

Target-aware 3D Molecular Generation Based on Guided Equivariant Diffusion

Corresponding Author: Professor Qiaoyu Hu

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors devised a target-aware molecule generation method (DiffGui) using a guided diffusion model. They integrated the guidance of bond diffusion and molecular properties to enhance the structure rationality, the binding affinity and drug-like attributes of the generated molecules. Besides de novo drug design, they also implemented lead optimization based on known molecular fragments by employing novel sampling techniques. It looks interesting as it is very important to minimize the creation of molecules with illogical structures and inferior properties. However, to improve the quality of this manuscript and clarify any confusion on my part, I would like to pose several questions:

1. It is well known that the amount of data on membrane proteins is quite limited in the PDBbind database. In light of this, I am curious about the performance of DiffGui in generating ligand molecules within membrane proteins. If the performance is not satisfactory, what measures can the authors take to enhance it?

2. The previous generation models mostly utilize CrossDocked database as training set. Despite containing a huge amount of data, the CrossDocked database is constructed by docking ligands into multiple similar binding pockets. I agree with the authors to utilize protein-ligand co-crystal structures to train the deep learning model. However, the question arises why they do not use the BindingMoad database, which offers more than 40,000 structures, exceeding the number available in PDBbind.

3. The authors considered the properties of binding affinity, synthetic accessibility, drug likeness within their model. As is commonly recognized, the ADMET properties play a pivotal role in the development of pharmaceuticals. Might it be feasible to integrate these ADMET properties into the methodology of molecule generation?

4. In the section "Quality of generated molecules", please discuss more on why the model shows better performance on the PDBbind dataset than the CrossDocked dataset.

5. In the section "Quality of generated molecules" line 268-271, the author mentioned that "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 also commonly observed in methods using diffusion models, such as TargetDiff, DiffSBDD, and PMDM." Please discuss the reasons why diffusion models would cause this issue?

6. The clarity of the images in the article file is not high. Please upload images with higher resolution. Please increase font sizes for legibility in Figure 1b.

(Remarks on code availability)

The software package is well-organized to some extend.

Reviewer #2

(Remarks to the Author)

Summary:

This work develops a new diffusion model to generate drug-like small molecules. The main contribution is to consider the diffusion of atom-atom bonds in the diffusion process to make the generated molecules more realistic. The method is compared with several existing methods, yielding good performance. However, there some concerns with the evaluation that need to be addressed.

1) Major comments:

The proposed method is compared with some existing methods using some evaluation methods. But quite a few commonly used evaluation metrics are missing. Please add the following metrics into the evaluation: negative log-likelihood, atom stability, molecular stability, uniqueness, PoseBusters-validated score (PB-validity), and RDKit-valid whenever possible.

The proposed method is compared with several existing methods. But many useful methods are not considered in the comparison. Please also compare your method with the following methods: GDM, GDM-aug, EDM, Bridge, GraphLDM, GeoLDM, GCDM, and G-Schnet.

2) Minor comments:

For $q(x_t | x_{t-1}, p, c)$, please define the variables such as p, c clearly near the formula so that it is easy for readers to read when it first appears.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

The manuscript by Hu et al. describes the development of a computational tool (DefGui) to propose target-aware ligands using a diffusion-based model explicitly taking into account both atom positions and bond generation. The method has been benchmarked on two public datasets (PDBbind, CrossDocked) and applied to drug design scenarios aimed at growing seed fragments into full ligands, and at generating ligands for pocket mutants.

The basic concept of the methodology is clearly interesting as applying the diffusion process to bonds and not only atoms is supposed to lead to more valid molecules in terms of molecular topology. However, as for many studies of that kind currently flourishing in the literature, the full study is not accompanied by a single experiment to support the predictions. Therefore, I think the manuscript is not suitable as it is, for publication in a journal like Nature Communications with a broad authorship. I do believe that experimental validations (lacking here) must absolutely be part of any report describing a novel AI-based drug design method. Moreover, the manuscript is full of claims that are purely speculative (computational) and not supported by any experimental data. Here are my major and minor concerns that should be addressed in a revised version.

MAJOR CONCERNS

1. Throughout the text, authors are confusing the notion of 'binding affinity' with a simple docking score. I should here recall that docking score are very poor proxies for binding affinities (see Pansar et al. *Molecules* 2018 Jul 30;23(8):1899) and that this confusion must be avoided. By no way can a docking score be labelled as "binding affinity." In absence of experimental binding affinity data, authors cannot claim that their method is generating ligands with increased binding affinity

2. Page 4: Authors seem to ignore that the available chemical space for virtual screening is no more restricted to "the historical knowledge of previously investigated molecules". Today, about 3 trillion "make on-demand" drug-like compounds, synthesizable from commercial building blocks and hundreds of robust 1-3 steps organic chemistry reactions, can be screened. using either a ligand-based or a structure-based paradigm. For a recent review, see Sadybekov et al. *Nature* volume 616, pages 673–685 (2023).

3. The Introductory section beginning line 129 is far too long and should find a better place in either the Discussion or the Conclusion sections.

4. The validity of the generated ligands has been not checked in enough details to guarantee that the poses are really valid. A gold standard check for generative models is to check for validity of PoseBusters filters (Buttenschoen et al. *Chemical Sciences*, 2024, 15, 3130) that investigates chemical validity, intra and inter-molecular validity. The JS divergence and rmsd criteria, proposed by the authors, do not suffice notably for being sure that the proposed poses are not free of high local and global strain energies that would preclude for strong binding. The higher JS divergence observed between reference and generated molecules (Table 1) for the dihedral angles makes me suspicious about the internal strain energy of generated compounds.

In addition to the PoseBusters check, authors should also compare generated and reference ligands with respect to simple 2D similarity (e.g. using Morgan-2 fingerprints), 3D pharmacophore overlap (e.g. using OpenEye's ROCS) and similarity of protein-ligand interaction fingerprints. These verifications permit to ascertain that the proposed designs are not trivial and truly respect key pharmacophoric groups interacting with key binding site residues.

5. Contrarily to authors' claims I do not think that the set of 4 targets used for the de novo design section are really diverse. Three of them are protein kinases and exhibit very similar binding pockets. A greater diversity would be preferable by adding

for example a G protein-coupled receptor, a nuclear hormone receptor, an ion channel or a transporter.

MINOR CONCERNS

Figures 5 and 6 are difficult to read mainly because the plots are too small. Moreover, the legends should clearly precise the reason for the selection of these ligands. Do they correspond to the solutions with best docking scores?

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Authors have appropriately addressed reviewers' comments in the first round.

It is encouraging to notice the incorporation of wet-lab validations. The synthesis and enzymatic inhibition assays for RSK4 provide experimental confirmation of the generated molecules' biological relevance.

No additional questions to post.

(Remarks on code availability)

It says that the reviewer's browser is not compatible with the code platform that Private link above directs to.

The reviewer accessed the code through the GitHub link provided in the manuscript. <https://github.com/QiaoyuHu89/DiffGui>

Overall, the code is well-organized with a detailed README file. It is relatively straightforward to have the DiffGui deployed.

Reviewer #2

(Remarks to the Author)

The authors addressed one of my two main comments well. However, the authors did not address the following main comment satisfactorily:

"The proposed method is compared with several existing methods. But many useful methods are not considered in the comparison. Please also compare your method with the following methods: GDM, GDM-aug, EDM, Bridge, GraphLDM, GeoLDM, GCDM, and G-Schnet."

The authors did not compare their method with the methods in the comment above because they think all the methods above can only do unconditional molecule generation, which is different from the pocket-conditioned molecule generation that their method aims to do. However, at least one method above - GCDM can perform both unconditional and pocket-conditioned molecule generation. Therefore, the authors should add a comparison with GCDM to show the advance that the method in this work can make.

(Remarks on code availability)

The code is well organized.

Reviewer #3

(Remarks to the Author)

He revised manuscript by Hu et al. partially addressed some of my original concerns but I am still not satisfied with the proposed experimental validation that is very weak to me for the following reasons:

- Predictions are made to generate yet another kinase inhibitors (RSK4 inhibitors) of very weak potency (dose response curves are even incomplete to me). Since the PDBbind dataset the method is trained is already biased by a high number of kinase-inhibitor complexes, I am not convinced at all by these two weak examples. By comparison, many nanomolar kinase inhibitors have been rationally designed from scratch (e.g. Beroza et al. Nat Commun. 2022 Oct 28;13(1):6447; Eguida et al. J Med Chem. 2022 Oct 27;65(20):13771-13783).

- The two structures are not original at all since the two molecular scaffolds can be found in ChEMBL (623 compounds for the scaffold of compound 1, 55 compounds for the scaffold of compound 2), mostly for already described kinase inhibitors (CDK4, JAK2, etc.). I fear that the program is just memorizing known data but not genuinely discover novel chemotypes. Authors should provide a much more consistent validation with respect to the target (please not a target family already widely investigated in PDBbind), number and potency of designed/synthesized/tested ligands.

