Journal of Cheminformatics

METHODOLOGY

Open Access

Activity cliff-aware reinforcement learning for *de novo* drug design

Xiuyuan Hu[1], Guoqing Liu[2], Yang Zhao[1] and Hao Zhang[1*]

Abstract

The integration of artificial intelligence (AI) in drug discovery offers promising opportunities to streamline and enhance the traditional drug development process. One core challenge in *de novo* molecular design is modeling complex structure-activity relationships (SAR), such as activity cliffs, where minor molecular changes yield significant shifts in biological activity. In response to the limitations of current models in capturing these critical discontinuities, we propose the Activity Cliff-Aware Reinforcement Learning (ACARL) framework. ACARL leverages a novel activity cliff index to identify and amplify activity cliff compounds, uniquely incorporating them into the reinforcement learning (RL) process through a tailored contrastive loss. This RL framework is designed to focus model optimization on high-impact regions within the SAR landscape, improving the generation of molecules with targeted properties. Experimental evaluations across multiple protein targets demonstrate ACARL's superior performance in generating high-affinity molecules compared to existing state-of-the-art algorithms. These findings indicate that ACARL effectively integrates SAR principles into the RL-based drug design pipeline, offering a robust approach for *de novo* molecular design

Scientific contribution Our work introduces a machine learning-based drug design framework that explicitly models activity cliffs, a first in AI-driven molecular design. ACARL's primary technical contributions include the formulation of an activity cliff index to detect these critical points, and a contrastive RL loss function that dynamically enhances the generation of activity cliff compounds, optimizing the model for high-impact SAR regions. This approach demonstrates the efficacy of combining domain knowledge with machine learning advances, significantly expanding the scope and reliability of AI in drug discovery.

Keywords AI for drug design, Activity cliff, Reinforcement learning, Contrastive loss

Introduction

The application of artificial intelligence (AI) in drug discovery has generated considerable enthusiasm for its potential to accelerate the traditionally lengthy and costly process of identifying effective drug molecules [1, 2]. *De novo* molecular design, where novel compounds are computationally generated to meet specific biological properties, is a particularly challenging domain within drug discovery due to its reliance on dealing with intricate structure-activity relationships (SAR) [3–6]. Despite recent progress, many AI-driven molecular design algorithms struggle to account for a crucial pharmacological phenomenon known as the activity cliff-a scenario where minor structural changes in a molecule lead to significant, often abrupt shifts in biological activity [7] (Fig. 1). Activity cliffs hold substantial value in medicinal chemistry, as understanding these discontinuities in SAR can guide the design of molecules with enhanced efficacy [8]. However, conventional molecular generation models

*Correspondence:
Hao Zhang
haozhang@tsinghua.edu.cn
[1] Department of Electronic Engineering, Tsinghua University, Beijing, China
[2] Microsoft Research AI for Science, Beijing, China

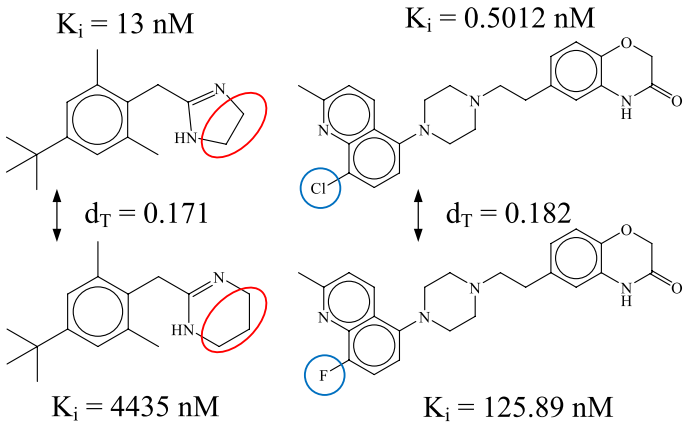

Fig. 1 Two examples of pairs of activity cliff compounds in the activity dataset on the 5HT1B target [9]. They illustrate that minor differences in molecular structure may result in orders-of-magnitude variations in biological activity, a phenomenon of particular interest in the field of pharmaceutical research

largely overlook this phenomenon, treating activity cliff compounds as statistical outliers rather than leveraging them as informative examples within the design process.

To address this gap, we introduce Activity Cliff-Aware Reinforcement Learning (ACARL), a novel framework specifically designed to incorporate activity cliffs into the *de novo* drug design process. ACARL enhances AI-driven molecular design by embedding domain-specific SAR insights directly within the reinforcement learning (RL) paradigm, targeting high-impact regions in molecular space for optimized drug candidate generation. The core innovations of ACARL lie in two key contributions:

- Activity Cliff Index (ACI): We propose a quantitative metric for detecting activity cliffs within molecular datasets. The ACI captures the intensity of SAR discontinuities by comparing structural similarity with differences in biological activity, enabling the model to systematically identify compounds that exhibit activity cliff behavior. This metric provides a novel tool to measure and incorporate discontinuities in SAR, bridging a longstanding gap in *de novo* molecular design.
- Contrastive Loss in RL: ACARL introduces a contrastive loss function within the RL framework that actively prioritizes learning from activity cliff compounds. By emphasizing molecules with substantial SAR discontinuities, the contrastive loss shifts the model's focus toward regions of high pharmacological significance. This unique approach contrasts with traditional RL methods, which often equally weigh all samples, and enhances ACARL's ability to generate molecules that align with complex SAR patterns seen in real-world drug targets.

To validate ACARL, we conducted comprehensive experiments targeting three biologically relevant proteins, demonstrating that our method surpasses current state-of-the-art algorithms in generating molecules with both high binding affinity and diverse structures. The experimental outcomes underscore the practical potential of ACARL in drug discovery, showcasing its ability to model SAR complexity in molecular generation more effectively than baseline approaches.

By addressing the limitations of existing *de novo* drug design frameworks, ACARL exemplifies a new approach in AI for drug discovery, where the integration of SAR-specific insights allows for more targeted and effective molecular design.

Related works

Activity cliffs in drug discovery

A quantitative depiction of activity cliffs involves two aspects: molecular similarity and activity. Two common criteria for molecular similarity are utilized; one is computed by Tanimoto similarity between molecular structure descriptors [10], and the other employs matched molecular pairs (MMPs) [11], defined as two compounds differing only at a single site (substructure). The biological activity of a molecule, also known as potency, is typically measured by the inhibitory constant (K_i) [12]. The ChEMBL database [13] contains millions of such activities, each of which records the binding affinity of a molecule against a protein target. Moreover, the relationship between the binding free energy ΔG obtained from docking software [14] and K_i is:

$$\text{Docking score}: \ \Delta G = RT \ln K_i, \tag{1}$$

where R is the universal gas constant (1.987 cal· K^{-1}· mol^{-1}), and T is the temperature (298.15 K). A lower K_i indicates a higher activity, as does the docking score.

