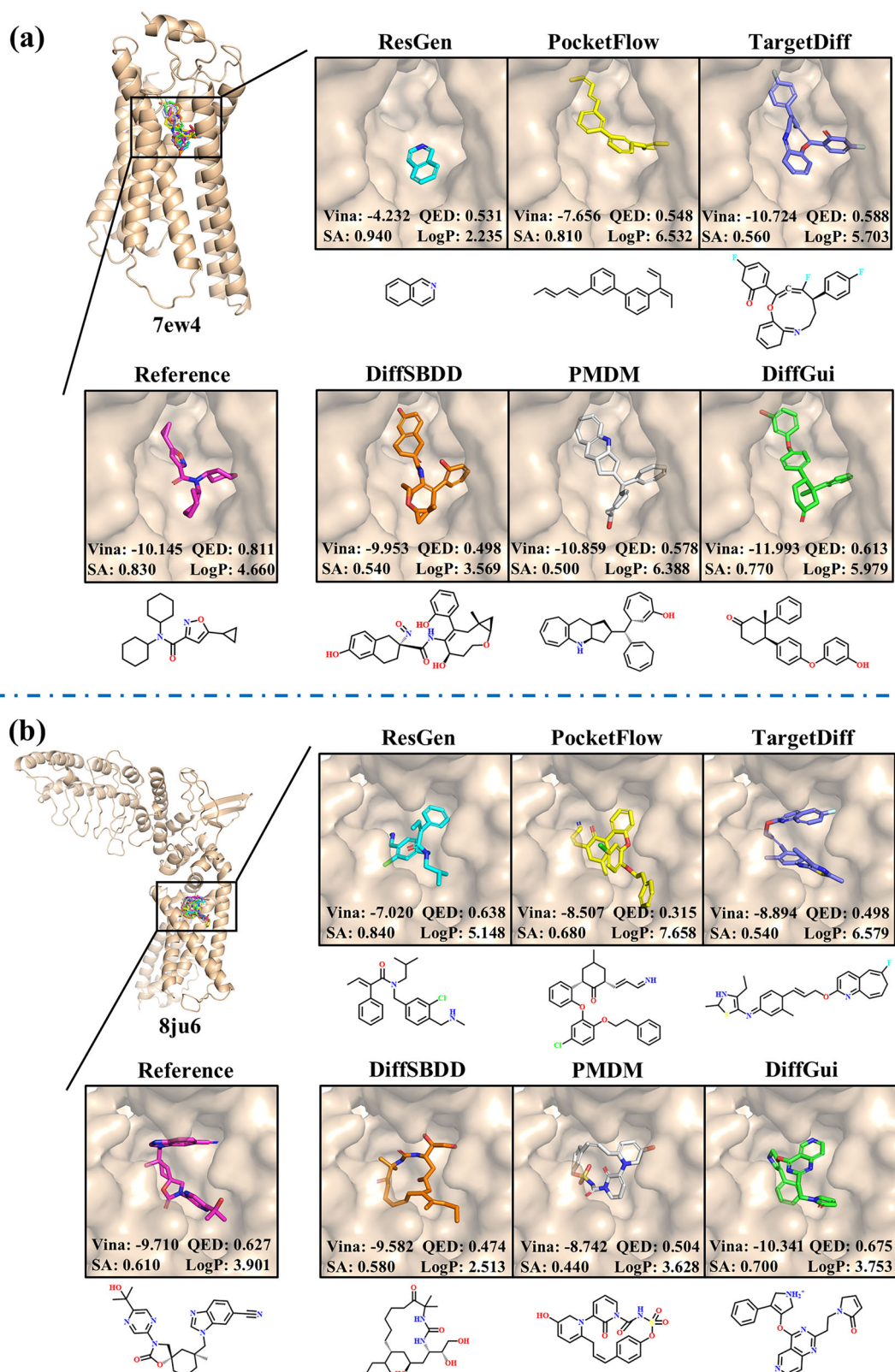


Fig. 6 | Molecules generated for targets outside the PDBbind dataset. The protein targets of 7ew4 (a) and 8ju6 (b) are G protein-coupled receptor and ion channel, respectively. The visualized molecules have the best docking scores among those PB-valid.

kinase 4, PDBid 6g77), a protein structure not included in the PDBbind dataset. Only two simple molecules are selected because of their rapid and straightforward synthesis. As illustrated in Supplementary Fig. 7a, b, despite their structural differences, both Compound 1 and Compound 2 demonstrate potent inhibitory activity in the HTRF assay, with

IC_{50} values of approximately 215.0 nM and 111.1 nM, respectively, highlighting their potential as lead compounds for further development. The binding modes reveal that both compounds interact with key residues (K105, D153, L155, and K221) in the binding pocket of RSK4.

