

<https://doi.org/10.1038/s42003-024-06746-w>

# Enabling target-aware molecule generation to follow multi objectives with Pareto MCTS



Yaodong Yang<sup>1</sup>, Guangyong Chen<sup>2</sup>✉, Jinpeng Li<sup>1</sup>, Junyou Li<sup>2</sup>, Odin Zhang<sup>3</sup>, Xujun Zhang<sup>3</sup>,  
Lanqing Li<sup>2</sup>, Jianye Hao<sup>4</sup>✉, Ercheng Wang<sup>2</sup>✉ & Pheng-Ann Heng<sup>1</sup>

Target-aware drug discovery has greatly accelerated the drug discovery process to design small-molecule ligands with high binding affinity to disease-related protein targets. Conditioned on targeted proteins, previous works utilize various kinds of deep generative models and have shown great potential in generating molecules with strong protein-ligand binding interactions. However, beyond binding affinity, effective drug molecules must manifest other essential properties such as high drug-likeness, which are not explicitly addressed by current target-aware generative methods. In this article, aiming to bridge the gap of multi-objective target-aware molecule generation in the field of deep learning-based drug discovery, we propose ParetoDrug, a Pareto Monte Carlo Tree Search (MCTS) generation algorithm. ParetoDrug searches molecules on the Pareto Front in chemical space using MCTS to enable synchronous optimization of multiple properties. Specifically, ParetoDrug utilizes pretrained atom-by-atom autoregressive generative models for the exploration guidance to desired molecules during MCTS searching. Besides, when selecting the next atom symbol, a scheme named ParetoPUCT is proposed to balance exploration and exploitation. Benchmark experiments and case studies demonstrate that ParetoDrug is highly effective in traversing the large and complex chemical space to discover novel compounds with satisfactory binding affinities and drug-like properties for various multi-objective target-aware drug discovery tasks.

The rational design of molecules to act as clinical drugs remains a significant challenge in biopharmaceutical research, especially concerning the attainment of favorable physiochemical and pharmacological properties. In support of such endeavors, target-based drug discovery aims to identify small-molecule ligands that exhibit high affinity and specificity for a particular protein pocket structure<sup>1</sup>. Traditionally, target-based drug discovery has been approached through either high-throughput experimental methods or virtual screening of extensive chemical databases<sup>2,3</sup> targeted at specific biomolecular targets<sup>4,5</sup>. Subsequently, the screening of bioanalytical indicators through elaborate clinical experiments is conducted to evaluate drug-like properties. This pursuit contributes to the conventional 10-year drug development cycle and staggering research and development costs of approximately 2.8 billion USD, coupled with a remarkably high failure rate. The predetermined selection of compounds for screening further constrains the exploration of chemical space, tethering it to historical knowledge derived from previously investigated molecules. This ultimately leads to a fervent industry focus on popular drug targets, resulting in the challenge that

the molecules selected through screening are unable to avoid patent restrictions. In contrast, recent advancements in target-aware molecule generation, particularly the development of generative models trained on extensive datasets, present a promising paradigm shift. These models, rooted in deep learning, offer an innovative approach to expedite ligand discovery and optimization. They achieve this by generating entirely novel and diverse molecules capable of binding to a specified protein target, starting from scratch<sup>6</sup>. This transformative approach holds great potential to overcome the limitations associated with traditional methods, offering a more efficient and expansive exploration of the entire chemical space.

Since the first inception of an autoencoder model conditioned on targeted proteins in 2018<sup>7</sup>, there has been rapid progress in deep learning-based target-aware molecule generation methods. Various works take advantage of conditional generative models, such as the autoencoder<sup>7–9</sup>, generative adversarial network<sup>10</sup>, and diffusion model<sup>11,12</sup>, to infer entire molecules through a one-time feedforward process, incorporating binding site information as input. Moreover, to enhance structural representation,

<sup>1</sup>Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China. <sup>2</sup>Zhejiang Lab, Hangzhou, China. <sup>3</sup>Zhejiang University, Hangzhou, China. <sup>4</sup>Noah's Ark Lab, Huawei, Shenzhen, China. ✉e-mail: [gychen@zhejianglab.com](mailto:gychen@zhejianglab.com); [jianye.hao@tju.edu.cn](mailto:jianye.hao@tju.edu.cn); [wangec@zju.edu.cn](mailto:wangec@zju.edu.cn)

convolutional neural networks<sup>7</sup> and graph convolutional networks<sup>8</sup> are employed. In the meantime, some approaches utilize voxelized representations<sup>10</sup> or atomic density grids<sup>13</sup> to characterize compound-receptor complexes. Another pivotal category of deep learning-based target-aware drug discovery involves autoregressive generative models, which predict the next atom (and its position) sequentially conditioned on the molecular fragment and binding site information. To model the conditioned intermediate context, diverse network architectures like transformers<sup>14,15</sup>, recurrent neural networks<sup>16,17</sup>, and flow models<sup>18</sup> are introduced as the context encoder. Additionally, graph neural networks<sup>18–20</sup> are widely utilized to extract chemical and geometrical features of ligands and protein pockets. However, these efforts are not yet integrated into mainstream drug discovery practices, and a significant obstacle lies in the inherent multi-objective optimization nature of drug discovery<sup>21</sup>. Beyond strong binding affinity to the targeted protein, drug molecules must exhibit other desirable properties, such as high drug-likeness and low toxicity. Presently, existing deep learning-based target-aware generative methods predominantly focus on the single objective of optimizing binding affinity. The multi-objective nature of drug molecules, with sometimes conflicting demands, necessitates ongoing development of novel multi-objective target-aware drug discovery techniques to enhance the overall success rates of drug discovery.

Conversely, numerous studies have explored the domain of general multi-objective drug discovery. Certain approaches, such as MolGPT<sup>22</sup>, fall within the ligand-based methodology, aspiring to generate novel compounds with favorable physicochemical properties. However, these methods fall short in incorporating protein information, thus lacking assurance that the generated molecules can effectively bind to specified protein targets. Concurrently, other methodologies like MCMG<sup>21</sup>, RationaleRL<sup>23</sup>, MolSearch<sup>24</sup>, and GENERA<sup>25</sup> aim to optimize not only the binding affinity objective but also other property objectives. Specifically, these methodologies leverage optimization techniques such as reinforcement learning<sup>26</sup> and genetic algorithms<sup>27</sup> to enhance the binding affinity objective predicted by machine learning-based or simulation-based docking score functions. However, a notable drawback is their failure to explicitly incorporate target protein information when constructing generative models. The absence of protein information renders the optimization of the binding affinity objective inefficient, and the resulting generative models from these target-scoring-based methods cannot be readily generalized to other protein targets. In contrast to ligand-based and target-scoring-based approaches, a recent development is CProMG<sup>28</sup>, designed to generate molecules that meet multiple property constraints with an enhanced representation of protein structure information. CProMG treats this task as a multi-constraint molecule generation problem, with each property constraint set to exceed a predefined threshold. However, CProMG does not attempt to maximize molecule properties through optimization techniques for a comprehensive exploration of the chemical space. A more in-depth discussion is provided in the Discussion section.

Similar challenges also exist in natural language generation tasks, where models predicting the next token often express unintended behaviors, such as making up facts, generating biased or toxic text, or not following user instructions. To address this issue, OpenAI focuses on fine-tuning approaches to align language models. Specifically, they employ reinforcement learning from human feedback (RLHF) to fine-tune GPT-3<sup>29</sup> to follow a broad class of written instructions<sup>30</sup>. In contrast to the fuzzy, hard-to-quantify human values in natural language tasks, we can explicitly calculate multiple molecular metrics in the context of drug development.

In this study, we explore the use of an autoregressive Pareto Monte Carlo Tree Search (MCTS) generation algorithm named ParetoDrug for the design of drug molecules to address the existing gap in multi-objective target-aware drug discovery within the domain of deep learning-based drug discovery. This algorithm effectively facilitates the simultaneous optimization of multiple molecule properties. In its operation, ParetoDrug first explores molecules on the Pareto Front within the chemical space. It

achieves this by maintaining a global pool comprising Pareto optimal molecules, each of which is not surpassed by another molecule in the same pool across every property objective. During the exploration process, ParetoDrug leverages existing pretrained autoregressive target-aware molecule generation models to guide the search for the next atom symbol, facilitating the identification of molecules with high binding affinity to protein targets. Additionally, in the selection of the next atom symbol, ParetoDrug introduces a scheme named ParetoPUCT. This scheme is designed to balance the exploration of chemical space and the exploitation of the pretrained autoregressive generative model. Through these strategies, ParetoDrug owns the ability to generate molecules with multiple desirable properties, including binding affinity. Computational evaluations on the benchmark dataset and case studies, including multi-objective target-aware drug discovery tasks for known drugs (e.g., Tropifexor and Copanlisib), a multi-target drug discovery task for HIV-related disease targets, and a multi-target multi-objective drug discovery task for a dual-inhibitor Lapatinib, demonstrate the high effectiveness of ParetoDrug. The algorithm exhibits proficiency in discovering small-molecule drug candidates possessing multiple required properties, particularly including binding affinities to specified protein targets.

## Results

In this section, we first conduct the experiments on a benchmark to demonstrate ParetoDrug's remarkable ability to generate molecules with multiple desired properties including the binding affinity and drug-like properties when compared with various baselines. Meanwhile, we also give the statistical analysis of the generated molecules of ParetoDrug. Then we use ParetoDrug to perform the case studies for the multi-objective target-aware drug discovery task, multi-target drug discovery task, and multi-target multi-objective drug discovery task respectively. In these case studies, ParetoDrug is able to generate the Pareto Dominate molecules over the known drug ligands in terms of the specified molecule property objectives, which exhibits the promising molecule discovery potential of ParetoDrug.

### Benchmark experiments

In the benchmark experiments, we follow the settings as Qian et al.<sup>15</sup> where there are 100 protein targets sampled from the public database of protein-ligand pairs BindingDB<sup>31</sup> as the test set. For each test protein target, we generate 10 candidate molecules for evaluation. All 1000 candidate molecules are evaluated by a set of molecule property metrics, and the scores are averaged for an overall comparison. Please refer to Supplementary Information A and B for a detailed experimental and hyperparameter setup. We use several important metrics to evaluate the generated molecules, including docking score, uniqueness, LogP, QED, SA score, and NP-likeness described as follows.

- **Docking score.** Binding energy is regarded as a general indicator to describe the binding affinity between molecule ligands and target proteins. Specifically, we utilize a free and widely used tool called smina<sup>32</sup> to compute the binding affinity. We use the negative value of the output by smina as the docking score. The higher the docking score is, the better the molecule is docked into the target protein.
- **Uniqueness.** Drug design models should be able to generate different molecules conditioning on different target proteins. The higher the uniqueness value is, the more sensitive the model is to the specified target protein. This metric is computed as follows:

$$\text{Uniqueness}(\%) = \frac{\#(\text{Set}(\cup_{S_p \in \mathbb{S}_p} \text{Set}(M_{S_p})))}{\#(\cup_{S_p \in \mathbb{S}_p} \text{Set}(M_{S_p}))} \times 100\%, \quad (1)$$

where  $\mathbb{S}_p$  indicates the set of test proteins,  $M_{S_p}$  denotes the collection of generated molecules for the target protein  $S_p \in \mathbb{S}_p$ ,  $\#$  counts the number of molecules, and Set is an operator to remove the repeated molecules in the given set.

**Table 1 | Average metric scores of generated molecules of each method ( $n = 1000$  molecules) on the sampled 100 test proteins**

Type	Methods	Docking score (†)	Uniqueness (†)	LogP	QED (†)	SA (†)	NP (†)
Reference	Known ligands	9.8	–	2.2	0.5	3.3	–1.0
Generative model	LiGANN <sup>10</sup>	6.7	94.7%	2.9	0.6	3.0	–1.1
	SBMolGen <sup>37</sup>	7.7	100%	2.6	0.7	2.8	–1.2
	SBDD-3D <sup>18</sup>	7.7	99.3%	1.5	0.6	4.0	0.3
	Pocket2Mol <sup>20</sup>	8.1	99.8%	2.0	0.7	3.0	–0.2
	CProMG <sup>28</sup>	8.2	26.9%	1.9	0.8	2.9	–0.9
	BeamLmsr <sup>15</sup>	8.5	98.1%	4.0	0.5	2.7	–1.0
	TargetDiff <sup>12</sup>	8.6	100%	2.9	0.5	5.2	0.6
RL/MCTS	REINVENT 4 <sup>39</sup>	9.3 ± 0.1	98.7%	3.3 ± 0.0	0.8 ± 0.0	2.7 ± 0.0	0.7 ± 0.0
	AlphaDrug <sup>15</sup>	12.0 ± 0.1	99.9%	5.4 ± 0.1	0.4 ± 0.0	2.7 ± 0.0	–0.9 ± 0.0
	ParetoDrug	10.9 ± 0.1	99.9%	4.3 ± 0.1	0.6 ± 0.0	2.4 ± 0.0	–0.4 ± 0.0

The unit of docking score is kcal · mol<sup>–1</sup>.