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)

The authors have addressed my comments well.

(Remarks on code availability)

The code in the CodeOcean looks good.

Reviewer #3

(Remarks to the Author)

Authors have addressed my main concern regarding novel experimental data to support predictions.

The revised version is suitable for publication.

(Remarks on code availability)

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <https://creativecommons.org/licenses/by/4.0/>



華東師範大學
EAST CHINA NORMAL UNIVERSITY

Dear Reviewers,

We are deeply grateful for your recognition on our work and the invaluable, constructive feedback you have provided to our manuscript entitled “**Target-aware 3D Molecular Generation Based on Guided Equivariant Diffusion**”, which was submitted to Nature Communications with Manuscript ID: **NCOMMS-24-67247-T**. We have thoroughly considered all of your comments and revised our manuscript in accordance with your suggestions. The modifications have been highlighted in red within the revised manuscript. We sincerely hope that this updated version is acceptable at this time.

Presented below is a detailed summary of your comments and our responses.



Comments from Reviewer #1 and Responses:

The authors devised a target-aware molecule generation method (DiffGui) using a guided diffusion model. They integrated the guidance of bond diffusion and molecular properties to enhance the structure rationality, the binding affinity and drug-like attributes of the generated molecules. Besides de novo drug design, they also implemented lead optimization based on known molecular fragments by employing novel sampling techniques. It looks interesting as it is very important to minimize the creation of molecules with illogical structures and inferior properties. However, to improve the quality of this manuscript and clarify any confusion on my part, I would like to pose several questions:

- 1. It is well known that the amount of data on membrane proteins is quite limited in the PDBbind database. In light of this, I am curious about the performance of DiffGui in generating ligand molecules within membrane proteins. If the performance is not satisfactory, what measures can the authors take to enhance it?*

Response: Thank you for your valuable feedback. It is true to state that the PDBbind database contains a limited number of complexes (not exceeding 10) featuring transmembrane proteins. Consequently, we assessed our model using two transmembrane proteins, namely S1PR3 (PDBid 7ew4) and TRPV4 (PDBid 8ju6), and visualized the generated ligands. As displayed in Figure 6 and Table 7, the molecules produced by our method exhibit the best docking scores, the lowest binding free energies, and favorable molecular properties, thereby validating the excellent generalization performance of our model. To further enhance the model's generative capability towards the membrane proteins, it is advisable to increase the quantity of membrane proteins within the training set. This expansion can be achieved by either gathering data from literatures/patents or accessing internal databases of other institutions.

The corresponding figures, tables, and descriptions have been integrated into the revised manuscript (lines 367-368, 371-372, 374-377, 379-385 on page 14).

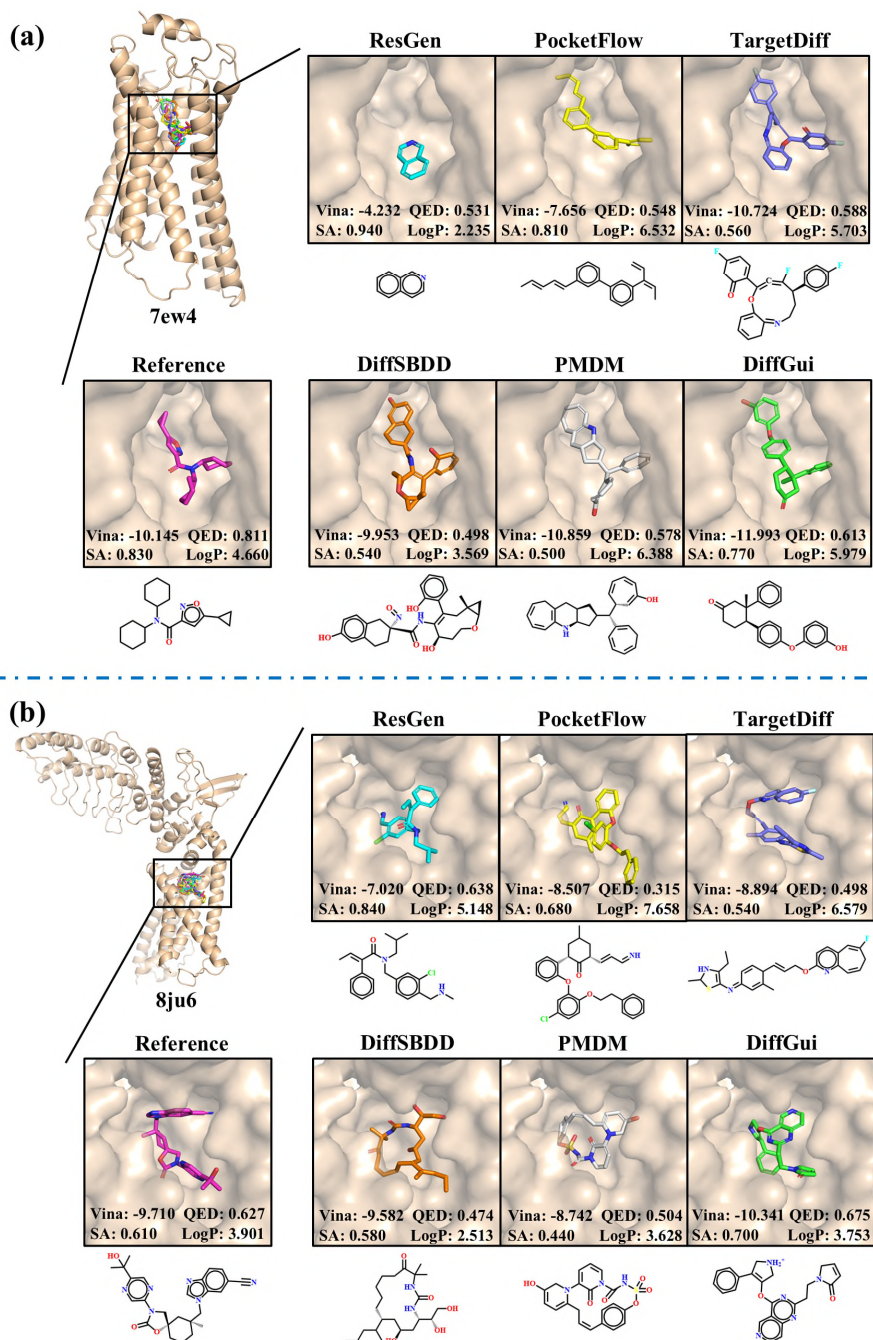


Fig 6. Molecules generated for targets outside the PDBbind dataset. The protein targets of 7ew4 and 8ju6 are G protein-coupled receptor and ion channel, respectively. The visualized molecules have the best docking scores among those PB-valid.

Table 7. The binding free energies ΔG (kcal/mol) calculated by MMGBSA for generated molecules. Bold fonts indicate the best results.

PDBId	Ref.	ResGen	PocketFlow	TargetDiff	DiffSBDD	PMDM	DiffGui
-------	------	--------	------------	------------	----------	------	---------



1w51	-80.30	-26.00	-49.57	-49.30	-21.39	-44.62	-90.98
3ctj	-84.36	-23.44	-53.25	-30.24	-40.65	-50.26	-72.43
7ew4	-63.84	-29.24	-44.22	-60.30	-19.63	-19.44	-67.94
8ju6	-64.43	-50.98	-72.32	-42.59	-41.78	-34.09	-73.22

2. *The previous generation models mostly utilize CrossDocked database as training set. Despite containing a huge amount of data, the CrossDocked database is constructed by docking ligands into multiple similar binding pockets. I agree with the authors to utilize protein-ligand co-crystal structures to train the deep learning model. However, the question arises why they do not use the BindingMoad database, which offers more than 40,000 structures, exceeding the number available in PDBbind.*

Response: Thank you for your insightful question. The BindingMoad dataset¹ comprises 41,409 structures with affinity coverage for only 15,223 complexes, which are all involved in the PDBbind database. Our method necessitates affinity data during the training process to facilitate the effective generation of high-affinity ligands. This is the reason why we chose the PDBbind over the BindingMoad dataset. But we must acknowledge that a greater amount of data generally leads to better performance for deep learning models. In future endeavors, we can employ prediction methods²⁻⁵ to estimate the binding affinity of ligands that lack experimental affinity data. However, it should be noted that the accuracy of these predictions cannot be guaranteed, and it is always preferable to utilize true labels over estimated ones.