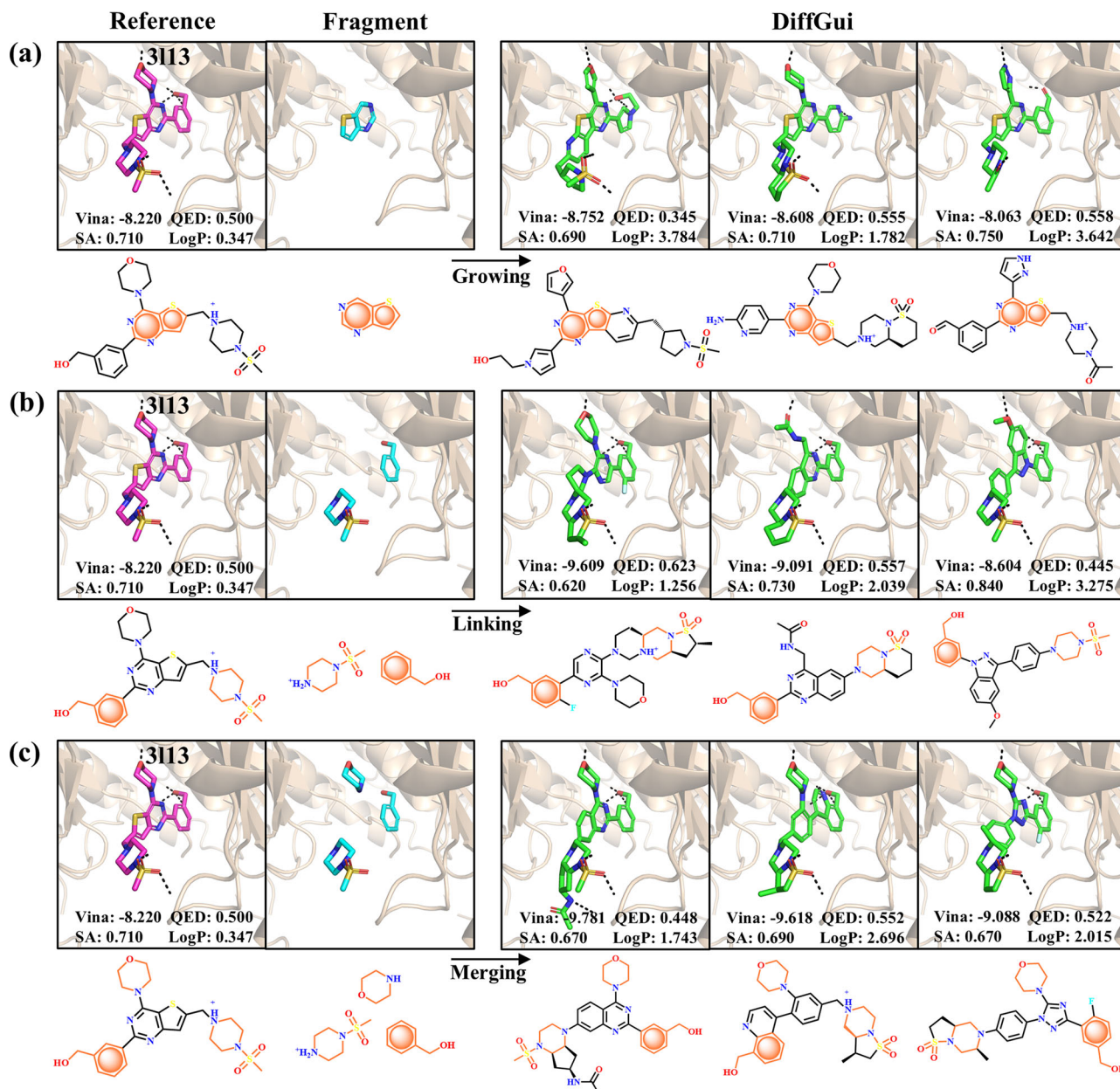


Fig. 7 | Lead optimization by fragment denoising method on PDBid 3l13. a Fragment growing, **(b)** Fragment linking - two fragments, **(c)** Fragment merging - three fragments. The seed fragments are highlighted in orange.

Lead optimization based on fragments

In drug discovery, lead optimization is a critical task to refine the existing lead compounds for improved affinities and drug-like properties. Based on the sub-structures or fragments of known drug candidates, fragment growing and scaffold hopping are two effective strategies to perform lead optimization. Fragment growing expands the small fragments into the complete molecules by adding functional groups or larger sub-structures. Scaffold hopping replaces the core structure of the lead compound by an alternative core with the purpose of enhancing its biological activity and potency. We enable our model to implement the above task by adopting two sampling methodologies - fragment denoising and fragment conditioning. Fragment denoising manually diffuses the fixed fragment at every step, subsequently denoising it along with the remaining part from the previous step. For the next iteration, the denoised fixed fragment is discarded, while the other part is retained. Fragment conditioning inputs the fixed fragment at every step as an additional condition, and the complete

molecule is obtained by denoising the fixed fragment and the denoised remaining part at the last step. For more details of these two sampling techniques, please refer to the Methods section.

As illustrated in Fig. 7, we applied the fragment denoising method on PDBid 3l13 to develop potential inhibitors based upon the structure of the active ligand. The protein target of 3l13 is phosphoinositide-3-kinase (PI3K), an enzyme involved in numerous cellular functions, such as cell growth, proliferation, differentiation, motility, survival, and intracellular trafficking⁵⁶. It plays a crucial role in the PI3K/AKT/mTOR signaling pathway, and dysregulation of PI3K signaling is often associated with various diseases, including cancer, diabetes, and autoimmune disorders. Thus, the inhibitors of PI3K can be used to treat certain cancers and inflammatory conditions. In Fig. 7a, thienopyrimidine (highlighted in orange) is fixed as a seed fragment to generate molecules via fragment growing. The generated ligands successfully replicate most of the interactions found in the native protein-ligand complex, specifically with Lys802, Ala805, Asp841, Tyr867, Val882, and

Asp964. However, the interacting functional groups in these ligands are distinct from the original ones. For instance, the morpholine group is substituted by furan and pyrazole. The piperazine sulfonyl group is replaced by pyrrolidine sulfonyl, cyclized piperazine sulfonyl, and piperazine carbonyl groups. This illustrates that DiffGui can not only learn the interaction patterns in protein-ligand complexes, but also assimilate the structural information of numerous chemical groups. In Fig. 7b, c, two and three fragments are provided, respectively, to conduct scaffold hopping (fragment linking or merging), where the core scaffold of the active ligand is transformed into alternative scaffolds. The Vina, QED, and SA scores of the generated molecules are either better than or at least competitive with those of the reference compound. Therefore, taking into account the overall performance, our method effectively achieves the goal of lead optimization, whether through fragment growing or scaffold hopping strategies.

The fragment conditioning method is applied to PDBid 6e23, with the results visualized in Supplementary Fig. 8. The protein target of 6e23 is WD repeat-containing protein 5 (WDR5), a crucial protein involved in chromatin remodeling and gene expression regulation. It belongs to the WD-repeat protein family and plays essential roles in various cellular processes, including embryonic development, stem cell pluripotency, and cancer progression. Its dysregulation is associated with several diseases, making WDR5 an important target for therapeutic interventions and research in epigenetics and cancer biology⁵⁷. As depicted in Supplementary Fig. 8, the molecules generated by DiffGui through fragment growing, linking, and merging exhibit either superior or comparable QED and SA scores compared to those of the reference ligand. However, the estimated binding affinities of these molecules are not higher than the reference, thereby elucidating that the fragment conditioning method may not be suitable for direct use without retraining on the specialized dataset. It forcibly injects the condition of a fixed fragment at every denoising step, whereas the DiffGui model is not trained on the combined data of the fixed fragment and the denoised remaining part. This would lead to an inconsistency problem between the training and the sampling processes.

We further verified the effectiveness of lead optimization of DiffGui through wet-lab experiments on a non-kinase target, dihydroorotate dehydrogenase (DHODH). From the structures of 4zmg and 4ls1 (PDBid), the lead optimization is conducted on the basis of the fixed fragments (highlighted in orange in Supplementary Fig. 7c, d). As a result, the optimized molecules (Compounds 3 and 4) exhibit enhanced potency against DHODH, with IC_{50} values decreasing from 8.02 μ M to 4.27 μ M and from 32.20 nM to 10.45 nM, respectively. Compound 3 features a thiazole ring modified with both carboxylic acid and methyl substituents. The carboxylic acid moiety engages in a salt bridge interaction with R136, while the methyl group inserts into a hydrophobic sub-pocket formed by residues V134, V143, and Y356. Compared to the original ligand, Compound 4 involves the conversion of a carboxylic acid to a hydroxamic acid, which extends the hydrogen-bonding network to incorporate additional interactions with Q47 and T360. A fluorine atom is evolved on the benzene ring to occupy the hydrophobic sub-pocket identified previously. And, a methyl group is introduced on the linker moiety to fill a distinct hydrophobic cavity surrounded by L46, A55, and L58.