The application of machine learning (ML) techniques in drug discovery poses significant challenges due to activity cliffs. In recent years, a multitude of quantitative structure-activity relationship (QSAR) models have been proposed to predict the bioactivity of molecules, yet they often make mistakes when applied to activity cliff compounds. Research has shown that the prediction performance of descriptor-based, graph-based, and sequence-based ML methods significantly deteriorates when dealing with activity cliff molecules [15]. Similarly, studies indicate that ML models tend to generate analogous predictions for structurally similar molecules, which is accurate for most cases but fails when predicting the potency of activity cliff compounds due to their statistical underrepresentation [16]. Further evidence suggests that neither enlarging the training set size nor increasing model complexity improves predictive accuracy for these challenging compounds [17]. Additionally, experimental findings indicate that existing QSAR models exhibit low sensitivity toward activity cliffs [18]. Moreover, experimental evidence provided by [18] suggests that existing QSAR models demonstrate low sensitivity towards activity cliffs. In conclusion, current ML techniques encounter difficulties when addressing the discontinuity in SAR.

Although activity cliffs have received attention in molecular property prediction, no efforts have been made thus far to consider this pivotal pharmacological problem in machine learning algorithms for drug design. Moreover, current prevalent benchmarks and oracles (i.e., scoring functions) for drug design fail to accurately emulate the objectives of real-world drug design. Their most frequent flaw is a lack of discontinuity in their scoring functions, that is, activity cliffs. For instance, this defect is evident in goal-directed molecular design tasks in the GuacaMol benchmark [19], as well as frequently utilized oracles such as LogP, DRD2 [20], JNK3, and GSK3β [21]. This obstructs the evolution of practically meaningful drug design algorithms. In contrast, structure-based docking software has been proven to reflect activity cliffs authentically [22, 23], leading to calls for the use of docking in the evaluation of drug design algorithms, as opposed to simpler scoring functions [24, 25].

Machine learning based drug design

Various advanced machine learning models and algorithms have been employed in drug design, including variational autoencoders (VAE) [26, 27], generative adversarial networks (GAN) [28], flow models [29], diffusion models [30], autoregressive models [31], genetic algorithm (GA) [32, 33], Bayesian optimization [34], active learning [35], and reinforcement learning (RL). Notably, methods utilizing RL and recurrent neural networks (RNN) to generate 1D SMILES (simplified molecular input line entry system) [36] strings are considered to exhibit the highest competitive edge [37].

RL-based methods for drug design Reinforcement learning is a machine learning paradigm that enables an agent to learn policies by interacting with its environment, thereby maximizing cumulative rewards. It is well-suited for the task of *de novo* drug design, which often lacks labeled data. Specifically, molecular scoring functions are considered as the environment providing feedback. 1D RL-based methods are typically used to train autoregressive generative models, guiding them toward generating molecules with high property scores [20, 38–40]. On the other hand, 2D methods treat the addition or modification of atoms, bonds, and rings as RL actions, facilitating the generation of satisfactory molecular graphs [41–43].

Language modeling for drug design Transformer models have recently made disruptive progress in the field of NLP, with the success of generative language models raising expectations for its application in other areas. In particular, transformer-based language models trained on chemical languages (primarily utilizing SMILES notation) have demonstrated potential in drug design tasks [44–47].

Methodology

In this section, we provide a comprehensive explanation of the ACARL algorithm that we have developed. We begin by mathematically delineating the *de novo* drug design problem. Subsequently, we introduce the notion of an activity cliff index, which enables the identification of activity cliff compounds, and then present an RL framework built upon a transformer decoder for molecular generation. We leverage the weighed loss to augment them during the RL fine-tuning process.

Problem definition

The drug-like chemical space, denoted as $\mathcal{S}$, is extremely vast, containing approximately 10^{33} synthesizable molecular structures [48].

A molecular scoring function f maps a molecule $x \in \mathcal{S}$ to a real value:

$$f : \mathcal{S} \to \mathbb{R}, \tag{2}$$

and this value represents a type of physical, chemical, or biological property of the molecule. For example, the docking score of a compound represents its binding affinity against a certain target. Typically, it is impracticable to express f via algebraic formulations.

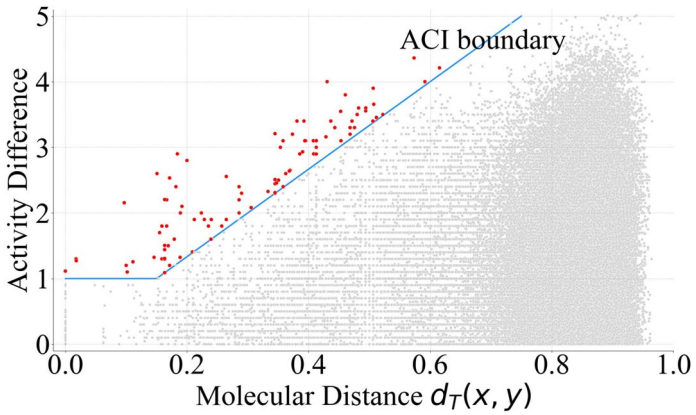

Fig. 2 The relationship between biological activity differences and molecular distances. Each data point corresponds to a pair of molecules from the 5HT1B target activity dataset. It is evident that although molecules with small molecular distances generally exhibit similar biological activities, there also exists a considerable part of counterexamples above the blue lines, i.e., activity cliffs

De novo drug design can be formulated using a combinatorial optimization (CO) problem, where the primary objective entails the discovery of molecular structures that maximize (or minimize) the scoring function f, i.e.

$$\arg\max_{x\in\mathcal{S}} f(x) \quad \text{or} \quad \arg\min_{x\in\mathcal{S}} f(x). \tag{3}$$

Identification of activity cliff compounds

Evidence of activity cliffs The distribution of biological activity differences vs. pair-wise Tanimoto distances of molecular pairs is depicted in Fig. 2, where biological activities are quantified using $pK_i = \log_{10} K_i$.[1] The red points indicate pairs of activity cliff molecules, and the blue line marks the activity cliff index boundary. We can observe that although molecules with small molecular distances typically display similar biological activities, there is also a negligible proportion of exceptions above the blue line, known as activity cliffs.

