

# DrugMCTS: a drug repurposing framework combining multi-agent, RAG and Monte Carlo Tree Search

Zerui Yang 

Yuwei Wan

Yinqiao Li  
Linqi Song

Yudai Matsuda

Tong Xie

June 2025

## Abstract

Recent advances in large language models have demonstrated considerable potential in scientific domains such as drug discovery. However, their effectiveness remains constrained when reasoning extends beyond the knowledge acquired during pre-training. Conventional approaches, such as fine-tuning or retrieval-augmented generation, face limitations in either imposing high computational overhead or failing to fully exploit structured scientific data. To overcome these challenges, we propose DrugMCTS, a novel framework that synergistically integrates RAG, multi-agent collaboration, and Monte Carlo Tree Search for drug repurposing. The framework employs five specialized agents tasked with retrieving and analyzing molecular and protein information, thereby enabling structured and iterative reasoning. Without requiring domain-specific fine-tuning, DrugMCTS empowers Qwen2.5-7B-Instruct to outperform Deepseek-R1 by over 20%. Extensive experiments on the DrugBank and KIBA datasets demonstrate that DrugMCTS achieves substantially higher recall and robustness compared to both general-purpose LLMs and deep learning baselines. Our results highlight the importance of structured reasoning, agent-based collaboration, and feedback-driven search mechanisms in advancing LLM applications for drug discovery.

## 1 Introduction

Large language models (LLMs) have demonstrated remarkable capabilities across a wide range of domains, including question answering, logical reasoning, and knowledge-intensive tasks such as mathematics and code generation. These models are increasingly being explored for applications in scientific fields, particularly in drug discovery [1]. However, when confronted with problems that lie beyond their pre-training knowledge or inherent reasoning abilities, such as predicting novel drug-target interactions,

their performance may fall short of expectations [2].

As interest grows in applying general-purpose large models to scientific domains, various approaches have been proposed to address the aforementioned limitations. Among them, fine-tuning on domain-specific datasets has become a widely adopted paradigm [3] [4]. However, this approach comes with several notable drawbacks. First, fine-tuning demands substantial computational resources, and most current methods are tailored to specific domains. Extending these models to new domains typically requires additional training, which is computationally inefficient and often impractical for many users.

Moreover, scientific knowledge is inherently dynamic, constantly evolving with new discoveries. This necessitates continuous updates to the model’s internal knowledge base, which in turn introduces challenges such as increased training costs and the risk of catastrophic forgetting [5]. These issues significantly limit the scalability and practicality of fine-tuning-based approaches in real-world scientific applications.

To mitigate the limitations of fine-tuning, an alternative paradigm, retrieval-augmented generation (RAG), has gained increasing attention [6]. In this framework, external agents are employed to retrieve relevant information from scientific literature and databases, which is then used to augment the reasoning capabilities of LLMs [7]. This approach eliminates the need for fine-tuning and enables access to up-to-date knowledge, making it particularly appealing for fast-evolving domains like drug discovery.

However, significant challenges remain. As noted in prior work [38], Data in scientific domain can be broadly categorized into two types: scientific data and general-purpose data. Most existing RAG-based systems heavily rely on the latter due to its compatibility with general-purpose LLMs [8], while largely overlooking the value of structured, authoritative scientific data, such as molecular structures and protein sequences, which is often cleaner and more informa-

tive than generic text. For instance, in drug-target interaction prediction tasks, some approaches only provide knowledge graphs or textual descriptions as context, completely omitting structural information about proteins or molecules [10]. This omission undermines both the reliability of the model’s predictions and the interpretability of its decision-making process.

Some approaches incorporate domain-specific models during the reasoning phase to better understand scientific data [9]. While effective in certain contexts, this strategy inherits many of the limitations of the fine-tuning paradigm. Notably, when the input data deviates from the distribution seen during training, the performance of these models can degrade significantly. For example, deep learning models for drug-target interaction prediction may experience accuracy drops of over 20% when encountering previously unseen molecule-protein pairs [11].

On the other hand, general-purpose data sources often contain irrelevant or even erroneous information [12], necessitating preprocessing steps such as filtering and cleaning before use. However, LLMs may inadvertently discard useful content due to incomplete contextual understanding or internal biases. As highlighted in [1], drug discovery should ideally be an iterative process driven by feedback and refinement. Yet, most current approaches rely on single-step inference without mechanisms for error correction or knowledge updating, limiting their robustness and adaptability [36] [37].

To address the aforementioned limitations, we propose DrugMCTS, a novel drug discovery algorithm based on RAG, multi-agent collaboration, and Monte Carlo Tree Search (MCTS) [34]. Our system comprises five specialized agents:

- Retrieval Agent: Identifies and gathers potentially relevant molecules.
- Molecule-Analysis Agent: Evaluates the physicochemical and pharmacological properties of the query molecule.
- Molecule-Selection Agent: Filters out molecules with limited therapeutic potential or low relevance.
- Interaction-Analysis Agent: Interprets the molecular mechanisms underlying drug-target interactions.
- Decision Agent: Integrates all available evidence and generates final recommendations.

With the guidance of MCTS, our lightweight Qwen7b-based model [40] achieves performance exceeding that of Deepseek-R1 [33] and GPT-4o-mini [32], surpassing it by approximately 20% on key benchmarks. We also conduct ablation studies demonstrating that removing any component of our pipeline results in a performance drop of 2–10%. Our key contributions include:

- a. We introduce an end-to-end drug discovery framework that enables Qwen2.5-7B-Instruct to outperform much larger models Deepseek-R1. The method does not require any domain-specific fine-tuned models; instead, it leverages external knowledge to enhance reasoning and decision-making.
- b. We propose a systematic workflow that integrates scientific data, hybrid scientific-general data, and general-purpose data, leveraging the strengths of each type. This pipeline not only improves the interpretability of model decisions but also establishes a standardized framework applicable beyond drug-target interaction tasks.
- c. By incorporating a feedback mechanism through Monte Carlo Tree Search, our framework enables the model to iteratively refine its understanding, filter noisy data, and autonomously identify the most valuable information for decision-making.

## 2 Method

### 2.1 Overview

Our framework, DrugMCTS, as shown in Figure 1, takes a query molecule  $M_{qm}$  as input and processes it through a series of specialized agents to output proteins that can potentially interact with  $M_{qm}$ . The process begins with the Retrieval Agent, which queries databases based on the molecular structure of the  $M_{qm}$  to identify similar molecules, thereby generating a pool of candidate molecules  $M_{cm}$ . Next, the Molecule-Analysis Agent retrieves chemical properties of the query molecule via API calls and generates a detailed report  $R_{qm}$ . This report is then passed to the Molecule-Selection Agent, which filters the  $M_{cm}$  based on the analysis, producing a refined list of reference molecules  $M_{rm}$ . The Interaction-Analysis Agent identifies proteins  $P_{rp}$  that can interact with the reference molecules  $M_{rm}$  by querying relevant databases. For each identified protein-molecule pair, the agent retrieves the corresponding binding pocket data  $D_{bp}$  from the Protein Data Bank (PDB) [13] using API calls and python libraries. Additionally,

it fetches literature descriptions  $L_{rp}$  related to the proteins from public repositories such as PubMed. It then compiles all this information,  $M_{rm}$ ,  $P_{rp}$ ,  $P_{bp}$  and  $L_{rp}$ , into a comprehensive interaction report  $R_{ia}$  for each molecule-protein pair.

Finally, the Decision Agent integrates all available information, including the  $M_{qm}$ ,  $R_{qm}$ ,  $M_{rm}$ ,  $D_{bp}$ , and  $R_{ia}$ , to select the most promising target protein.

