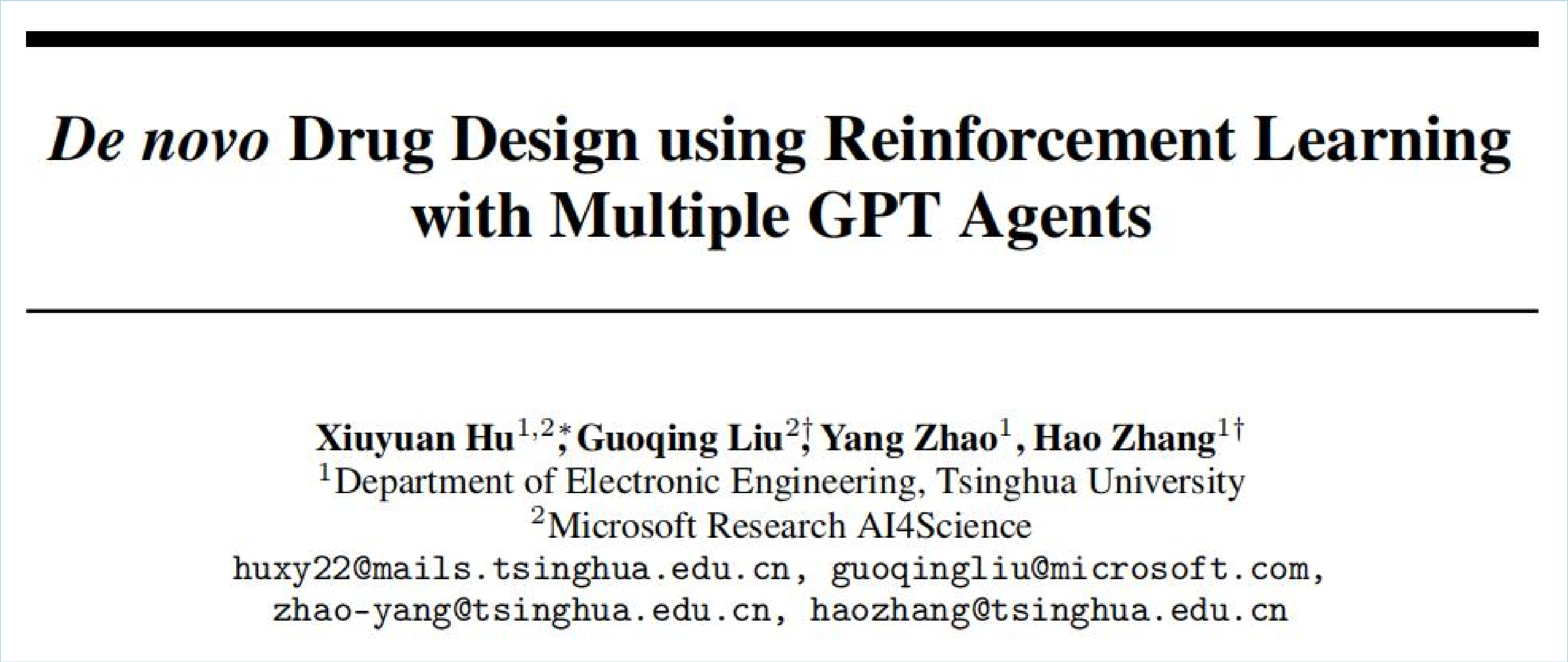
# Our paper



## *De novo* Drug Design using Reinforcement Learning with Multiple GPT Agents

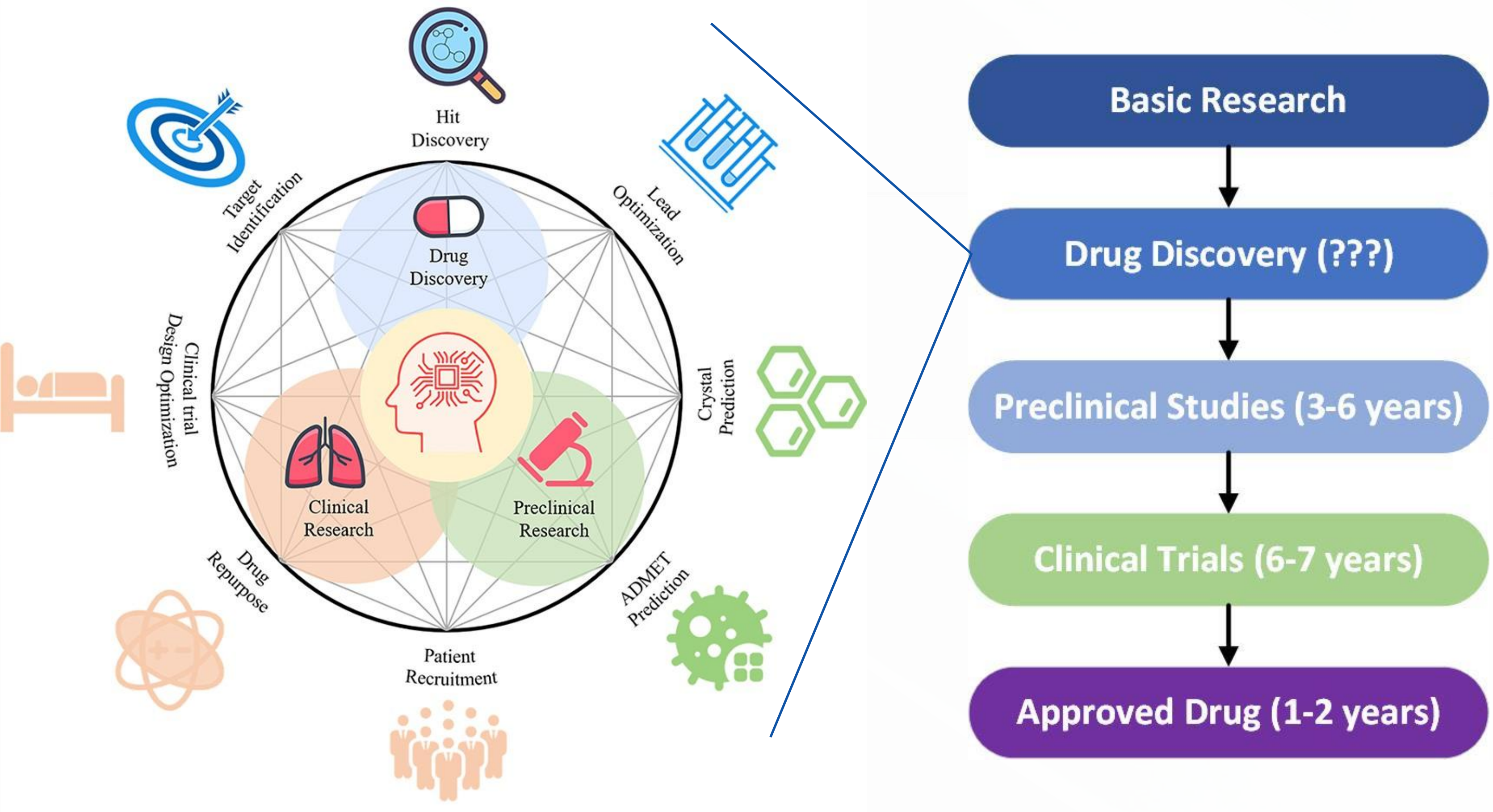
Xiuyuan Hu[1,2]*, Guoqing Liu[2]†, Yang Zhao[1], Hao Zhang[1]†
[1]Department of Electronic Engineering, Tsinghua University
[2]Microsoft Research AI4Science
huxy22@mails.tsinghua.edu.cn, guoqingliu@microsoft.com,
zhao-yang@tsinghua.edu.cn, haozhang@tsinghua.edu.cn

- Accepted by NeurIPS 2023 as a poster
- Code:
  https://github.com/HXYfighter/MolRL-MGPT

# Background: AIDD

CADD (Computer-aided Drug Development) / AI for Science

**AIDD: AI for Drug Development**



Basic Research

Drug Discovery (???)

Preclinical Studies (3-6 years)