The 95% confidence intervals for the property mean values of RL/MCTS methods are given. We conducted the two-sided *T* test to compare ParetoDrug and AlphaDrug, and the *p* values for Docking score, LogP, QED, SA, and NP are 3.4e–52, 1.1e–58, 8.0e–128, 3.6e–20, and 2.0e–62, respectively. We conducted the two-sided *T* test to compare ParetoDrug and REINVENT 4, and the *p* values for Docking score, LogP, QED, SA, and NP are 1.3e–110, 1.4e–104, 9.3e–265, 2.2e–21, and 8.5e–217, respectively.

- **LogP.** A large LogP value indicates the substance is lipophilic, while a small LogP value means it is easy to dissolve in water. According to Ghose filter<sup>33</sup>, the LogP value of a druggable molecule should range from –0.4 to +5.6.
- **QED.** This score measures the drug-likeness and ranges from 0 to 1. A higher QED score indicates that a molecule is more likely to be a potential drug-like compound, with the desired molecular properties such as hydrogen bond acceptor, hydrogen bond donor, and polar molecular surface area<sup>34</sup>.
- **SA score.** The synthetic accessibility (SA) score indicates how difficult one molecule is to synthesize, which is calculated based on a combination of fragment contributions and a complexity penalty<sup>35</sup>. The range of the estimated SA metric is from 1 (easy to make) to 10 (very difficult to make).
- **NP-likeness.** Natural products play an important role in the history of drug discovery. Many drugs are natural products and their derivatives. The higher the score is, the more likely the molecule is to be a natural product. The calculated NP-likeness is typically in the range from –5 to 5<sup>36</sup>.

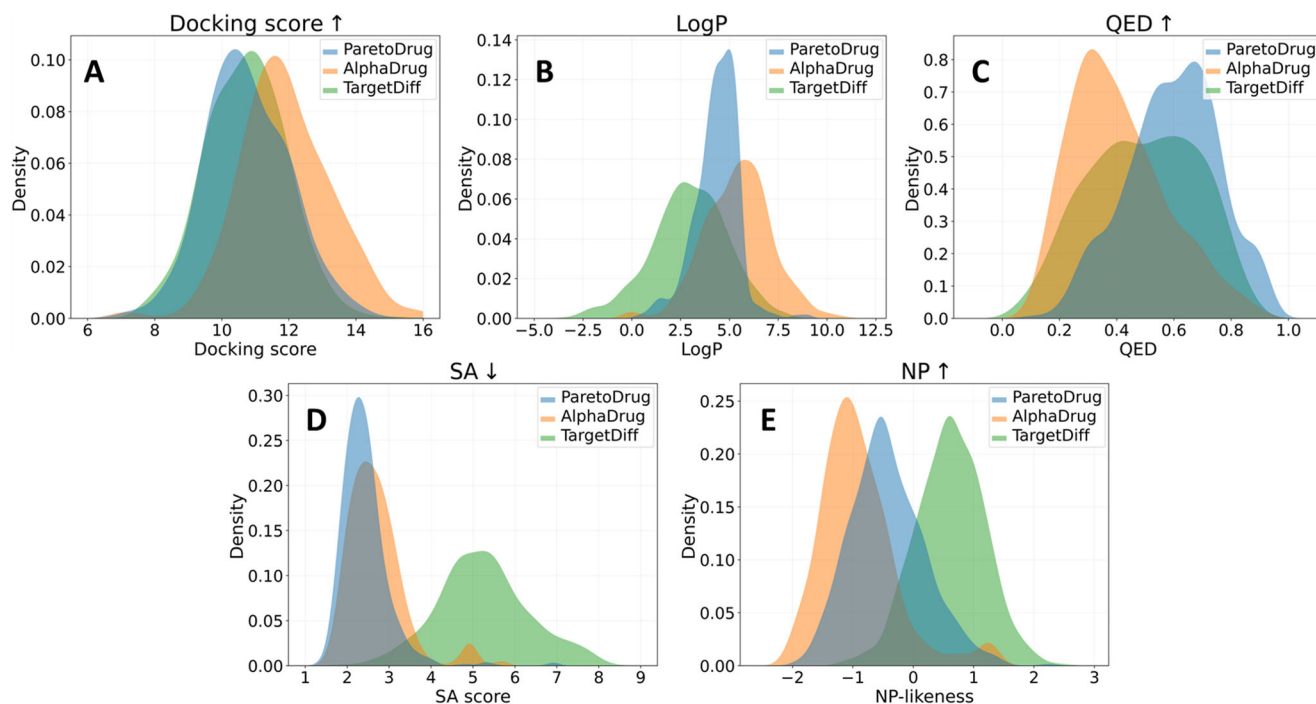
The reported results of “Known ligands”, SBMolGen, LiGANN, SBDD-3D, and BeamLmsr are from AlphaDrug<sup>15</sup>. The “Known ligands” indicates the original molecules binding to protein targets in the database. The results of LiGANN<sup>10</sup> were collected on the web-based application provided in the original paper. SBMolGen<sup>37</sup> is developed from ChemTS<sup>38</sup> for target-specific molecular generation. The results of SBDD-3D<sup>18</sup> were based on the released codes and trained model published by the authors. BeamLmsr applies the beam search on the pretrained Lmsr Transformer<sup>15</sup>. The beam size of BeamLmsr is set at 10 to collect 10 molecules for each test protein target. Besides the above representative baselines, we also test three recent advanced methods. The first is Pocket2Mol<sup>20</sup>, which uses the equivariant generative network and autoregressive sampling scheme to generate three-dimensional molecules. For Pocket2Mol, we utilize the official codes and trained model for sampling molecules. The second is TargetDiff<sup>12</sup>, which develops a three-dimensional equivariant diffusion model to sample molecules. For TargetDiff, we also use the officially released trained model and codes for sampling. We keep the sample numbers of Pocket2Mol and TargetDiff at 100 for each test protein, which is the default configuration to ensure the quality of generated molecules. To make a fair comparison with other methods, for each test protein target, we randomly select 10 molecules from the generated 100 molecules of Pocket2Mol and TargetDiff for the evaluation. The third is CProMG<sup>28</sup>, which proposes a multi-constraint autoregressive model to

generate small molecules with controllable properties. We use the official codes and default configurations of CProMG to generate 10 molecules for each test protein with the pretrained CProMG-VQSLT model, which is trained to control multiple property metrics including the docking score, LogP, QED, and SA score that are evaluated here.

Besides the above basic generative models, there also emerges another kind of fundamental approach that integrates the powerful MCTS-based searching technique to better control the molecule generation procedure of the pretrained autoregressive generative models with the simulation feedback, and AlphaDrug and the proposed ParetoDrug fall into this kind. For AlphaDrug<sup>15</sup> which utilizes MCTS with the pretrained Lmsr Transformer model to generate molecules based on given protein targets, we run the official codes and set iteration times (IT) at 150 when selecting the next atom symbol in MCTS. For ParetoDrug which conducts Pareto MCTS with the same pretrained Lmsr Transformer model, we also set IT at 150 and let it optimize all objectives (docking score, LogP, QED, SA score, and NP-likeness) synchronously except the unoptimizable Uniqueness, which is a statistic metric for all generated molecules. In addition, we set the metric value of LogP as 1 if the molecule’s LogP value is in the range of [–0.4, 5.6], and 0 otherwise. After each Pareto MCTS, ParetoDrug obtains a global pool of Pareto optimal molecules. We choose the molecule with the largest reward vector summation value from the pool, which means this molecule has top rankings in each property metric. When testing, we collect 10 generated molecules for each test protein target from AlphaDrug and ParetoDrug.

Additionally, we compare a multi-objective drug discovery algorithm REINVENT 4<sup>39</sup> while its generation model is not conditioned on the protein information. It uses a reinforcement learning algorithm to generate optimized molecules compliant with a user-defined property profile defined as a multi-component score. We let REINVENT 4 optimize the docking score, LogP, QED, SA, and NP while setting their weights in the property profile all at 0.2. For each test protein target, we collected 10 molecules with the highest multi-component scores during the training process of REINVENT 4.

The results are shown in Table 1 and the direction of the arrow in the table means a better property score. The 95% confidence intervals for property scores of RL/MCTS are included. As we see, in terms of the docking score, ParetoDrug demonstrates superiority over all baselines except AlphaDrug. However, AlphaDrug is a single-objective target-aware drug discovery method that only optimizes the binding affinity. As AlphaDrug and ParetoDrug have the same iteration budgets (IT=150) for each atom symbol in sequence but ParetoDrug needs to optimize multiple objectives including the binding affinity, it is expected that ParetoDrug has a lower docking score than AlphaDrug. Meanwhile, although the docking score of ParetoDrug decreases slightly, other metrics including QED, SA score, and NP-likeness are improved significantly compared with AlphaDrug.



**Fig. 1 | The molecular property distributions of generated molecules ( $n = 1000$  molecules) by ParetoDrug, AlphaDrug, and TargetDiff respectively. **A** The docking score ( $\text{kcal} \cdot \text{mol}^{-1}$ ) distributions of each method. **B** The LogP value**

distributions of each method. **C** The QED value distributions of each method. **D** The SA score distributions of each method. **E** The NP-likeness distributions of each method.

Notably, QED changes from 0.4 to 0.6 (50% improvement) while NP-likeness changes from -0.9 to -0.4 (55.6% improvement). For the special LogP metric, although the average LogP value of AlphaDrug falls into the druggable molecule range, only 52.7% generated molecules of AlphaDrug satisfy the LogP range constraint if tested individually. On the contrary, 96.5% (83.1% improvement over AlphaDrug) generated molecules of ParetoDrug satisfy the LogP range constraint. These impressive results demonstrate that ParetoDrug is able to address the multi-objective target-aware drug discovery task by discovering novel compounds that possess multiple satisfactory properties including the binding affinity. On the other hand, we observe that the pretrained autoregressive Lmsr Transformer with beam search (BeamLmsr) cannot generate molecules with higher docking scores than the most recent TargetDiff. But with MCTS replacing beam search, AlphaDrug greatly boosts Lmsr Transformer's performance to find molecules with stronger binding affinity than BeamLmsr even with the same docking time budgets<sup>15</sup>. Furthermore, ParetoDrug proposes the multi-objective Pareto MCTS to replace the MCTS used in AlphaDrug. With the same iteration times, ParetoDrug significantly improves multiple molecule properties compared with AlphaDrug while maintaining the docking score at the same level. Additionally, when compared with the multi-constraint conditional generation method CProMG, ParetoDrug has advantages in docking score, Uniqueness, SA score, and NP-likeness. In addition, the Uniqueness of CProMG is only 26.9% as it generates the same molecules for different protein targets, which is undesirable in de novo target-aware drug discovery tasks. Lastly, for REINVENT 4 which does not belong to the kind of target-aware drug discovery methods, we could see although it achieves superior performance in some metrics such as QED and NP, its docking score is much lower than ParetoDrug as it does not encode the protein-ligand prior to its generation model. This also indicates the importance of incorporating the protein target information into the molecule generation process as in the generative target-aware drug discovery methods.