To enhance decision-making, we employ MCTS during inference time scaling. Unlike the retrieval agent, other agents generate multiple answers in a single invocation. These answers are evaluated using the Upper Confidence Bound applied to Trees (UCT) algorithm (Equation 1) [19] to select and expand optimal nodes until reaching the leaf node. We use a self-confidence scoring method for leaf nodes and backpropagate scores accordingly. The final selection of proteins is determined by majority voting, choosing the top N proteins that appear most frequently. This approach allows our model to autonomously select high-quality data without additional fine-tuning, effectively achieving data cleaning through feedback mechanisms.

$$\text{UCT} = \frac{W_i}{n_i} + c\sqrt{\frac{\ln N}{n_i}} \quad (1)$$

## 2.2 Data Processing Pipeline

In prior work [38], data in scientific domain has been categorized into two types: scientific data and general data. Scientific data includes highly specialized information such as molecular formulas and amino acid sequences, characterized by its authoritative nature, well-structured format, and cleanliness. These datasets are typically sourced from professional databases like the PDB [13], where data storage formats adhere to strict standards. However, LLMs often struggle to interpret such structured data directly due to its complexity, necessitating the use of domain-specific models like ESM2 [14] or ChemBERTa [15] for processing.

Conversely, general data encompasses more diverse and less standardized sources, such as research papers, which are easier for LLMs to understand but may contain noise and inaccuracies. Leveraging the strengths of both data types while mitigating their weaknesses is crucial for effective drug discovery.

Based on the original work, we propose a new data type called hybrid scientific-general data. Tools like RDKit [16] can accept scientific inputs, such as SMILES representations of molecules, and output structured yet textually described data, including chiral centers, functional groups, and Murcko scaffolds,

packaged as Python dictionaries. Similarly, PLIP [17] methods can process PDB files to extract pocket information and return it in tabular form. These hybrid data outputs retain the authority and well-structured nature of scientific data while being more interpretable by LLMs due to their textual explanations.

Our workflow involves transforming both molecular and protein data through three stages: from scientific data to hybrid scientific-general data, and finally to general data. This structured pipeline ensures that our model benefits from the precision and reliability of scientific data while maintaining interpretability and ease of integration with LLMs. By systematically leveraging these different data types, we not only improve the accuracy and robustness of our predictions but also enhance the explainability of our model’s decision-making process.

## 2.3 Formulating Drug Discovery as a Tree Search Problem

Given that MCTS is a well-established algorithm widely adopted and thoroughly explained in various studies [18], we provide only an overview here. The MCTS algorithm consists of four main steps: selection, expansion, simulation, and backpropagation.

In each rollout, the process begins at the root node and proceeds by selecting a leaf node using the UCT algorithm. The selected leaf node is then expanded by generating one or more child nodes. This selection-expansion cycle continues iteratively until reaching an end node. Upon reaching this terminal node, a predefined scoring rule is applied to evaluate its quality. The score is then backpropagated up the tree to update the scores and visit counts of all nodes along the path. After completing a pre-defined number of rollouts, we obtain a series of candidate solutions from which the final answer is chosen.

## 2.4 Action Space

The action space refers to the set of possible actions that can be performed during each expansion phase. Similar to chain-of-thought (CoT) reasoning, these actions are sequential and interdependent; each subsequent action builds upon the results of the previous one. Therefore, coordination among different agents is essential for coherent execution.

Our framework includes six distinct actions corresponding to five specialized agents. Except for the retrieval agent, which does not invoke any model, all other agents utilize the same LLM without requiring additional fine-tuned models.

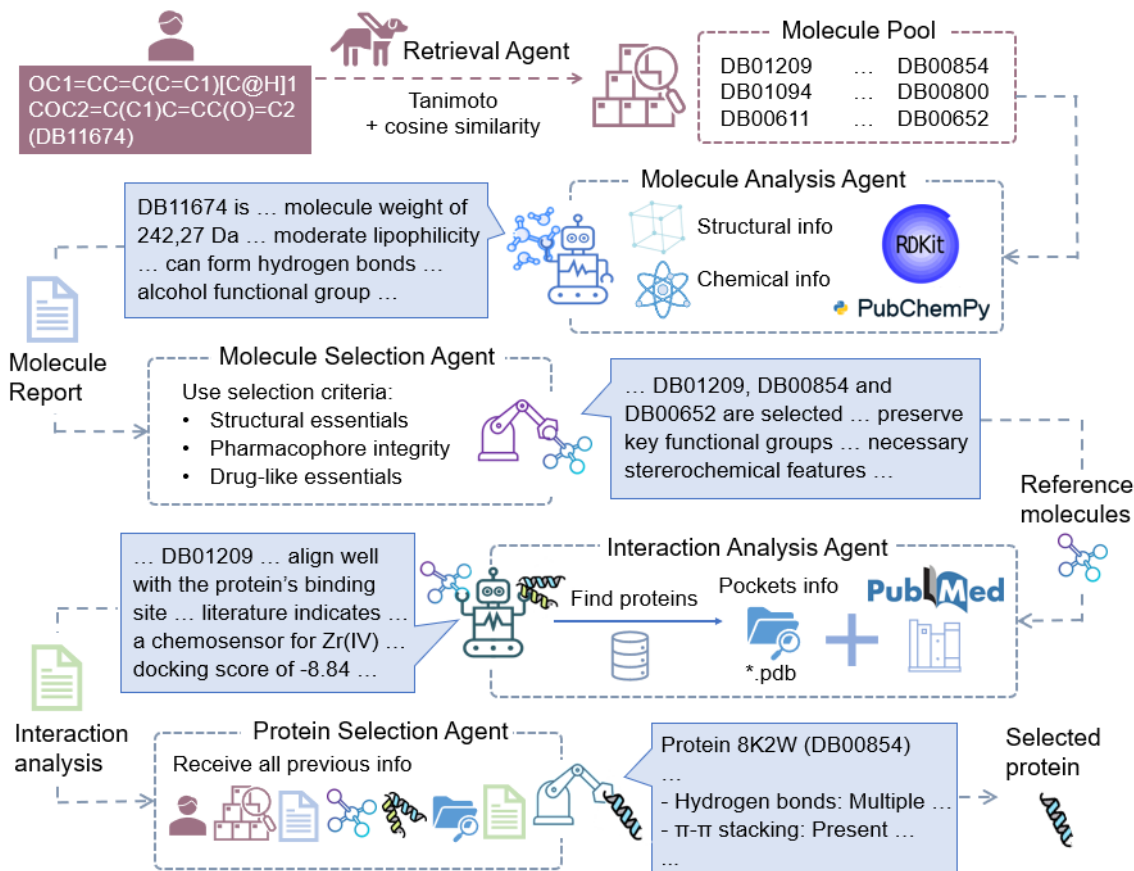


Figure 1: Workflow of DrugMCTS

**$A_1$  Retrieval Action.** Upon receiving the query molecule, the Retrieval agent queries databases to identify molecules structurally similar to the query molecule. We employ two similarity metrics: the Tanimoto coefficient [29] and the cosine similarity based on the last hidden state computed by ChemBERTa. For each metric, we retrieve the top-10 most similar molecules, merge the results, and remove duplicates to form the initial candidate molecules  $M_{cm}$ . The proteins can interact with  $M_{cm}$  are also retrieved, forming candidate proteins  $P_{cp}$ . These molecules serve as inputs to the root node of the Monte Carlo tree search.

Since the similarity scores are deterministic, the resulting molecule pool remains consistent, satisfying the requirement that the root node must be unique.

$$\text{sim}_C(M_{qm}, M_i) = \frac{\mathbf{h}_{qm} \cdot \mathbf{h}_i}{\|\mathbf{h}_{qm}\| \cdot \|\mathbf{h}_i\|} \quad (2)$$

$$M_{cm} = \text{dedup} \left( \text{Top}_{10}(\text{sim}_T(M_{qm})) \cup \text{Top}_{10}(\text{sim}_C(M_{qm})) \right) \quad (3)$$

**$A_2$  Molecule Analysis Action.** Understanding molecular properties is crucial for predicting molecular-target interactions. While general-purpose LLMs can perform some analysis using SMILES representations, their interpretations are often incomplete and prone to errors, especially when experimental data, such as hydrophobicity, is required.