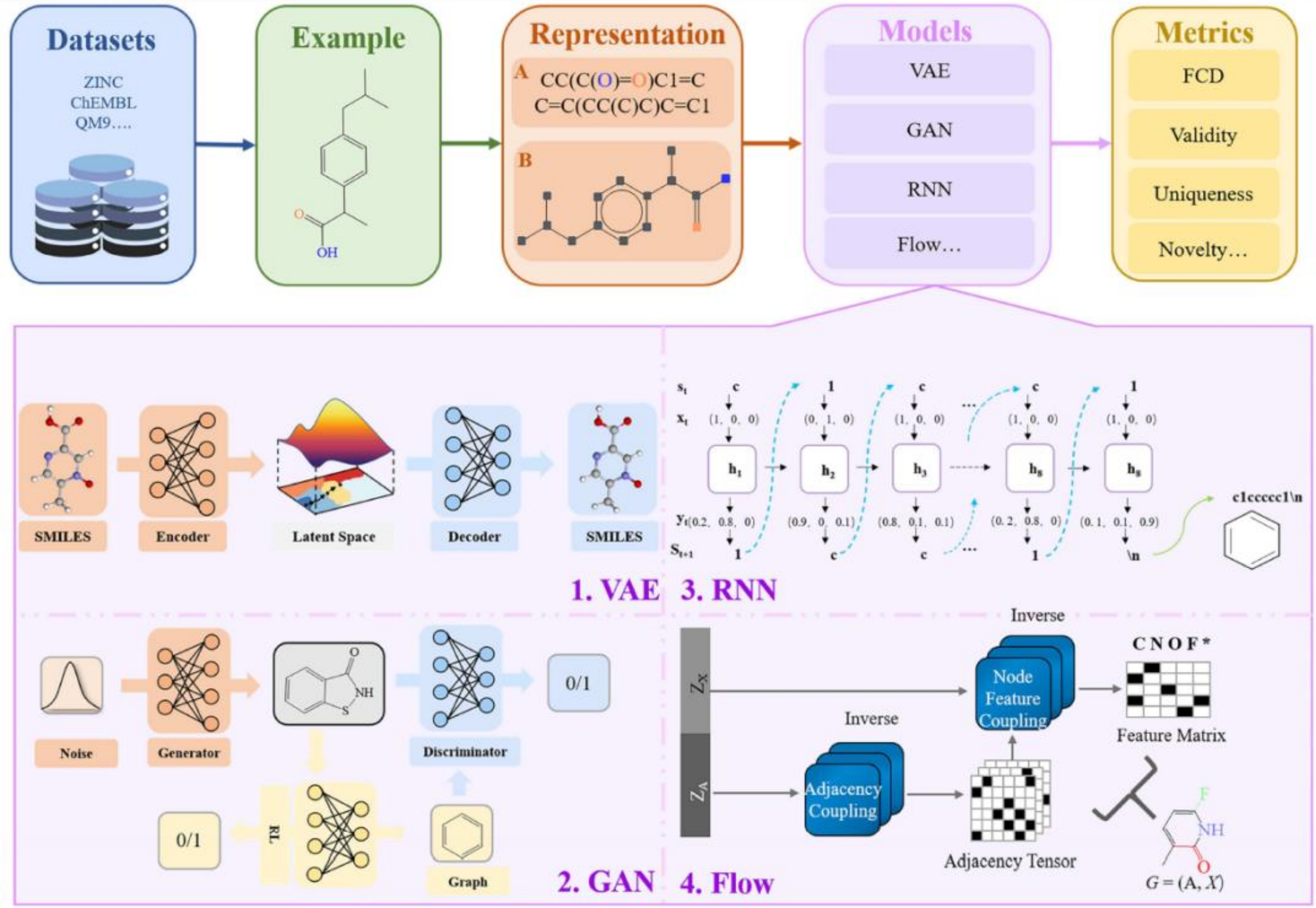
Clinical Trials (6-7 years)

Approved Drug (1-2 years)

Yu Cheng, Yongshun Gong, Yuansheng Liu, Bosheng Song, Quan Zou, Molecular design in drug discovery: a comprehensive review of deep generative models, *Briefings in* Bioinformatics, 6(6), 2021.

# Background: *De novo* Drug Design

## *De novo* Drug Design: Molecular Generation

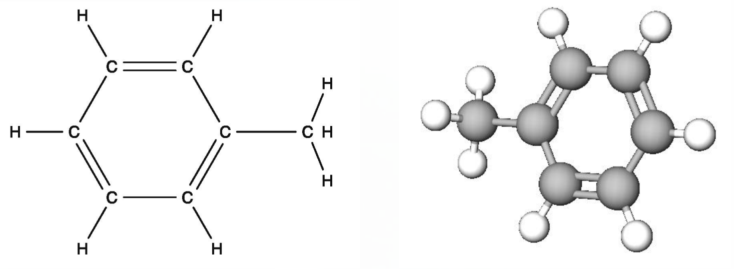


# Related Works: RL-based Molecular Generation

- Reinforcement learning (RL) is the most widely-used technique in molecular generation.
- Basic idea:
  - Actions: adding atoms / bonds / substructures
  - Rewards: property scores