Molecular generation for mutated targets

As DiffGui is a pocket-aware molecule generation model, we conducted experiments on wild-type and mutated targets to investigate its sensitivity to subtle variations within the pocket structure. We chose KRAS^{G12D} (PDBid 7rpz) as an example. KRAS^{G12D} is a specific mutation of the KRAS gene, where the glycine (G) at position 12 is replaced by aspartic acid (D). The G12D mutation leads to the constitutive activation of the KRAS protein, which drives the uncontrolled cell growth and division, contributing to cancer development and progression.

Due to its significant role in oncogenesis, KRAS^{G12D} is a critical target for cancer research and therapeutic development⁵⁸. The distinct binding patterns of the native ligand (MRTX1133) and the generated molecules for KRAS^{G12D} and its mutants are shown in Supplementary Fig. 9. MRTX1133 optimally fills the switch II pocket and extends three substituents to form noncovalent interactions with seven key residues Asp12, Glu62, Tyr64, Arg68, As69, His95, and Tyr96, resulting in a K_D of 0.2 pM⁵⁸. This exceptionally high binding affinity is also evidenced by its remarkably low Vina score (−12.877). The protein mutants include both single-point and multi-point mutations. The single-point mutations transfer each of the key residues to alanine. In multi-point mutations, Asp12Glu62Ala converts both Asp12 and Glu62 to alanine. ‘Interact’ refers to the protein in which all interacting residues are mutated to alanine, while ‘Pocket-mu’ denotes the protein where all pocket residues (residues within 10 Å of the reference ligand) are mutated to alanine.

For the wild-type protein, the generated molecule maintains the hydrogen bond with Arg68 and the salt bridges with Asp12/Glu62. As a comparison, in the Asp12Ala and Glu62Ala mutants, the relevant salt bridge disappears, and the electronegative groups, such as carboxylic acid or methyl phenyl ether, replace the positive amine groups at the corresponding positions. Since Tyr64Ala mutates Tyr64 to alanine, the π - π interaction between the resulting molecule and the residue at position 64 is absent. A phenyl group is developed to occupy the space of Arg68 in Arg68Ala, while the hydroxy group that forms a hydrogen bond with Asp69 is missing in Asp69Ala. In the mutants of His95Ala and Tyr96Ala, the aromatic system of generated molecules is extended to occupy the position of the original residue, forming stronger π - π interaction with Tyr96 and His95, respectively. However, these interactions are weaker in the wild-type protein due to the close distance between these two residues. Additionally, the salt bridges with Asp12 are reserved in the Asp69Ala and His95Ala mutation systems. Therefore, the single-point mutations of key residues significantly affect the chemical structures of molecules generated by DiffGui. Although the generated ligands for the mutants exhibit relatively higher docking scores, their QED, SA, and LogP values demonstrate notable improvements over the reference ligand.

The multi-point mutations exert more profound influences on the generated molecules as they further alter the protein pocket environment. In the Asp12Glu62Ala mutant, the salt bridges with residues at 12 and 62 positions both vanish, whereas the hydrogen bonds with Tyr64, Arg68, and His95 are conserved. The ‘Interact’ mutation system with seven key residues mutated facilitates the production of a molecule with several hydrophobic groups, like isopropyl and isopentyl. This variation not only extinguishes most electrostatic interactions in the complex but also increases the LogP value of the molecule to 5.665, indicating lower binding affinity and higher hydrophobicity of the ligand. The pocket mutation system (‘Pocket-mu’) further decreases the estimated binding affinity (Vina score −5.730) of the resulting molecule, leading to a loss of specificity in binding with the pocket of KRAS^{G12D}. In conclusion, the mutation experiments fully testify to the sensitivity of the DiffGui model to the delicate changes within the pocket environment. Although the generated molecules do not bind as tightly as the reference ligand because of its extremely high affinity and the residue mutations, they exhibit higher drug-likeness and synthetic accessibility, which may be attributed to the property label guidance utilized during the generation process.

Discussion

Generating 3D potent molecules inside the protein pockets is of great significance, yet it remains a challenging task. In this study, we propose a novel guided diffusion model, DiffGui, to generate ligand molecules for any given protein target. By integrating bond diffusion and property guidance into the diffusion process, DiffGui enables the simultaneous generation of atoms and bonds in molecules, which exhibit high

structural rationality and desirable molecular properties. To incorporate bond diffusion with atom diffusion, we apply distinct noise schedules to atoms and bonds, thus effectively capturing the dependencies between interatomic distances and bond types. Moreover, property labels of molecules are injected into the atomic features, transforming the training process into a blend of conditional and unconditional modeling frameworks, and guiding the inference procedure to yield molecules with anticipated attributes.

Experimental results validate that DiffGui attains SOTA performance on the PDBbind dataset and competitive results on the Cross-Docked dataset. It greatly improves the quality of generated molecules, which more closely resemble the reference molecules in terms of the distributions of bond length, bond angle, dihedral angle, and ring percentage. Besides, we can conclude that DiffGui excels at generating novel and diverse molecules that bind more tightly to disease-relevant targets while preserving preferable drug-likeness. Through the adoption of specialized sampling algorithms, DiffGui is capable of performing lead optimization via fragment growing and scaffold hopping strategies, highlighting the versatility and applicability of our method for downstream drug design tasks. Furthermore, the effectiveness of DiffGui has also been validated by wet-lab experiments. Case studies on wild-type and mutated KRAS proteins indicate that DiffGui can not only reproduce the favorable interaction patterns presented in the reference complex, but also detect subtle variations in the protein environment. In summary, DiffGui deeply comprehends the geometric constraints and molecular interactions in the protein-ligand complexes and possesses enhanced generalization capability for new targets.

In future work, we aim to develop target-aware molecular generation techniques based on fragments^{59,60}, which represent more reliable and synthesizable molecular sub-structures. In addition, we will delve into more sophisticated noise schedules and guidance strategies to further improve the performance of deep generative models. The intricate dynamics of the protein-ligand complex, along with other key molecular properties such as pharmacokinetic/pharmacodynamic profile, toxicity, and metabolism, will also be thoroughly considered. Overall, our objective is to greatly boost the success rate and efficiency of drug discovery and development with the assistance of AI technologies.

Methods

Task definition

Let \mathbf{p} denote the protein pocket and \mathbf{x} denote the 3D ligand. A ligand molecule with N atoms can be represented as $\mathbf{x} = \{a_i, r_i, b_{ij}\}_{i,j=1}^N$, where $a_i \in \{0, 1\}^{N_a}$ is the atom types, $r_i \in \mathbb{R}^3$ is the atom coordinates, and $b_{ij} \in \{0, 1\}^{N_b}$ is the chemical bonds. We select ten atom types, including nine real atom types (C, N, O, S, P, F, Cl, Br, and I) and one dummy absorbing type. In addition, we identify five chemical bond types, consisting of four real bond types (single, double, triple, and aromatic bonds) and one dummy absorbing type, which also indicates no bond⁴⁶. In this paper, we denote the molecular properties as $\mathbf{c} \in \mathbb{R}^{N_c}$ and focus on five specific properties, namely binding affinity, QED, SA, LogP, and TPSA. Let superscript t represent the latent variables at timestep t ($t = 0, 1, \dots, T$) and $\mathbf{x}^0 = \mathbf{x}$. In a word, the task of conditional molecular generation here is to produce a series of \mathbf{x} given \mathbf{p} and \mathbf{c} .