To effectively harness the activity cliffs, it is imperative to establish a quantitative approach for their identification. To measure the "smoothness" of the function f [49] defined on the discrete set S, we intuitively define an activity cliff index (**ACI**):

$$\mathrm{ACI}(x, y; f) := \frac{|f(x) - f(y)|}{d_T(x, y)}, \; x, y \in \mathcal{S}, \tag{4}$$

where d_T is the Tanimoto distance between the ECFPs of pairs of molecules. It is worth mentioning that ACI is a generalized form of the structure-activity landscape index (SALI) [50]. While SALI is specifically defined for activity values such as K_i, ACI extends this concept to measure any molecular property, including the docking score, which is the primary focus of this study. This generalization makes ACI more versatile for evaluating molecular properties beyond activity values.

Additionally, an alternative activity cliff criteria of MMP is not chosen for two primary reasons. First, MMP relies on a self-defined set of substructure replacements, which lacks general applicability. Second, it represents a binary variable, making it impossible to set a threshold within the algorithm.

Based on ACI, a pair of molecules is considered as an activity cliff if it satisfies two conditions: 1) the absolute value of the difference in property scores exceeds a certain threshold α_1; 2) the ACI surpasses a predetermined threshold α_2, as mathematically expressed in Eq. 8.

An RL framework with contrastive Loss for Drug Design

Following the paradigm of "pre-training + fine-tuning", ACARL designs drug candidates based on SMILES sequences, where reinforcement learning techniques are employed to iteratively fine-tune the pre-trained transformer model, steering it towards generating compounds with desired properties. The pre-trained model can sample molecules approximately uniformly over the known drug space, and we fine-tune it to generate molecular sequences with higher property scores[2] measured by f.

Specifically, in the RL framework, we follow the policy gradient loss function to update the model [20], which has been proven competitive in various molecular generation scenarios [51, 52]. It penalizes the square error between the log-likelihood of generating a molecule $x \in \mathcal{S}$ and its reward $R(x)$ via the policy gradient algorithm [53].

[1] The boundary $\alpha_1 = 1$ and $\alpha_2 = 6.67$ are chosen from [9].

[2] Without loss of generality, we presume that higher scores correspond to superior properties.

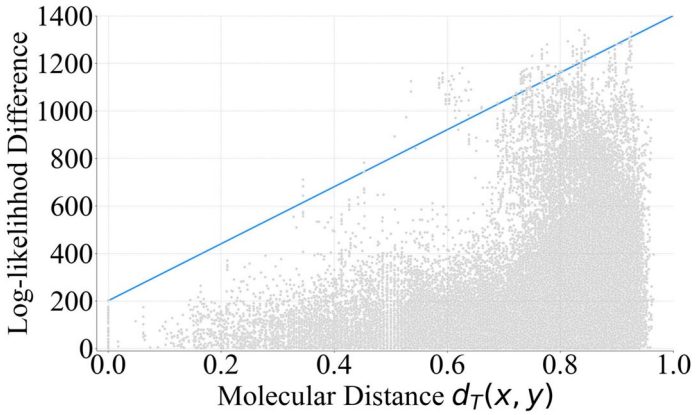

Fig. 3 The relationship between molecules' log-likelihood differences sampled from current language models and molecular distances. Each data point corresponds to a molecule pair from the 5HT1B target activity dataset. The result indicates that the chemical language model tends to assign similar generation probability to similar molecules, regardless of whether the pair is an activity cliff or not

$$\begin{aligned} L(x;\theta) &= [R(x) - \log \pi(x;\theta)]^2 \\ &= [\log \pi(x;\theta_{\text{pretrain}}) + \sigma \cdot f(x) \\ &\quad - \log \pi(x;\theta)]^2, \end{aligned} \tag{5}$$

in which $\pi(x;\cdot)$ represents the likelihood of sequence x being autoregressively generated by a language model, θ refers to the parameters of the RL agent, and θ_{pretrain} signifies the pre-trained parameters. The reward function R is composed of two components: one is the log-likelihood of sequence x generated by the pre-trained model, ensuring the agent retains its fundamental knowledge about S; the other is the scoring function f that rewards the agent for sampling molecules with superior properties, where σ is a hyper-parameter.

Contrastive loss At the k-th RL optimization step, the agent generates a batch of n molecules $\{x_1^{(k)}, x_2^{(k)}, ..., x_n^{(k)}\}$ for loss computation and model updating. Our objective is to optimize the log-probability distribution $\log \pi(x;\theta)$ to be close to the reward distribution $R(x)$, where $x \in S$. A traditional loss function is equally weighted across samples in the current step:

$$L(\theta^{(k)}) = \frac{1}{n}\sum_{i=1}^{n} L(x_i^{(k)};\theta^{(k)}). \tag{6}$$

Motivated by the imbalanced sample importance, balancing the loss function [54] and prioritized experience replay [55] have drawn great attention in recent years, which particularly apply to our scenario since we mainly focus on certain regions in the SAR. Moreover, in *de novo* drug design, the limited data accentuates the significance of sample efficiency. Therefore, we propose to utilize a contrastive RL loss with previous samples:

$$L_w(\theta^{(k)}) = \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{k} w_{ij}} \sum_{i=1}^{n}\sum_{j=1}^{k} w_{ij} L(x_i^{(j)};\theta^{(k)}), \tag{7}$$

where $w_{ij} \geq 0$ represents the weight coefficient of the i-th molecule sampled in the j-th step. This loss function serves as the basis for the augmentation of certain samples in the RL process.

All molecules generated during the RL process (m steps) are recorded in $\mathcal{M}_{all} = \{x_i^{(k)} : i = 1, ..., n; k = 1, ..., m\}$, which also serves as the output of the algorithm. They will be used to facilitate subsequent pharmaceutical research and development endeavours.

Augmenting activity cliffs in drug design

Based on the definition of the activity cliff compounds in section 3.2 and the RL framework with contrastive loss introduced in section 3.3, we will introduce how to incorporate these identified activity cliffs into the RL fine-tuning process.

Insights from medicinal chemistry In real-world drug design, activity cliffs often encompass crucial information about the SAR, and analysis of them indeed highly benefits drug discovery [56]. For instance, a type of PROTAC with high *in vivo* antitumor efficacy is discovered by the subtle modification of just one atom, which dramatically affects the degradation activities [57]. This indicates a promising lead for developing new chemotherapies targeting KRAS mutants.