3. *The authors considered the properties of binding affinity, synthetic accessibility, drug likeness within their model. As is commonly recognized, the ADMET properties play a pivotal role in the development of pharmaceuticals. Might it be feasible to integrate these ADMET properties into the methodology of molecule generation?*

Response: Thank you for your insightful suggestion regarding the integration of ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) properties into our methodology for molecule generation. In our current model, we focused on binding affinity, synthetic accessibility, and drug-likeness properties, which are essential for initial screening and selection of potential drug candidates. Nevertheless, we fully recognize the critical significance of ADMET properties and their impact on the overall pharmaceutical development process.

Although integrating ADMET properties into our methodology is theoretically feasible, we encounter implementation barriers primarily stemming from the scarcity of experimentally validated ADMET data for many compounds in the dataset. Additionally, directly incorporating ADMET values into the diffusion process could further escalate the complexity of our model, which might adversely affect both its generative performance and the interpretability of the underlying mechanisms.

To overcome these limitations, we propose to develop a reinforcement learning (RL)-based framework in subsequent research endeavors. RL represents a robust machine learning paradigm that



enables the optimization of multi-objective functions through iterative policy refinement guided by environmental feedback. Within this framework, our initial step is to build a reward function that evaluates the predicted⁶⁻⁸ or experimental ADMET data. Subsequently, the RL algorithm would be employed to iteratively optimize the molecular generation policy, thereby facilitating the discovery of compounds that maximize the reward function.

4. *In the section “Quality of generated molecules”, please discuss more on why the model shows better performance on the PDBbind dataset than the CrossDocked dataset.*

Response: We appreciate your suggestion to discuss more on why the model performs better on the PDBbind dataset than the CrossDocked dataset. This is primarily because our model was trained specifically on the PDBbind dataset, which allows it to capture the underlying patterns and features within this dataset more effectively. As a result, when evaluated on the PDBbind dataset, the model is able to leverage the knowledge it has acquired to make more accurate predictions.

The CrossDocked dataset differs from PDBbind in various aspects such as data distribution, complex structure, and binding affinity range. Due to the absence of 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.

The relevant descriptions have been added in the revised manuscript (lines 235-244 on page 9).

5. *In the section “Quality of generated molecules” line 268-271, the author mentioned that “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 also commonly observed in methods using diffusion models, such as TargetDiff, DiffSBDD, and PMDM.” Please discuss the reasons why diffusion models would cause this issue?*

Response: Thank you for raising this important question. The tendency of diffusion models to produce a higher frequency of seven-membered rings can be attributed to their underlying architecture and inference process. 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 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.

The relevant explanations have been incorporated into the revised manuscript (lines 283-295 on page 11).



華東師範大學
EAST CHINA NORMAL UNIVERSITY

6. The clarity of the images in the article file is not high. Please upload images with higher resolution. Please increase font sizes for legibility in Figure 1b.

Response: Thank you for pointing out this problem. Regarding the clarity of the images, please note that the original images in the submitted file are of high resolution. However, during the format conversion from Word to PDF on submission system, the resolution of the images appears to be compromised. We understand that this may affect the legibility of the content. To address your concerns, we will upload the high-resolution images as separate files if our article is accepted for publication. Additionally, the font sizes in Figure 1b are increased for better legibility.

7. The software package is well-organized to some extent.

Response: Thanks for the reviewer's positive comments.



Comments from Reviewer #2 and Responses:

This work develops a new diffusion model to generate drug-like small molecules. The main contribution is to consider the diffusion of atom-atom bonds in the diffusion process to make the generated molecules more realistic. The method is compared with several existing methods, yielding good performance. However, there some concerns with the evaluation that need to be addressed.

- 1. The proposed method is compared with some existing methods using some evaluation methods. But quite a few commonly used evaluation metrics are missing. Please add the following metrics into the evaluation: negative log-likelihood, atom stability, molecular stability, uniqueness, PoseBusters-validated score (PB-validity), and RDKit-valid whenever possible.*

Response: Thank you for bringing up this crucial point. We recognize the significance of the additional evaluation metrics mentioned in assessing the generative abilities of various methods. Therefore, we have calculated these metrics and added them in Tables 3 and S8. Our method, DiffGui, demonstrates superior performance over other methods in terms of atom stability, molecular stability, PB-validity⁹, 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.

The associated descriptions and tables have been incorporated into the revised manuscript (lines 297-302 on page 11) and the supporting information file.

Table 3. Basic molecular metrics of generated molecules for the PDBbind dataset. Bold fonts indicate the best results.

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

Table S8. Basic molecular metrics of generated molecules for the CrossDocked dataset. Bold fonts indicate the best results.

Metric	ResGen	PocketFlow	TargetDiff	DiffSBDD	PMDM	DiffGui
Atom Stab. (↑)	0.8580	0.8631	0.9483	0.8770	0.8750	0.9584
Mol. Stab. (↑)	0.1377	0.1472	0.3727	0.2050	0.4958	0.5362
PB-valid. (↑)	0.9420	0.9542	0.9556	0.7323	0.6767	0.9769



RDKit-valid. ($\uparrow$)	0.9347	0.9561	0.9036	0.9808	0.9822	0.9894
Novel. ($\uparrow$)	0.9585	0.9328	0.9902	0.9860	0.9873	0.9798
Uniq. ($\uparrow$)	0.9230	0.8447	0.9869	0.9838	0.9797	0.9371
Sim. Ref. ($\downarrow$)	0.3872	0.4018	0.5239	0.4705	0.4904	0.3824
Inter. Sim. ($\uparrow$)	0.5381	0.5155	0.6689	0.5519	0.4099	0.6829

2. *The proposed method is compared with several existing methods. But many useful methods are not considered in the comparison. Please also compare your method with the following methods: GDM, GDM-aug, EDM, Bridge, GraphLDM, GeoLDM, GCDM, and G-Schnet.*

Response: Thank you for your constructive feedback and suggestion on our manuscript. Regarding your concern about the comparative analysis between our proposed framework and the methods of GDM¹⁰, GDM-aug¹⁰, EDM¹⁰, Bridge¹¹, GraphLDM¹², GeoLDM¹², GCDM¹³, and G-Schnet¹⁴, we would like to emphasize that these established methods are primarily designed for either unconditional molecular generation or property-based molecular design. In contrast, our framework specifically focuses on protein pocket-conditioned molecular generation, which represents a distinct paradigm with fundamentally different objectives and constraints. Therefore, a direct comparison between our method and the above approaches would be inappropriate due to the fundamental discrepancies in the underlying objectives.

Moreover, the baseline methods selected for our comparative analysis represent the state-of-the-art in molecular generation research, encompassing both autoregressive architectures (ResGen¹⁵, PocketFlow¹⁶) and diffusion-based frameworks (TargetDiff¹⁷, DiffSBDD¹⁸, and PMDM¹⁹). These carefully chosen benchmarks collectively capture the spectrum of mainstream methodological approaches in the field. We believe that this comprehensive comparison can provide a robust evaluation of our method against the latest and most relevant techniques.

3. *For $\frac{p}{c}(\frac{p}{c} - 1, \frac{p}{c}, \frac{p}{c})$, please define the variables such as p , c clearly near the formula so that it is easy for readers to read when it first appears.*

Response: Thank you for your suggestion that contributes to enhance the readability of our work. While the variables p and c have already been defined in the Task Definition part of the Methods section, we understand that providing additional definitions near the formula can greatly benefit the readers, especially when they encounter these variables for the first time. Therefore, we have revised the text (lines 171-172 on page 7) to include the clear definitions of p and c in proximity to the formula.



Comments from Reviewer #3 and Responses:

The manuscript by Hu et al. describes the development of a computational tool (DefGui) to propose target-aware ligands using a diffusion-based model explicitly taking into account both atom positions and bond generation. The method has been benchmarked on two public datasets (PDBbind, CrossDocked) and applied to drug design scenarios aimed at growing seed fragments into full ligands, and at generating ligands for pocket mutants.

The basic concept of the methodology is clearly interesting as applying the diffusion process to bonds and not only atoms is supposed to lead to more valid molecules in terms of molecular topology. However, as for many studies of that kind currently flourishing in the literature, the full study is not accompanied by a single experiment to support the predictions. Therefore, I think the manuscript is not suitable as it is, for publication in a journal like Nature Communications with a broad authorship.

I do believe that experimental validations (lacking here) must absolutely be part of any report describing a novel AI-based drug design method. Moreover, the manuscript is full of claims that are purely speculative (computational) and not supported by any experimental data. Here are my major and minor concerns that should be addressed in a revised version.

- 1. I do believe that experimental validations (lacking here) must absolutely be part of any report describing a novel AI-based drug design method.*

Response: Thank you so much! We agree with the reviewer's assertion that the experimental validations should be accompanied with the AI-based drug design method to substantiate its effectiveness. Therefore, we conducted wet-lab experiments on molecules generated for RSK4 (ribosomal S6 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 Figure S7, despite their structural differences, both Compound 1 and Compound 2 demonstrate potent inhibitory activity in the HTRF assay, with IC_{50} values of approximately 1.2 μ M and 5.3 μ M, respectively, highlighting their potential as lead compounds. The binding modes reveal that both compounds interact with key residues (K105, D153, L155, and T215) in the binding pocket of RSK4.