- SMILES-based RL
  - SMILES——Most popular 1D string representation of molecules
  - Reinvent: a deep reinforcement learning framework for training RNN to generate SMILES

- Graph-based RL



SMILES: Cc1ccccc1

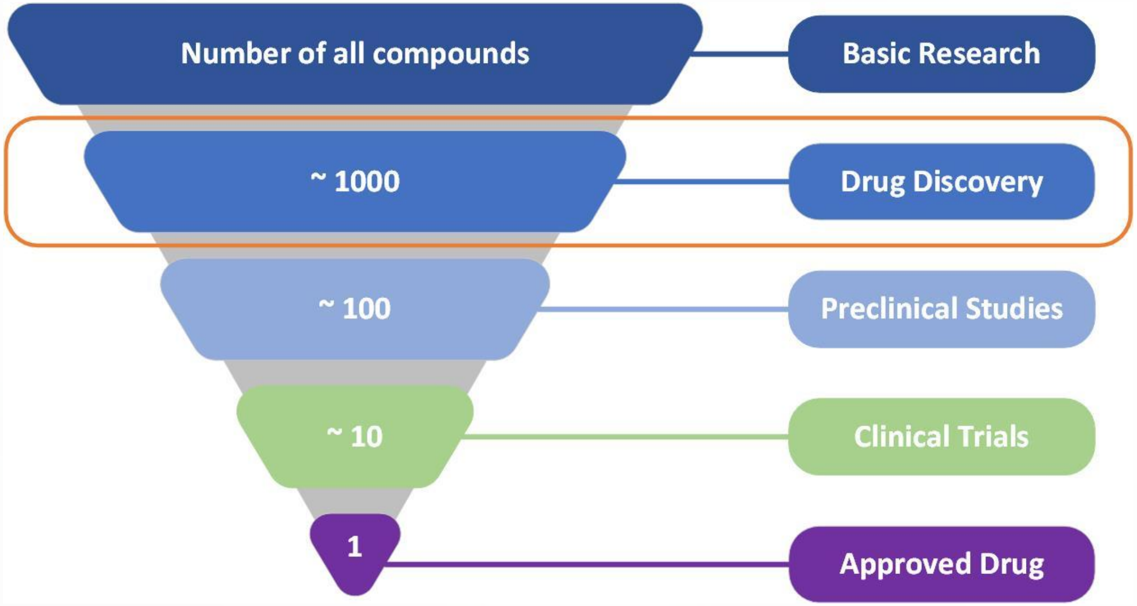
# Related Works: Transformers in Molecular Generation

- Transformer has obtained a great success in NLP
- Generative Pre-Trained Transformer (GPT) has achieved a breakthrough in machine conversation

- Transformers has also been applied to the chemical language:
  - MolGPT
  - Chemformer
  - TamGent
  - ……

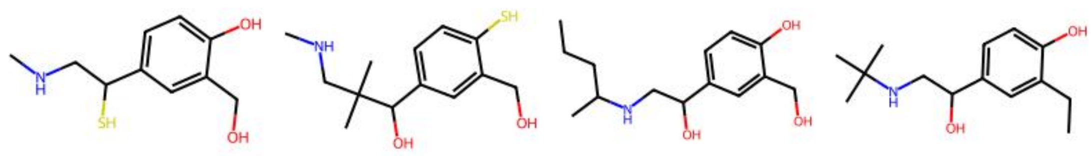
# Background: Diversity in Drug Development

- For one design objective (e.g. a protein target), we hope to design a set of diverse candidates with desirable properties
- Due to: the gap between *in silico* scores and *in vivo properties*
- Diverse candidates can greatly improve the possibility of success of downstream drug development

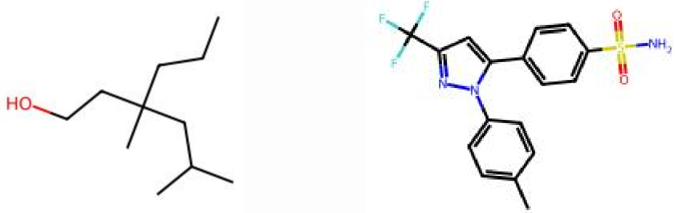


# Our motivation: to promote the diversity in DD

- Previous works tend to generate a set of highly similar molecular structures
- Similar:



- Dissimilar:

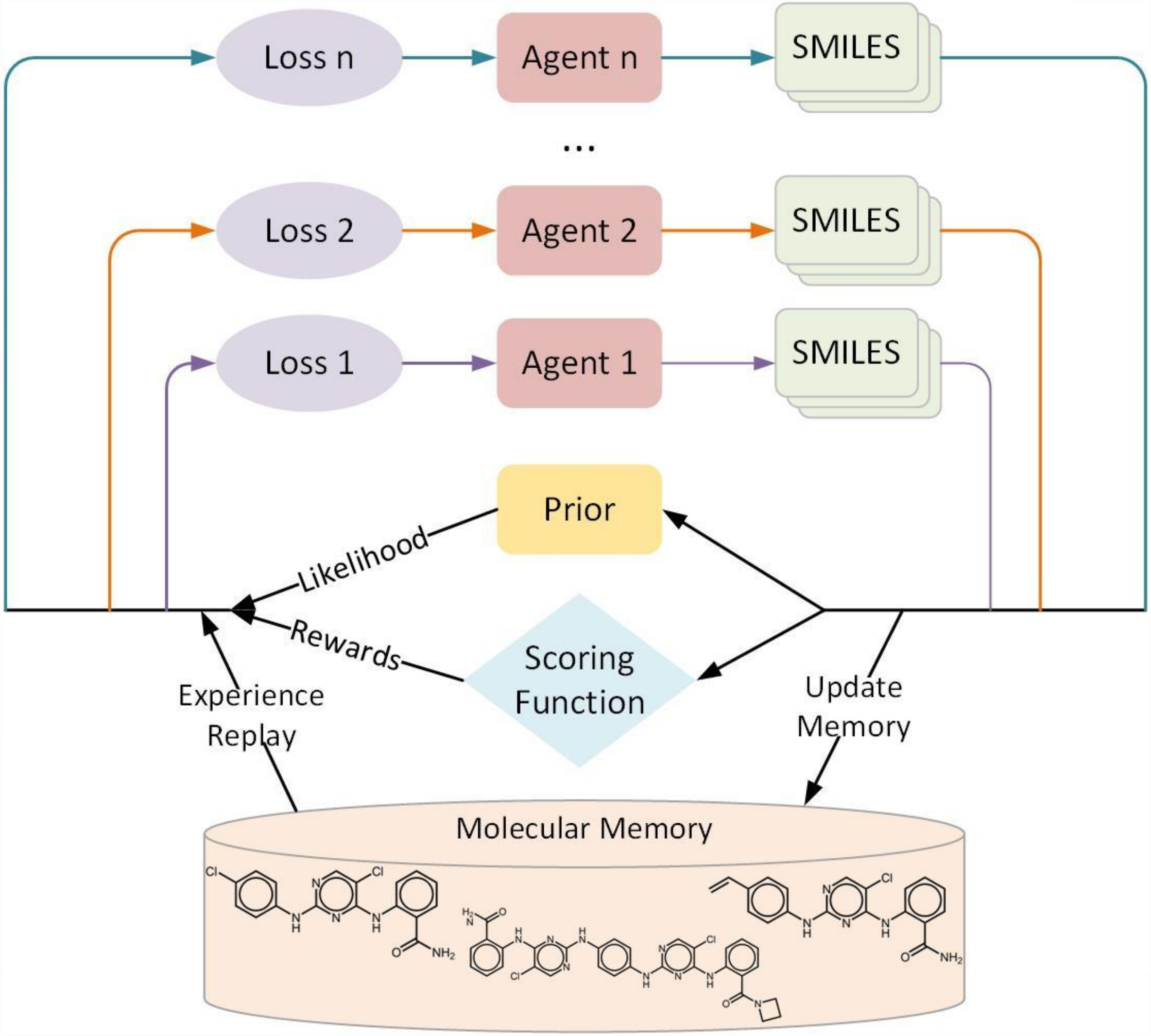


- Molecular similarity / diversity can be measure by molecular distances
- Our motivation：