To address this limitation, the Molecule Analysis (MA) agent first utilizes RDKit and PubChemPy APIs [20] to extract a set of structural and physicochemical properties ( $C_{q,s}$  &  $C_{q,phy}$ ) for the query molecule  $M_{qm}$  by calling RDKit and PubChemPy APIs. These properties include:

- Structural features: chiral centers, scaffolds, and functional groups.

- Physicochemical properties: molecular weight, lipophilicity (logP), polar surface area (PSA), hydrogen bond donors/acceptors, rotatable bonds, and heavy atom counts.

Based on this structured and quantified information, the agent then generates a comprehensive molecular analysis report  $R_{qm}$ .

$$R_{qm} = \text{MA Agent}(M_{qm}, C_{q,s}, C_{q,phy}) \quad (4)$$

**A<sub>3</sub> Molecule Selection Action.** Research has shown that the quality of retrieved information significantly impacts the accuracy of model-generated answers. Excessive irrelevant information can negatively affect the model’s performance. Although the retrieved molecules share structural similarities with the query molecule, they may not necessarily provide useful insights for drug discovery tasks. Thus, the Molecule Selection (MS) agent filters the molecule pool, based on structural similarity to the query molecule, pharmacophore integrity, and drug-like essentials, to generate reference molecules  $M_{rm}$ . The reference proteins  $P_{rp}$  are obtained by selecting the proteins that can interact with  $M_{rm}$  from  $P_{cp}$ . To ensure that these reference molecules are thoroughly characterized, this action also invokes the **A<sub>2</sub> Molecule Analysis Action**. Specifically, the MA agent is called to retrieve the structural and physicochemical properties ( $C_{c,s}$  &  $C_{c,phy}$ ) of  $M_{cm}$  but without generating reports.

$$M_{rm} = \text{MS Agent}(M_{qm}, R_{qm}, M_{cm}, C_{c,s}, C_{c,phy}) \quad (5)$$

**A<sub>4</sub> Interaction Analysis Action.** In this step, we aim to analyze potential interactions between the  $M_{rm}$  and  $P_{cp}$ . A major challenge lies in interpreting protein structures from amino acid sequences, which general-purpose LLMs struggle to handle due to their complexity. To overcome this, we adopt the methodology from DrugRealign [22], utilizing Python’s PLIP library to extract binding pocket information  $D_{bp}$  from PDB files and present it in textual format. In contrast to the original tabular representation which can be difficult for LLMs to parse, we reformat each entry into a descriptive paragraph to improve interpretability. Additionally, we retrieve relevant scientific literature  $L_{rp}$  from PubMed [21] to provide contextual support for interaction analysis by Interaction Analysis (IA) agent.

$$R_{ia} = \text{IA Agent}(M_{rm}, P_{rp}, D_{bp}, L_{rp}) \quad (6)$$

**A<sub>5</sub> Protein Selection Action.** At this stage, the Decision Agent synthesizes all available information, including the  $M_{qm}$ ,  $R_{qm}$ ,  $M_{rm}$ ,  $P_{cp}$ ,  $D_{bp}$ , and

$R_{ia}$ . Based on this integrated knowledge, the agent selects the most promising target protein  $P_s$  from the full list of candidates.

$$P_s = \text{Decision Agent}(M_{qm}, R_{qm}, M_{rm}, P_{rp}, D_{bp}, R_{ia}) \quad (7)$$

**A<sub>6</sub> End Action.** This action does not involve any agent invocation. When the model selects the final protein, the End Action is executed. Upon encountering an end node, the MCTS algorithm terminates further expansion, evaluates the end node’s score, and backpropagates the updated values to all nodes along the traversal path, concluding the current rollout.

## 2.5 Reward Calculation

Predicting molecular-protein binding affinity typically involves two primary approaches: molecular docking methods [23] and deep learning-based methods [11].

**Molecular Docking Methods.** Molecular docking methods, such as AutoDock Vina [23], are widely used but suffer from notable limitations. First, they require accurate three-dimensional (3D) structural information of the target proteins, which is often unavailable or unreliable for many biologically relevant molecules. Second, these methods are computationally intensive, limiting their scalability in large-scale drug discovery tasks. Given these limitations, there has been a growing interest in developing more efficient alternatives.

**Deep Learning-Based Methods.** Deep learning models have emerged as a promising solution due to their superior computational efficiency [39]. However, their performance is highly dependent on the training dataset. When predicting samples that deviate significantly from the training data, the accuracy of these models can drop by over 30%, limiting their applicability across diverse scenarios.

### 2.5.1 Self-Consistency Score.

To address these challenges, our work adopts an alternative reward calculation method known as the self-consistency score [28]. This approach involves querying the model multiple times with the same question and selecting the most frequently occurring answer  $p^*$  as the final output. The frequency of this answer serves as the **relative reward**. However, this method introduces a potential issue: if all candidate proteins in one rollout exhibit strong binding affinity with the query molecule, while in another rollout none do, the calculated relative rewards could still be similar despite the stark differences in actual binding

affinities. To mitigate this limitation, we introduce an absolute reward mechanism.

**Absolute Reward.** The absolute reward is computed by inputting the  $P_s$ ,  $D_{bp}$ ,  $L_{rp}$ ,  $M_{qm}$ , and  $R_{qm}$  into a decision-making model. This model evaluates whether there is a significant interaction between the protein and the query molecule. The frequency of affirmative responses ("yes") is then used as the absolute reward.

**Final Reward Calculation.** The final reward for each rollout is calculated as the average of the relative reward and the absolute reward. This combined approach ensures that both the consistency and the strength of the predicted interactions are taken into account, providing a more robust evaluation metric.

$$R_{\text{relative}}(p^*) = \frac{\text{Number of times } p^* \text{ is selected}}{\text{Total number of selections}} \quad (8)$$

$$R_{\text{absolute}}(p^*) = \frac{\text{Number of "yes" responses}}{k} \quad (9)$$

$$R_{\text{final}}(p^*) = \frac{R_{\text{relative}}(p^*) + R_{\text{absolute}}(p^*)}{2} \quad (10)$$

## 3 Experiments

### 3.1 Datasets and Metrics

We utilized two datasets, DrugBank [31] and KIBA [30], which were processed to include a total of 788 entries from DrugBank and 626 entries from KIBA. Each entry consists of a molecule as input and its corresponding interacting proteins as output. The number of ground truth interactions per entry varies. To evaluate model performance, we used recall, defined as the ratio of correctly predicted proteins to the total number of ground truth proteins. The detailed information of data processing pipeline can be found in Appendix Section A.1

$$\text{Recall} = \frac{|\{\text{proteins predicted correctly}\}|}{|\{\text{all ground truth proteins}\}|} \quad (11)$$

### 3.2 Settings

For the retrieval phase, we employed ESM2\_t33.650M\_UR50D to compute cosine similarity. For all other inference stages, we used Qwen2.5-7B-Instruct. Specifically, each search process involved 12 rounds of rollouts. Except for the protein selection and end actions, which generate

only one node during expansion, all other actions generated four nodes with distinct answers per expansion. The temperature was set to 0.8. Both relative and absolute rewards were computed by generating four responses per rollout, also with a temperature of 0.8.

### 3.3 Baselines

We established three sets of baselines to compare our model’s performance:

**General Models. (GM)** We selected GPT-4o-mini and Deepseek-R1 as general-purpose models. In this group, we provided the models with minimal information: the SMILES representation of the query molecule, reference molecules, and candidate proteins along with their pocket types.