Next, we conduct the statistical analysis with kernel density estimate<sup>40</sup>, which is analogous to a histogram but endowed with benefits such as smoothness and continuity. The property distributions of molecules

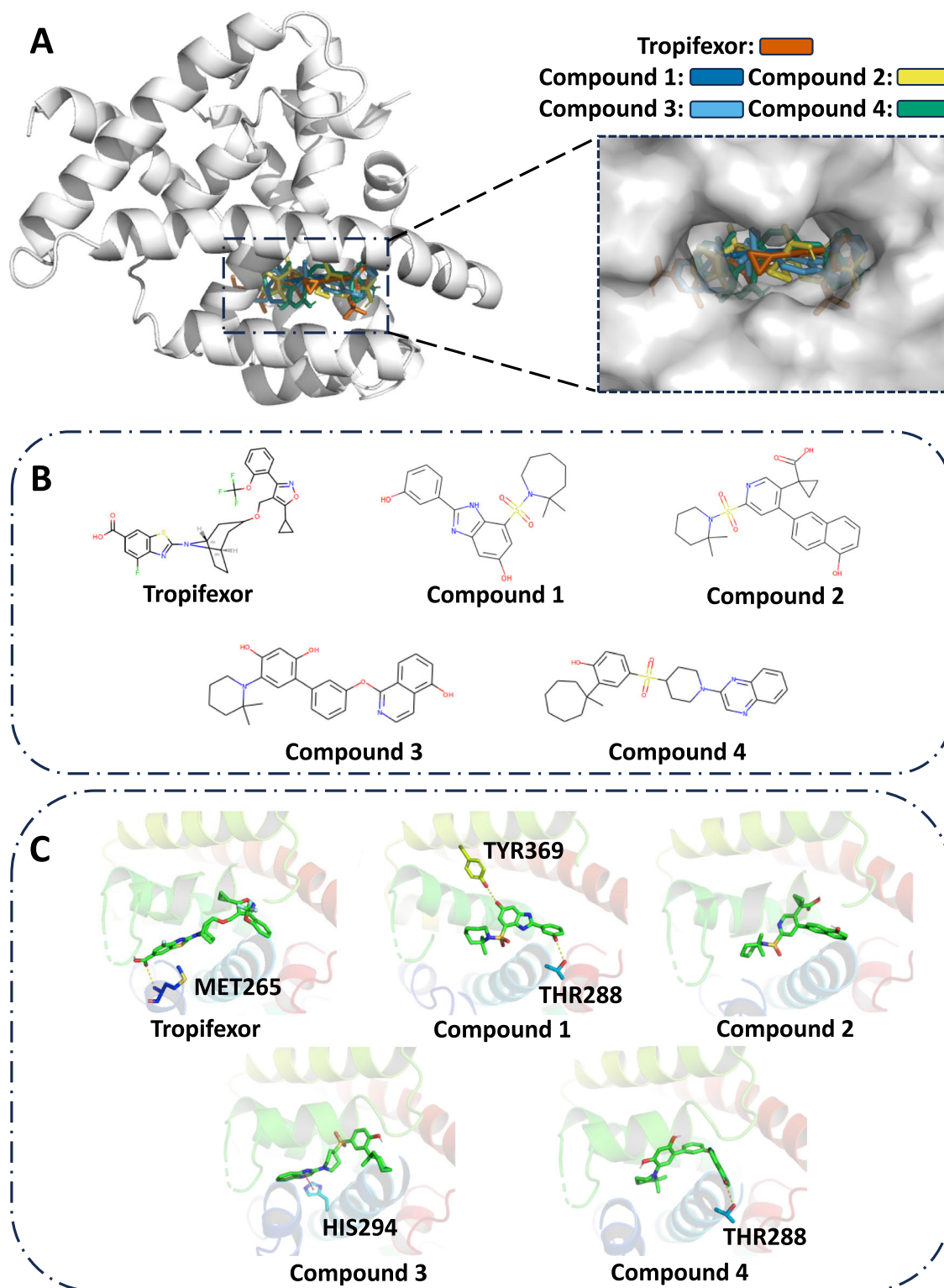
generated by TargetDiff, AlphaDrug, and ParetoDrug are shown in Fig. 1. For TargetDiff, here we use 10 molecules with the highest docking scores among the generated 100 molecules for each test protein to make an aligned comparison. We can see that although the docking score distributions of the three methods are similar while AlphaDrug is slightly better, other property distributions present differently. For LogP, ParetoDrug satisfies the range constraint of  $[-0.4, 5.6]$  while TargetDiff tends to generate more molecules with LogP values below the lower bound and AlphaDrug tends to generate more molecules with LogP values above the upper bound. Besides, ParetoDrug is able to generate more molecules with high QED values than the other two methods especially when QED is larger than 0.8 that molecules are very likely to be potential initiators of a drug candidate. Meanwhile, TargetDiff's molecules are with significantly higher SA values than ParetoDrug, which means that TargetDiff's molecules are much harder to synthesize. These statistical findings demonstrate that ParetoDrug has better molecule distributions than AlphaDrug and TargetDiff when taking multiple properties into account. More comparisons of computational efficacy, score distributions of a specific target, and the diversity of generated molecules between ParetoDrug and other methods could be referred to Supplementary Information C (and Supplementary Table 1), D (and Supplementary Fig. 1), and E (and Supplementary Table 2).

### Case studies for multi-objective target-aware drug discovery

Here we use two case studies of disease protein targets to show the molecule discovery ability of ParetoDrug for the multi-objective target-aware drug discovery tasks. The molecule objectives optimized by ParetoDrug are the docking score, LogP, QED, SA score, and NP-likeness. Additionally, the binding affinity is further validated by MM-GBSA<sup>41,42</sup>, which is a more accurate metric than docking scores but computationally expensive. For the analysis of the protein-ligand interactions, we use PLIP<sup>43</sup> and detailed can be referred to Supplementary Information F.

**Case 1: targeting FXR.** Non-alcoholic fatty liver disease (NAFLD) is defined as the excessive and abnormal intracellular accumulation of lipids in the liver, primarily in the form of triglycerides<sup>44,45</sup>. Currently,





**Fig. 2 | Static structural analysis of ligands binding to FXR (PDB ID: 7D42).** Tropifexor is an agonist of FXR and Compounds 1–4 are found by ParetoDrug. Hydrogen bonds are displayed in yellow dashed lines and  $\pi$ - $\pi$  interactions are in red.

**A** Solvent-accessible surfaces of the binding pocket of FXR for Tropifexor and Compounds 1 to 4. **B** Chemical structures of Tropifexor and Compounds 1–4. **C** The binding poses of Tropifexor and Compounds 1–4.

NAFLD has been the most common cause of chronic liver disease, especially in Western countries, and the estimated prevalence of NAFLD is approximately 30% in the general population<sup>46,47</sup>. One of the best-known drugs for NAFLD is Tropifexor, which acts as an agonist of the farnesoid X receptor (FXR). The structural basis of Tropifexor as a potent and selective agonist of FXR is shown in Fig. 2A (PDB ID: 7D42)<sup>48</sup>. In this

case study, we use ParetoDrug to discover potential drug molecules with desired computational properties for FXR. Using ParetoDrug, we collect 10 molecules and find four Pareto Dominate molecules compared with Tropifexor. The chemical structures of Tropifexor and the discovered ligands by ParetoDrug for FXR are shown in Fig. 2B. Table 2 shows the property metrics of different ligands. Our ParetoDrug model discovers

**Table 2 | Metrics of generated molecules for protein target FXR**

Ligands	Docking score (†)	LogP	QED (†)	SA (‡)	NP (†)	MM-GBSA score (‡)
Tropifexor	12.3	7.3	0.21	4.6	-0.91	-57.80 ± 6.69
Compound 1	12.5	4.0	0.60	2.9	-0.37	-55.11 ± 5.61
Compound 2	12.3	5.4	0.51	2.8	-0.75	-60.30 ± 4.27
Compound 3	12.5	6.6	0.32	2.8	-0.04	-62.09 ± 4.72
Compound 4	13.0	4.7	0.55	3.1	-0.02	-64.82 ± 4.64

The unit of docking score and MM-GBSA score ( $n = 3$  simulations for MM-GBSA) is kcal · mol<sup>-1</sup>. The 95% confidence intervals of MM-GBSA scores are included. We conducted the two-sided *T* test to compare the MM-GBSA scores of each compound with Tropifexor. The *p* value is 0.5784 for Compound 1, 0.5713 for Compound 2, 0.3622 for Compound 3, and 0.1662 for Compound 4. These *p* values (>0.05) indicate Compound 1–4 have the same level of binding free energies as the known drug Tropifexor.

multiple ligands that outperform Tropifexor on all the optimized properties. Especially, the SA scores of the new ligands are much lower than Tropifexor, which means that they are easier to synthesize. We also run AlphaDrug and TargetDiff to collect molecules for FXR, however, no Pareto Dominate molecule over Tropifexor is found for the two methods. For example, the best molecule from AlphaDrug (with the most number of better properties than Tropifexor) has the Docking score at 12.9, LogP at 4.1, QED at 0.3, SA at 2.6, and NP at -1.59. Compared with it, Compound 4 generated by ParetoDrug has 3 better properties (Docking score, QED, and NP) and 1 worse property (SA). Meanwhile, the best molecule from TargetDiff (with the most number of better properties than Tropifexor) has the Docking score at 12.8, LogP at 4.9, QED at 0.48, SA at 7.7, and NP at 1.01. Compared with it, Compound 4 generated by ParetoDrug has 3 better properties (Docking score, QED, and SA) and 1 worse property (NP). As shown in Fig. 2C, the docked poses and interactions of these four discovered compounds are quite different compared with Tropifexor. More specifically, one hydrogen bond forms between Tropifexor and the amino-acid residue MET265. At the same time, Compounds 1, 3, and 4 with new scaffolds form new hydrogen bonds with other residues (Compound 1 with THR288 and TYR369, Compound 3 with HIS294, and Compound 4 with THR288) while no hydrogen bond forms between Compound 2 and FXR.

Besides the docking score, MM-GBSA rescoring based on molecular dynamics simulations is used to further computationally validate the discovered compounds<sup>41</sup>. MM-GBSA uses molecular mechanics with generalized Born surface area to determine highly potential inhibitors for targets. The detailed settings of MM-GBSA are provided in Supplementary Information G. In this case, we use MM-GBSA to further validate the generated Pareto optimal molecules with promising docking scores. As shown in Table 2, the MM-GBSA scores indicate that the generated molecules have the same level of binding free energies as the known drug Tropifexor. Surprisingly, although two hydrogen bonds formed between Compound 1 and FXR, the MM-GBSA scores show no significant difference. Possibly because hydrogen bonding interaction is ignored in the MM-GBSA calculation.

**Case 2: targeting PI3K- $\gamma$ .** Follicular lymphoma (FL) is a systemic neoplasm of the lymphoid tissue displaying germinal center B-cell differentiation, which belongs to a cancer that involves certain types of white blood cells known as lymphocytes. FL represents 5% of all hematological neoplasms and about 20–25% of all new non-Hodgkin lymphoma diagnoses in Western countries<sup>49</sup>. One of the best-known drugs for FL is Copanlisib, which has been shown to affect the survival and spread of cancerous B-cells. The structural basis of the PI3K- $\gamma$  related to FL in complex with Copanlisib is shown in Fig. 3A (PDB ID: 5G2N)<sup>50</sup>. Here we use ParetoDrug to discover potential drug molecules with desired computational properties for PI3K- $\gamma$ . We collect 10 molecules and find one Pareto Dominate molecule (Compound 5) compared with Copanlisib (Fig. 3B). As shown in Fig. 3C, Compound 5 found by ParetoDrug has

three hydrogen bonds with surrounding residues (VAL882, ASP836, and LYS833) and two  $\pi$ - $\pi$  stackings with surrounding residues (TRP812 and TYR867). Notably, the hydrogen bonds to VAL882 and LYS833 as well as  $\pi$ - $\pi$  stacking to TYR867 also appear in Copanlisib's docking interactions. Meanwhile, Table 3 shows the computational metric values of Copanlisib and Compound 5. Compound 5 is better than Copanlisib in terms of the optimized molecule metric objectives. However, the MM-GBSA scores indicate that the binding strength of Compound 5 decreases compared with Copanlisib (-46.48 kcal · mol<sup>-1</sup> vs. -55.51 kcal · mol<sup>-1</sup>). The possible reason is that hydrogen bonding and  $\pi$ - $\pi$  interactions are not considered in the energy terms of MM-GBSA. We also run AlphaDrug and TargetDiff to collect molecules for PI3K- $\gamma$  and no Pareto Dominate molecule over Copanlisib is found by the two methods. For example, the best molecule from AlphaDrug (with the most number of better properties than Copanlisib) has the Docking score at 12.4, LogP at 4.35, QED at 0.41, SA at 5.0, and NP at 1.29. Meanwhile, the best molecule from TargetDiff (with the most number of better properties than Copanlisib) has the Docking score at 11.6, LogP at 3.2, QED at 0.56, SA at 3.9, and NP at 0.18. Although with some good properties, the two top molecules from AlphaDrug and TargetDiff cannot dominate the drug Copanlisib with worse SA.