**To promote molecular diversity in drug design**

# Our approach: MolRL-MGPT

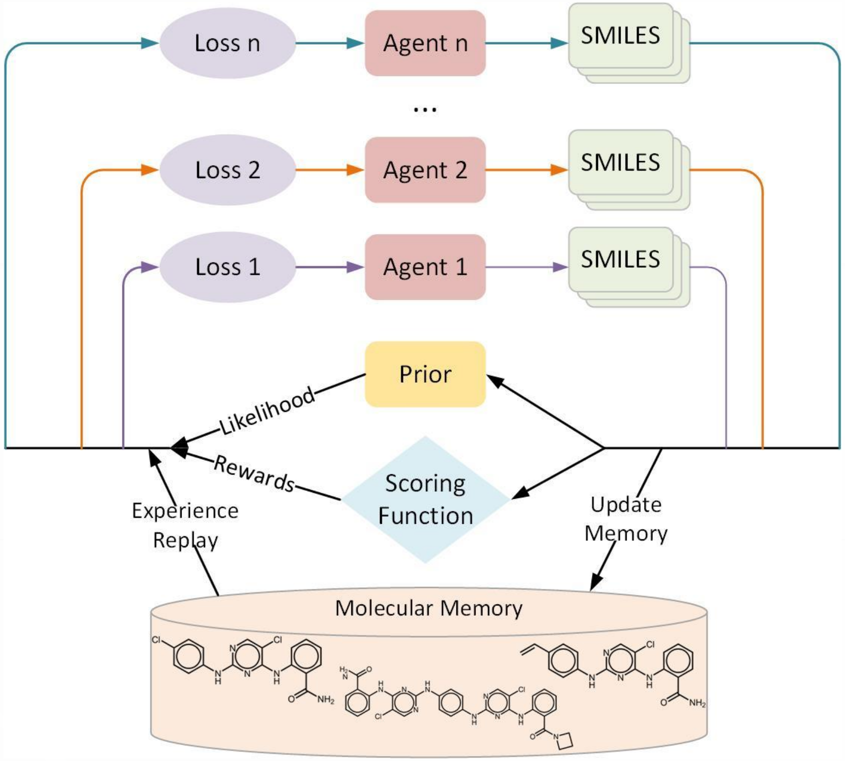
MolRL-MGPT: **Mol**ecular design using **R**einforcement Learning with **M**ultiple **GPT** agents



# Our approach: MolRL-MGPT

- Using the pre-trained weights of the prior model to initialize all the *n* agents

- In each iteration, agents are updated in order:
  - Each agent generate a batch of SMILES strings
  - Update the memory, experience replay
  - Calculate loss by Prior, scoring function and other agents; update the agent



# Loss Functions

1-st agent:

$$L_1(x;\Theta_1) = [\log P(x)_{\text{Prior}} - \log P(x)_{\text{Agent}_1} + \sigma_1 \cdot s(x)]^2$$

*k*-th agent:

$$\begin{aligned} L_k(x;\Theta_k) =& L_1(x;\Theta_k) - \sigma_2 \sum_{j=1}^{k-1} s(x) \cdot |\log P(x)_{\text{Agent}_k} - \log P(x)_{\text{Agent}_j}| \\ =& [\log P(x)_{\text{Prior}} - \log P(x)_{\text{Agent}_k} + \sigma_1 \cdot s(x)]^2 \\ & \boxed{- \sigma_2 \sum_{j=1}^{k-1} s(x) \cdot |\log P(x)_{\text{Agent}_k} - \log P(x)_{\text{Agent}_j}|} \end{aligned}$$