The corresponding figures and descriptions have been incorporated into the revised manuscript (lines 436-444 on page 16) and the supporting information file.

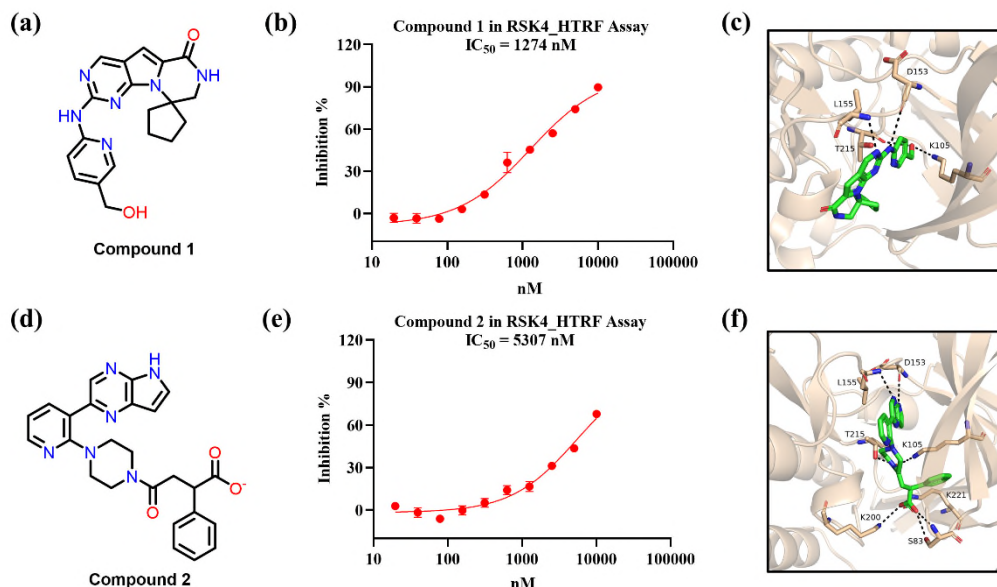


Fig S7. Wet-lab experimental validation of DiffGui. (a) Chemical structure of Compound 1, (b) Dose-activity curve of Compound 1 in the RSK4 kinase assay, (c) The binding mode of Compound 1 with RSK4, (d) Chemical structure of Compound 2, (e) Dose-activity curve of Compound 2 in the RSK4 kinase assay, (f) The binding mode of Compound 2 with RSK4.

2. Throughout the text, authors are confusing the notion of 'binding affinity' with a simple docking score. I should here recall that docking score are very poor proxies for binding affinities (see Pansar et al. *Molecules* 2018 Jul 30;23(8):1899) and that this confusion must be avoided. By no way can a docking score be labelled as "binding affinity." In absence of experimental binding affinity data, authors cannot claim that their method is generating ligands with increased binding affinity.

Response: Thank you for pointing out this important problem. In order to eliminate any ambiguity between the notions of binding affinity and docking score, we have revised the manuscript, replacing instances of binding affinity with either "estimated binding affinity" or "docking score", wherever the former term could potentially cause misunderstanding. In addition, we calculated the binding free energies of the generated molecules (top-ranked by docking score and PB-validity) using the MMGBSA method²⁰ for targets of 1w51 (beta-secretase), 3ctj (tyrosine kinase), 7ew4 (G protein-coupled receptor), and 8ju6 (ion channel), which cover diverse types of proteins. As shown in Table 7, while the binding free energy (ΔG) 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. For wet-lab experimental validation, please kindly refer to the response of Question 1.

The corresponding descriptions and table have been added in the revised manuscript (lines 379-385 on page 14).



Table 7. The binding free energies ΔG (kcal/mol) calculated by MMGBSA for generated molecules. Bold fonts indicate the best results.

PDBId	Ref.	ResGen	PocketFlow	TargetDiff	DiffSBDD	PMDM	DiffGui
1w51	-80.30	-26.00	-49.57	-49.30	-21.39	-44.62	-90.98
3ctj	-84.36	-23.44	-53.25	-30.24	-40.65	-50.26	-72.43
7ew4	-63.84	-29.24	-44.22	-60.30	-19.63	-19.44	-67.94
8ju6	-64.43	-50.98	-72.32	-42.59	-41.78	-34.09	-73.22

3. *Page 4: Authors seem to ignore that the available chemical space for virtual screening is no more restricted to "the historical knowledge of previously investigated molecules". Today, about 3 trillion "make on-demand" drug-like compounds, synthesizable from commercial building blocks and hundreds of robust 1-3 steps organic chemistry reactions, can be screened, using either a ligand-based or a structure-based paradigm. For a recent review, see Sadybekov et al. Nature volume 616, pages 673-685 (2023).*

Response: We appreciate your valuable comment in enhancing the quality of our manuscript. Having thoroughly examined the review paper, we revised our description (lines 74-81 on page 4) as below.

“Yet, it is computationally expensive to search the physical libraries that involve $10^6\sim10^7$ molecules or the virtual on-demand libraries that contain $10^{10}\sim10^{15}$ molecules, let alone the massive chemical space ($10^{60}\sim10^{100}$) of potential pharmacologically active molecules^{21, 22}. 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 approximations²³. 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 selection²⁴.”

4. *The Introductory section beginning line 129 is far too long and should find a better place in either the Discussion or the Conclusion sections.*

Response: Thank you for your perceptive suggestion. We removed the paragraph from the Introduction section and integrated it into the Results and Discussion section (lines 173-185 on page 7).

5. *The validity of the generated ligands has been not checked in enough details to guarantee that the poses are really valid. A gold standard check for generative models is to check for validity of PoseBusters filters (Buttenschoen et al. Chemical Sciences, 2024, 15, 3130) that investigates chemical validity, intra and inter-molecular validity. The JS divergence and rmsd criteria, proposed by the authors, do not suffice notably for being sure that the proposed poses are not free of high local and global strain energies that would preclude for strong binding. The higher JS divergence observed between reference and generated molecules (Table 1) for the dihedral angles*



makes me suspicious about the internal strain energy of generated compounds.

Response: Thank you for your valuable remark. To address your concerns, we have calculated the validities of PoseBusters filters (PB-valid.)⁹ for generated ligands and added them in Tables 3 and S8. Comparing to other methods, DiffGui achieves the best values (0.9537 and 0.9769) in this metric for both the PDBbind and CrossDocked datasets. The molecules produced by DiffGui also possess the highest validity score (RDKit-valid.) computed by the RDKit program. Moreover, in comparison to other baselines, DiffGui demonstrates the lowest average JS divergence for the dihedral angles on the PDBbind dataset, while achieving near-optimal values on the CrossDocked dataset (Table 1). Taking all these factors into account, it is reasonable to conclude that the ligand poses generated by our method are more valid.

The related texts and tables have been integrated into the revised manuscript (lines 297-300 on page 11) and the supporting information file.

Table 3. Basic molecular metrics of generated molecules for the PDBbind dataset. Bold fonts indicate the best results.

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

Table S8. Basic molecular metrics of generated molecules for the CrossDocked dataset. Bold fonts indicate the best results.

Metric	ResGen	PocketFlow	TargetDiff	DiffSBDD	PMDM	DiffGui
Atom Stab. (↑)	0.8580	0.8631	0.9483	0.8770	0.8750	0.9584
Mol. Stab. (↑)	0.1377	0.1472	0.3727	0.2050	0.4958	0.5362
PB-valid. (↑)	0.9420	0.9542	0.9556	0.7323	0.6767	0.9769
RDKit-valid. (↑)	0.9347	0.9561	0.9036	0.9808	0.9822	0.9894
Novel. (↑)	0.9585	0.9328	0.9902	0.9860	0.9873	0.9798
Uniq. (↑)	0.9230	0.8447	0.9869	0.9838	0.9797	0.9371
Sim. Ref. (↓)	0.3872	0.4018	0.5239	0.4705	0.4904	0.3824
Inter. Sim. (↑)	0.5381	0.5155	0.6689	0.5519	0.4099	0.6829



6. *In addition to the PoseBusters check, authors should also compare generated and reference ligands with respect to simple 2D similarity (e.g. using Morgan-2 fingerprints), 3D pharmacophore overlap (e.g. using OpenEye's ROCS) and similarity of protein-ligand interaction fingerprints. These verifications permit to ascertain that the proposed designs are not trivial and truly respect key pharmacophoric groups interacting with key binding site residues.*

Response: We appreciate your insightful suggestion to include the evaluation metrics of 2D similarity, 3D pharmacophore overlap, and similarity of protein-ligand interaction fingerprints. The corresponding results are listed in Tables 3 and S8, and presented in Figures S4 and S5. The 2D similarity between generated molecules and reference ligands is calculated using Morgan-2 fingerprints. The 3D pharmacophore overlap and similarity of protein-ligand interaction fingerprints are computed by the Schrödinger's Maestro program. As demonstrated in Tables 3 and S8, our method 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. As displayed in Figure S4, the molecules generated by DiffGui for 1w51 (beta-secretase) and 3ctj (tyrosine kinase) possess the highest number of 3D pharmacophore overlap. However, for 7ew4 (GPCR) and 8ju6 (ion channel) in Figure S5, 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.