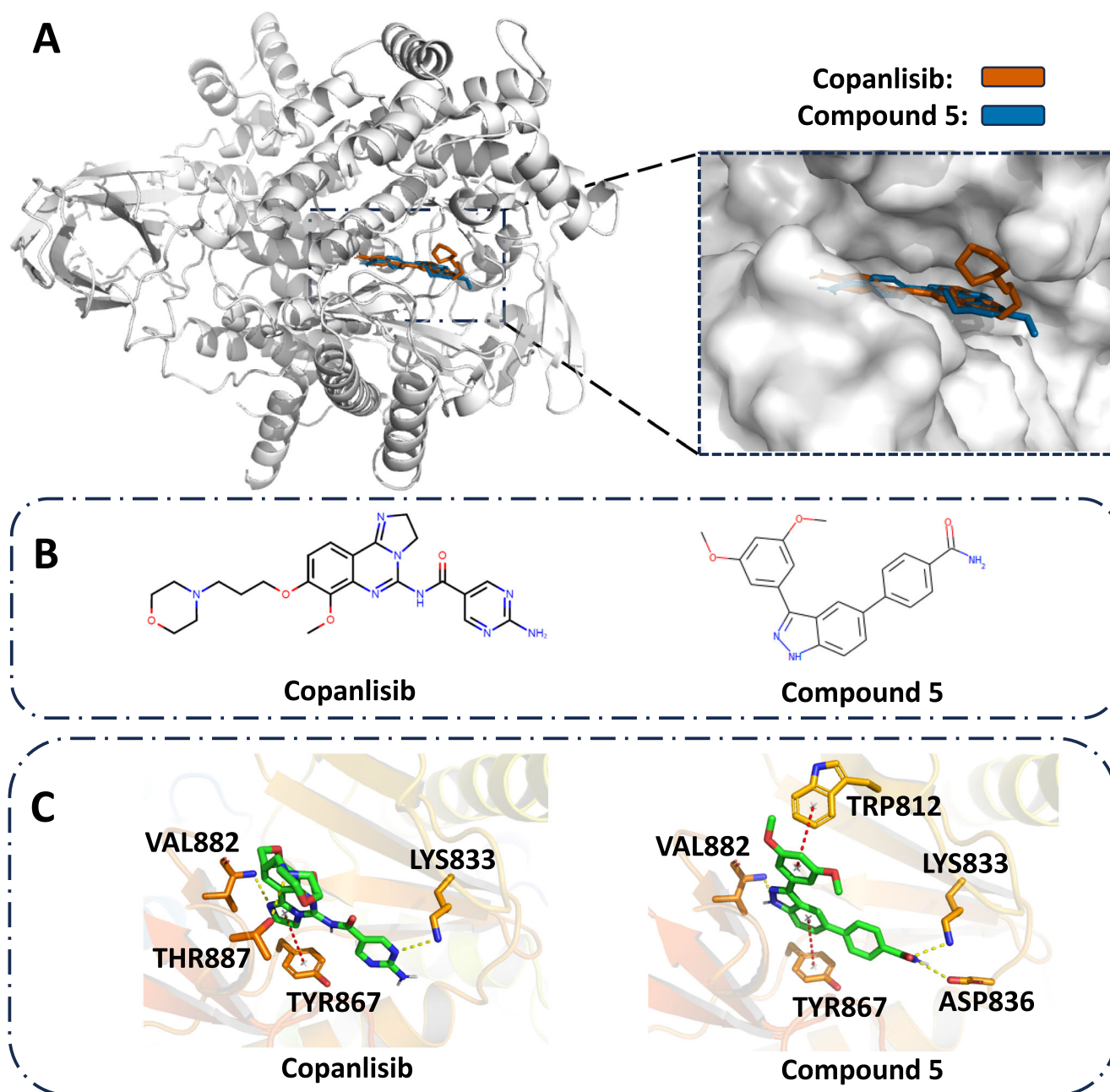
While the *in silico* computational metrics of molecules discovered by ParetoDrug show promise in comparison to existing drugs, it is crucial to acknowledge that these molecules are still far from being drugs. Drug discovery is an extremely complicated process, and the current metrics for molecules cannot perfectly reflect the physicochemical properties required for a compound to be a drug. Nevertheless, we clearly see ParetoDrug's promising potential in addressing multi-objective target-aware drug discovery tasks.

### Case study for multi-target drug discovery

Multi-target drug discovery can be considered a special case of multi-objective drug discovery where each protein target is going to be regarded as an objective to optimize. Until now, the study of multi-target target-aware drug discovery remains underexplored as it is challenging to consider the information of multiple protein targets at the same time to derive one ligand that could bind to all these given targets. Meanwhile, previous generative target-aware drug discovery works mainly focus on the single-target situation as there lack data sets to train the multi-target conditioned generative models. In this case study, we use ParetoDrug to perform a multi-target target-aware drug discovery task to design dual-functional inhibitors for both the HIV protease (HIV-PR) and HIV reverse transcriptase (HIV-RT). ParetoDrug is slightly modified to be compatible with this kind of task, and details are given in the Method section.

The crystal structures of HIV-PR and HIV-RT used here are 3A2O and 4G1Q<sup>51</sup>. Both structures are complexes with potent inhibitors solved at high resolution. We compare with LigBuilder V3<sup>51</sup>, the first *de novo* multi-target drug design program, and the variants of Pocket2Mol and TargetDiff extended by combining with screening. There are three different strategies in LigBuilder V3, including multi-target *de novo* design, multi-target growing, and multi-target linking. The best molecules for each strategy from the original paper are reported here. For the variants of Pocket2Mol and TargetDiff, we use each method to generate 100 molecules for each target. Then we use *smina* to screen the generated 200 molecules of each method to find the best molecule that has the best docking scores for both targets. We call the variants as Pocket2Mol-screen and TargetDiff-screen, and report the best molecules of each variant.

Results of both the docking scores and MM-GBSA scores for each method's best molecule are shown in Table 4. For ParetoDrug, we additionally report another two top molecules (Compound 7 and Compound 8) which are almost the same good as Compound 6 in terms of the docking scores. As shown in Table 4, the promising docking scores and MM-GBSA scores of Compounds 6–8 demonstrate that they are potential strong dual inhibitors to the given two protein targets in this task. Additionally, Fig. 4 shows Compounds 6–8 and their docking poses and interactions with HIV-



**Fig. 3 | Static structural analysis of ligands binding to PI3K- $\gamma$  (PDB ID: 5G2N).** Copanlisib is a drug binding to PI3K- $\gamma$  and Compound 5 is found by ParetoDrug. Hydrogen bonds are displayed in yellow dashed lines and  $\pi$ - $\pi$  interactions are in red.

**A** Solvent-accessible surfaces of the binding pocket of PI3K- $\gamma$  for Copanlisib and Compound 5. **B** Chemical structures of Copanlisib and Compound 5. **C** The binding poses of Copanlisib and Compound 5.

PR (PDB ID: 3A2O) and HIV-RT (PDB ID: 4G1Q). Interestingly, the similar structures of Compounds 6 to 8 in Fig. 4B indicate that ParetoDrug found a chemical subspace of strong inhibitors for both the HIV-PR and HIV-RT targets.

#### Case study for multi-target multi-objective drug discovery

We have shown ParetoDrug's promising ability for both the multi-objective target-aware drug discovery and multi-target drug discovery tasks separately. Naturally, a more attractive and challenging task is the multi-target multi-objective drug discovery task where the generated molecules need to bind to a given set of protein targets while manifesting other desired computational molecule properties. To the best of our knowledge, there are no published works yet to specifically address this kind of task. Here we use Lapatinib as a case study to evaluate ParetoDrug's ability for the multi-target multi-objective drug discovery task. Lapatinib is a dual tyrosine kinase

inhibitor that interrupts the EGFR pathway and inhibits HER4/ErbB4 Kinase for the treatment of breast cancer<sup>52,53</sup>. 1XKK and 3BBT are respectively the PDB IDs of crystal structures for Lapatinib binding to EGFR<sup>54</sup> and the HER4/ErbB4 kinase<sup>55</sup>. In this task, we configure ParetoDrug to bind to both protein targets while optimizing LogP, QED, SA, and NP metrics synchronously with IT at 150. We compare the generated molecules in the global Pareto pool with the known dual-inhibitor drug Lapatinib. Impressively, plenty of Pareto Dominate molecules over Lapatinib are discovered by ParetoDrug, as shown in Fig. 5. Furthermore, the QED values of Compound 12 and Compound 14 are greater than 0.8, which indicates the two molecules are potential initiators for a drug.

#### Discussion

In this work, we divide multi-objective drug discovery methods into three kinds based on whether they use the target protein information. The



**Table 3 | Metrics of the generated molecule for protein target PI3K-γ**

Ligands	Docking score (†)	LogP	QED (†)	SA (‡)	NP (†)	MM-GBSA score (†)
Copanlisib	9.6	0.7	0.55	3.1	-1.06	-55.51 ± 1.75
Compound 5	10.3	4.0	0.55	2.2	-0.74	-46.48 ± 5.41

The unit of docking score and MM-GBSA score ( $n = 3$  simulations for MM-GBSA) is kcal · mol<sup>-1</sup>. The 95% confidence intervals of MM-GBSA scores are included. We conducted the two-sided  $T$  test to compare the MM-GBSA scores of Compound 5 with Copanlisib, and the  $p$  value is 0.0358.

**Table 4 | Docking and MM-GBSA scores of the generated molecules by baselines and ParetoDrug for protein targets HIV-PR and HIV-RT**

Methods		Docking score (†)		MM-GBSA score (‡)	
		HIV-PR	HIV-RT	HIV-PR	HIV-RT
LigBuilder V3	de novo	12.0	8.4	-30.38 ± 5.49	-47.82 ± 2.88
	growing	10.6	8.1	-25.18 ± 6.09	-38.95 ± 10.07
	linking	13.2	10.8	-38.04 ± 2.40	-59.36 ± 0.96
Pocket2Mol-screen		16.4	12.4	-35.03 ± 17.89	-60.87 ± 1.31
TargetDiff-screen		19.0	11.7	-44.59 ± 16.37	-66.36 ± 5.90
Compound 6 (ParetoDrug)		20.7	12.9	-67.78 ± 7.29	-78.51 ± 4.09
Compound 7 (ParetoDrug)		20.5	13.0	-61.37 ± 9.99	-86.30 ± 17.10
Compound 8 (ParetoDrug)		20.4	13.1	-53.64 ± 6.51	-77.23 ± 3.70

The unit of docking score and MM-GBSA score ( $n = 3$  simulations for MM-GBSA) is kcal · mol<sup>-1</sup>. The 95% confidence intervals of MM-GBSA scores are included. We conducted the  $T$  test with the alternative hypothesis that the MM-GBSA score on HIV-PR of Compound 6 is less than molecules from other baselines. All  $p$  values are less than 0.05. It is the same for HIV-RT.

representative works of each kind of method and their comparisons are provided in Table 5. The first kind is the ligand-based method such as MolGPT<sup>22</sup>, which utilizes the conditional transformer to generate molecules that satisfy multiple inputting property constraints. However, the ligand-based method does not consider protein information and thus cannot guarantee to generate molecules with high binding affinity to a given protein target.