To encourage deviation between agents
--Diverse search

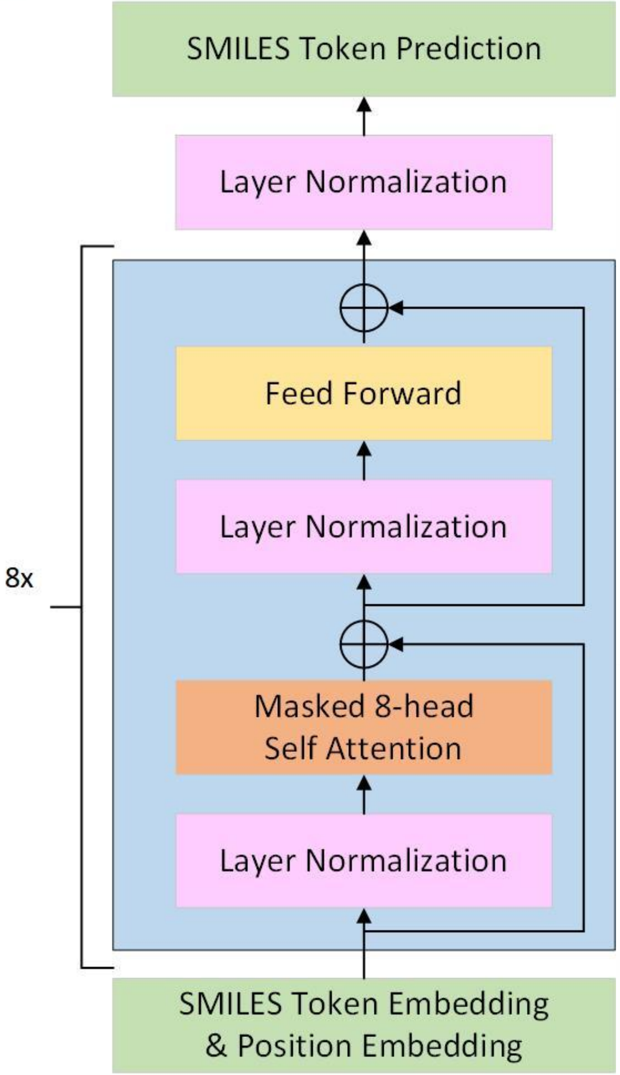
# Pre-training

- Mini version of GPT-2
- 6.4M parameters

- Training dataset: ChEMBL (2M), ZINC-100M
- Data augmentation:
  SMILES randomization

- Unsupervised learning

- Results: valid ratio > 98%



# Experiments: GuacaMol benchmark

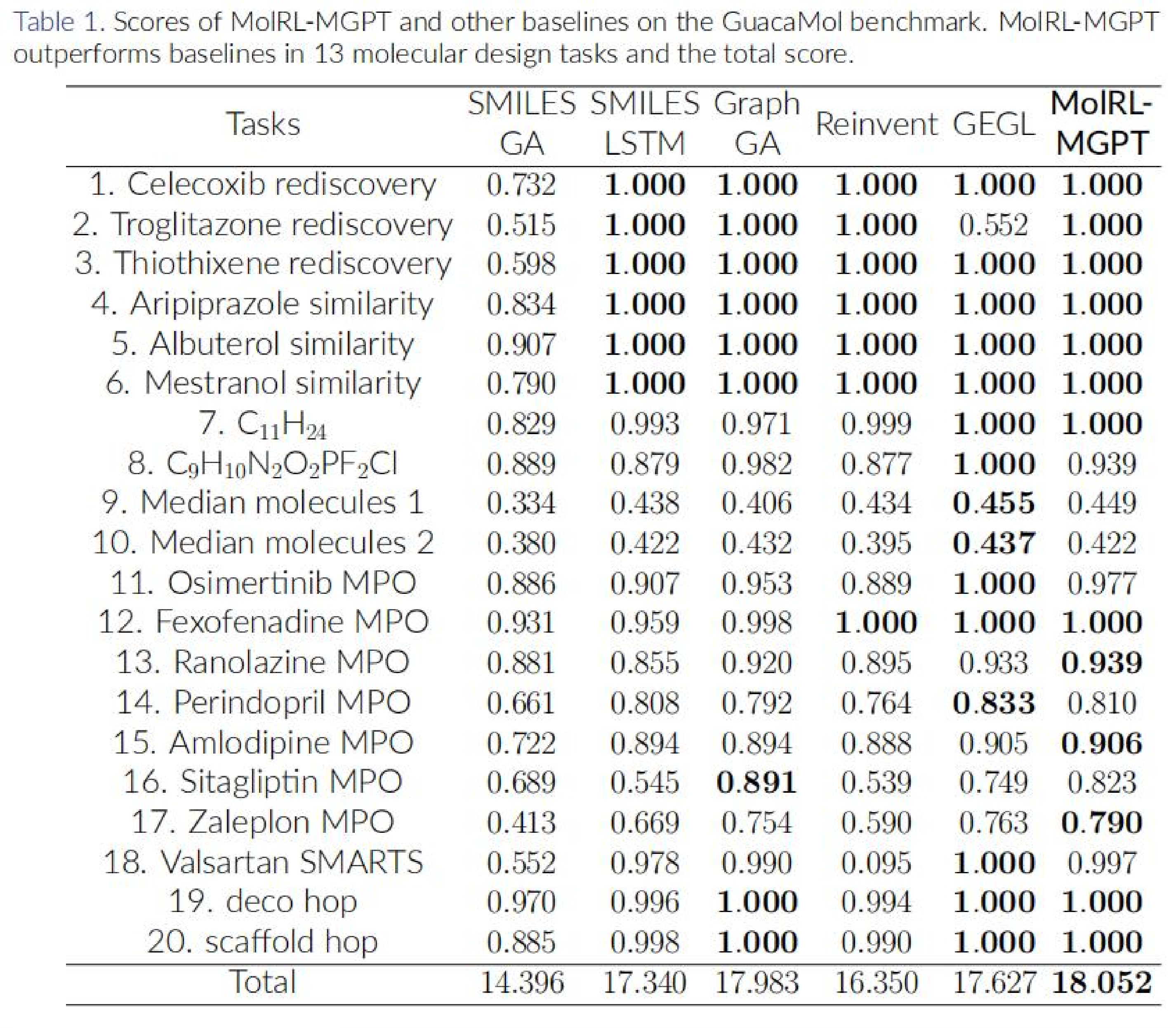


Table 1. Scores of MolRL-MGPT and other baselines on the GuacaMol benchmark. MolRL-MGPT outperforms baselines in 13 molecular design tasks and the total score.