Overview of DiffGui model

Unlike the pure diffusion model³⁶, DiffGui is a conditional guided diffusion model, where the protein pocket and desired properties guide the molecule generation process. Thus, we aim to model the $p_\theta(\mathbf{x}|\mathbf{p}, \mathbf{c})$ to determine the distribution of ligands that can bind to any given protein pocket while possessing the desired properties. Formally, DiffGui is a latent variable model represented as

$p_\theta(\mathbf{x}^0|\mathbf{p}, \mathbf{c}) = \int p_\theta(\mathbf{x}^{0:T}) d\mathbf{x}^{1:T}$, where \mathbf{x}^t for $t = 1, \dots, T$ is a sequence of latent variables with the same dimensionality as the data $\mathbf{x}^0 \sim p(\mathbf{x}^0|\mathbf{p}, \mathbf{c})$. As shown in Fig. 1, the proposed DiffGui framework consists of a forward diffusion process and a reverse generative process, both defined as Markov chains. The forward process (Eq. 1) progressively perturbs the data into a stationary distribution, while the reverse process (Eq. 2) gradually denoises the samples back towards the data distribution with a network parameterized by θ :

$$q(\mathbf{x}^{1:T}|\mathbf{x}^0, \mathbf{p}, \mathbf{c}) = \prod_{t=1}^T q(\mathbf{x}^t|\mathbf{x}^{t-1}, \mathbf{p}, \mathbf{c}) \quad (1)$$

$$p_\theta(\mathbf{x}^{0:T-1}|\mathbf{x}^T, \mathbf{p}, \mathbf{c}) = \prod_{t=1}^T p_\theta(\mathbf{x}^{t-1}|\mathbf{x}^t, \mathbf{p}, \mathbf{c}) \quad (2)$$

Since our goal is to produce 3D molecules inside the protein pocket, the model needs to generate continuous atom coordinates, discrete atom and bond types, while preserving SE(3)-equivariance throughout the entire generative process. In the following sections, we will elaborate on how we construct the diffusion process, parameterize the generative process, and implement the classifier-free guidance of molecular properties.

Molecular diffusion process

Building on recent progress in learning continuous atom coordinates and discrete atom or bond types with diffusion models^{33,46}, we employ a Gaussian distribution \mathcal{N} to model continuous atom coordinates and a categorical distribution \mathcal{C} to model discrete atom or bond types. The forward diffusion process is formulated as follows:

$$q(r_i^t|r_i^{t-1}, \mathbf{p}, \mathbf{c}) = \mathcal{N}\left(r_i^t \middle| \sqrt{1 - \beta^t} r_i^{t-1}, \beta^t \mathbf{I}\right) \quad (3)$$

$$q(a_i^t|a_i^{t-1}, \mathbf{p}, \mathbf{c}) = \mathcal{C}(a_i^t | (1 - \beta^t) a_i^{t-1} + \beta^t \mathbb{I}_k) \quad (4)$$

$$q(b_{ij}^t|b_{ij}^{t-1}, \mathbf{p}, \mathbf{c}) = \mathcal{C}(b_{ij}^t | (1 - \beta^t) b_{ij}^{t-1} + \beta^t \mathbb{I}_{k'}) \quad (5)$$

where $\beta^t \in [0, 1]$ is the pre-defined noise scaling schedule, $\mathbf{I} \in \mathbb{R}^{3 \times 3}$ is the identity matrix, and \mathbb{I}_k represents a one-hot vector with a one at the k -th position and zeros elsewhere. For the atom coordinates, we gradually add scaled standard Gaussian noise. For the atom or bond types, we increase the probability mass on the k -th or k' -th type, ensuring that these types are gradually perturbed toward the desired types during the forward process. We refer to it as the absorbing type because it functions by gradually assimilating all atom or bond types into this specific category⁴⁶.

Denoting $\alpha^t = 1 - \beta^t$ and $\bar{\alpha}^t = \prod_{s=1}^t \alpha^s$, a desirable feature of the diffusion process is the ability to calculate the noisy data distribution $q(\mathbf{x}^t|\mathbf{x}^0, \mathbf{p}, \mathbf{c})$ of timestep t in closed-form:

$$q(r_i^t|\mathbf{x}^0, \mathbf{p}, \mathbf{c}) = \mathcal{N}\left(r_i^t \middle| \sqrt{\bar{\alpha}^t} r_i^0, (1 - \bar{\alpha}^t) \mathbf{I}\right) \quad (6)$$

$$q(a_i^t|\mathbf{x}^0, \mathbf{p}, \mathbf{c}) = \mathcal{C}\left(a_i^t \middle| \bar{\alpha}^t a_i^0 + (1 - \bar{\alpha}^t) \mathbb{I}_k\right) \quad (7)$$

$$q(b_{ij}^t|\mathbf{x}^0, \mathbf{p}, \mathbf{c}) = \mathcal{C}\left(b_{ij}^t \middle| \bar{\alpha}^t b_{ij}^0 + (1 - \bar{\alpha}^t) \mathbb{I}_{k'}\right) \quad (8)$$

As $t \rightarrow T$, we get $q(r_i^t|\mathbf{x}^0, \mathbf{p}, \mathbf{c}) \rightarrow \mathcal{N}(0, \mathbf{I})$, $q(a_i^t|\mathbf{x}^0, \mathbf{p}, \mathbf{c}) \rightarrow \mathbb{I}_k$, and $q(b_{ij}^t|\mathbf{x}^0, \mathbf{p}, \mathbf{c}) \rightarrow \mathbb{I}_{k'}$ according to Eqs. 6–8. This suggests that the atom coordinates approximately approach the standard Gaussian distribution for large T , while the atom and bond types place all probability mass on the absorbing types when $t = T$. These

distributions, known as prior distributions, will serve as the initial distributions for the reverse process.

Since bond types in molecules are closely related to atom distances and atom types, applying the same noise schedule to bond types as to atom types and positions may lead to inconsistencies in the noised data distribution. Hence, we assign different β^t values for atoms (types and positions) and bond types, ensuring that the information level $\bar{\alpha}^t$ of bond types decays to zero much faster than that of atoms during the diffusion process. In the first stage, the atoms are only marginally perturbed, and the model pays more attention to disrupt the bond types. In the second stage, almost all real bonds have been removed, and the model concentrates solely on the perturbation of atoms. This approach allows the model to avoid learning bond types when atom distances have obviously deviated from the canonical bond lengths.