**General Models with RAG. (GM + RAG)** Many existing studies on drug repurposing using LLMs either employ divergent methodological formulations or are not open-sourced, which complicates comparative evaluation. Our approach adopts the framework of the DrugReAlign [22] method, with modifications and enhancements tailored to our specific formulation. To ensure equitable comparison, we incorporate not only protein pocket information but also incorporate  $C_{q,s}$  and  $C_{q,phy}$  as additional features. The models used in this group remained GPT-4o-mini and Deepseek-R1.

**Deep Learning Models. (DL Models)** We trained four deep learning models: AttentionDTA [25], GraphDTA [24], DeepConv.DTI [26], and Perceiver CPI [27], on both DrugBank and KIBA datasets. We extracted data involving the query molecules as test sets and used the remaining data for training. All four models achieved over 70% accuracy on the test sets. During testing, we applied majority voting to select the final answer by averaging the scores from the four models and choosing the top-k proteins, where k corresponds to the length of the ground truth.

### 3.4 Results

The experimental results (Figure 1) indicate that general-purpose LLMs (GPT-4o-mini and Deepseek-R1) achieved relatively low recall scores of only 12.59%–16.19% on the DrugBank dataset when operating in a zero-shot setting. However, when incorporating molecular structural features and chemical properties via RAG-based prompting, model performance decreased, with GPT-4o-mini dropping to 15.19% and Deepseek-R1 significantly falling to 12.59%. This suggests that the inclusion of poten-

tially irrelevant or misleading information through retrieval can negatively impact the reasoning capabilities of general-purpose LLMs.

Among the deep learning baselines, the ensemble of four specialized models (AttentionDTI, GIN-ConvNet, DeepConv\_DTI, Perceiver CPI) achieved a recall score of 23.64%, representing an 84% improvement over the best-performing general-purpose model. On the KIBA dataset, the same ensemble attained a recall score of 26.45%, further demonstrating its effectiveness in capturing drug-target interactions.

Our proposed method, DrugMCTS, significantly outperformed all baseline approaches. Using a base TopK strategy, DrugMCTS achieved a recall of 44.66% on DrugBank and 42.24% on KIBA. This represents improvements of approximately 88.9% and 31.7% over the best deep learning baselines, respectively. Furthermore, our dynamic adjustment strategy (TopK+3) boosted performance to 55.34% on DrugBank and 49.24% on KIBA, marking maximum improvements of 330% and 91.4% over the general-purpose models. It is worth noting that although the other three methods showed improved performance on the KIBA dataset, our method exhibited a slight drop. However, according to Appendix Table 2, this apparent improvement in other baseline settings is largely due to an increased ratio of ground truth candidates among the total options. As shown in Appendix Table 1, compared to DrugBank, the KIBA dataset contains a larger number of candidate proteins per drug, which increases the difficulty of selecting the correct  $P_{cp}$  after molecule selection. This observation indirectly highlights the importance of effective filtering mechanisms.

These results strongly underscore the superiority of dynamic decision-making mechanisms, such as those used in DrugMCTS, over traditional static prediction methods like deep learning models. They also highlight the limitations of general-purpose large models in zero-shot settings for drug discovery tasks, especially when retrieval-augmented prompting introduces noise or irrelevant context. This further emphasizes the importance of structured reasoning, domain-specific knowledge integration, and information filtering in such applications.

Furthermore, to demonstrate the interpretability and transparency of our approach, we provide a detailed case study in the Appendix Section A.2. It illustrates how the model predicts an interaction between Equol and the CXC chemokine receptor 3, including the complete step-by-step reasoning process during a specific MCTS rollout.

## 4 Ablation Studies

### 4.1 Settings

In our ablation studies, we aim to investigate several key aspects:

- Whether the MCTS algorithm can improve model accuracy.
- The effectiveness of our proposed data processing pipelines for scientific data, a hybrid of scientific and general data, and general data.
- The efficacy of the combined relative and absolute reward calculation method.

To address these questions, we conducted the following experiments:

- **$S_1$  Baseline Setup.** Provide only the query molecule, all proteins in the protein pool, and the types of their pocket. Do not use the MCTS algorithm.
- **$S_2$  Enhanced Information (EI) Setup.** On top of the baseline setup, add the detailed pockets information and literature information for all proteins and structural and chemical properties of the query molecule.
- **$S_3$  Molecule Analysis Exclusion (MAE).** Conduct the MCTS process while excluding the molecule analysis action.
- **$S_4$  Interaction Analysis Exclusion (IAE).** Conduct the MCTS process while excluding the interaction analysis action.
- **$S_5$  Dual Exclusion (DE).** Conduct the MCTS process while excluding both the molecule analysis and interaction analysis actions.
- **$S_6$  Relative Reward (RR) Only.** During the MCTS process, compute only the relative rewards without considering the absolute rewards.

### 4.2 Ablation Studies Results

The ablation study results (Figure 2) clearly demonstrate the effectiveness of our proposed framework components. First, comparing  $S_1$  (Baseline Setup) with  $S_2$  (Enhanced Information Setup) shows that providing richer contextual information, including detailed pocket features, literature descriptions, structural and chemical properties, does improve performance to some extent (e.g., from 12.85% to 15.86% on

Table 1: Performance comparison on DrugBank and KIBA datasets

Model	Size	Dynamic Update	DrugBank	KIBA
<i>General Models (GM)</i>				
GPT-4o-mini	~8B	×	0.1552	0.2580
Deepseek-R1	37B <sup>a</sup>	×	0.1619	0.2645
<i>GM + RAG</i>				
GPT-4o-mini	~8B	✓	0.1519	0.2252
Deepseek-R1	37B <sup>a</sup>	✓	0.1259	0.2173
<i>DL Models</i>				
DL models	8M-12M	×	0.2364	0.3216
<b>DrugMCTS (Ours)</b>				
Selection = GT count	7B	✓	0.4466	0.4224
Selection = GT + 3	7B	✓	0.5534	0.4924

<sup>a</sup>Activation-aware model size. GT = Ground Truth count. Dynamic update: ✓ Yes, × No.

Table 2: Performance comparison on DrugBank and KIBA datasets using Qwen7b (Top-k/Top-k+3 accuracy).

Setup	$S_1$ Baseline	$S_2$ EI	$S_3$ MAE	$S_4$ IAE	$S_5$ DE	$S_6$ RR	DrugMCTS
DrugBank	0.1285	0.1586	0.3879/0.4677	0.3946/0.5119	0.3472/0.3617	0.4320/0.5527	<b>0.4466/0.5534</b>
KIBA	0.2284	0.2452	0.3772/0.4352	0.3846/0.4491	0.3189/0.3264	0.4193/0.4861	<b>0.4224/0.4924</b>

DrugBank). However, the most significant improvement is observed when the MCTS algorithm is introduced in combination with these enhancements. Experimental settings that incorporate MCTS ( $S_3$ – $S_6$  and Final Result) consistently achieve much higher performance than  $S_1$  or  $S_2$ , indicating that while richer input representations are beneficial, it is the MCTS-based reasoning process that plays the central role in boosting model accuracy.

Second, by analyzing  $S_3$  (Molecule Analysis Exclusion),  $S_4$  (Interaction Analysis Exclusion), and  $S_5$  (Dual Exclusion), we observe a consistent drop in performance when either or both of the analysis modules are removed. For instance, on the DrugBank dataset, removing molecule analysis alone leads to a decrease from 44.66% (Final Result) to 38.79%, while removing interaction analysis results in a drop to 39.46%. The dual exclusion further reduces performance to 34.72%, demonstrating that each data processing step contributes meaningfully to the overall effectiveness of the system. This supports our hypothesis that the proposed hybrid data processing pipeline, incorporating both molecular and interaction-level analy-

ses, is essential for capturing comprehensive contextual information.