The second kind is the target-scoring-based method that employs a docking scoring function to predict the binding affinity of the generated molecules to the given protein target. In this way, although the target-scoring-based method also does not explicitly consider the target protein information, the binding affinity scores could be optimized by optimization techniques such as reinforcement learning and genetic algorithm. For example, Wang et al., propose MCMG<sup>21</sup> to combine conditional transformer, knowledge distillation, and reinforcement learning to generate molecules that satisfy multiple constraints including binding to targets such as GSK3β and JNK3. However, MCMG does not incorporate the target protein information into the generation process of its model and needs to design a reward that is a linear combination of each metric. Recently, MolSearch<sup>24</sup> is proposed to use multi-objective MCTS to generate molecules based on molecule fragments. However, MolSearch is a pure search-based method with predefined massive rules to modify molecules and also does not consider the target protein information. Furthermore, REINVENT 4<sup>39</sup> uses a reinforcement learning algorithm to generate optimized molecules compliant with a user-defined property profile defined as a multi-component score. However, it is not a generative target-aware method although it optimizes multiple objectives while treating the binding affinity as a standard optimizing objective. The lack of specific protein information in these target-

scoring-based methods makes the optimization of binding affinity objective inefficient and the trained models cannot be generalized to other target proteins. Therefore, this kind of method is different from the mainstream target-aware molecule generation in that the protein-ligand interactions are modeled in the molecule generation process. Meanwhile, most of them are only evaluated on several case studies which limits the assessment of their generality on various target proteins. For example, RationaleRL<sup>23</sup>, MCMG<sup>21</sup>, and MolSearch<sup>24</sup> optimize the molecule's binding affinity to GSK3β and JNK3, which is predicted by random forest models pretrained on the data sets<sup>56</sup> that contain samples of positive and negative compounds to the GSK3β and JNK3 targets, and are not available for most protein targets.

The third kind of method is the multi-objective target-aware molecule generation, which models the protein-ligand interactions to generate the molecules with high binding affinity to the inputting protein target. Recently, CProMG<sup>28</sup> is proposed to use the conditional multi-constraint autoregressive framework to generate molecules owning desired property constraints in a controllable manner. However, the ability of CProMG largely depends on the quality of data used to train the model and it does not involve an optimization process for a comprehensive searching in the chemical space. Compared with CProMG, ParetoDrug does not use a multi-constraint generative model. Instead, ParetoDrug employs the Pareto MCTS to optimize multiple objectives synchronously by searching desired molecules with the guidance of the pretrained autoregressive molecule generative model. Also as shown in the benchmark experiment, ParetoDrug achieves better multi-objective metrics of the generated molecules when compared with CProMG on all the property objectives except QED.

In conclusion, in this work, we propose ParetoDrug to fulfill the gap of multi-objective target-aware drug discovery in the field of deep learning-based drug discovery. ParetoDrug is an autoregressive Pareto MCTS algorithm that integrates the pretrained autoregressive generative model to search desired multi-objective molecules in an atom-by-atom way with the help of Pareto MCTS. We perform the evaluation of ParetoDrug on a standard benchmark setting with various baselines. The benchmark results show that ParetoDrug achieves multiple satisfactory molecule properties including binding affinity while previous single-objective methods cannot. We further conduct the case studies of the multi-objective target-aware drug discovery tasks for two known drugs, the multi-target drug discovery task for HIV-related disease targets, and the multi-target multi-objective drug discovery task for a dual inhibitor. In these case studies, new molecules discovered by ParetoDrug exhibit high potentials that Pareto Dominate the known drugs of the disease targets on all required property objectives. In conclusion, ParetoDrug demonstrates its ability to handle the challenging multi-objective target-aware drug discovery tasks and its superiority in searching in the large and complex chemical space for novel compounds that possess multiple promising properties including binding affinity.

For future work, on the one hand, making ParetoDrug compatible with more recent advanced autoregressive molecule generative models such as the Diffusion model is highly promising. On the other hand, extending ParetoDrug into the multi-objective design of protein, polypeptide, and nucleic acid drugs also holds significant potential.

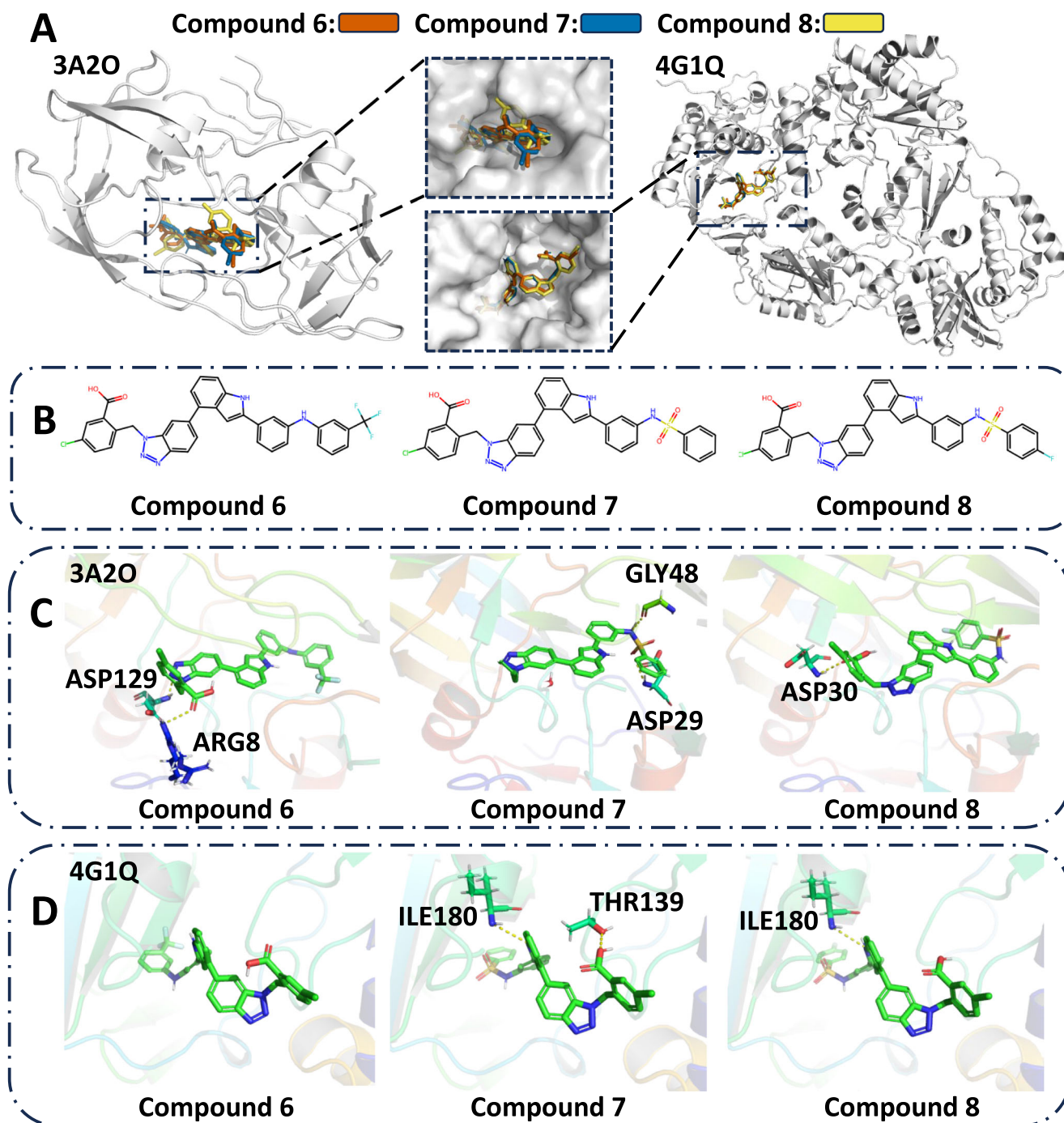
## Methods

In this section, we first formulate the target-aware drug discovery task as a Markov decision process. Then we introduce the concepts of Pareto Dominate and Pareto Front in the multi-objective optimization domain. Finally, we propose the framework of ParetoDrug designed for the multi-objective target-aware drug discovery task and the multi-target target-aware drug discovery task.

## Problem definition

Target-aware molecule generation can be formulated as a Markov decision process (MDP)<sup>57</sup> given that the next atom to be chosen only depends on the





**Fig. 4** | Static structural analysis of ligands binding to both the HIV-PR (PDB ID: 3A2O) and HIV-RT (PDB ID: 4G1Q). Compounds 6–8 are found by ParetoDrug. Hydrogen bonds are displayed in yellow dashed lines. **A** Solvent-accessible surfaces

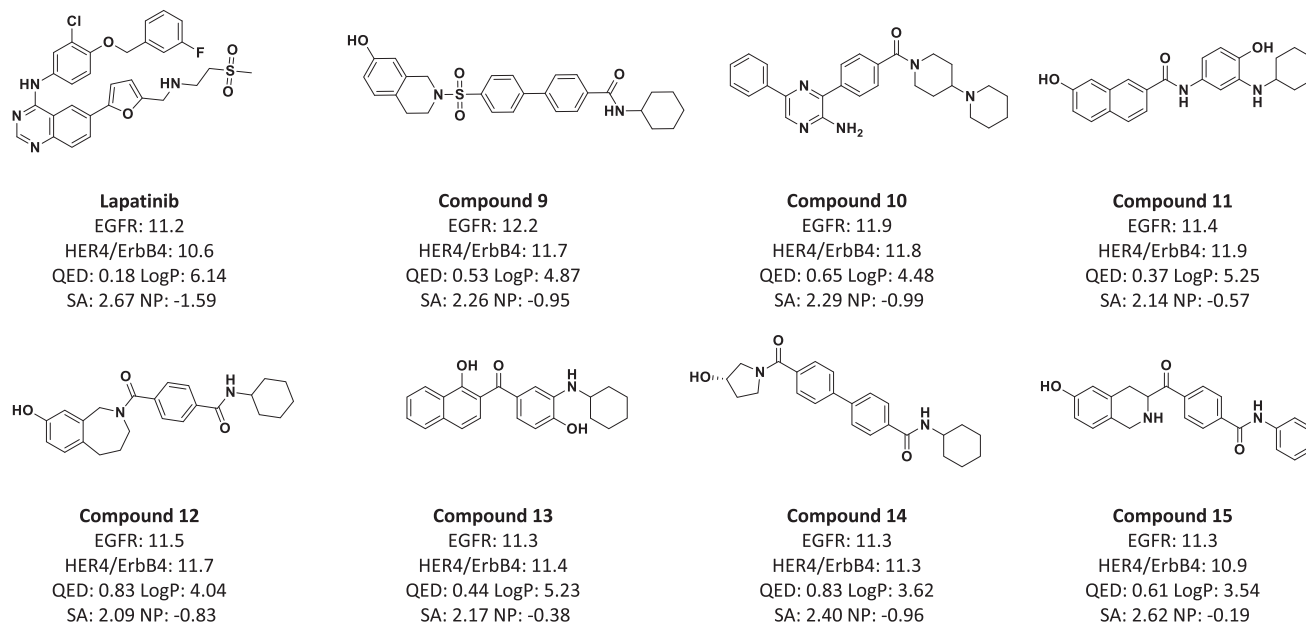
of the binding pockets of HIV-PR (left) and HIV-RT (right) for Compounds 6–8. **B** Chemical structures of Compounds 6–8. **C** The binding poses of Compounds 6 to 8 with HIV-PR. **D** The binding poses of Compounds 6–8 with HIV-RT.

generated molecule fragment and the protein target. The MDP can be defined as  $M = (S, A, P, R)$  where  $S$  denotes the set of states that describe the current molecule fragment and the protein,  $A$  denotes the set of actions that indicate the chosen atom symbol to be added to the current molecule fragment, and  $P: S \times A \rightarrow S$  is the state transition function where the molecule fragment incorporates the chosen atom symbol to grow up to a new molecule fragment.  $R: S \rightarrow \mathbb{R}^d$  is the reward function based on the current state. In target-aware molecule generation, the reward to evaluate the generated molecule is usually available at the terminal state, which is a typical sparse-reward setting. If  $d > 1$ , multiple reward objectives are considered such as strong binding affinity, high drug-likeness, and low toxicity

in drug discovery. The goal is to take the action that maximizes the expected episodic reward  $\bar{R}(s, a)$ , which can be approximated under repeated rollouts<sup>58</sup> as

$$\bar{R}(s, a) = \frac{1}{N(s, a)} \sum_{j=1}^{N(s)} \mathbb{I}_j(s, a) r^j(s), \quad (2)$$

where  $N(s)$  denotes the rollout times starting from state  $s$  and  $N(s, a)$  is the times that action  $a$  has been taken from state  $s$ .  $\mathbb{I}_j(s, a)$  is an indicator function with value 1 if action  $a$  is selected from state  $s$  at the  $j$ th rollout



**Fig. 5 | Property metric values of Lapatinib and its Pareto Dominate molecules (Compounds 9 to 15) found by ParetoDrug.** EGFR means the ligand's docking score to EGFR (PDB ID: 1XKK) while HER4/ErbB4 means the ligand's docking score to HER4/ErbB4 kinase (PDB ID: 3BBT), and the unit for the two docking scores is kcal · mol<sup>-1</sup>.