Parameterization of molecular generative process

The generative process, conversely, aims to reconstruct the original molecule \mathbf{x}^0 from the initial noise \mathbf{x}^T . To achieve this, we approximate the reverse distribution using a neural network parameterized by θ :

$$p_{\theta}(r_i^{t-1}|\mathbf{x}^t, \mathbf{p}, \mathbf{c}) = \mathcal{N}(r_i^{t-1}|\mu_{\theta}(\mathbf{x}^t, t, \mathbf{p}, \mathbf{c}), \beta^t \mathbf{I}) \quad (9)$$

$$p_{\theta}(a_i^{t-1}|\mathbf{x}^t, \mathbf{p}, \mathbf{c}) = \mathcal{C}(a_i^{t-1}|a_{\theta}(\mathbf{x}^t, t, \mathbf{p}, \mathbf{c})) \quad (10)$$

$$p_{\theta}(b_{ij}^{t-1}|\mathbf{x}^t, \mathbf{p}, \mathbf{c}) = \mathcal{C}(b_{ij}^{t-1}|b_{\theta}(\mathbf{x}^t, t, \mathbf{p}, \mathbf{c})) \quad (11)$$

where μ_{θ} , a_{θ} and b_{θ} are all neural networks. An essential characteristic that a neural network should possess for modeling 3D molecules is E(3)-equivariance, i.e., the network's outputs should be equivariant under any 3D transformation, such as rotation, translation, and reflection. There exist different ways to parameterize μ_{θ} , a_{θ} and b_{θ} , and in this case, we choose to predict \mathbf{x}^t by the above neural networks. Drawing inspiration from MolDiff⁴⁶ that utilized an E(3)-equivariant network to update atom and bond representations through message passing algorithms, we propose modeling the intricate interactions between ligand and protein atoms using an SE(3)-equivariant GNN:

$$\left\{ \hat{r}_i^{t-1}, \hat{a}_i^{t-1}, \hat{b}_{ij}^{t-1} \right\}_{i,j} = \phi_{\theta}(\mathbf{x}^t, t, \mathbf{p}, \mathbf{c}) = \phi_{\theta}\left(\left\{ r_i^t, a_i^t, b_{ij}^t \right\}_{i,j}, t, \mathbf{p}, \mathbf{c}\right) \quad (12)$$

Formally, given an input protein-ligand complex $\{\mathbf{p}, \mathbf{x}\} = \{a_i, r_i, b_{ij}\}_{i,j=1}^N$ (we overload the notation N to denote the number of atoms in the protein-ligand complex and omit timestep t for simplicity), we construct a complete graph in which vertices represent the atoms and all vertices are connected. Let $v_i \in \mathbb{R}^d$ and $e_{ij} \in \mathbb{R}^d$ denote the hidden representations for vertex i and edge $\langle i, j \rangle$, respectively. The input vertex features comprise one-hot encodings of atom types, while the input edge features are one-hot encodings of bond types. The updates for the vectors v_i , e_{ij} , and the coordinates r_i are then defined as follows:

$$\tilde{e}_{ij} \leftarrow \phi_d(e_{ij}, \|r_i - r_j\|_2) \quad (13)$$

$$v_i \leftarrow \text{Linear}(v_i) + \sum_j \phi_v(v_j, \tilde{e}_{ij}, t) \quad (14)$$

$$e_{ij} \leftarrow \sum_k \phi_e(v_k, \tilde{e}_{ki}, t) + \sum_k \phi_e(v_k, \tilde{e}_{jk}, t) + \text{Linear}(v_i) + \text{Linear}(v_j) + \text{Linear}(\tilde{e}_{ij}) \quad (15)$$

$$r_i \leftarrow r_i + \sum_j \phi_r(v_i, v_j, \tilde{e}_{ij}, t) \frac{r_i - r_j}{\|r_i - r_j\|_2^2} \cdot \mathbf{M}_{\text{ligand}} \quad (16)$$

where $\text{Linear}(\cdot)$ represents linear transformations of the inputs, and $\phi_d, \phi_v, \phi_e, \phi_r$ are neural networks composed of different multilayer perceptrons (MLPs). $\mathbf{M}_{\text{ligand}}$ is the ligand mask, which ensures that the coordinates of protein atoms are not updated. The final atom features v_i and bond features e_{ij} are fed into a multi-layer perceptron and a softmax function to obtain \hat{a}_i and \hat{b}_{ij} , respectively. The main difference between our proposed model and MolDiff lies in our introduction of the protein pocket framework, which deviates from the unconditional molecular generation used in MolDiff. In the generative process, we keep protein-related information fixed to enable the pocket-conditioned molecular generation.

Classifier-free guidance of molecular properties

Guided sampling has emerged as a critical strategy in the development of molecular diffusion models that are capable of generating samples adhering to desired properties \mathbf{c} . In this work, we adopt classifier-free guidance⁴⁷ to explicitly incorporate conditional signals. Distinct from classifier guidance⁴⁸, which necessitates the inclusion of an additional classifier, classifier-free guidance streamlines the model architecture by directly integrating guided signals into the training phase, thereby offering enhanced control and flexibility.

Formally, DiffGui consists of an unconditional model $\phi_{\theta}(\mathbf{x}^t, t, \mathbf{p}, \emptyset)$ and a property-conditional model $\phi_{\theta}(\mathbf{x}^t, t, \mathbf{p}, \mathbf{c})$. On one hand, the unconditional model is trained on 3D structures of protein-ligand complexes without property labels. On the other hand, the property-conditional model has access to both molecular properties and the corresponding protein-ligand complex \mathbf{x}^t at each timestep t . During the reverse generative process, we utilize a hyperparameter γ to modulate the strength of classifier-free guidance from the conditional model, so that

$$\tilde{\mathbf{x}}^{t-1} = (1 + \gamma)\phi_{\theta}(\mathbf{x}^t, t, \mathbf{p}, \mathbf{c}) - \gamma\phi_{\theta}(\mathbf{x}^t, t, \mathbf{p}, \emptyset) \quad (17)$$