Third, regarding the reward mechanism, the comparison between  $S_6$  (Relative Reward Only) and the full reward setting (Final Result) shows that the combined use of relative and absolute rewards does not lead to a significant improvement in performance. One possible explanation is that prior steps, including candidate protein and molecule selection, have already filtered out most irrelevant options, leaving a refined set of high-quality reference proteins. As a result, the likelihood of encountering scenarios where none of the candidates interact with the query molecule becomes rare, reducing the added value of the absolute reward component.

In summary, these findings confirm the importance of the MCTS algorithm, the multi-step data processing pipeline, and the overall design of the retrieval-augmented reasoning framework in achieving strong performance in molecular-target interaction prediction.



### 4.3 Computation Overhead Results

For inference time scaling, the discussion typically revolves around two key aspects: the trade-off between model performance improvement and additional computational overhead. Therefore, in this section, we first analyze the impact of different rollout numbers on model performance and then compare the performance-overhead profile with baseline models.

Our analysis (Figure 2 and Figure 3) reveals that when the number of rollouts increases from 8 to 12, both Top-K and Top-K+3 metrics exhibit significant improvements across the two datasets. However, further increasing the rollout count from 12 to 24 only yields a notable gain in the Top-K+3 metric on the KIBA dataset, while other scenarios show either marginal or even negative improvements. Consequently, to balance computational cost and model performance, we ultimately adopt rollout=12 for our experiments. Compared to baseline models, our approach not only achieves the highest recall scores but also demonstrates superior cost efficiency, as evidenced by its position on the Pareto front.

## 5 Conclusion

Our DrugMCTS framework revolutionizes drug repurposing by integrating multi-agent collaboration (five specialized agents), hybrid data processing (scientific  $\rightarrow$  hybrid  $\rightarrow$  general), and Monte Carlo Tree Search to enable the lightweight Qwen2.5-7B-Instruct model to outperform Deepseek-R1 by more than 20% recall on DrugBank/KIBA datasets. The system achieves 55.34% recall via dynamic Top-K+3 selection, validated by 1,221 experimental interactions and case studies like Equol-CXCR3 binding (docking score: -8.4 kcal/mol). This work establishes a template for LLM-powered scientific discovery beyond drug-target prediction.

## 6 Limitations

While DrugMCTS demonstrates significant improvements over baseline models, several limitations highlight opportunities for further optimization:

- Despite achieving more than 20% recall gains over Deepseek-R1 (Table 2), the absolute performance (55.34% recall) suggests untapped optimization potential. The plateau in gains beyond 12 rollouts indicates diminishing returns from current MCTS configurations.
- Current predictions primarily leverage PDB-derived binding pocket data (Section 2.4), omitting higher-order biological context. Future work may augment the framework with knowledge graph or Pathway activation score.
- The combined relative/absolute reward system (Eq. 10) yields only around 1% improvement over relative-only rewards, suggesting the necessities of a more effective reward system.

## References

- [1] Ye, G., Cai, X., Lai, H., Wang, X., Huang, J., Wang, L., Liu, W. & Zeng, X. Drugassist: A large language model for molecule optimization. *Briefings In Bioinformatics*. **26**, bbae693 (2025)
- [2] Zheng, Y., Koh, H., Ju, J., Nguyen, A., May, L., Webb, G. & Pan, S. Large language models for scientific discovery in molecular property prediction. *Nature Machine Intelligence*. pp. 1-11 (2025)
- [3] Zhang, W., Wang, Q., Kong, X., Xiong, J., Ni, S., Cao, D., Niu, B., Chen, M., Li, Y., Zhang, R. & Others Fine-tuning large language models for chemical text mining. *Chemical Science*. **15**, 10600-10611 (2024)
- [4] Van Herck, J., Gil, M., Jablonka, K., Abrudan, A., Anker, A., Asgari, M., Blaiszik, B., Buffo, A., Choudhury, L., Corminboeuf, C. & Others Assessment of fine-tuned large language models for real-world chemistry and material science applications. *Chemical Science*. **16**, 670-684 (2025)
- [5] Nguyen, C., Achille, A., Lam, M., Hassner, T., Mahadevan, V. & Soatto, S. Toward understanding catastrophic forgetting in continual learning. *ArXiv Preprint ArXiv:1908.01091*. (2019)
- [6] Zhang, P., Peng, X., Han, R., Chen, T. & Ma, J. Rag2Mol: Structure-based drug design based on Retrieval Augmented Generation. *Briefings In Bioinformatics*. **26**, bbae265 (2025)
- [7] Che, X., Zhao, Y., Liu, Q., Yu, F., Gao, H. & Zhang, L. CSstep: Step-by-step exploration of the chemical space of drug molecules via multi-agent and multi-stage reinforcement learning. *Chemical Engineering Science*. pp. 122048 (2025)
- [8] Song, K., Trotter, A. & Chen, J. Llm agent swarm for hypothesis-driven drug discovery. *ArXiv Preprint ArXiv:2504.17967*. (2025)

- [9] Inoue, Y., Song, T., Wang, X., Luna, A. & Fu, T. Drugagent: Multi-agent large language model-based reasoning for drug-target interaction prediction. *ICLR 2025 Workshop On Machine Learning For Genomics Explorations*. (2025)
- [10] Lee, N., De Brouwer, E., Hajiramezanali, E., Biancalani, T., Park, C. & Scalia, G. RAG-Enhanced Collaborative LLM Agents for Drug Discovery. *ArXiv Preprint ArXiv:2502.17506*. (2025)
- [11] Yang, Z., Li, Y., Matsuda, Y. & Song, L. mHMG-DTI: A Drug-Target Interaction Prediction Framework Combining Modified Hierarchical Molecular Graphs and Improved Convolutional Block Attention Module. *Trends And Applications In Knowledge Discovery And Data Mining: PAKDD 2025 Workshops, ADUR, FairPC, GLFM, PM4B And RAFDA, Sydney, NSW, Australia, June 10–13, 2025, Proceedings*. **15835** pp. 191 (2025)
- [12] Hutter, J., Rau, D., Marx, M. & Kamps, J. Lost but not only in the middle: Positional bias in retrieval augmented generation. *European Conference On Information Retrieval*. pp. 247-261 (2025)
- [13] Bank, P. Protein data bank. *Nature New Biol.* **233**, 10-1038 (1971)
- [14] Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y. & Others Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*. **379**, 1123-1130 (2023)
- [15] Chithrananda, S., Grand, G. & Ramsundar, B. ChemBERTa: large-scale self-supervised pretraining for molecular property prediction. *ArXiv Preprint ArXiv:2010.09885*. (2020)
- [16] Landrum, G. Rdkit documentation. *Release*. **1**, 4 (2013)
- [17] Salentin, S., Schreiber, S., Haupt, V., Adasme, M. & Schroeder, M. PLIP: fully automated protein–ligand interaction profiler. *Nucleic Acids Research*. **43**, W443-W447 (2015)
- [18] Li, B., Zhang, J., Fan, J., Xu, Y., Chen, C., Tang, N. & Luo, Y. Alpha-sql: Zero-shot text-to-sql using monte carlo tree search. *ArXiv Preprint ArXiv:2502.17248*. (2025)
- [19] Couëtoux, A., Hoock, J., Sokolovska, N., Teytaud, O. & Bonnard, N. Continuous upper confidence trees. *Learning And Intelligent Optimization: 5th International Conference, LION 5, Rome, Italy, January 17-21, 2011. Selected Papers 5*. pp. 433-445 (2011)
- [20] Swain, M. PubChemPy documentation. *PubChemPy Documentation*. (2014)
- [21] White, J. PubMed 2.0. *Medical Reference Services Quarterly*. **39**, 382-387 (2020)
- [22] Wei, J., Zhuo, L., Fu, X., Zeng, X., Wang, L., Zou, Q. & Cao, D. DrugReAlign: a multisource prompt framework for drug repurposing based on large language models. *BMC Biology*. **22**, 226 (2024)
- [23] Huey, R., Morris, G., Forli, S. & Others Using AutoDock 4 and AutoDock vina with AutoDockTools: a tutorial. *The Scripps Research Institute Molecular Graphics Laboratory*. **10550**, 1000 (2012)
- [24] Nguyen, T., Le, H., Quinn, T., Nguyen, T., Le, T. & Venkatesh, S. GraphDTA: predicting drug–target binding affinity with graph neural networks. *Bioinformatics*. **37**, 1140-1147 (2021)
- [25] Zhao, Q., Xiao, F., Yang, M., Li, Y. & Wang, J. AttentionDTA: prediction of drug–target binding affinity using attention model. *2019 IEEE International Conference On Bioinformatics And Biomedicine (BIBM)*. pp. 64-69 (2019)
- [26] Lee, I., Keum, J. & Nam, H. DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences. *PLoS Computational Biology*. **15**, e1007129 (2019)
- [27] Nguyen, N., Jang, G., Kim, H. & Kang, J. Perceiver CPI: a nested cross-attention network for compound–protein interaction prediction. *Bioinformatics*. **39**, btac731 (2023)
- [28] Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A. & Zhou, D. Self-consistency improves chain of thought reasoning in language models. *ArXiv Preprint ArXiv:2203.11171*. (2022)
- [29] Bajusz, D., Rácz, A. & Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?. *Journal Of Cheminformatics*. **7** pp. 1-13 (2015)