**Table 5 | Comparison of different multi-objective drug discovery methods**

Method	Target-based type	Multi-objective type	Objectives
MolGPT <sup>22</sup>	Ligand-based	constraint	QED, SA, LogP, TPSA <sup>72</sup>
MolSearch <sup>24</sup>	Target-scoring-based	optimization	QED, SA, docking <sup>a</sup> (GSK3β, JNK3)
MCMG <sup>21</sup>	Target-scoring-based	constraint & optimization	QED, SA, docking <sup>b</sup> (GSK3β, JNK3, DRD2)
RationaleRL <sup>23</sup>	Target-scoring-based	constraint & optimization	QED, SA, docking <sup>c</sup> (GSK3β, JNK3)
GENERA <sup>25</sup>	Target-scoring-based	optimization	docking <sup>d</sup> (ACE2)
REINVENT 4 <sup>39</sup>	Target-scoring-based	optimization	User-defined objectives including docking <sup>e</sup>
CProMG <sup>28</sup>	Target-aware	constraint	QED, SA, LogP, TPSA, docking <sup>f</sup>
ParetoDrug	Target-aware	optimization	QED, SA, LogP, NP-likeness, docking <sup>g</sup>

<sup>a</sup>Docking scores of GSK3β and JNK3 are predicted by pretrained random forest models<sup>23</sup>.

<sup>b</sup>Docking scores of GSK3β and JNK3 are predicted by pretrained random forest models<sup>23</sup>. The docking score of DRD2 is predicted by a support vector machine classifier with a Gaussian kernel<sup>73</sup>.

<sup>c</sup>Docking scores of GSK3β and JNK3 are predicted by pretrained random forest models<sup>23</sup>.

<sup>d</sup>Docking scores of ACE2 are predicted by PLANTS<sup>74</sup> and Glide<sup>75</sup>.

<sup>e</sup>In this study, we configure the optimization objectives of REINVENT 4 as QED, SA, LogP, NP-likeness, and Docking score. The docking score of a given protein target (not limited to specific targets) is predicted by smina<sup>32</sup>.

<sup>f</sup>Docking score of a given protein target (not limited to specific targets) is predicted by Autodock Vina<sup>76</sup>.

<sup>g</sup>Docking score of a given protein target (not limited to specific targets) is predicted by smina<sup>32</sup>.

round, 0 otherwise.  $r^j(s)$  is the final reward to evaluate the final generated molecule at the terminal state for the  $j$ th rollout round starting from state  $s$ . A larger  $\bar{R}(s, a)$  value indicates a higher expected reward by taking action  $a$  from state  $s$ .

### Multi-objective optimization

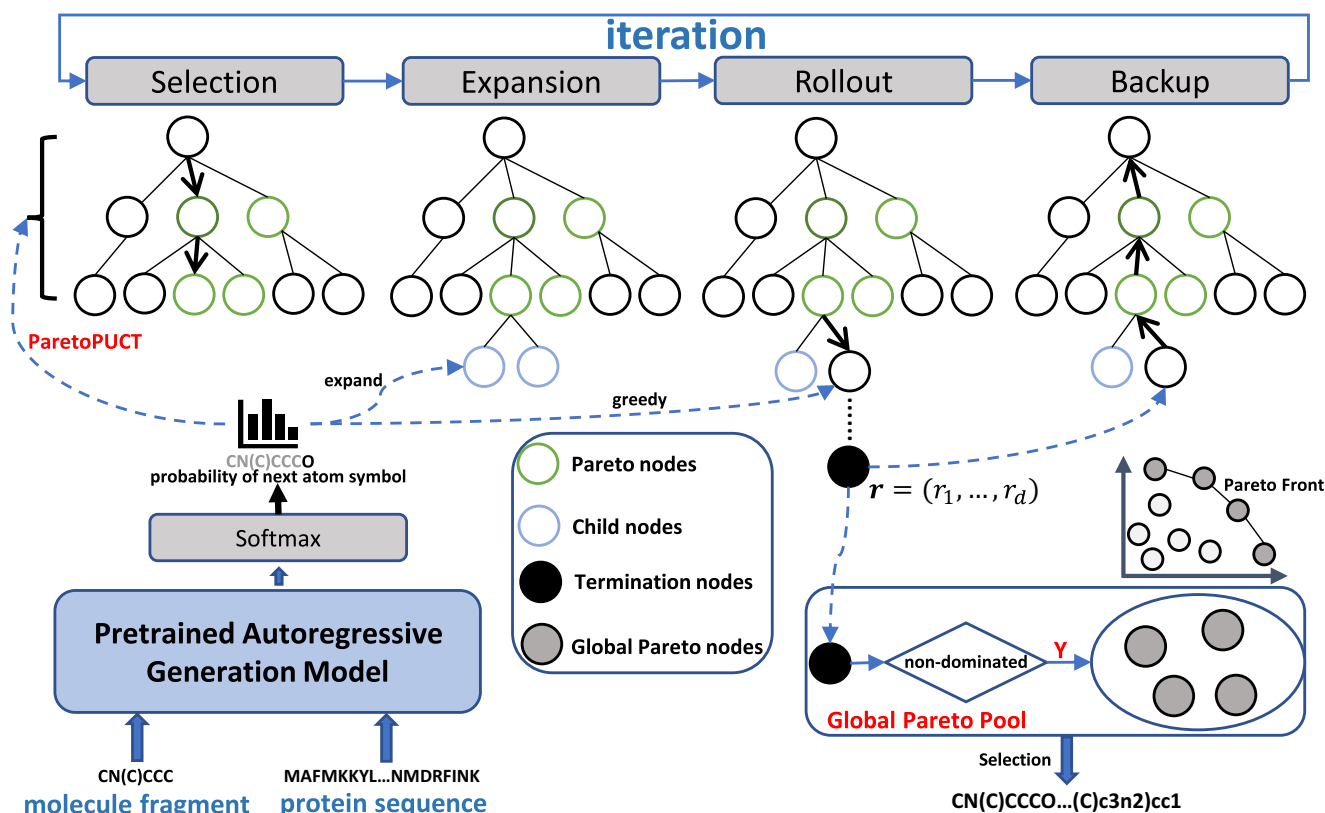
Multi-objective optimization (also known as Pareto optimization) is concerned with optimization problems involving more than one objective function to be optimized simultaneously<sup>29</sup>, which has been applied in many fields. In multi-objective optimization, there does not typically exist a feasible solution that maximizes all objective functions at the same time. Therefore, attention is paid to Pareto optimal solutions<sup>60</sup>, which cannot be improved in any of the objectives without degrading at least one of the other objectives. In mathematical terms, a feasible vector  $\mathbf{X} \in \mathbb{R}^d$  is said to Pareto Dominate another vector  $\mathbf{X}' \in \mathbb{R}^d$  is defined as below<sup>61</sup>.

**Definition 1.** Pareto Dominate. Given two vectors  $\mathbf{X} = (x_1, \dots, x_d)$  and  $\mathbf{X}' = (x'_1, \dots, x'_d)$ ,  $\mathbf{X}$  is said to dominate  $\mathbf{X}'$ , i.e.,  $\mathbf{X} \succ \mathbf{X}'$  if and only if  $x_i \geq x'_i, \forall i = 1, \dots, d$ .  $\mathbf{X}$  is said to strictly dominate  $\mathbf{X}'$ , i.e.,  $\mathbf{X} \succ \mathbf{X}'$  if and only if  $\mathbf{X} \succ \mathbf{X}'$  and  $\exists i$  such that  $x_i > x'_i$ .

A vector  $\mathbf{X}^* \in \mathbb{R}^d$  is called Pareto optimal if there does not exist another vector that Pareto Dominates it. The set of Pareto optimal vectors  $\mathbb{X}^*$ , called Pareto Front, is defined as below.

**Definition 2.** Pareto Front. Given a set of vectors  $\mathbb{X} \subset \mathbb{R}^d$ , the non-dominant set  $\mathbb{X}^* \in \mathbb{X}$  is defined as  $\mathbb{X}^* = \{\mathbf{X} \in \mathbb{X} : \nexists \mathbf{X}' \in \mathbb{X} \text{ s.t. } \mathbf{X}' \succ \mathbf{X}\}$ .

In drug molecule design, the optimization or constraint of multiple properties is a pervasive requirement. For instance, for a new drug to be successful, it must simultaneously be potent, bioavailable, safe, and synthesizable, with these properties being often competing<sup>62</sup>. As Pareto



**Fig. 6 | The framework of ParetoDrug.** The pretrained autoregressive generative model gives the probability distribution of the next atom symbol. The ParetoPUCT balances the exploration and exploitation when searching for the next atom symbol

in the Selection step. ParetoDrug maintains a global pool of Pareto optimal molecules, which are updated during Pareto MCTS.

optimization is capable of discovering a set of solutions that reveal trade-offs among objectives and relies on no prior measure of the importance of competing objectives, it is believed as the most robust approach to multi-objective drug discovery<sup>62</sup>. Next, we introduce how to utilize the concepts of Pareto Dominate and Pareto Front to construct the ParetoDrug framework for multi-objective target-aware drug discovery with the help of MCTS and the pretrained autoregressive generative model.

### ParetoDrug

To solve the challenging multi-objective target-aware drug discovery task, which requires generating molecules with multiple desired properties including the strong binding affinity to specified protein targets, we propose an autoregressive Pareto MCTS generation algorithm called ParetoDrug. First, ParetoDrug employs existing pretrained autoregressive generative models to provide exploration guidance toward desired molecules during searching. Based on both the protein context and intermediate molecule fragment, the pretrained autoregressive model predicts the probability of the next atom symbol to be added to the current molecule fragment. Second, with exploration guidance from the pretrained autoregressive model, ParetoDrug performs Pareto MCTS to progressively find Pareto optimal molecules with multiple desired properties. Third, to achieve the exploration-exploitation balance during searching, we propose the ParetoPUCT selection criterion to determine the next atom symbol in the Selection step of Pareto MCTS. Through these three key components, ParetoDrug is able to generate high-quality molecules for multi-objective target-aware drug discovery tasks. The overall framework of ParetoDrug is shown in Fig. 6. Next, we explain the details including Pareto MCTS, ParetoPUCT, and how to extend ParetoDrug into the case of multi-target target-aware molecule generation.

### Pareto MCTS

In MCTS, pretrained neural networks based on the expert data could be used for the guidance of action selection<sup>63</sup> and this idea has been extended into single-objective target-aware drug discovery<sup>15</sup>. Similarly, to enable the exploration guidance to desired molecules with strong binding affinity to specified protein targets, ParetoDrug employs an existing pretrained autoregressive generative model<sup>15</sup> to predict the next atom symbol given the protein target and current molecule fragment. The protein target is represented by the amino acid sequence. The molecule fragment is based on SMILES<sup>64</sup>, which describes molecules with short ASCII strings. The autoregressive generative model includes a protein encoder based on the protein's amino acid sequence and a ligand molecule decoder. At each step, the protein encoder receives the target protein sequence and outputs the protein embedding into the ligand decoder. Next, the ligand decoder predicts the probability of the next atom symbol based on both the protein embedding and intermediate molecule fragment from the last step. This autoregressive generative model is pretrained on the protein-ligand data set and used in MCTS.

When generating molecules with the pretrained autoregressive generative model, although we could obtain the ligand molecule in a greedy manner by taking the next atom symbol with the maximum probability, it is prone to be stuck in a local optimum due to the unpredictable complexity of the chemical space. At the same time, the predicted atom with the maximum probability does not mean that it must be in the optimal molecule that satisfies multiple required properties as the pretrained model is not optimized for these properties. To address the above difficulties for the multi-objective target-aware drug discovery, we propose ParetoDrug, which employs the Pareto MCTS to enable a synchronous optimization of multiple properties together with the help of a pretrained autoregressive model for selection guidance. Next, we introduce Pareto MCTS.