Training

In the training stage, we add noise to the data and train the neural network to recover \mathbf{x}^{t-1} from \mathbf{x}^t by optimizing the predicted distributions $p_{\theta}(\mathbf{x}^{t-1}|\mathbf{x}^t, \mathbf{p}, \mathbf{c})$ to approximate the true posterior $q(\mathbf{x}^{t-1}|\mathbf{x}^0, \mathbf{x}^t, \mathbf{p}, \mathbf{c})$, which can be derived from Eqs. 3–8. The loss functions are defined as follows:

$$L^{t-1} = L_{\text{pos}}^{t-1} + \lambda_1 L_{\text{atom}}^{t-1} + \lambda_2 L_{\text{bond}}^{t-1} \quad (18)$$

$$L_{\text{pos}}^{t-1} = \frac{1}{N} \sum_i \|r_i^{t-1} - \mu_{\theta}(\mathbf{x}^t, t, \mathbf{p}, \mathbf{c})_i\|_2^2 \quad (19)$$

$$L_{\text{atom}}^{t-1} = \frac{1}{N} \sum_i D_{\text{KL}}[q(a_i^{t-1}|\mathbf{x}^0, \mathbf{x}^t, \mathbf{p}, \mathbf{c}) \| p_{\theta}(a_i^{t-1}|\mathbf{x}^t, \mathbf{p}, \mathbf{c})] \quad (20)$$

$$L_{\text{bond}}^{t-1} = \frac{1}{N} \sum_i D_{\text{KL}}[q(b_{ij}^{t-1}|\mathbf{x}^0, \mathbf{x}^t, \mathbf{p}, \mathbf{c}) \| p_{\theta}(b_{ij}^{t-1}|\mathbf{x}^t, \mathbf{p}, \mathbf{c})] \quad (21)$$

where λ_1 and λ_2 are pre-defined constants. We randomly sample a timestep t and optimize the neural networks by minimizing the total loss L^{t-1} . Supplementary Algorithm 1 describes the training process with classifier-free guidance in detail.

Sampling

To generate new molecules, we first sample \mathbf{x}^T from the prior distributions $p(\mathbf{x}^T)$ and then iteratively sample from $p_\theta(\mathbf{x}^{t-1}|\mathbf{x}^t, \mathbf{p}, \mathbf{c})$ to gradually remove noise. The prior distribution $p(\mathbf{x}^T)$ is the standard Gaussian distribution $\mathcal{N}(0, \mathbf{I})$ for atom positions, along with the categorical distributions for atom and bond types, where all probability mass is assigned to the absorbing type. Supplementary Algorithm 2 describes the sampling process with classifier-free guidance in detail.

Two novel sampling methodologies (fragment denoising and fragment conditioning) are proposed to perform lead optimization based on known fragments. Fragment denoising (Eqs. 22 and 23) first diffuses the fixed fragment \mathbf{x}_f^0 at each step to obtain the hidden information \mathbf{x}_f^t , which is then combined with the rest of the ligand \mathbf{x}_r^t to accomplish one-step denoising. In the next denoising step, the denoised fragment part \mathbf{x}_f^{t-1} is discarded and the corresponding information is retained through a step of forward diffusion. The final molecule is produced by denoising from timestep 1 to 0. Fragment conditioning (Eq. 24) consistently inputs the fixed fragment \mathbf{x}_f^0 at each step as context, which also includes information of protein pocket and molecular properties. The final molecule is created by denoising the fixed fragment \mathbf{x}_f^0 and the denoised remaining part \mathbf{x}_r^t at the last step.

$$q(\mathbf{x}_f^t|\mathbf{x}_f^0, \mathbf{p}, \mathbf{c}) = \mathcal{N}(\mathbf{x}_f^t | \sqrt{\alpha^t} \mathbf{x}_f^0, (1 - \alpha^t) \mathbf{I}) \quad (22)$$

$$\mathbf{x}_f^{t-1}, \mathbf{x}_r^{t-1} = \phi_\theta(\mathbf{x}_f^t, \mathbf{x}_r^t, t, \mathbf{p}, \mathbf{c}) \quad (23)$$

$$\mathbf{x}_f^{t-1}, \mathbf{x}_r^{t-1} = \phi_\theta(\mathbf{x}_f^0, \mathbf{x}_r^t, t, \mathbf{p}, \mathbf{c}) \quad (24)$$

Datasets

We utilize the PDBbind dataset for training/testing and the CrossDocked dataset only for testing. The PDBbind dataset^{61,62} is a collection of experimentally determined three-dimensional structures of biomolecular complexes archived in Protein Data Bank (PDB), accompanied by binding affinity data (K_d , K_i or IC_{50}). The current 2020 version provides 23,496 biomolecular complexes, of which 19,443 are protein-ligand complexes. For this dataset, we employ 17.3 K complexes for training, 1.8 K complexes for validation and 0.1 K complexes for testing. The CrossDocked dataset⁶³ originally contains 22.5 million poses of ligands docked into multiple similar binding pockets across the PDB. Following the previous work²⁵, we refine the dataset by choosing binding poses with root mean square deviation (RMSD) less than 1.0 Å and split the refined data based on a threshold of less than 30% protein sequence identity. This results in 100,000 protein-ligand pairs for the training set and 100 pairs for the test set. We do not train DiffGui on the CrossDocked dataset due to the absence of affinity data; instead, we only evaluate it on the CrossDocked test set. To ensure a fair comparison, we retrain the baseline models on the PDBbind dataset when evaluating them on this dataset. And we compare DiffGui with the original baselines (trained on the CrossDocked dataset) to assess the generalization capability of our approach. We sample 100 molecules for each protein in the test set to perform evaluation.

Evaluation metrics

We adopt a wide range of metrics to evaluate the quality of generated molecules: (1) **Atom Stability** refers to the proportion of atoms that possess the correct valencies. (2) **Molecular Stability** refers to the proportion of molecules in which all constituent atoms are stable. (3) **PB-validity** is computed using the PoseBusters tool⁶⁴ to check whether the generated molecular conformations have reasonable geometries, including standard bond lengths, appropriate bond angles, and the absence of steric clashes. (4) **RDKit-validity** measures the proportion of generated molecules that pass the basic test of RDKit program. (5)

Novelty is the ratio of generated molecules that are not present in the training dataset. (6) **Uniqueness** represents the ratio of distinct molecules within all generated molecules. (7) **Similarity** is computed by comparing the Morgan-2 fingerprints of generated molecules and reference ligands. (8) **Interaction Similarity** represents the similarity of protein-ligand interaction fingerprints. (9) **Binding Free Energy** is calculated by the MMGBSA method⁵⁵. (10) **Docking Score** is estimated by the AutoDock Vina⁵⁴ program, which reports three types of scores - Vina Score, Vina Min, and Vina Dock. Vina Score represents the score of the directly generated ligand pose. Vina Min computes the score after local energy minimization, and Vina Dock provides the best possible score after redocking. (11) **Jensen-Shannon (JS) divergence** measures the similarity between the generated and reference distributions of bond lengths, bond angles, and dihedral angles. (12) **RMSD** is the root mean square deviation of heavy atoms between aligned conformations. (13) **QED** stands for quantitative estimation of drug-likeness combining multiple molecular properties. (14) **SA** signifies the synthetic accessibility that measures the difficulty of synthesizing organic molecules. (15) **LogP** is the octanol-water partition coefficient that assesses a compound's lipophilicity, and it indicates how well the compound dissolves in fats compared to water. (16) **TPSA** is the abbreviation of topological polar surface area, which predicts the molecule's ability to interact with biological membranes and its overall bioavailability. The JS divergence and RMSD are defined as follows:

$$JS(P||Q) = \frac{1}{2} D_{KL}(P||M) + \frac{1}{2} D_{KL}(Q||M) \quad (25)$$

$$RMSD(R, \tilde{R}) = \min_{\Phi} \sqrt{\left(\frac{1}{n} \sum_{i=1}^n \|\Phi(R_i) - \tilde{R}_i\|^2 \right)} \quad (26)$$

where P and Q are two probability distributions, M ($M = 1/2(P + Q)$) is a mixture distribution of P and Q , n is the number of heavy atoms, Φ is an alignment function that aligns two conformations by rotation and translation, R and \tilde{R} are generated conformation and optimized/predicted conformation, respectively.