Limitation of current models As investigated in Fig. 3, the blue line signifies a trend that molecule pairs with lesser Tanimoto distances are generated by the transformer model with smaller absolute differences in log-likelihood.

Notably, we can find that the likelihood of similar molecules from the transformer model is always similar, regardless of whether the pair is an activity cliff or not. Such continuity is naturally expected for typical deep networks that satisfy Lipschitz continuity conditions [49]. When considering the loss function in Eq. 5, it is possible for the property scores of two activity cliff compounds to differ significantly, but the generation probabilities may be relatively closer due to the constraints of Lipschitz conditions. This presents a challenge in fitting both molecules concurrently. Without special attention, the network is likely to regard activity cliffs as outliers.

Algorithm 1 ACARL

Input: θ_{pre} and $f(\cdot)$
Create $\mathcal{M}_{all}, \mathcal{M}_{AC}$, and initialize the RL agent θ with θ_{pre}
for $k = 1$ **to** m **do**
(1) Sample $\{x_1^{(k)}, x_2^{(k)}, ..., x_n^{(k)}\}$ using $\theta^{(k)}$
(2) Score the samples using $f(\cdot)$ and identify activity cliffs
(3) Update the samples and scores to $\mathcal{M}_{all}, \mathcal{M}_{AC}$
(4) Calculate the contrastive loss $L_w(\theta^{(k)})$ with augmentation of activity cliffs
(5) Update θ using $L_w(\theta^{(k)})$
end for
Output: $\mathcal{M}_{all}$

ACARL algorithm As illustrated in Algorithm 1, we propose the algorithm ACARL intended to augment activity cliff compounds in the RL framework by assigning higher weights to them. Two memories for molecules are constructed: $\mathcal{M}_{all}$ records all molecules sampled during the RL process sorted by property scores, and $\mathcal{M}_{AC}$ documents molecules that are identified as activity cliffs. At each RL step, all sampled molecules and their scores are updated into $\mathcal{M}_{all}$, and those that meet the following criteria are updated into $\mathcal{M}_{AC}$:

$$\mathbb{I}(|f(x) - f(y)| > \alpha_1) \cdot \mathbb{I}(\text{ACI}(x, y; f) > \alpha_2), \tag{8}$$

where $\mathbb{I}$ is the indicator function. The identification of activity cliffs is performed by comparing each newly generated molecule with all molecules in $\mathcal{M}_{all}$. This remains computationally feasible because fewer than 5% of molecular pairs require Tanimoto similarity calculations after filtering based on the absolute difference between property scores. Furthermore, over 10^5 molecular pairwise distances can be computed per second, ensuring the efficiency of the ACARL framework.

The weights in the loss $L_w(\theta^{(k)})$ are determined by:

$$\begin{aligned} w_{ij} = \mathbb{I}(j = k) &+ \text{Bern}(\frac{s_1}{t}) \cdot \mathbb{I}(x_i^{(j)} \in \text{top}_t(\mathcal{M}_{all})) \\ &+ \text{Bern}(\frac{s_2}{|\mathcal{M}_{AC}|}) \cdot \mathbb{I}(x_i^{(j)} \in \mathcal{M}_{AC}), \end{aligned} \tag{9}$$

where t, s_1 and s_2 denote integer hyper-parameters, $\text{Bern}(p)$ refers to a Bernoulli variable with the success rate p, and $\text{top}_t(\mathcal{M}_{all})$ contains the top-t highest-scoring molecules in $\mathcal{M}_{all}$. In this way, we augment the high-scoring compounds as well as activity cliff molecules. Here, we choose to augment all the high-scoring compounds instead of only molecules on the higher-scoring side of activity cliffs under the belief that for the sparsely distributed high-scoring molecules, there is likely to be potential activity cliffs near them that are not sampled.

Implementation

Molecule scoring To normalize different scoring functions for facilitating the application of a uniform algorithm to various drug design objectives and to mitigate the interference caused by invalid strings, we stipulate that the f used in the RL framework maps the molecular properties onto the range of [0, 1], with a higher score indicating superior properties. For invalid samples, the value taken by f is -1. For many scoring functions, including docking, it is necessary to devise transformations to convert the original scores of the molecules to fit within the [0, 1] range.

Transformer pre-trained on ChEMBL We train an autoregressive generative pre-trained transformer (GPT) decoder model from scratch for the purpose of generating chemical molecules. Following [44], our model employs a miniature version of the GPT-2 architecture [58], containing approximately 6.4M parameters and capable of producing sequences composed of up to 128

tokens. The ChEMBL dataset [13], comprised of roughly two million small molecule drugs in the form of SMILES strings, is utilized. After filtering out structures containing ions and excessively lengthy sequences, we employ unsupervised learning to train the transformer model, resulting in over 98% of the generated SMILES strings corresponding to valid molecules. This pre-trained model serves to initialize the RL agent with an identical architecture, and concurrently, it anchors the fine-tuned policy to the pre-trained one by the loss function.

Experiments

Experimental setup

Many *in silico* oracles and benchmarks mimicking drug discovery objectives are impractical in real-world drug discovery, especially since they do not exhibit the activity cliff phenomenon [23, 25]. Therefore, we choose molecular docking as the scoring function in *de novo* drug design against protein targets. Specifically, Quick Vina 2 [59] is employed to perform docking calculations, which is a commonly used docking software with high efficiency and accuracy.

Targets We choose three protein targets of high pharmaceutical value following [25] to design small molecule drug candidates against each of them:

(1) 5HT1B [60]: 5-hydroxytryptamine receptor 1B, which is associated with several neurological disorders, including depression, anxiety, aggressive behavior, and substance abuse.
(2) 5HT2B [61]: 5-hydroxytryptamine receptor 2B, which is often of concern in the treatment of heart valve disease and psychiatric disorders.
(3) ACM2 [62]: muscarinic acetylcholine receptor M2, which plays a significant role in the regulation of the heart rate and smooth muscle function.

Baselines We compare our algorithm with eight state-of-the-art baselines on molecular design: Reinvent [20], JT-VAE [26], GCPN [41], Reinvent 2.0 [63], MARS [64], GFlowNet [29], and LIMO [65]. For all the baselines, we adopt the default settings provided in their official codebases. Additionally, we report the corresponding results of the ChEMBL database [13, 25].