Pareto MCTS<sup>65</sup> extends basic MCTS<sup>66</sup> to optimize multiple objectives, which adopts a tree structure to perform simulation iterations and estimates action values to guide searching. The Pareto MCTS procedure consists of four steps per iteration:

- **Selection.** Each iteration starts from the current root node  $a_t$  and the best child is recursively selected until a leaf node  $a_{t+l}$  after  $l$  selections, i.e., a node that has not been expanded or terminated, is reached. For each selection  $t \in [1, l]$ , we need a selection criterion to determine which child node is the best to be chosen. This criterion balances between exploitation and exploration to avoid being trapped in local optimums and is given in Eq. (6).
- **Expansion.** Given a selected leaf node  $a_{t+l}$ , the probability  $P(a|C_{t+l})$  for each expandable atom symbol  $a \in A$  is computed by the pretrained autoregressive generative model.  $C_{t+l} = \{S_p, m_{t+l}\}$  is the state context with the target protein sequence  $S_p$  and the current simulated intermediate molecule fragment  $m_{t+l} = a_1 \cdots a_t a_{t+1} \cdots a_{t+l}$ . Here  $A$  is the legal action space, i.e. the SMILES vocabulary of molecules, under the given state context. The expanded child nodes of  $a_{t+l}$  are added to the tree and initialized immediately.
- **Rollout.** The value of the reached leaf node  $a_{t+l}$  is evaluated by a fast rollout. From the leaf node, MCTS recursively generates the next state until termination and receives the reward of the final molecule at the termination state. During the rollout, each atom symbol is selected in a greedy manner according to the predicted probability given by the pretrained autoregressive network until a terminal symbol  $a_{t+L}$  is generated or the tree reaches a maximum depth. The path from the initial atom symbol to the terminal atom symbol forms a complete molecule  $m = a_1 \cdots a_t a_{t+1} \cdots a_{t+L}$ . The reward  $\mathbf{r}$  of the final molecule  $m$  is then evaluated based on the molecule property metrics. Specifically, the binding affinity is computed by the docking function  $f(S_p, m)$  such as  $\text{smiNA}^{32}$ . The reward  $\mathbf{r}$  is calculated as defined in Eq. (3) by normalizing the property metric in each dimension.
- **Backup.** The reward is backpropagated along the visited nodes to update their statistics until the root node. The detailed updating process for tree nodes is elaborated in Eq. (4) with the defined reward vector  $\mathbf{r}$  for nodes.

When performing Pareto MCTS, we maintain a global pool of all the Pareto optimal molecules found so far to represent the molecule Pareto Front as defined by Definition 2. We update the global pool of Pareto optimal molecules by adding newly generated Pareto optimal molecules and removing invalid ones if they are Pareto Dominated by the global pool's new coming molecules. The molecule comparison is based on the reward vector  $\mathbf{r}$  defined as follows. For each generated molecule in the rollout with the property metric vector  $\mathbf{h} = (h_1, \dots, h_d) \in \mathbb{R}^d$ , the reward vector  $\mathbf{r} = (r_1, \dots, r_d) \in \mathbb{R}^d$  of this molecule is defined as

$$r_i = \frac{1}{N_p} \sum_{k=1}^{N_p} \mathbb{I}[h_i \geq h_i^k], \forall i = 1, \dots, d, \quad (3)$$

where  $N_p$  is the number of Pareto optimal molecules in the global pool,  $h_i^k$  is the  $i$ th property metric of the  $k$ th Pareto optimal molecule, and  $h_i$  is the  $i$ th property metric of the current generated molecule to be compared.  $\mathbb{I}$  is the indicator function with value 1 if the condition  $h_i \geq h_i^k$  is satisfied, 0 otherwise. The calculation of reward  $\mathbf{r}$  treats each dimension separately, regardless of their scale difference, which gains an advantage over methods that aggregate all dimensions into one score using predefined weights<sup>24</sup>. With the reward vector, the Backup step is performed as

$$N_a \leftarrow N_a + 1, \mathbf{W}_a \leftarrow \mathbf{W}_a + \mathbf{r}, a \leftarrow \text{parent of } a, \quad (4)$$

where  $N_a$  is the total times that node  $a$  has been selected and  $\mathbf{W}_a$  is the cumulative reward vector of node  $a$ . For each selection  $t \in [1, l]$ , the statistics

of node  $a_{t+t}$  are updated by adding the reward vector of the node  $a_{t+l}$ 's rollout result to  $\mathbf{W}_a$  and increasing the visiting times  $N_a$  by 1.

## ParetoPUCT

The most important step of MCTS is the Selection step where a criterion is needed to select the next child node by comparing all child nodes. The most commonly used criterion is the upper confidence bound<sup>67</sup> in which a child node is selected to maximize

$$U = \frac{W_a}{N_a} + \sqrt{\frac{2 \ln N}{N_a}}, \quad (5)$$

where  $N$  is the total times of iterations and  $N_a$  is the times of node  $a$  being selected.  $U$  is a scalar used to select the best child node with the largest value. However, in the multi-objective target-aware drug discovery, the reward becomes a vector and  $U$  is not applicable for the comparison of vectors. At the same time, we also want to utilize the pretrained autoregressive generative model to provide the exploration guidance in the chemical space<sup>15,68</sup> when selecting the next child node in the Selection step.

Therefore, here we propose ParetoPUCT that extends the scalar predictor upper confidence bound applied to trees (PUCT) selection criterion<sup>69</sup> with the concepts of Pareto Dominate and Pareto Front into a vectorial selection criterion for the multi-objective MCTS<sup>65</sup>. At each selection  $t$ , we first compute a selection score vector for each candidate child node as

$$\mathbf{U}_p(C_{t+t-1}, a) = \frac{\mathbf{W}_a}{N_a} + cP(a|C_{t+t-1}) \frac{\sqrt{N}}{1 + N_a}, \quad (6)$$

where  $c$  is a constant that controls the degree of exploration. Here  $\mathbf{W}_a$  is the cumulative reward vector for node  $a$ . The  $\frac{\sqrt{N}}{1 + N_a}$  part guides MCTS to initially prefer to visit the nodes with a low number of visits. At the same time, the  $P(a|C_{t+t-1})$  part tends to visit the atom nodes that are probably to produce a molecule with strong binding affinity to the protein target indicated by the pretrained autoregressive generative model. Furthermore, the  $\frac{\mathbf{W}_a}{N_a}$  makes ParetoDrug exploit the nodes with multiple high property metrics while  $c$  balances the exploitation and exploration. As the  $\mathbf{U}_p$  score is in the vector from for each child node  $a$ , to determine which child node to be selected, ParetoPUCT constructs a Pareto Front for those child nodes that are not Pareto Dominated by other child nodes by comparing their  $\mathbf{U}_p$  score vectors. Each child node in the resulting Pareto Front cannot be replaced by a better child node and thus becomes the candidate node for the selection. Finally, ParetoPUCT selects a node from the Pareto Front of the candidate child nodes uniformly at random.

## Modifications of ParetoDrug for multi-target target-aware molecule generation

As the multi-target target-aware molecule generation involves multiple protein targets, we modify the ParetoPUCT node selection criterion to handle multiple predictions from the pretrained autoregressive generative model for different protein targets. Therefore, we propose the Multi-Target ParetoPUCT (M-ParetoPUCT) defined as

$$\mathbf{U}_{mp} = \frac{\mathbf{W}_a}{N_a} + cf(P_1(a|C_{1,t+t-1}), \dots, P_m(a|C_{m,t+t-1})) \frac{\sqrt{N}}{1 + N_a}, \quad (7)$$

where there are  $m$  predictions for the next node  $a$  and  $c$  is a constant to control the exploration degree. For the prediction fusing function  $f(P_1(a|C_{1,t+t-1}), \dots, P_m(a|C_{m,t+t-1}))$ , each prediction is a distribution of the next atom symbol with the inputting of the molecule's SMILES string representation<sup>64</sup> and specified target protein's amino acid sequence. Here  $P_i(a|C_{i,t+t-1})$  is the neural network prediction for the  $i$ th target pre-trained on the protein-ligand data set. As the distributions are on the



same action set, we use the mean-pooling operation for  $f$  as

$$f(P_1(a|C_{1,\tau+t-1}), \dots, P_m(a|C_{m,\tau+t-1})) = \frac{\sum_{i=1}^m P_i(a|C_{i,\tau+t-1})}{m}, \forall a \in A. \quad (8)$$

This mean-pooling operation keeps the probabilities of the possible next atom symbols for each protein target. Meanwhile, it enhances the probabilities of the next atom symbol if it is predicted to be preferred by all the given protein targets.

After a leaf node  $a_{\tau+l}$  is selected, we need to expand it. The probability for each expandable atom symbol is computed the same as Eq. (8) from multiple predictions of the pretrained autoregressive model on multiple protein targets. Each child node  $a$  of  $a_{\tau+l}$  is initialized to  $\{N_a = 0, \mathbf{W}_a = \mathbf{0}, \frac{\sum_{i=1}^m P_i(a|C_{i,\tau+t-1})}{m}\}$ .

### Statistics and reproducibility

Data manipulation and processing analyses were conducted using the packages Python (version 3.7), Biopython (version 1.79), Pandas (version 1.3.4), MMseqs2 (version 13.45111), RDKit (version 2020.09.5), PyTorch (version 1.13.1), and Openbabel (version 3.1.1). We used PyMOL (version 2.6.0a0) to analyze the protein structures. We used AMBER22 package to calculate the MM-GBSA scores (Supplementary Information G). We used PLIP 2021 to analyze protein-ligand Interactions Supplementary Information F. We use smina (version 2020.12.10) to calculate the docking score. The molecule property distributions are drawn by Matplotlib (version 3.4.3) and Seaborn (version 0.12.2), where the function “kdeplot” is called for kernel density estimate. The  $T$  test and  $p$  value calculation in this article are conducted with SciPy (version 1.9.1).

### Data availability

The training and testing data of the autoregressive generative model used in ParetoDrug is processed from [BindingDB](#) and [PDBbind](#), and is the same as in AlphaDrug and could be obtained from <https://github.com/CMACH508/AlphaDrug>. For the multi-objective target-aware drug discovery and multi-target drug discovery case studies, the PDB and ligand files of 7D42, 5G2N, 3A2O, 4G1Q, 1XKK, and 3BBT are downloaded from [RCSB Protein Data Bank](#).

### Code availability

The source code of this study is publicly available from the GitHub repository: <https://github.com/CNDOTA/ParetoDrug>. We also provide the Google Colab version of [ParetoDrug](#), which could be directly run online. The numerical source data for graphs and charts in this article is provided in Figshare with DOI<sup>70</sup>. We also deposit the data and codes of ParetoDrug in Figshare with DOI<sup>71</sup>.