Baselines

We compare our proposed method DiffGui with the following molecular generation methods: (1) ResGen³⁰, an autoregressive generative model built on the principle of parallel multiscale modeling; (2) PocketFlow³¹, a structure-based autoregressive framework with chemical knowledge explicitly considered; (3) GCDM³², a geometry-complete diffusion model for 3D molecule generation; (4) TargetDiff³³, an initial attempt to produce target-aware 3D molecules using the diffusion model; (5) DiffSBDD³⁴, a diffusion-based framework to generate novel ligands by an inpainting-based sampling approach; (6) PMDM³⁵, a pocket-aware generative method that incorporates a dual diffusion strategy and the cross-attention mechanism.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw protein-ligand complex structural data are available in the public database of PDBbind (<https://www.pdbbind-plus.org.cn/>) and CrossDocked (<http://bits.csb.pitt.edu/files/crossdock2020/>). The PDB files used in this study are available in the PDB dataset (<https://www.rcsb.org/>) under accession code: PDBid [https://www.rcsb.org/structure/PDBid]. A video of the generation trajectory of a molecule in the pocket of the WDR5 protein (WIN site) is also available as Supplementary Video 1. Source data are provided in this paper.

Code availability

The processed data and source code of this work are publicly available at GitHub: <https://github.com/QiaoyuHu89/DiffGui> and Zenodo⁶⁵.

References

- Maxmen, A. Busting the billion-dollar myth: How to slash the cost of drug development. *Nature* **536**, 388–390 (2016).
- Morgan, S., Grootendorst, P., Lexchin, J., Cunningham, C. & Greyson, D. The cost of drug development: A systematic review. *Health Policy* **100**, 4–17 (2011).
- Wouters, O. J., McKee, M. & Luyten, J. Estimated research and development investment needed to bring a new medicine to market, 2009–2018. *JAMA* **323**, 844–853 (2020).
- Jorgensen, W. L. Efficient drug lead discovery and optimization. *Acc. Chem. Res.* **42**, 724–733 (2009).
- Hughes, J. P., Rees, S., Kalindjian, S. B. & Philpott, K. L. Principles of early drug discovery. *Br. J. Pharmacol.* **162**, 1239–1249 (2011).
- Goodkin, H. P. & Kapur, J. The impact of diazepam's discovery on the treatment and understanding of status epilepticus. *Epilepsia* **50**, 2011–2018 (2009).
- Wainwright, M. The mystery of the plate: Fleming's discovery and contribution to the early development of penicillin. *J. Med. Biogr.* **1**, 59–65 (1993).
- Ajjarapu, S. M., Tiwari, A., Ramteke, P. W., Singh, D. B. & Kumar, S. Chapter 15 - ligand-based drug designing. in *Bioinformatics: Methods and Applications*. (Academic Press, 2022).
- Senior, A. W. et al. Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).
- Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
- Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
- Wang, X., Song, K., Li, L. & Chen, L. Structure-based drug design strategies and challenges. *Curr. Top. Med. Chem.* **18**, 998–1006 (2018).
- Anderson, A. C. The process of structure-based drug design. *Chem. Biol.* **10**, 787–797 (2003).
- Ferreira, L. G., Dos Santos, R. N., Oliva, G. & Andricopulo, A. D. Molecular docking and structure-based drug design strategies. *Molecules* **20**, 13384–13421 (2015).
- Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **432**, 862–865 (2004).
- McInnes, C. Virtual screening strategies in drug discovery. *Curr. Opin. Chem. Biol.* **11**, 494–502 (2007).
- Kirkpatrick, P. & Ellis, C. Chemical space. *Nature* **432**, 823–823 (2004).
- Reymond, J. L. & Awale, M. Exploring chemical space for drug discovery using the chemical universe database. *ACS Chem. Neurosci.* **3**, 649–657 (2012).
- Lyu, J., Irwin, J. J. & Shoichet, B. K. Modeling the expansion of virtual screening libraries. *Nat. Chem. Biol.* **19**, 712–718 (2023).
- Sadybekov, A. V. & Katritch, V. Computational approaches streamlining drug discovery. *Nature* **616**, 673–685 (2023).
- Atz, K., Grisoni, F. & Schneider, G. Geometric deep learning on molecular representations. *Nat. Mach. Intell.* **3**, 1023–1032 (2021).
- Bronstein, M. M., Bruna, J., Cohen, T. & Veličković, P. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. Preprint at <https://doi.org/10.48550/arXiv.2104.13478> (2021).
- Li, F. et al. DeepPROTACs is a deep learning-based targeted degradation predictor for PROTACs. *Nat. Commun.* **13**, 7133 (2022).
- Peng, X. et al. Pocket2Mol: Efficient molecular sampling based on 3D protein pockets. In *Proceedings of the 39th International Conference on Machine Learning* **162**, 17644–17655 (PMLR, 2022).
- Luo, S., Guan, J., Ma, J. & Peng, J. A 3D generative model for structure-based drug design. In *Advances in Neural Information Processing Systems* **34**, 6229–6239 (NeurIPS, 2021).
- Liu, M., Luo, Y., Uchino, K., Maruhashi, K. & Ji, S. Generating 3D molecules for target protein binding. In *Proceedings of the 39th International Conference on Machine Learning* **162**, 13912–13924 (PMLR, 2022).
- Ragoza, M., Masuda, T. & Koes, D. R. Generating 3D molecules conditional on receptor binding sites with deep generative models. *Chem. Sci.* **13**, 2701–2713 (2022).
- Li, F., Hu, Q., Zhou, Y., Yang, H. & Bai, F. DiffPROTACs is a deep learning-based generator for proteolysis targeting chimeras. *Brief. Bioinform.* **25**, bbae358 (2024).
- Skalic, M., Varela-Rial, A., Jiménez, J., Martínez-Rosell, G. & De Fabritiis, G. LigVoxel: Inpainting binding pockets using 3D-convolutional neural networks. *Bioinformatics* **35**, 243–250 (2018).
- Zhang, O. et al. ResGen is a pocket-aware 3D molecular generation model based on parallel multiscale modelling. *Nat. Mach. Intell.* **5**, 1020–1030 (2023).
- Jiang, Y. et al. PocketFlow is a data-and-knowledge-driven structure-based molecular generative model. *Nat. Mach. Intell.* **6**, 326–337 (2024).
- Morehead, A. & Cheng, J. Geometry-complete diffusion for 3D molecule generation and optimization. *Commun. Chem.* **7**, 150 (2024).
- Guan, J. et al. 3D equivariant diffusion for target-aware molecule generation and affinity prediction. In *International Conference on Learning Representations (ICLR)*, 2023.
- Schneuing, A. et al. Structure-based drug design with equivariant diffusion models. *Nat. Comput. Sci.* **4**, 899–909 (2024).
- Huang, L. et al. A dual diffusion model enables 3D molecule generation and lead optimization based on target pockets. *Nat. Commun.* **15**, 2657 (2024).
- Jonathan, H., Ajay, J. & Pieter, A. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems* **33**, 6840–6851 (NeurIPS, 2020).
- Garcia Satorras, V., Hoogeboom, E. & Welling, M. E(n) equivariant graph neural networks. In *Proceedings of the 38th International Conference on Machine Learning* **139**, 9323–9332 (PMLR, 2021).
- Garcia Satorras, V., Hoogeboom, E., Fuchs, F., Posner, I. & Welling, M. E(n) equivariant normalizing flows. In *Advances in Neural Information Processing Systems* **34**, 4181–4192 (NeurIPS, 2021).
- Fuchs, F., Worrall, D., Fischer, V. & Welling, M. SE(3)-Transformers: 3D roto-translation equivariant attention networks. In *Advances in Neural Information Processing Systems* **33**, 1970–1981 (NeurIPS, 2020).
- O'Boyle, N. M. et al. Open babel: An open chemical toolbox. *J. Cheminform.* **3**, 33 (2011).
- Guan, J. et al. DecompDiff: Diffusion models with decomposed priors for structure-based drug design. In *Proceedings of the 40th International Conference on Machine Learning* **202**, 11827–11846 (PMLR, 2023).
- Tian, S. et al. The application of in silico drug-likeness predictions in pharmaceutical research. *Adv. Drug Del. Rev.* **86**, 2–10 (2015).
- Ursu, O., Rayan, A., Goldblum, A. & Oprea, T. I. Understanding drug-likeness. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 760–781 (2011).
- Bonnet, P. Is chemical synthetic accessibility computationally predictable for drug and lead-like molecules? A comparative assessment between medicinal and computational chemists. *Eur. J. Med. Chem.* **54**, 679–689 (2012).
- Ulrich, N., Goss, K.-U. & Ebert, A. Exploring the octanol-water partition coefficient dataset using deep learning techniques and data augmentation. *Commun. Chem.* **4**, 90 (2021).
- Peng, X., Guan, J., Liu, Q. & Ma, J. MolDiff: Addressing the atom-bond inconsistency problem in 3D molecule diffusion generation.