Ablation studies In addition to the default configuration of ACARL, we also conduct experiments on its four variants: (1) RL-base, the RL baseline without the contrastive loss; (2) ACARL-rand, a variant which randomly replay the same number of samples as the default setting from $\mathcal{M}_{all}$; (3) ACARL-high, another variant in which only higher-scoring molecules from $\mathcal{M}_{AC}$ are sampled and augmented; (4) ACARL-low, the final variant where solely lower-scoring molecules from $\mathcal{M}_{AC}$ are sampled and augmented.

Metrics To quantitatively assess the performance of each approach on each of the three targets, we report the best (Top-1) docking score (kcal/mol) and the average of the 100 best docking scores (Top-100) in each generated set of molecules. In addition, we also report the internal diversity (IntDiv) [66] of the "top-100" molecules.

Molecular design with high biological activities

Docking scores typically present as negative values, with lower numbers corresponding to better biological activity. To normalize them, we utilize the following transformation function:

$$f_{\mathrm{d}}(x) = \frac{1}{1 + 10^{\lambda \cdot [\mathrm{Dock}(x) - (\beta_u + \beta_l)/2]/(\beta_h - \beta_l)}} \in [0, 1], \tag{10}$$

where β_u and β_l respectively represent the upper and lower bounds of typical docking score values. They are set to -8 and -12 kcal/mol, respectively. The term λ represents a hyper-parameter which is fixed at 0.25 in our study.

For each target, ACARL is executed for a total of 1,000 RL steps with a batch size set at 128. The hyper-parameters are: $\alpha_1 = 0.5, \alpha_2 = 2, t = 100, s_1 = 20, s_2 = 20$, in which α_1 and α_2 are selected based on the boundaries in Fig. 2 and Eq. (1), and the performance is not sensitive to the values of t, s_1, and s_2, as shown in the supplementary material. The running process for each target takes less than 50 h on a single NVIDIA A100 GPU and 64 CPU cores.

Table 1 presents a comprehensive display of the numerical results of our experiments. The outcomes indicate that ACARL surpasses all the existing state-of-the-art baselines in the three docking-based *de novo* drug design tasks, particularly represented by the best Top-1 and Top-100 mean docking scores, with maintaining a comparable diversity level.

Furthermore, ACARL outperforms its four variants, of which the first and second ones indicate that activity cliff molecules are indeed more important for drug design than ordinary molecules. The third and fourth variants suggest that compounds from both higher-scoring and lower-scoring sides of the activity cliffs are beneficial for drug design.

Figure 4 illustrates the change curves of the number of activity cliffs detected and the batch-wise mean score during the operation of ACARL. The results suggest that prior to the identification of any activity cliffs, the model's enhancement progresses relatively slowly. While following the utilization of activity cliffs, there is a rapid increase in the mean score of sampled compounds,

Table 1 Experimental results of designing molecules with high biological activities. The IntDiv represents the internal diversity of the Top-100 molecules. Each value is the median of the results run under five different seeds except the Dataset

	5HT1B			5HT2B			ACM2		
	Top-1 (↓)	Top-100 (↓)	IntDiv (↑)	Top-1 (↓)	Top-100 (↓)	IntDiv (↑)	Top-1 (↓)	Top-100 (↓)	IntDiv (↑)
Dataset	− 14.4	− 11.3	0.819	−14.7	− 10.9	0.782	− 12.7	− 10.9	0.808
Reinvent	− 12.2	− 10.5	0.676	− 12.0	− 10.1	0.661	− 13.5	− 11.8	0.592
JT-VAE	− 8.5	− 6.7	0.862	− 8.8	− 6.5	0.850	− 9.4	− 6.8	0.823
GCPN	− 13.0	− 11.1	0.655	− 12.8	− 10.7	0.643	− 14.8	− 12.0	0.587
Reinvent 2.0	− 12.1	− 10.2	0.708	− 12.3	− 10.3	0.692	− 13.4	− 11.6	0.641
MARS	− 14.6	− 12.5	0.532	− 14.1	− 11.9	0.547	− 15.2	− 12.8	0.519
GFlowNet	− 10.8	− 8.4	0.770	− 10.7	− 8.4	0.754	− 12.2	− 9.7	0.698
LIMO	− 15.2	− 11.9	0.591	− 14.7	− 11.5	0.618	− 15.9	− 14.3	0.502
RL-base	− 13.4	− 11.6	0.625	− 13.3	− 11.1	0.620	− 14.6	− 13.3	0.551
ACARL-rand	− 13.4	− 11.4	0.633	− 13.2	− 11.0	0.619	− 14.7	− 13.3	0.540
ACARL-high	− 14.9	− 12.8	0.582	− 14.6	− 11.7	0.597	− 16.4	− 14.0	0.515
ACARL-low	− 13.5	− 11.7	0.601	− 13.7	− 11.5	0.624	− 14.9	− 13.2	0.536
ACARL	− 15.6	− 13.0	0.563	− 15.0	− 12.2	0.589	− 17.0	− 14.6	0.525

which gradually approaches 1. This may reveal the reason for the better performance of ACARL.

Figure 5 gives an example of an activity cliff pair of compounds identified by ACARL. After the RL process, the difference in their generation log-likelihood of the agent is 162.45, which is large enough to distinguish them. However, that of the pre-trained transformer model is only 31.37. This is strong evidence that the model is indeed better at characterizing activity cliffs by our augmentation.

Molecular design with multi-property objectives

In real-world drug discovery, in addition to the biological activity of molecules, there are other properties related to downstream development that should be considered. Here we combine docking with two other commonly used oracles, QED (quantitative estimate of drug-likeness) [67] and SA (synthetic accessibility) [68], to establish multi-property objectives (MPO) for molecular design. QED (↑) ranges in [0, 1], and SA (↓) ranges in [1, 10].