Received: 3 April 2024; Accepted: 16 August 2024;

Published online: 02 September 2024

### References

- Anderson, A. C. The process of structure-based drug design. *Chem. Biol.* **10**, 787–797 (2003).
- Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **432**, 862–865 (2004).
- Gorgulla, C. et al. An open-source drug discovery platform enables ultra-large virtual screens. *Nature* **580**, 663–668 (2020).
- Blundell, T. L. Structure-based drug design. *Nature* **384**, 23–26 (1996).
- Macarron, R. et al. Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* **10**, 188–195 (2011).
- Cheng, Y., Gong, Y., Liu, Y., Song, B. & Zou, Q. Molecular design in drug discovery: a comprehensive review of deep generative models. *Brief. Bioinforma.* **22**, bbab344 (2021).
- 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 (2019).
- Aumentado-Armstrong, T. Latent molecular optimization for targeted therapeutic design ArXiv:1809.02032 [cs, q-bio] (2018).
- Long, S., Zhou, Y., Dai, X. & Zhou, H. Zero-Shot 3D Drug design by sketching and generating. In Oh, A. H., Agarwal, A., Belgrave, D. & Cho, K. (eds.) *Advances in Neural Information Processing Systems* (2022).
- Skalic, M., Sabbadin, D., Sattarov, B., Sciabola, S. & De Fabritiis, G. From target to drug: Generative modeling for the multimodal structure-based ligand design. *Mol. Pharmaceutics* **16**, 4282–4291 (2019).
- Schneuing, A. et al. Structure-based drug design with equivariant diffusion models ArXiv:2210.13695 [cs, q-bio] (2022).
- Guan, J. et al. 3D Equivariant diffusion for target-aware molecule generation and affinity prediction. In *The Eleventh International Conference on Learning Representations* (2023).
- 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).
- Grechishnikova, D. Transformer neural network for protein-specific de novo drug generation as a machine translation problem. *Sci. Rep.* **11**, 321 (2021).
- Qian, H., Lin, C., Zhao, D., Tu, S. & Xu, L. AlphaDrug: Protein target specific de novo molecular generation. *PNAS Nexus* **1**, pgac227 (2022).
- Xu, M., Ran, T. & Chen, H. De novo molecule design through the molecular generative model conditioned by 3D information of protein binding sites. *J. Chem. Inf. Model.* **61**, 3240–3254 (2021).
- Zhang, J. & Chen, H. De novo molecule design using molecular generative models constrained by ligand-protein interactions. *J. Chem. Inf. Model.* **62**, 3291–3306 (2022).
- Liu, M., Luo, Y., Uchino, K., Maruhashi, K. & Ji, S. Generating 3D molecules for target protein binding. In Chaudhuri, K. et al. (eds.) *Proceedings of the 39th International Conference on Machine Learning*, **162**, 13912–13924 (2022).
- Luo, S., Guan, J., Ma, J. & Peng, J. A 3D generative model for structure-based drug design. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P. S. & Vaughan, J. W. (eds.) *Advances in Neural Information Processing Systems*, **34**, 6229–6239 (2021).
- 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 (2022).
- Wang, J. et al. Multi-constraint molecular generation based on conditional transformer, knowledge distillation and reinforcement learning. *Nat. Mach. Intell.* **3**, 914–922 (2021).
- Bagal, V., Aggarwal, R., Vinod, P. K. & Priyakumar, U. D. MolGPT: Molecular generation using a transformer-decoder model. *J. Chem. Inf. Model.* **62**, 2064–2076 (2022).
- Jin, W., Barzilay, D. & Jaakkola, T. Multi-objective molecule generation using interpretable substructures. In *Proceedings of the 37th International Conference on Machine Learning*, 119 of *Proceedings of Machine Learning Research*, 4849–4859 (2020).
- Sun, M. et al. MolSearch: Search-based multi-objective molecular generation and property optimization. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 4724–4732 (2022).
- Lamanna, G. et al. GENERA: A combined genetic/deep-learning algorithm for multiobjective target-oriented de novo design. *J. Chem. Inf. Model.* **63**, 5107–5119 (2023).
- Sutton, R. S. & Barto, A. G. *Reinforcement learning: an introduction*. Adaptive computation and machine learning (MIT Press, Cambridge, Mass, 1998).
- Mitchell, M. *An introduction to genetic algorithms*. Complex adaptive systems (MIT Press, Cambridge, Mass, 1996).

28. Li, J.-N., Yang, G., Zhao, P.-C., Wei, X.-X. & Shi, J.-Y. CProMG: Controllable protein-oriented molecule generation with desired binding affinity and drug-like properties. *Bioinformatics* **39**, i326–i336 (2023).
29. Brown, T. et al. Language Models are Few-Shot Learners. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F. & Lin, H. (eds.) *Advances in Neural Information Processing Systems*, **33**, 1877–1901 (2020).
30. Ouyang, L. et al. Training language models to follow instructions with human feedback. In Koyejo, S. et al. (eds.) *Advances in Neural Information Processing Systems*, **35**, 27730–27744 (2022).
31. Liu, T., Lin, Y., Wen, X., Jorissen, R. N. & Gilson, M. K. BindingDB: A web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic Acids Res.* **35**, D198–D201 (2007).
32. Koes, D. R., Baumgartner, M. P. & Camacho, C. J. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J. Chem. Inf. Model.* **53**, 1893–1904 (2013).
33. Ghose, A. K., Viswanadhan, V. N. & Wendoloski, J. J. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A Qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.* **1**, 55–68 (1999).
34. Bickerton, G. R., Paolini, G. V., Besnard, J., Muresan, S. & Hopkins, A. L. Quantifying the chemical beauty of drugs. *Nat. Chem.* **4**, 90–98 (2012).
35. Ertl, P. & Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminf.* **1**, 8 (2009).
36. Ertl, P., Roggo, S. & Schuffenhauer, A. Natural product-likeness score and its application for prioritization of compound libraries. *J. Chem. Inf. Model.* **48**, 68–74 (2008).
37. Ma, B. et al. Structure-based de novo molecular generator combined with artificial intelligence and docking simulations. *J. Chem. Inf. Model.* **61**, 3304–3313 (2021).
38. Yang, X., Zhang, J., Yoshizoe, K., Terayama, K. & Tsuda, K. ChemTS: an efficient python library for de novo molecular generation. *Sci. Technol. Adv. Mater.* **18**, 972–976 (2017).
39. Loeffler, H. H. et al. Reinvent 4: Modern AI-driven generative molecule design. *J. Cheminf.* **16**, 20 (2024).
40. Chen, Y.-C. A tutorial on kernel density estimation and recent advances. *Biostat. Epidemiol.* **1**, 161–187 (2017).
41. Rastelli, G., Rio, A. D., Degliesposti, G. & Sgobba, M. Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J. Comput. Chem.* **31**, 797–810 (2010).
42. Wang, E. et al. End-point binding free energy calculation with MM/PBSA and MM/GBSA: Strategies and applications in drug design. *Chem. Rev.* **119**, 9478–9508 (2019).
43. Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F. & Schroeder, M. PLIP: fully automated protein–ligand interaction profiler. *Nucleic Acids Res.* **43**, W443–W447 (2015).
44. Angulo, P. Nonalcoholic fatty liver disease. *N. Engl. J. Med.* **346**, 1221–1231 (2002).
45. Adams, L. A. et al. The natural history of nonalcoholic fatty liver disease: A population-based cohort study. *Gastroenterology* **129**, 113–121 (2005).
46. Bellentani, S., Bedogni, G., Miglioli, L. & Tiribelli, C. The epidemiology of fatty liver. *Eur. J. Gastroenterol. Hepatol.* **16**, 1087–1093 (2004).
47. Browning, J. D. et al. Prevalence of hepatic steatosis in an urban population in the United States: Impact of ethnicity. *Hepatology* **40**, 1387–1395 (2004).
48. Jiang, L. et al. Structural basis of tropifexor as a potent and selective agonist of farnesoid X receptor. *Biochem. Biophys. Res. Commun.* **534**, 1047–1052 (2021).
49. Carbone, A. et al. Follicular lymphoma. *Nat. Rev. Dis. Prim.* **5**, 83 (2019).
50. Scott, W. J. et al. Discovery and SAR of Novel 2,3-Dihydroimidazo[1,2- c]quinazoline PI3K Inhibitors: Identification of Copanlisib (BAY 80-6946). *ChemMedChem* **11**, 1517–1530 (2016).
51. Yuan, Y., Pei, J. & Lai, L. LigBuilder V3: A multi-target de novo drug design approach. *Front. Chem.* **8**, 142 (2020).
52. Higa, G. M. & Abraham, J. Lapatinib in the treatment of breast cancer. *Expert Rev. Anticancer Ther.* **7**, 1183–1192 (2007).
53. El-Gamal, M. I. et al. A review of HER4 (ErbB4) kinase, its impact on cancer, and its inhibitors. *Molecules* **26**, 7376 (2021).
54. Wood, E. R. et al. A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib). *Cancer Res.* **64**, 6652–6659 (2004).
55. Qiu, C. et al. Mechanism of activation and inhibition of the HER4/ ErbB4 kinase. *Structure* **16**, 460–467 (2008).
56. Li, Y., Zhang, L. & Liu, Z. Multi-objective de novo drug design with conditional graph generative model. *J. Cheminf.* **10**, 33 (2018).
57. Bellman, R. A Markovian decision process. *J. Math. Mech.* **6**, 679–684 (1957).
58. Gelly, S. & Silver, D. Monte-Carlo tree search and rapid action value estimation in computer Go. *Artif. Intell.* **175**, 1856–1875 (2011).
59. Miettinen, K. Nonlinear multiobjective optimization. No. 12 in International series in operations research & management science (1999).
60. Luc, D. T. Pareto Optimality. In Pardalos, P. M., Chinchuluun, A., Pardalos, P. M., Migdalas, A. & Pitsoulis, L. (eds.) *Pareto Optimality, Game Theory And Equilibria*, **17**, 481–515 (2008).
61. Wang, W. & Sebag, M. Multi-objective Monte-Carlo Tree Search. In *Proceedings of the Asian Conference on Machine Learning*, **25**, 507–522 (Singapore Management University, Singapore, 2012).
62. Fromer, J. C. & Coley, C. W. Computer-aided multi-objective optimization in small molecule discovery. *Patterns* **4**, 100678 (2023).
63. Silver, D. et al. Mastering the game of Go without human knowledge. *Nature* **550**, 354–359 (2017).
64. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28**, 31–36 (1988).
65. Chen, W. & Liu, L. Pareto Monte Carlo Tree Search for Multi-Objective Informative Planning. In *Proceedings of Robotics: Science and Systems* (Freiburg/Breisgau, Germany, 2019).
66. Browne, C. B. et al. A Survey of Monte Carlo tree search methods. *IEEE Trans. Comput. Intell. AI Games* **4**, 1–43 (2012).
67. Auer, P. Using confidence bounds for exploitation-exploration trade-offs. *J. Mach. Learn. Res.* **3**, 397–422 (2003).
68. Schrittwieser, J. et al. Mastering Atari, Go, chess and shogi by planning with a learned model. *Nature* **588**, 604–609 (2020).
69. Rosin, C. D. Multi-armed bandits with episode context. *Ann. Math. Artif. Intell.* **61**, 203–230 (2011).
70. Yang, Y. ParetoDrug numerical source data [https://figshare.com/articles/dataset/ParetoDrug\\_numerical\\_source\\_data/26304124](https://figshare.com/articles/dataset/ParetoDrug_numerical_source_data/26304124) (2024).
71. Yang, Y. ParetoDrug codes and data [https://figshare.com/articles/dataset/ParetoDrug\\_codes\\_and\\_data\\_zip/26309932](https://figshare.com/articles/dataset/ParetoDrug_codes_and_data_zip/26309932) (2024).
72. Prasanna, S. & Doerksen, R. Topological polar surface area: A useful descriptor in 2D-QSAR. *Curr. Med. Chem.* **16**, 21–41 (2009).
73. Olivecrona, M., Blaschke, T., Engkvist, O. & Chen, H. Molecular de-novo design through deep reinforcement learning. *J. Cheminf.* **9**, 48 (2017).
74. Korb, O., Stützle, T. & Exner, T. E. PLANTS: Application of ant colony optimization to structure-based drug design. In *Ant Colony Optimization and Swarm Intelligence*, **4150**, 247–258 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2006).
75. Friesner, R. A. et al. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **47**, 1739–1749 (2004).
76. 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).

## Acknowledgements

The work described in this article was supported by a grant from the Hong Kong Innovation and Technology Fund (Project No. ITS/241/21) and a grant from the National Natural Science Foundation of China (22377111).

## Author contributions

Pheng-Ann Heng and Guangyong Chen conceived the research topic. Yaodong Yang designed and developed the method and carried out the computational benchmark experiments. Odin Zhang and Ercheng Wang provided the necessary domain background and helped with experiments. Yaodong Yang, Guangyong Chen, and Ercheng Wang designed and performed the case studies. Odin Zhang, Junyou Li, and Xujun Zhang contributed to the computational analysis. Yaodong Yang, Ercheng Wang, Jinpeng Li, Guangyong Chen, Lanqing Li, and Jianye Hao participated in the writing of the paper. Guangyong Chen, Ercheng Wang, Jianye Hao, and Pheng-Ann Heng supervised the work. All the co-authors participated in the discussions and agreed with the contents of this work.

## 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/s42003-024-06746-w>.

**Correspondence** and requests for materials should be addressed to Guangyong Chen, Jianye Hao or Ercheng Wang.

**Peer review information** *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Laura Rodríguez Pérez.

**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 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 licence, and indicate if changes were made. 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/4.0/>.

© The Author(s) 2024