The relevant tables, figures, and descriptions have been incorporated into the revised manuscript (lines 302-306 on page 11; lines 385-394 on page 14) and the supporting information file.

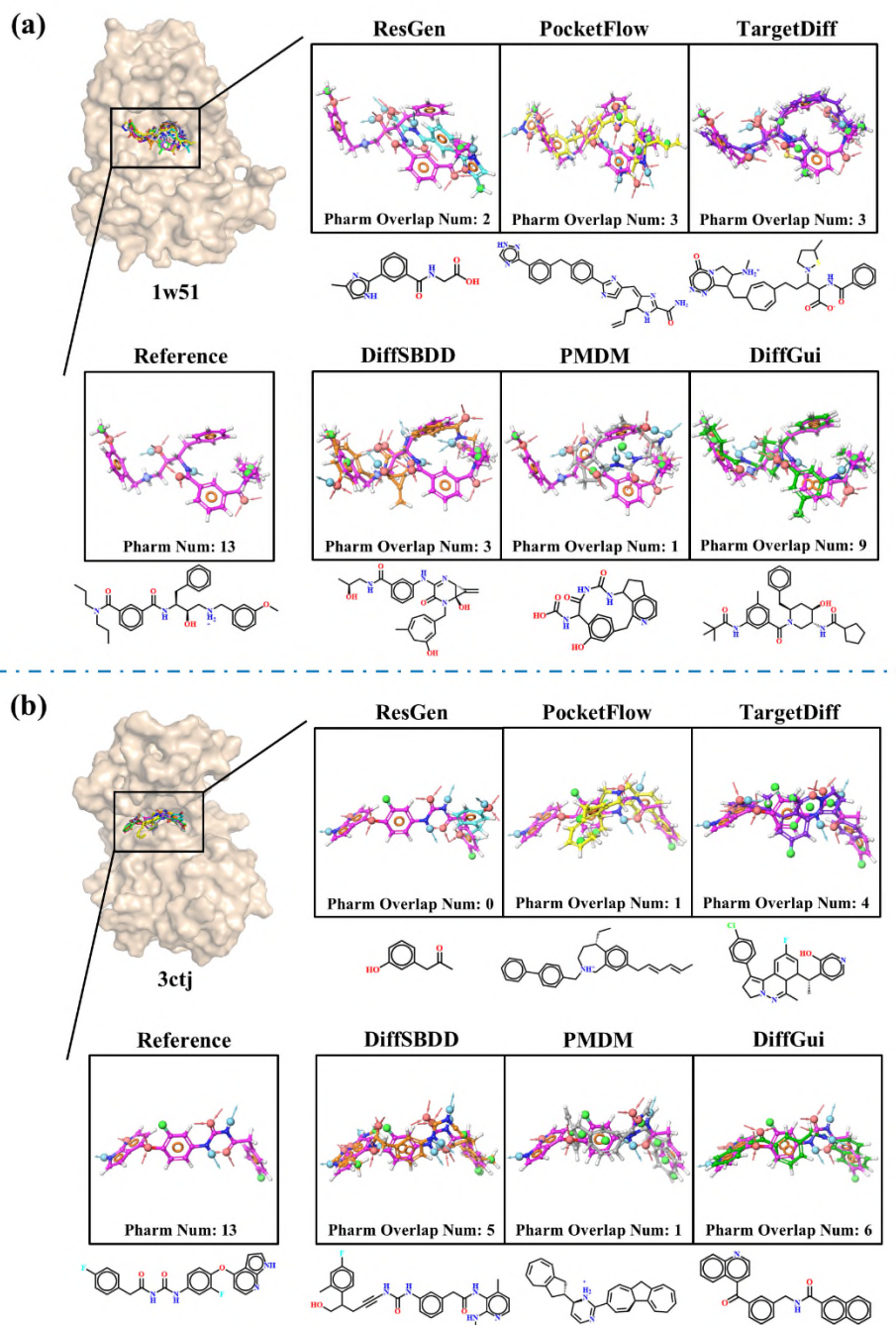


Fig S4. 3D pharmacophore overlaps of molecules generated for targets from the PDBbind test set. The protein targets of 1w51 and 3ctj are beta-secretase 1 and tyrosine kinase, respectively. The visualized molecules have the best docking scores among those PB-valid. Red ball represents hydrogen acceptor, light blue ball signifies hydrogen donor, blue ball denotes positive charge group, green ball represents hydrophobic group, and red circle indicates aromatic group.

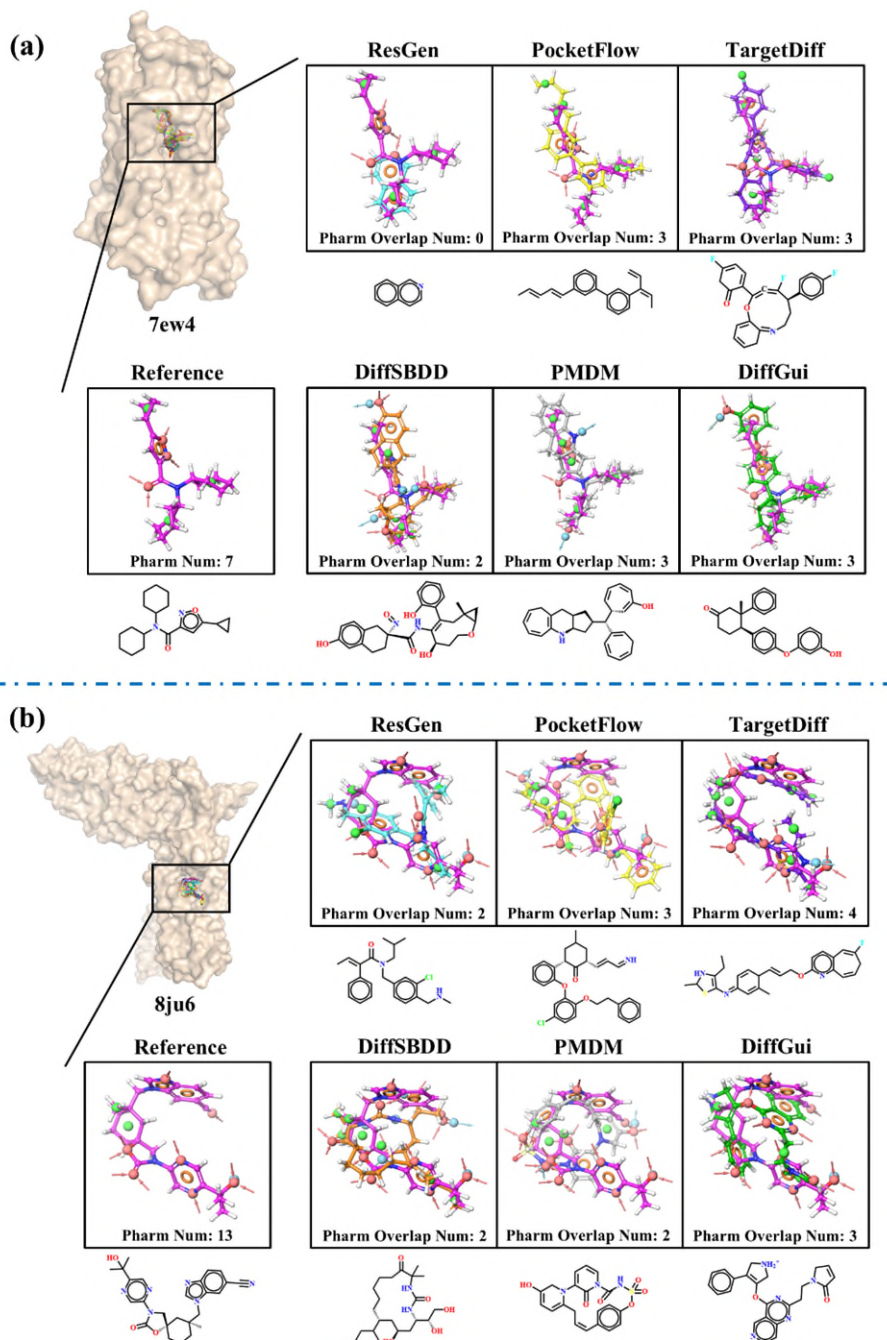


Fig S5. 3D pharmacophore overlaps of molecules generated for targets outside the PDBbind dataset. The protein targets of 7ew4 and 8ju6 are G protein-coupled receptor and ion channel, respectively. The visualized molecules have the best docking scores among those PB-valid. Red ball represents hydrogen acceptor, light blue ball signifies hydrogen donor, blue ball denotes positive charge group, green ball represents hydrophobic group, and red circle indicates aromatic group.