- [30] Tang, J., Szwajda, A., Shakyawar, S., Xu, T., Hintsanen, P., Wennerberg, K. & Aittokallio, T. Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. *Journal Of Chemical Information And Modeling*. **54**, 735-743 (2014)
- [31] Knox, C., Wilson, M., Klinger, C., Franklin, M., Oler, E., Wilson, A., Pon, A., Cox, J., Chin, N., Strawbridge, S. & Others DrugBank 6.0: the DrugBank knowledgebase for 2024. *Nucleic Acids Research*. **52**, D1265-D1275 (2024)
- [32] Hurst, A., Lerer, A., Goucher, A., Perelman, A., Ramesh, A., Clark, A., Ostrow, A., Welihinda, A., Hayes, A., Radford, A. & Others Gpt-4o system card. *ArXiv Preprint ArXiv:2410.21276*. (2024)
- [33] Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., Zhu, Q., Ma, S., Wang, P., Bi, X. & Others Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *ArXiv Preprint ArXiv:2501.12948*. (2025)
- [34] Chaslot, G. Monte-carlo tree search. (2010)
- [35] Yuan, S., Chan, H. & Hu, Z. Using PyMOL as a platform for computational drug design. *Wiley Interdisciplinary Reviews: Computational Molecular Science*. **7**, e1298 (2017)
- [36] Chen, Y., Yan, L., Sun, W., Ma, X., Zhang, Y., Wang, S., Yin, D., Yang, Y. & Mao, J. Improving Retrieval-Augmented Generation through Multi-Agent Reinforcement Learning. *ArXiv Preprint ArXiv:2501.15228*. (2025)
- [37] Edwards, C., Lu, Z., Hajiramezanali, E., Biancalani, T., Ji, H. & Scalia, G. MolCap-Arena: A Comprehensive Captioning Benchmark on Language-Enhanced Molecular Property Prediction. *ArXiv Preprint ArXiv:2411.00737*. (2024)
- [38] Zheng, Y., Koh, H., Yang, M., Li, L., May, L., Webb, G., Pan, S. & Church, G. Large language models in drug discovery and development: From disease mechanisms to clinical trials. *ArXiv Preprint ArXiv:2409.04481*. (2024)
- [39] Yang, Z., Shao, W., Matsuda, Y. & Song, L. iResNetDM: An interpretable deep learning approach for four types of DNA methylation modification prediction. *Computational And Structural Biotechnology Journal*. **23** pp. 4214-4221 (2024)
- [40] Team, Q. Qwen2 technical report. *ArXiv Preprint ArXiv:2412.15115*. (2024)

## A Appendix

### A.1 Dataset Construction

**Experimental Dataset.** We first extracted all molecules from the original dataset and computed both Tanimoto similarity and cosine similarity between each pair. For each molecule, we selected the top 10 most similar molecules based on each metric, merged the results, and removed duplicates to form the candidate molecule set  $M_{cm}$ . Each unique query molecule paired with its corresponding  $M_{cm}$  constitutes one problem instance. These problem instances were further filtered according to the following criteria:

- For each query molecule, the number of associated interacting proteins must be between 2 and 10.
- For each candidate molecule in  $M_{cm}$ , the number of associated interacting proteins must be between 2 and 4.
- The total number of candidate molecules per query must not exceed 15.

Table 3: Dataset Statistics Summary

Dataset	Processed Points	All Proteins	Ground Truth	Unique Molecules	All Molecules
DrugBank	788	22 508	1595	1304	7717
KIBA	626	23 849	1664	752	6219

We then extracted all proteins that interact with any molecule in  $M_{cm}$ , denoted as  $P_{cp}$  (as mentioned in Section 2.4). Additionally, we collected all proteins that directly interact with the query molecule. The intersection of these two protein sets was defined as the ground truth set. The ground truth set must satisfy the following constraints:

- Its size must be between 1 and 5.
- Its size must not exceed 70% of the size of the candidate protein set.

**Baseline Dataset.** This dataset is derived from the Experimental Dataset, with the following modifications:

- Only the query molecules, the candidate protein set  $P_{cp}$ , and the ground truth set are retained.
- All candidate molecules  $M_{cm}$  are removed.

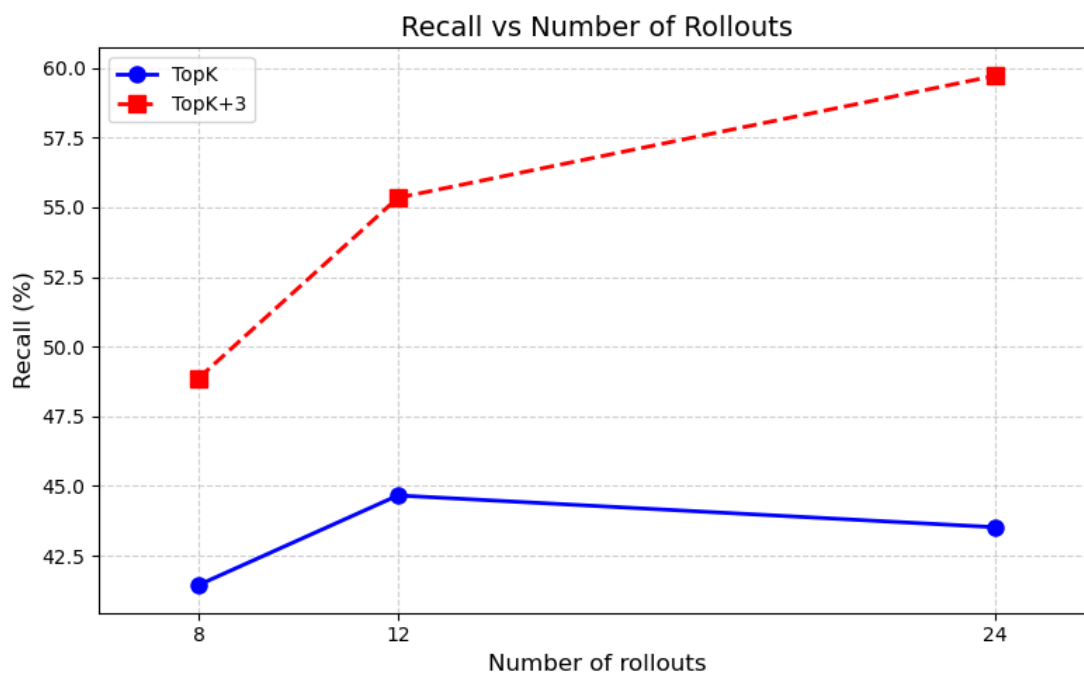
Table 4: Baseline Dataset Statistics

Dataset	Ground Truth	All Proteins	Ratio (%)
DB	1595	14 654	10.88
KIBA	1664	10 593	15.71