Our objective is to obtain drug candidates with desirable scores for all three properties. Therefore, we employ a linear combination of normalized oracles as the scoring function:

$$f_{\mathrm{MPO}}(x) = v_1 \cdot f_{\mathrm{d}}(x) + v_2 \cdot \mathrm{QED}(x) + v_3 \cdot \frac{10 - \mathrm{SA}(x)}{9}, \tag{11}$$

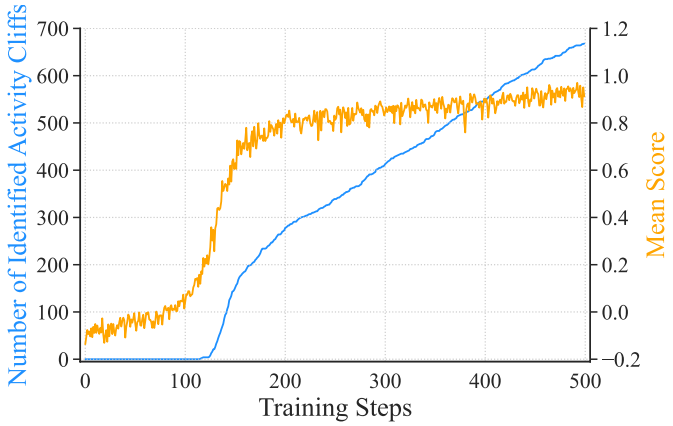

Fig. 4 The curves of the number of identified activity cliffs and the mean score of each batch of molecules during the RL process of ACARL (on the 5HT1B target)

where $\nu_1 + \nu_2 + \nu_3 = 1$. Here we set $(\nu_1, \nu_2, \nu_3) = (0.6, 0.2, 0.2)$ since the docking score is more crucial. For ACARL, we adopt the same hyper-parameter settings as in the previous section, except $\alpha_1 = 0.4$ and $\alpha_2 = 1.6$.

The molecule with the highest combined score generated by ACARL against each of the three targets is visualized in Fig. 6. The candidates have desirable docking and SA scores, with simple structures that bind well with the target pockets. The QED values are not particularly high, indicating that potential drugs of these three targets may not closely resemble existing drug molecules.

Admittedly, there is potential for refinement in the design of the combined scoring function. However, this MPO experiment primarily serves to demonstrate the flexibility of our algorithm in aligning with design objectives. This characteristic facilitates its convenient application in real-world drug design scenarios. Specifically, it allows medicinal chemists to cater to their unique requirements directly through the combination of scoring functions.

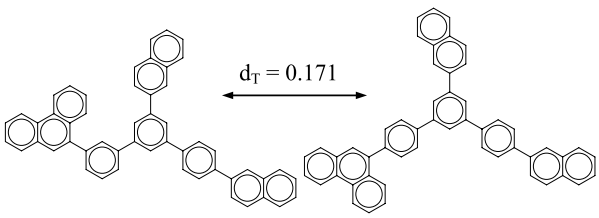

Fig. 5 An example of an activity cliff pair of compounds identified by ACARL during designing drug candidates against the 5HT1B target. The difference in their generation log-likelihood by the transformer agent significantly increases through ACARL

Conclusion and discussion

This paper presents Activity Cliff-Aware Reinforcement Learning (ACARL), a novel framework for *de novo* drug design that directly incorporates activity cliffs-a critical aspect of the structure-activity relationship (SAR)-into the molecular generation process. By leveraging an activity cliff index (ACI) and a contrastive reinforcement learning framework, ACARL addresses key limitations in existing drug design algorithms, which often overlook the discontinuities in SAR represented by activity cliffs. Experimental results demonstrate that ACARL generates diverse molecules with superior binding affinities, surpassing state-of-the-art baselines on three biologically relevant protein targets. These findings underscore ACARL's effectiveness in aligning AI-driven molecular design with practical pharmacological objectives.

The incorporation of activity cliffs into molecular design has significant implications for drug discovery, where capturing SAR discontinuities can lead to more potent and selective compounds. ACARL's focus on activity cliffs offers a pathway for generating molecules with high pharmacological relevance, particularly beneficial in early-stage discovery where identifying lead compounds with strong target affinity and diverse structures is critical. By embedding SAR-specific knowledge directly into the RL process, ACARL sets a new benchmark for integrating domain insights within machine learning frameworks, enhancing both the efficiency and robustness of molecular design.

Despite its advantages, ACARL has certain limitations that warrant further investigation. First, the evaluation is strongly reliant on docking, which, while capable of simulating activity cliff phenomena, still exhibits discrepancies with *in vivo* compound behavior. The development of improved *in silico* design objectives is needed, and we will pursue wet-lab experiments

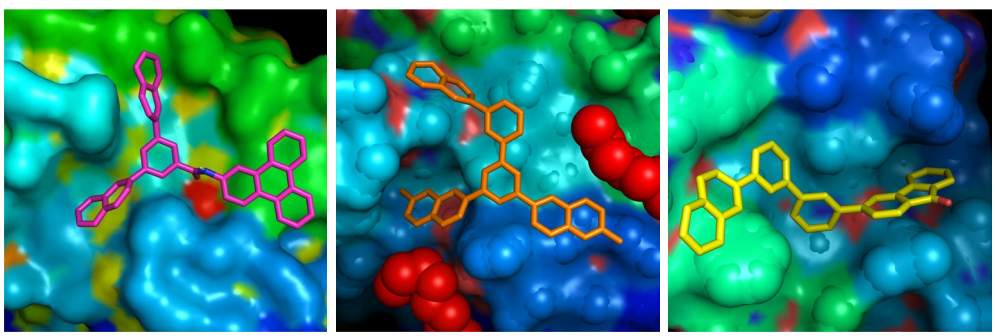

Fig. 6 The molecule with the best combined score against each of the three targets generated by ACARL. The (Docking, QED, SA) scores of them are: 5HT1B: (−13.3, 0.200, 2.312); 5HT2B: (−12.1, 0.201, 2.087), ACM2: (−12.9, 0.258, 1.980)

if conditions permit. Second, ACARL focuses on activity cliffs, yet this concept remains poorly defined, lacking a clear mechanistic understanding. Integrating more inherent knowledge about activity cliffs could refine the approach and improve interpretability. Lastly, while ACARL marks a substantial advancement, further exploration is necessary. Enhancing the algorithm's understanding of activity cliffs may benefit from incorporating specific structural information related to these cliffs. However, such an approach would require significant innovation in model architecture. We aspire to a future where AI can accurately model and comprehend complex SAR, potentially ushering in a new era of drug discovery and development.

In conclusion, ACARL represents a robust framework that bridges AI-driven molecular generation with domain-specific SAR insights. By focusing on critical pharmacological features such as activity cliffs, it demonstrates a practical and scalable approach to drug discovery. We encourage further research that combines scientific principles with machine learning advancements to continue enhancing the capabilities and applications of AI in drug design.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13321-025-01006-3.

Supplementary file 1.

Author contributions

Xiuyuan Hu wrote the main manuscript and all authors reviewed and revised the manuscript.

Funding

Not applicable.