7. Contrarily to authors' claims I do not think that the set of 4 targets used for the de novo design section are really diverse. Three of them are protein kinases and exhibit very similar binding



pockets. A greater diversity would be preferable by adding for example a G protein-coupled receptor, a nuclear hormone receptor, an ion channel or a transporter.

Response: Thank you for raising this important question. To address your concerns, except for beta-secretase (1w51) and kinase (3ctj), we generated 100 molecules per method for each target of S1PR3 (Sphingosine 1-Phosphate Receptor 3, a G protein-coupled receptor, PDBid 7ew4) and TRPV4 (Transient Receptor Potential Cation Channel Subfamily V Member 4, an ion channel, PDBid 8ju6). The molecules with best docking scores among those PB-valid are visualized. As displayed in Figures 5, 6, and Table 7, the ligands produced by our method manifest the best docking scores, the lowest binding free energies, and favorable molecular properties, thereby confirming the superiority of DiffGui.

Moreover, 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 Figure S6. These targets 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.

The associated figures and descriptions have been incorporated into the revised manuscript (lines 367-368, 371-372, 374-377, 379-385 on page 14).

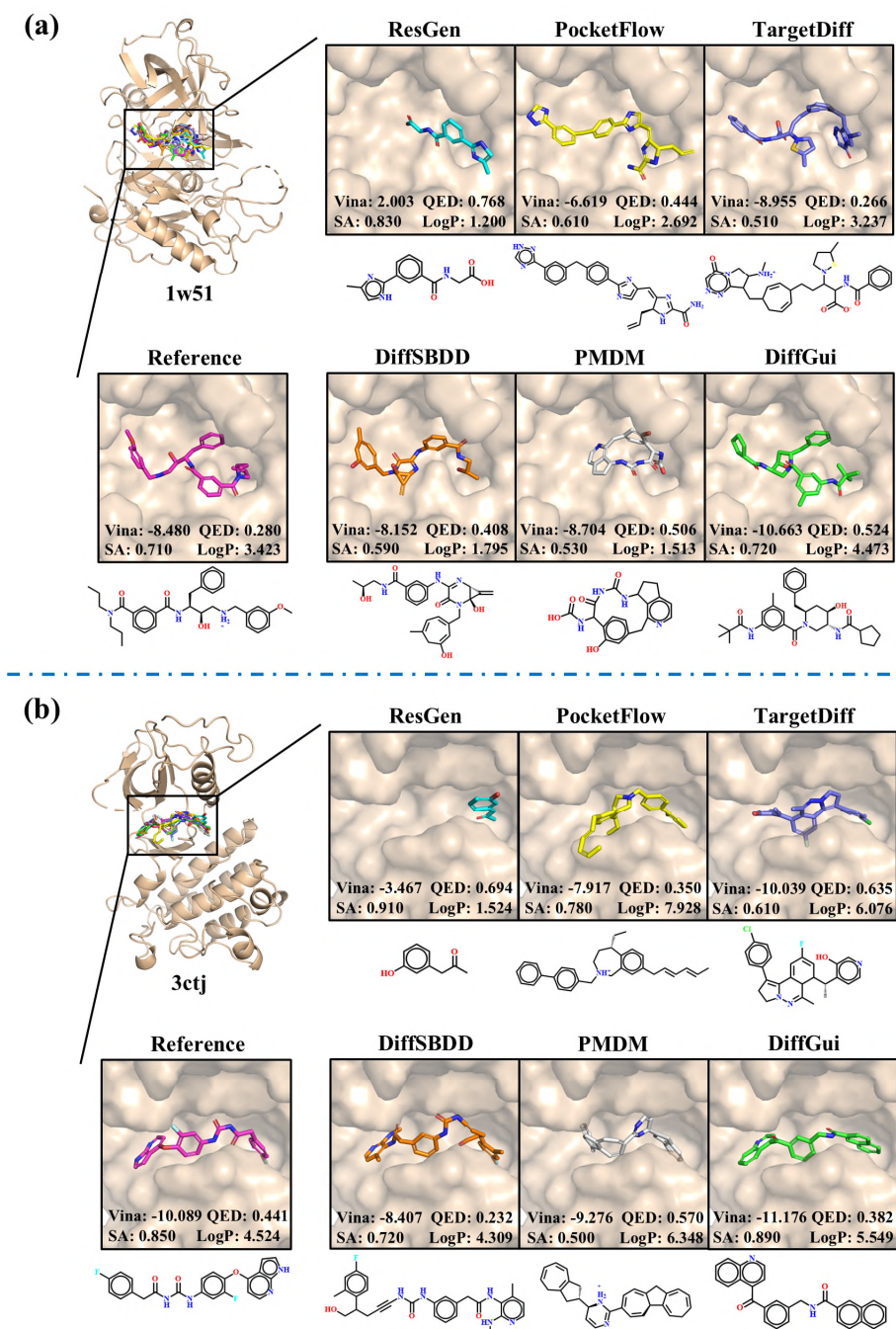


Fig 5. Molecules generated for targets from the PDBbind test set. The protein targets of 1w51 and 3ctj are beta-secretase 1 and tyrosine kinase, respectively. The visualized molecules have the best docking scores among those PB-valid.

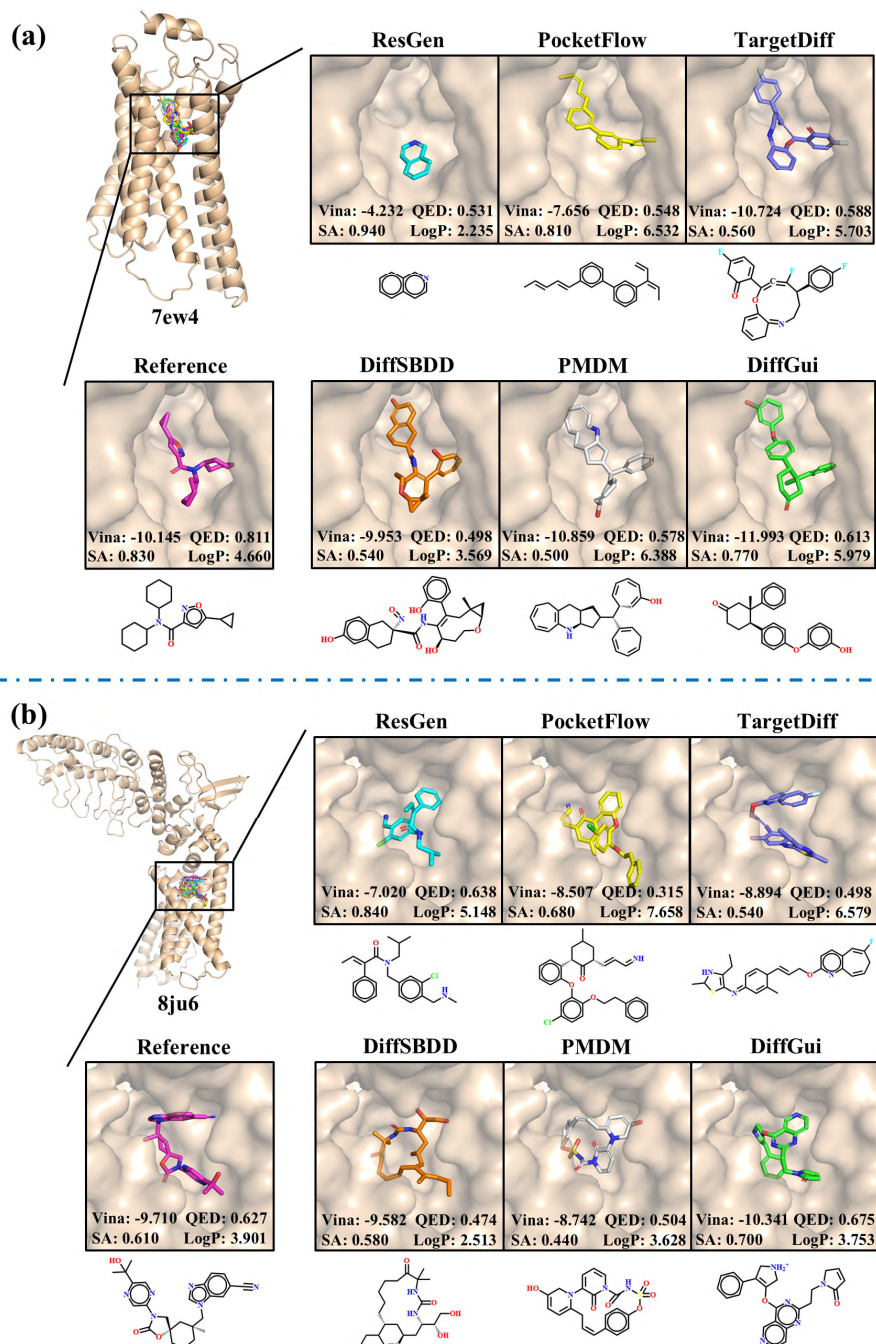


Fig 6. Molecules generated for targets outside the PDBbind dataset. The protein targets of 7ew4 and 8ju6 are G protein-coupled receptor and ion channel, respectively. The visualized molecules have the best docking scores among those PB-valid.