- In *Proceedings of the 40th International Conference on Machine Learning* **202**, 27611–27629 (PMLR, 2023).
47. Jonathan, H. & Salimans, T. Classifier-free diffusion guidance. Preprint at <https://doi.org/10.48550/arXiv.2207.12598> (2022).
 48. Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P. S. & Vaughan, J. W. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems* **34**, 8780–8794 (NeurIPS, 2021).
 49. Lin, J. Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory* **37**, 145–151 (1991).
 50. Kullback, S. & Leibler, R. A. On information and sufficiency. *Ann. Inst. Stat.* **22**, 79–86 (1951).
 51. Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **17**, 490–519 (1996).
 52. Riniker, S. & Landrum, G. A. Better informed distance geometry: Using what we know to improve conformation generation. *J. Chem. Inf. Model.* **55**, 2562–2574 (2015).
 53. Rappe, A. K., Casewit, C. J., Colwell, K. S., Goddard III, W. A. & Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **114**, 10024–10035 (1992).
 54. Trott, O. & Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
 55. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **10**, 449–461 (2015).
 56. Sutherland, D. P. et al. Discovery of (thienopyrimidin-2-yl)aminopyrimidines as potent, selective, and orally available Pan-PI3-kinase and dual Pan-PI3-kinase/mTOR inhibitors for the treatment of cancer. *J. Med. Chem.* **53**, 1086–1097 (2010).
 57. Aho, E. R. et al. Displacement of WDR5 from chromatin by a WIN site inhibitor with picomolar affinity. *Cell Rep.* **26**, 2916–2928 (2019).
 58. Wang, X. et al. Identification of MRTX1133, a noncovalent, potent, and selective KRAS^{G12D} inhibitor. *J. Med. Chem.* **65**, 3123–3133 (2022).
 59. Qiang, B. et al. Coarse-to-fine: A hierarchical diffusion model for molecule generation in 3D. In *Proceedings of the 40th International Conference on Machine Learning* **202**, 28277–28299 (PMLR, 2023).
 60. Podda, M., Bacciu, D. & Micheli, A. A deep generative model for fragment-based molecule generation. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics* **108**, 2240–2250 (PMLR, 2020).
 61. Wang, R., Fang, X., Lu, Y. & Wang, S. The PDBbind database: Collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J. Med. Chem.* **47**, 2977–2980 (2004).
 62. Wang, R., Fang, X., Lu, Y., Yang, C.-Y. & Wang, S. The PDBbind database: Methodologies and updates. *J. Med. Chem.* **48**, 4111–4119 (2005).
 63. Francoeur, P. G. et al. Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. *J. Chem. Inf. Model.* **60**, 4200–4215 (2020).
 64. Buttenschoen, M., Morris, G. M. & Deane, C. M. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chem. Sci.* **15**, 3130–3139 (2024).
 65. Hu, Q. et al. Target-aware 3D molecular generation based on guided equivalent diffusion. DiffGui v1.0 Zenodo, <https://doi.org/10.5281/zenodo.16193760> (2025).

Acknowledgements

We acknowledge the support from the HPC platform of the Innovation Center for AI and Drug Discovery at East China Normal University. We express our sincere gratitude to Prof. Yanyan Diao, Prof. Shiliang Li, and Prof. Lili Zhu for their advice on experimental target selection. This work has been supported by the National Key Research and Development Program of China (2022YFC3400501 K.Z.), the National Natural Science Foundation of China (82425104 H.L. and 82404518 D.L.), the Science and Technology Commission of Shanghai Municipality (24JS2830200 Q.H.), and the Shanghai Municipal Education Commission (2024AI01014 H.L.).

Author contributions

Q.H. constructed the model, conducted the experiments and wrote the initial manuscript. C.S. solved the code problem encountered and wrote the method section. H.H. conducted the wet-lab experimental validations and wrote the synthesis section. J.X. contributed to the figure drawing, code investigation and manuscript writing. D.L. collected the training data and revised the manuscript. W.Z. and S.S. computed the similarity of protein-ligand interaction fingerprints. K.Z. supervised the project, designed the experiments and improved the manuscript. H.L. conceived and directed the study.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-63245-0>.

Correspondence and requests for materials should be addressed to Qiaoyu Hu or Honglin Li.

Peer review information *Nature Communications* thanks Yuemin Bian and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025