Availability of data and materials

The code and data for ACARL is available at: https://github.com/HXYfighter/ACARL.

Declarations

Competing interests

The authors declare no competing interests.

Received: 27 October 2024 Accepted: 30 March 2025
Published online: 21 April 2025

References

1. Wang H, Fu T, Du Y, Gao W, Huang K, Liu Z, Chandak P, Liu S, Van Katwyk P, Deac A et al (2023) Scientific discovery in the age of artificial intelligence. Nature 620(7972):47–60
2. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A et al (2021) Highly accurate protein structure prediction with alphafold. Nature 596(7873):583–589
3. Du Y, Fu T, Sun J, Liu S (2022) Molgensurvey: A systematic survey in machine learning models for molecule design. arXiv preprint arXiv:2203.14500
4. Bajorath J, Chávez-Hernández AL, Duran-Frigola M, Fernández-de Gortari E, Gasteiger J, López-López E, Maggiora GM, Medina-Franco JL, Méndez-Lucio O, Mestres J et al (2022) Chemoinformatics and artificial intelligence colloquium: progress and challenges in developing bioactive compounds. J Cheminform 14(1):82
5. Deng J, Yang Z, Ojima I, Samaras D, Wang F (2022) Artificial intelligence in drug discovery: applications and techniques. Brief Bioinform 23(1):430
6. Hu X, Liu G, Yao Q, Zhao Y, Zhang H (2024) Hamiltonian diversity: effectively measuring molecular diversity by shortest hamiltonian circuits. J Cheminform 16(1):94
7. Maggiora GM (2006) On outliers and activity cliffs why qsar often disappoints. J Chem Inf Model 46(4):1535–1535
8. Stumpfe D, Hu H, Bajorath J (2019) Evolving concept of activity cliffs. ACS Omega 4(11):14360–14368
9. Zhang Z, Zhao B, Xie A, Bian Y, Zhou S (2023) Activity cliff prediction: Dataset and benchmark. arXiv preprint arXiv:2302.07541
10. Stumpfe D, Bajorath J (2012) Exploring activity cliffs in medicinal chemistry: miniperspective. J Med Chem 55(7):2932–2942
11. Hu X, Hu Y, Vogt M, Stumpfe D, Bajorath J (2012) Mmp-cliffs: systematic identification of activity cliffs on the basis of matched molecular pairs. J Chem Inf Model 52(5):1138–1145
12. Neubig RR, Spedding M, Kenakin T, Christopoulos A (2003) International union of pharmacology committee on receptor nomenclature and drug classification xxxviii. update on terms and symbols in quantitative pharmacology. Pharmacol Rev 55(4):597–606
13. Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, Magariños MP, Mosquera JF, Mutowo P, Nowotka M et al (2019) Chembl: towards direct deposition of bioassay data. Nucl Acids Res 47(D1):930–940
14. Trott O, Olson AJ (2010) Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31(2):455–461
15. Tilborg D, Alenicheva A, Grisoni F (2022) Exposing the limitations of molecular machine learning with activity cliffs. J Chem Inf Model 62(23):5938–5951
16. Janela T, Bajorath J (2023) Anatomy of potency predictions focusing on structural analogues with increasing potency differences including activity cliffs. J Chem Inf Model 63(22):7032–7044
17. Tamura S, Miyao T, Bajorath J (2023) Large-scale prediction of activity cliffs using machine and deep learning methods of increasing complexity. J Cheminform 15(1):4
18. Dablander M, Hanser T, Lambiotte R, Morris GM (2023) Exploring qsar models for activity-cliff prediction. J Cheminform 15(1):47
19. Brown N, Fiscato M, Segler MH, Vaucher AC (2019) Guacamol: benchmarking models for de novo molecular design. J Chem Inf Model 59(3):1096–1108
20. Olivecrona M, Blaschke T, Engkvist O, Chen H (2017) Molecular de-novo design through deep reinforcement learning. J Cheminform 9(1):1–14
21. Li Y, Zhang L, Liu Z (2018) Multi-objective de novo drug design with conditional graph generative model. J Cheminform 10(1):1–24
22. Husby J, Bottegoni G, Kufareva I, Abagyan R, Cavalli A (2015) Structure-based predictions of activity cliffs. J Chem Inf Model 55(5):1062–1076
23. Thomas M, Smith RT, O'Boyle NM, Graaf C, Bender A (2021) Comparison of structure-and ligand-based scoring functions for deep generative models: a gpcr case study. J Cheminform 13(1):1–20
24. Tripp A, Chen W, Hernández-Lobato JM (2022) An evaluation framework for the objective functions of de novo drug design benchmarks. In: ICLR2022 Machine Learning for Drug Discovery
25. Ciepliński T, Danel T, Podlewska S, Jastrzebski S (2023) Generative models should at least be able to design molecules that dock well: A new benchmark. J Chem Inf Model 63(11):3238–3247
26. Jin W, Barzilay R, Jaakkola T (2018) Junction tree variational autoencoder for molecular graph generation. In: International Conference on Machine Learning, pp. 2323–2332 PMLR
27. Kong X, Huang W, Tan Z, Liu Y (2022) Molecule generation by principal subgraph mining and assembling. Adv Neural Inf Process Syst 35:2550–2563
28. Guimaraes GL, Sanchez-Lengeling B, Outeiral C, Farias PLC, Aspuru-Guzik (2017) A Objective-reinforced generative adversarial networks