8. *Figures 5 and 6 are difficult to read mainly because the plots are too small. Moreover, the legends should clearly precise the reason for the selection of these ligands. Do they correspond to the solutions with best docking scores?*

Response: Thank you for your valuable suggestion aimed at enhancing the readability of our work.



We have divided the original Figure 5 into two separate figures (Figures 5 and 6), as shown above. Meanwhile, we added the protein-ligand complex at the top left corner of each subfigure to provide a comprehensive view of the complex. The figure legends are revised to illustrate the reason for the selection of these visualized ligands. They have the best docking scores among those PB-valid. The original Figure 6 (now Figure 7) has been enlarged, and the text within this figure has been bolded for enhanced clarity.

The associated figures have been incorporated into the revised manuscript.

References

1. Benson, M.L., Smith, R.D., Khazanov, N.A., Dimcheff, B., Beaver, J., Dresslar, P. et al. Binding MOAD, a high-quality protein-ligand database. *Nucleic Acids Res.* **36**, D674-678 (2008).
2. Öztürk, H., Özgür, A. & Ozkirimli, E. DeepDTA: Deep drug-target binding affinity prediction. *Bioinformatics* **34**, i821-i829 (2018).
3. Nguyen, T., Le, H., Quinn, T.P., Nguyen, T., Le, T.D. & Venkatesh, S. GraphDTA: Predicting drug-target binding affinity with graph neural networks. *Bioinformatics* **37**, 1140-1147 (2020).
4. Chen, L., Tan, X., Wang, D., Zhong, F., Liu, X., Yang, T. et al. TransformerCPI: Improving compound-protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. *Bioinformatics* **36**, 4406-4414 (2020).
5. Lu, W., Wu, Q., Zhang, J., Rao, J., Li, C. & Zheng, S. TankBind: Trigonometry-aware neural networks for drug-protein binding structure prediction. *bioRxiv*, 2022.2006.2006.495043 (2022).
6. Dong, J., Wang, N.-N., Yao, Z.-J., Zhang, L., Cheng, Y., Ouyang, D. et al. ADMETlab: A platform for systematic admet evaluation based on a comprehensively collected admet database. *J. Cheminform.* **10**, 29 (2018).
7. Xiong, G., Wu, Z., Yi, J., Fu, L., Yang, Z., Hsieh, C. et al. ADMETlab 2.0: An integrated online platform for accurate and comprehensive predictions of admet properties. *Nucleic Acids Res.* **49**, W5-W14 (2021).
8. Fu, L., Shi, S., Yi, J., Wang, N., He, Y., Wu, Z. et al. ADMETlab 3.0: An updated comprehensive online admet prediction platform enhanced with broader coverage, improved performance, api functionality and decision support. *Nucleic Acids Res.* **52**, W422-W431 (2024).
9. 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).
10. Hoogeboom, E., Satorras, V.c.G., Vignac, C. & Welling, M. Equivariant diffusion for molecule generation in 3D. In *Proceedings of the 39th International Conference on Machine Learning* **162**, 8867-8887 (PMLR, 2022).
11. Wu, L., Gong, C., Liu, X., Ye, M. & liu, q. Diffusion-based molecule generation with informative prior bridges. In *Advances in Neural Information Processing Systems* (NeurIPS, 2022).
12. Xu, M., Powers, A.S., Dror, R.O., Ermon, S. & Leskovec, J. Geometric latent diffusion models for 3D molecule generation. In *Proceedings of the 40th International Conference on Machine Learning* **202**, 38592-38610 (PMLR, 2023).
13. Morehead, A. & Cheng, J. Geometry-complete diffusion for 3D molecule generation and optimization. *Commun. Chem.* **7**, 150 (2024).
14. Gebauer, N., Gastegger, M. & Schütt, K. Symmetry-adapted generation of 3D point sets for the targeted discovery



- of molecules. In *Advances in Neural Information Processing Systems* **32** (NeurIPS, 2019).
15. Zhang, O., Zhang, J., Jin, J., Zhang, X., Hu, R., Shen, C. et al. ResGen is a pocket-aware 3D molecular generation model based on parallel multiscale modelling. *Nat. Mach. Intell.* **5**, 1020-1030 (2023).
 16. Jiang, Y., Zhang, G., You, J., Zhang, H., Yao, R., Xie, H. et al. PocketFlow is a data-and-knowledge-driven structure-based molecular generative model. *Nat. Mach. Intell.* **6**, 326-337 (2024).
 17. Guan, J., Qian, W.W., Peng, X., Su, Y., Peng, J. & Ma, J. 3D equivariant diffusion for target-aware molecule generation and affinity prediction. In *International Conference on Learning Representations* (ICLR, 2023).
 18. Schneuing, A., Du, Y., Harris, C., Jamasb, A., Igashov, I., Du, W. et al. Structure-based drug design with equivariant diffusion models. *Nat. Comput. Sci.* **4**, 899-909 (2024).
 19. Huang, L., Xu, T., Yu, Y., Zhao, P., Chen, X., Han, J. et al. A dual diffusion model enables 3D molecule generation and lead optimization based on target pockets. *Nat. Commun.* **15**, 2657 (2024).
 20. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **10**, 449-461 (2015).
 21. Kirkpatrick, P. & Ellis, C. Chemical space. *Nature* **432**, 823-823 (2004).
 22. Reymond, J.L. & Awale, M. Exploring chemical space for drug discovery using the chemical universe database. *ACS Chem. Neurosci.* **3**, 649-657 (2012).
 23. Lyu, J., Irwin, J.J. & Shoichet, B.K. Modeling the expansion of virtual screening libraries. *Nat. Chem. Biol.* **19**, 712-718 (2023).
 24. Sadybekov, A.V. & Katritch, V. Computational approaches streamlining drug discovery. *Nature* **616**, 673-685 (2023).



華東師範大學
EAST CHINA NORMAL UNIVERSITY

Dear Reviewers,

We are deeply grateful for your recognition on our work and the invaluable, constructive feedback you have provided to our manuscript entitled “**Target-aware 3D Molecular Generation Based on Guided Equivariant Diffusion**”, which was submitted to Nature Communications with Manuscript ID: **NCOMMS-24-67247-A**. We have thoroughly considered all of your comments and revised our manuscript in accordance with your suggestions. The modifications have been highlighted in red within the revised manuscript. We sincerely hope that this updated version is acceptable at this time.

Presented below is a detailed summary of your comments and our responses.



華東師範大學
EAST CHINA NORMAL UNIVERSITY

Comments from Reviewer #1 and Responses:

Authors have appropriately addressed reviewers' comments in the first round. It is encouraging to notice the incorporation of wet-lab validations. The synthesis and enzymatic inhibition assays for RSK4 provide experimental confirmation of the generated molecules' biological relevance. No additional questions to post.

It says that the reviewer's browser is not compatible with the code platform that Private link above directs to. The reviewer accessed the code through the GitHub link provided in the manuscript. <https://github.com/QiaoyuHu89/DiffGui>. Overall, the code is well-organized with a detailed README file. It is relatively straightforward to have the DiffGui deployed.

Response: Thank you for your positive comments!



Comments from Reviewer #2 and Responses:

The authors addressed one of my two main comments well. However, the authors did not address the following main comment satisfactorily:

"The proposed method is compared with several existing methods. But many useful methods are not considered in the comparison. Please also compare your method with the following methods: GDM, GDM-aug, EDM, Bridge, GraphLDM, GeoLDM, GCDM, and G-Schnet."

The authors did not compare their method with the methods in the comment above because they think all the methods above can only do unconditional molecule generation, which is different from the pocket-conditioned molecule generation that their method aims to do. However, at least one method above - GCDM can perform both unconditional and pocket-conditioned molecule generation. Therefore, the authors should add a comparison with GCDM to show the advance that the method in this work can make.

The code is well organized.

Response: Thank you for your important suggestion! We added a comparison with GCDM¹ in the revised manuscript. The main results are summarized below.