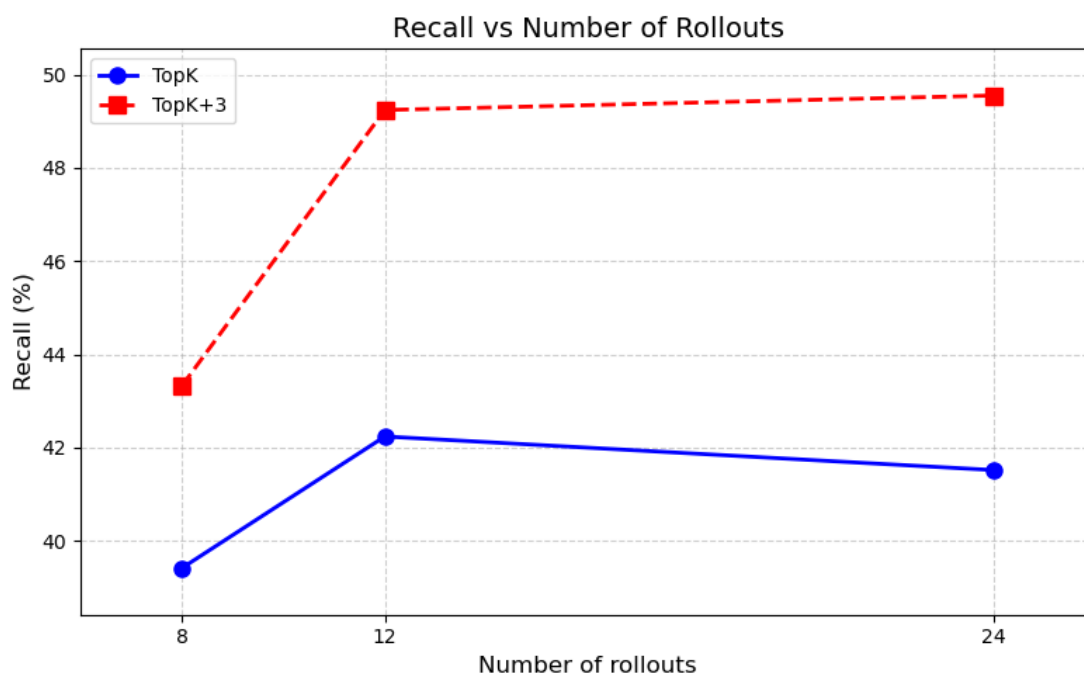
**Deep Learning Dataset.** From the original dataset, we extracted all data instances involving the query molecules from the Experimental Dataset to form the test set. The remaining data were used as the training set.

### A.2 Case Study

We manually selected a molecular-protein interaction with the highest self-consistency score that has never been previously reported: Equol (DrugBank ID: DB11674) and CXC chemokine receptor 3 (CXCR3, PDB ID: 8K2W). The binding affinity predicted by AutoDock Vina was -8.4 kcal/mol, indicating a strong potential interaction between the two. Visualization using PyMOL [35] revealed that Equol can bind within one of

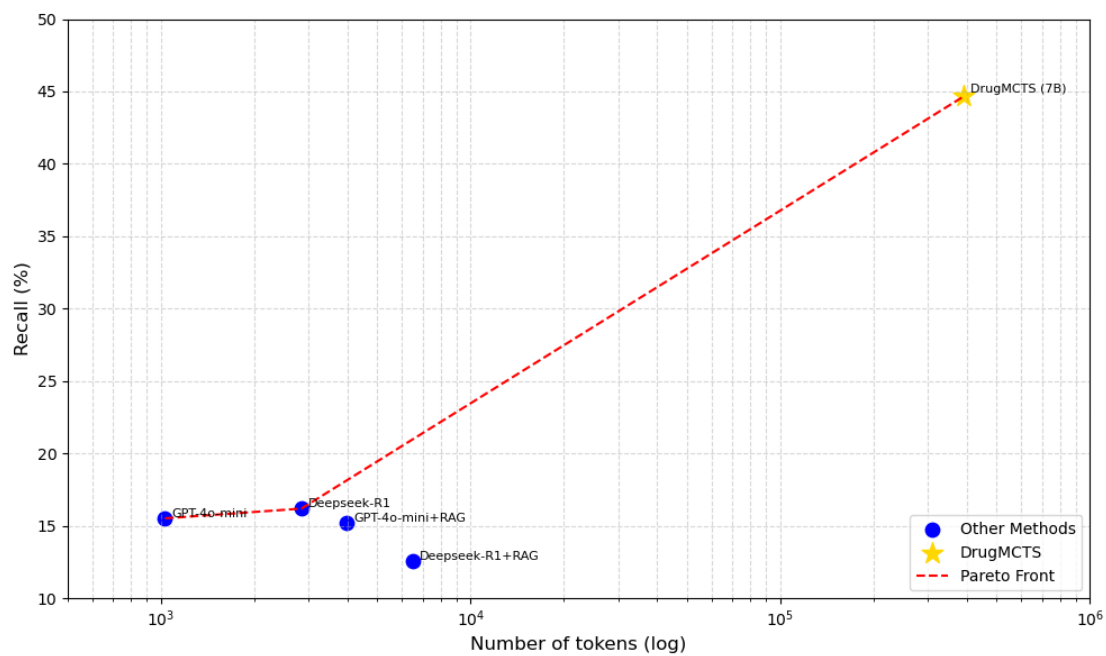


(a)

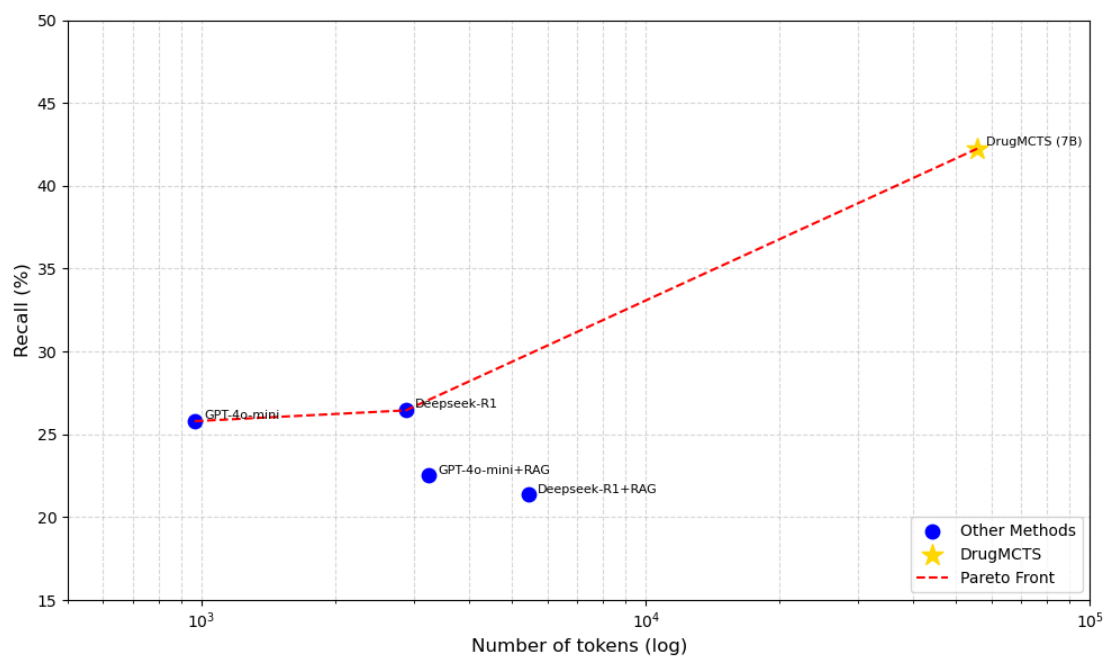


(b)