(organ) for sequence generation models. arXiv preprint arXiv:1705.10843
29. Bengio E, Jain M, Korablyov M, Precup D, Bengio Y (2021) Flow network based generative models for non-iterative diverse candidate generation. Adv Neural Inf Process Syst 34:27381–27394
30. Huang L, Zhang H, Xu T, Wong K-C (2023) Mdm: Molecular diffusion model for 3d molecule generation. Proc AAAI Conf Artif Intell 37(4):5105–5112
31. Segler MH, Kogej T, Tyrchan C, Waller MP (2018) Generating focused molecule libraries for drug discovery with recurrent neural networks. ACS Central Sci 4(1):120–131
32. Jensen JH (2019) A graph-based genetic algorithm and generative model/monte carlo tree search for the exploration of chemical space. Chem Sci 10(12):3567–3572
33. Fu T, Gao W, Coley CW, Sun J (2022) Reinforced genetic algorithm for structure-based drug design In: Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems NeurIPS
34. Moss H, Leslie D, Beck D, Gonzalez J, Rayson P (2020) Boss: Bayesian optimization over string spaces. Adv Neural Inf Process Syst 33:15476–15486
35. Kyro GW, Morgunov A, Brent RI, Batista VS (2024) Chemspaceal: An efficient active learning methodology applied to protein-specific molecular generation. J Chem Inf Model 64(3):653–665
36. Weininger D (1988) Smiles a chemical language and information system 1 introduction to methodology and encoding rules. J Chem Inf Comput Sci 28(1):31–36
37. Gao W, Fu T, Sun J, Coley C (2022) Sample efficiency matters: a benchmark for practical molecular optimization. Adv Neural Inf Process Syst 35:21342–21357
38. Blaschke T, Engkvist O, Bajorath J, Chen H (2020) Memory-assisted reinforcement learning for diverse molecular de novo design. J Cheminform 12(1):1–17
39. Wang J, Hsieh C-Y, Wang M, Wang X, Wu Z, Jiang D, Liao B, Zhang X, Yang B, He Q et al (2021) Multi-constraint molecular generation based on conditional transformer, knowledge distillation and reinforcement learning. Nat Mach Intell 3(10):914–922
40. Hu X, Liu G, Zhao Y, Zhang H (2023) De novo drug design using reinforcement learning with multiple gpt agents. In: Thirty-seventh Conference on Neural Information Processing Systems
41. You J, Liu B, Ying Z, Pande V, Leskovec J (2018) Graph convolutional policy network for goal-directed molecular graph generation. Adv Neural Inf Process Syst 1:31
42. Jin W, Barzilay R, Jaakkola T (2020) Multi-objective molecule generation using interpretable substructures. In: International Conference on Machine Learning (ICML), pp. 4849–4859. PMLR
43. Yang S, Hwang D, Lee S, Ryu S, Hwang SJ (2021) Hit and lead discovery with explorative rl and fragment-based molecule generation. Adv Neural Inf Process Syst 34:7924–7936
44. Bagal V, Aggarwal R, Vinod PK, Priyakumar UD (2022) Molgpt: Molecular generation using a transformer-decoder model. J Chem Inf Model 62(9):2064–2076
45. Irwin R, Dimitriadis S, He J, Bjerrum EJ (2022) Chemformer: a pre-trained transformer for computational chemistry. Mach Learn Sci Technol 3(1):015022
46. Wei L, Fu N, Song Y, Wang Q, Hu J (2023) Probabilistic generative transformer language models for generative design of molecules. J Cheminform 15(1):88
47. Aksamit N, Hou J, Li Y, Ombuki-Berman B (2024) Integrating transformers and many-objective optimization for drug design. BMC Bioinform 25(1):208
48. Polishchuk PG, Madzhidov TI, Varnek A (2013) Estimation of the size of drug-like chemical space based on gdb-17 data. J Comput-Aided Mol Des 27(8):675–679
49. Sohrab HH (2003) Basic Real Analysis, vol 231. Springer, Cham
50. Guha R, Van Drie JH (2008) Structure-activity landscape index: identifying and quantifying activity cliffs. J Chem Inf Model 48(3):646–658
51. Loeffler HH, He J, Tibo A, Janet JP, Voronov A, Mervin LH, Engkvist O (2024) Reinvent 4: Modern ai-driven generative molecule design. J Cheminform 16(1):20
52. Svensson HG, Tyrchan C, Engkvist O, Chehreghani MH (2023) Utilizing reinforcement learning for de novo drug design. arXiv preprint arXiv:2303.17615
53. Sutton RS, McAllester D, Singh S, Mansour Y (1999) Policy gradient methods for reinforcement learning with function approximation. Advances in neural information processing systems 12
54. Ren J, Zhang M, Yu C, Liu Z (2022) Balanced mse for imbalanced visual regression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 7926–7935
55. Schaul T, Quan J, Antonoglou I, Silver D (2016) Prioritized experience replay. In: International Conference on Learning Representations
56. Stumpfe D, Hu Y, Dimova D, Bajorath J (2014) Recent progress in understanding activity cliffs and their utility in medicinal chemistry: miniperspective. J Med Chem 57(1):18–28
57. Zhou Z, Zhou G, Zhou C, Fan Z, Cui R, Li Y, Li R, Gu Y, Li H, Ge Z et al (2023) Discovery of a potent, cooperative, and selective sos1 protac zz151 with in vivo antitumor efficacy in kras-mutant cancers. J Med Chem 66(6):4197–4214
58. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I et al (2019) Language models are unsupervised multitask learners. OpenAI Blog 1(8):9
59. Alhossary A, Handoko SD, Mu Y, Kwoh C-K (2015) Fast, accurate, and reliable molecular docking with quickvina 2. Bioinformatics 31(13):2214–2216
60. Wang C, Jiang Y, Ma J, Wu H, Wacker D, Katritch V, Han GW, Liu W, Huang X-P, Vardy E et al (2013) Structural basis for molecular recognition at serotonin receptors. Science 340(6132):610–614
61. Liu W, Wacker D, Gati C, Han GW, James D, Wang D, Nelson G, Weierstall U, Katritch V, Barty A et al (2013) Serial femtosecond crystallography of g protein-coupled receptors. Science 342(6165):1521–1524
62. Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, Weis WI, Okada T, Kobilka BK, Haga T et al (2012) Structure of the human m2 muscarinic acetylcholine receptor bound to an antagonist. Nature 482(7386):547–551
63. Blaschke T, Arús-Pous J, Chen H, Margreitter C, Tyrchan C, Engkvist O, Papadopoulos K, Patronov A (2020) Reinvent 2.0: an ai tool for de novo drug design. J Chem Inf Model 60(12):5918–5922
64. Xie Y, Shi C, Zhou H, Yang Y, Zhang W, Yu Y, Li L (2021) Mars: Markov molecular sampling for multi-objective drug discovery. In: International Conference on Learning Representations
65. Eckmann P, Sun K, Zhao B, Feng M, Gilson MK, Yu R (2022) Limo: Latent inceptionism for targeted molecule generation. In: International Conference on Machine Learning PMLR
66. Benhenda M (2018) Can ai reproduce observed chemical diversity? bioRxiv, 292177
67. Bickerton GR, Paolini GV, Besnard J, Muresan S, Hopkins AL (2012) Quantifying the chemical beauty of drugs. Nat Chem 4(2):90–98
68. Ertl P, Schuffenhauer A (2009) Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. J Cheminform 1(1):1–11

Publisher's Note