The JS divergences of C-C bond distance of GCDM approaches those of DiffGui, yet remain slightly higher for both PDBbind and CrossDocked datasets (Figures 2 and S1). However, GCDM performs poorly regarding the distance between all-atom pairs (Figures 3 and S2). The average JS divergences of bonds, angles, and dihedrals of GCDM are also higher than those of DiffGui (Table 1, Tables S1-S6).

For the evaluation of 3D geometry, ResGen and GCDM stand out as they reveal the two lowest median RMSD values between generated conformations and optimized/predicted conformations (Figures 4 and S3). 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 contrast, 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.

DiffGui demonstrates superior performance over GCDM in terms of basic molecular metrics such as atom stability, molecular stability, PB-validity, and RDKit-validity (Tables 3 and S8). Besides, DiffGui also exhibits lower 2D similarity and higher protein-ligand interaction similarity when compared to GCDM. As for molecular properties, DiffGui outperforms GCDM on Vina Scores and QED value (Tables 4 and S9). GCDM achieves a higher SA score than DiffGui, likely because it tends to generate smaller molecular bricks.

In brief, DiffGui achieves better performance than GCDM based on the comprehensive evaluation of all metrics. The associated figures, tables, and descriptions have been incorporated into the revised manuscript (lines 222, 225-227 on page 9; lines 262-264, 265-273 on page 10) and the supporting information file.



Comments from Reviewer #3 and Responses:

The revised manuscript by Hu et al. partially addressed some of my original concerns but I am still not satisfied with the proposed experimental validation that is very weak to me for the following reasons:

- Predictions are made to generate yet another kinase inhibitors (RSK4 inhibitors) of very weak potency (dose response curves are even incomplete to me). Since the PDBbind dataset the method is trained is already biased by a high number of kinase-inhibitor complexes, I am not convinced at all by these two weak examples. By comparison, many nanomolar kinase inhibitors have been rationally designed from scratch (e.g. Beroza et al. Nat Commun. 2022 Oct 28;13(1):6447; Eguida et al. J Med Chem. 2022 Oct 27;65(20):13771-13783).

- The two structures are not original at all since the two molecular scaffolds can be found in ChEMBL (623 compounds for the scaffold of compound 1, 55 compounds for the scaffold of compound 2), mostly for already described kinase inhibitors (CDK4, JAK2, etc.). I fear that the program is just memorizing known data but not genuinely discover novel chemotypes.

Authors should provide a much more consistent validation with respect to the target (please not a target family already widely investigated in PDBbind), number and potency of designed/synthesized/tested ligands.

Response: Thank you very much for your valuable comments! We would like to address the points raised through the following responses:

We recognized that the two previously identified RSK4 inhibitors demonstrated relatively weak activity (1.2 μM and 5.3 μM). Consequently, we generated a new series of RSK4 inhibitors by enhancing the affinity guidance and selected two ligands from top ten candidates for synthesis and biological evaluation. These two compounds (Compounds 1 and 2) exhibit improved inhibitory activity in the RSK4-HTRF assay, with IC_{50} values of approximately 215.0 nM and 111.1 nM (Figure S7a/b), respectively, highlighting their potential as lead compounds for further development. A similarity search in the ChEMBL database reveals that there are no compounds display more than 60% similarity to Compounds 1 and 2, confirming their structural novelty. Given the vast landscape of known chemical entities, we admit that it is very challenging to create completely new fragments, particularly when their size is small. Essentially, our model is designed to learn the fundamental protein-ligand interaction patterns while maintaining the ability to generate structural diversity. Although the model inevitably incorporates some scaffold features during training, its core strength lies in the intelligent assembly of molecular building blocks, generating intact novel ligands that are optimized for distinct protein binding pockets.

In addition to the kinase target, we performed lead optimization for a non-kinase target DHODH (dihydroorotate dehydrogenase). Compared to the 2,945 kinase-related entries, the PDBbind database includes only 139 dehydrogenase structures and no DHODH-specific complexes. From the structures of 4zmg and 4ls1 (PDBid), the lead optimization is conducted on the basis of the fixed fragments (highlighted in orange in Figure S7c/d). As a result, the optimized molecules (Compounds 3 and 4) exhibit enhanced potency against DHODH, with IC_{50} values decreasing from 8.02 μM to 4.27 μM 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.

In summary, we verified the effectiveness of DiffGui through wet-lab experiments on kinase (RSK4) and non-kinase (DHODH) targets. The corresponding figure, descriptions, chemical synthesis, and enzymatic activity assay have been incorporated into the revised manuscript (line 447 on page 16; lines 508-516 on page 18; lines 517-522 on page 19) and the supporting information file.

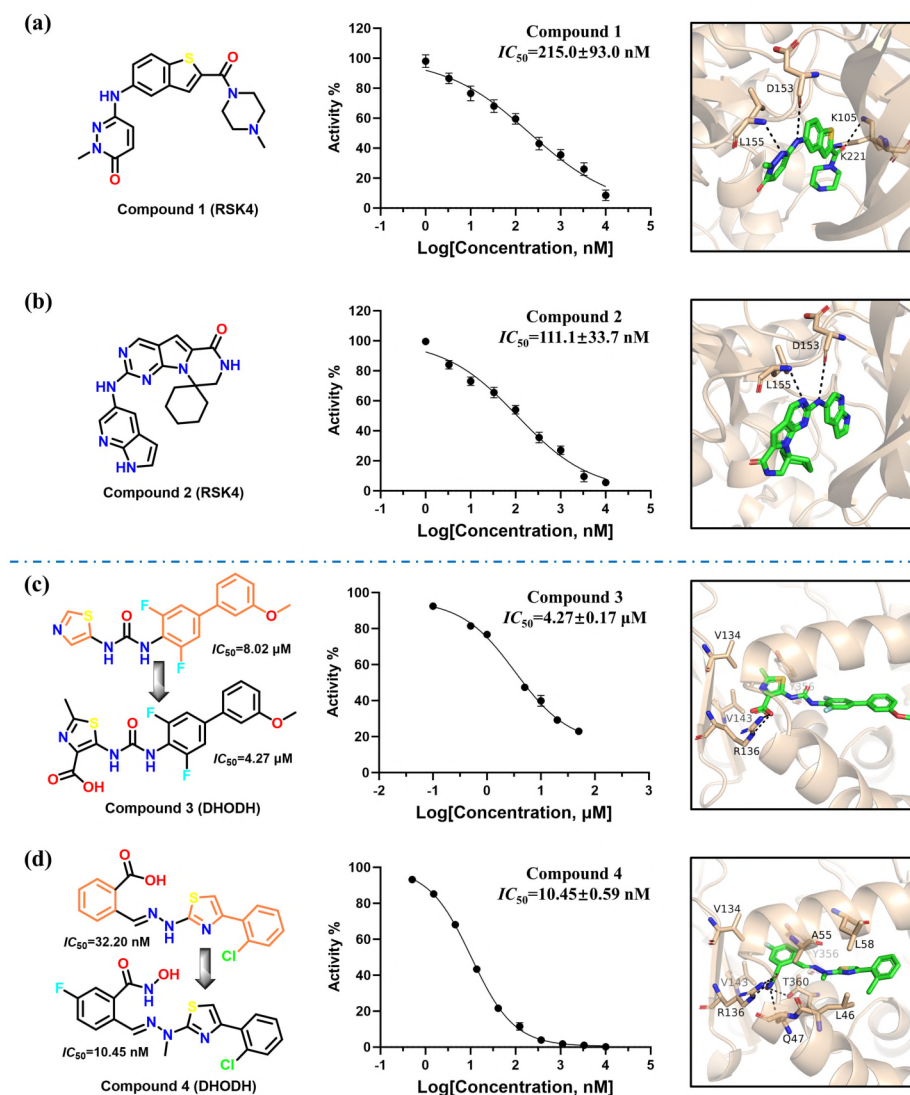


Fig S7. Wet-lab experimental validation of DiffGui. (a) Chemical structure of Compound 1, dose-activity curve of Compound 1 in the RSK4 kinase assay, and the predicted binding mode of Compound 1 with RSK4 (b) Chemical



華東師範大學
EAST CHINA NORMAL UNIVERSITY

structure of Compound 2, dose-activity curve of Compound 2 in the RSK4 kinase assay, and the predicted binding mode of Compound 1 with RSK4, **(c)** Chemical structure of Compound 3, dose-activity curve of Compound 3 in the DHODH assay, and the predicted binding mode of Compound 3 with DHODH, **(d)** Chemical structure of Compound 4, dose-activity curve of Compound 4 in the DHODH assay, and the predicted binding mode of Compound 4 with DHODH.

References

1. Morehead, A. & Cheng, J. Geometry-complete diffusion for 3D molecule generation and optimization. *Commun. Chem.* **7**, 150 (2024).