Figure 2: Number of rollouts vs Recall score on (a) DrugBank and (b) KIBA dataset



(a)



(b)

Figure 3: Number of tokens vs Recall score on (a) DrugBank and (b) KIBA dataset

Table 5: Class Distribution Statistics

Dataset	Split	Negative	Positive	Total
DrugBank	Train	14,787	12,428	27,215
	Test	1,960	4,111	6,071
KIBA	Train	62,553	17,350	79,903
	Test	31,643	4,804	36,447

the binding pockets of CXCR3 and form hydrogen bonds, as evidenced by the continuous red dots in the lower-right corner of Appendix Figure 1(b). This observation is consistent with the reasoning generated by the large language model during the molecule analysis, protein selection, and absolute reward calculation stages, thereby validating the effectiveness of our framework.

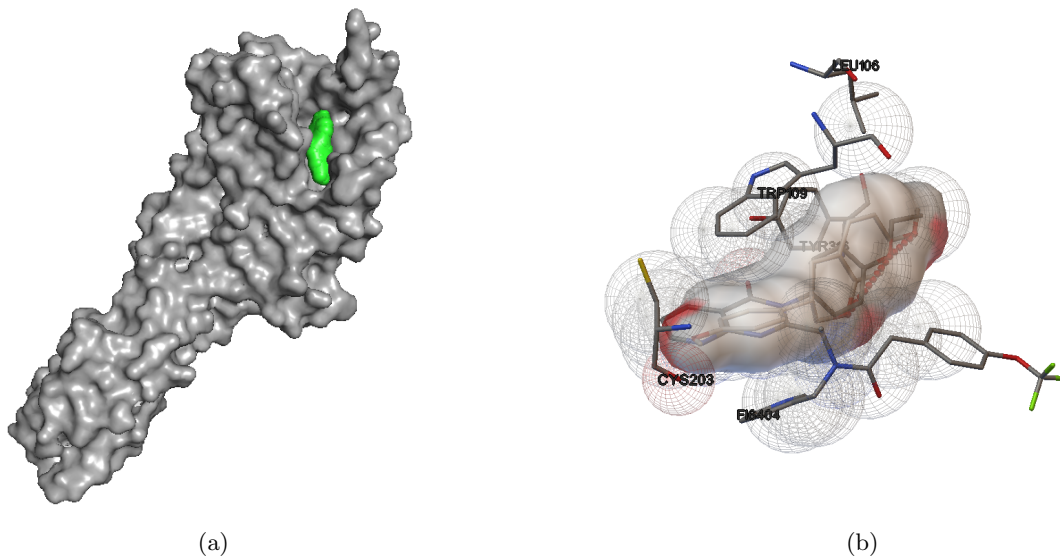


Figure 4: Protein-ligand docking results between (a) global binding site overview and (b) detailed interaction view, computed by AutoDock Vina with default scoring function and visualized using PyMOL.

### Molecule Analysis

{\n "chain\_of\_thought\_reasoning": "The molecule DB11674 is a complex organic compound with significant structural and chemical properties. ...., The stereochemistry at one chiral center adds complexity to its interaction profile and potentially its selectivity.",\n "summary": "DB11674 is an organic molecule with a molecular weight of 242.27 Da and moderate lipophilicity, suitable for membrane interaction. It has a polar surface area of 49.7 Å² and can form hydrogen bonds, indicating protein binding potential. The rigid structure, with one rotatable bond, suggests a fixed conformation. The core scaffold and alcohol functional group imply a specific pharmacophore. Stereochemical complexity at one chiral center adds selectivity considerations."}\n}

### Molecule Selection

{\n "chain\_of\_thought\_reasoning": "The selection criteria focus on maintaining structural essentials, pharmacophore integrity, and drug-like properties. DB01209, DB00854, and DB00652 are selected because they have a balanced molecular weight, moderate lipophilicity, and appropriate polar surface areas. They also preserve key functional groups such as amine and alcohol, which are present in the query molecule DB11674, and have the necessary stereochemical features to ensure pharmacophore integrity.",\n "reference\_molecule1": "DB01209",\n "reference\_molecule2": "DB00854",\n "reference\_molecule3": "DB00652"}\n}

### Interaction Analysis

{\n "chain\_of\_thought\_reasoning": "The molecule DB01209 has a core scaffold of c1ccc2c(c1)CC1CCCCC2C1 with functional groups of an amine and an alcohol. Given the hydrophobic interactions and hydrogen bonds between the ligand and the protein pocket (PDB ID: 8K2W), ....,\n "summary": "The molecule DB01209 interacts with the protein pocket through hydrophobic interactions and hydrogen bonding. Its core scaffold and functional groups align well with the protein's binding site, suggesting effective interaction. The literature indicates potential roles as a chemosensor for Zr(IV) and significant biological activity against cancer and bacteria. Specifically, the docking score of -8.84 kcal/mol indicates a strong binding affinity, supporting the hypothesis of a functional interaction."}\n}

### Protein Selection

{\n "chain\_of\_thought\_reasoning": "Based on the provided data, we need to evaluate the binding potential of each protein pocket for the query molecule (DB11674). We will use the Binding Affinity Triad criteria prioritized by quality of hydrogen-bond network, ....\n\n### Protein 8K2W (DB00854)\n\n- Hydrophobic interactions: Moderate to good.\n- Hydrogen bonds: Good, with multiple favorable angles.\n-  $\pi$ - $\pi$  stacking: Present, although less favorable than some other options.\n\n### Protein 8F7S (DB00854)\n\n- Hydrophobic interactions: Good.\n- Hydrogen bonds: Multiple and favorable.\n-  $\pi$ - $\pi$  stacking: Not available.....\n\nBased on the criteria, Protein 8K2W appears to have the best combination of binding interactions: moderate to good hydrophobic interactions, good hydrogen bonding, and favorable  $\pi$ - $\pi$  stacking. ....,\n "selected\_protein": "8K2W"}\n}

### Absolute Reward

{\n "chain\_of\_thought\_reasoning": "The analysis focuses on understanding whether the provided molecule and the protein can interact based on their chemical properties and binding characteristics.\n\n- \*\*Molecule Properties\*\*:\nDB11674 is a moderately sized compound with an alcohol group, one rotatable bond, and a single stereocenter. It has reasonable lipophilicity and polar surface area, suggesting potential for interaction.\n\n- \*\*Protein Pocket Characteristics\*\*:\nThe pocket shows hydrophobic interactions and hydrogen bonds, particularly involving amino acids like Leu, Ile, Phe, Tyr, and Trp. These interactions could facilitate binding for molecules with similar physicochemical properties.\n\n- \*\*Literature Review\*\*:\nThe literature mentions the protein (PDB ID: 2R3I and 4OZ5) being involved in molecular docking studies with compounds having different functionalities but shows no direct correlation with DB11674. The second source is uninformative.\n\nGiven the presence of hydrophobic and hydrogen bond interactions in the pocket and the potential of DB11674 to form such interactions due to its functional groups, a reaction is plausible.",\n "reaction\_possible": "y"}\n}

Figure 5: Answers generated by the model