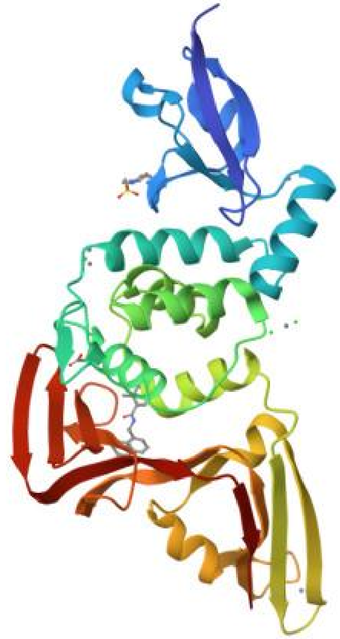
| Tasks | SMILES GA | SMILES LSTM | Graph GA | Reinvent | GEGL | MolRL-MGPT |
|---|---|---|---|---|---|---|
| 1. Celecoxib rediscovery | 0.732 | **1.000** | **1.000** | **1.000** | **1.000** | **1.000** |
| 2. Troglitazone rediscovery | 0.515 | **1.000** | **1.000** | **1.000** | 0.552 | **1.000** |
| 3. Thiothixene rediscovery | 0.598 | **1.000** | **1.000** | **1.000** | **1.000** | **1.000** |
| 4. Aripiprazole similarity | 0.834 | **1.000** | **1.000** | **1.000** | **1.000** | **1.000** |
| 5. Albuterol similarity | 0.907 | **1.000** | **1.000** | **1.000** | **1.000** | **1.000** |
| 6. Mestranol similarity | 0.790 | **1.000** | **1.000** | **1.000** | **1.000** | **1.000** |
| 7. $C_{11}H_{24}$ | 0.829 | 0.993 | 0.971 | 0.999 | **1.000** | **1.000** |
| 8. $C_9H_{10}N_2O_2PF_2Cl$ | 0.889 | 0.879 | 0.982 | 0.877 | **1.000** | 0.939 |
| 9. Median molecules 1 | 0.334 | 0.438 | 0.406 | 0.434 | **0.455** | 0.449 |
| 10. Median molecules 2 | 0.380 | 0.422 | 0.432 | 0.395 | **0.437** | 0.422 |
| 11. Osimertinib MPO | 0.886 | 0.907 | 0.953 | 0.889 | **1.000** | 0.977 |
| 12. Fexofenadine MPO | 0.931 | 0.959 | 0.998 | **1.000** | **1.000** | **1.000** |
| 13. Ranolazine MPO | 0.881 | 0.855 | 0.920 | 0.895 | 0.933 | **0.939** |
| 14. Perindopril MPO | 0.661 | 0.808 | 0.792 | 0.764 | **0.833** | 0.810 |
| 15. Amlodipine MPO | 0.722 | 0.894 | 0.894 | 0.888 | 0.905 | **0.906** |
| 16. Sitagliptin MPO | 0.689 | 0.545 | **0.891** | 0.539 | 0.749 | 0.823 |
| 17. Zaleplon MPO | 0.413 | 0.669 | 0.754 | 0.590 | 0.763 | **0.790** |
| 18. Valsartan SMARTS | 0.552 | 0.978 | 0.990 | 0.095 | **1.000** | 0.997 |
| 19. deco hop | 0.970 | 0.996 | **1.000** | 0.994 | **1.000** | **1.000** |
| 20. scaffold hop | 0.885 | 0.998 | **1.000** | 0.990 | **1.000** | **1.000** |
| Total | 14.396 | 17.340 | 17.983 | 16.350 | 17.627 | **18.052** |

# Experiments: SARS-CoV-2

- SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) caused the COVID-19 global pandamic.
- For this real-world drug design challenge, we select two crucial protein targets to design inhibitors:

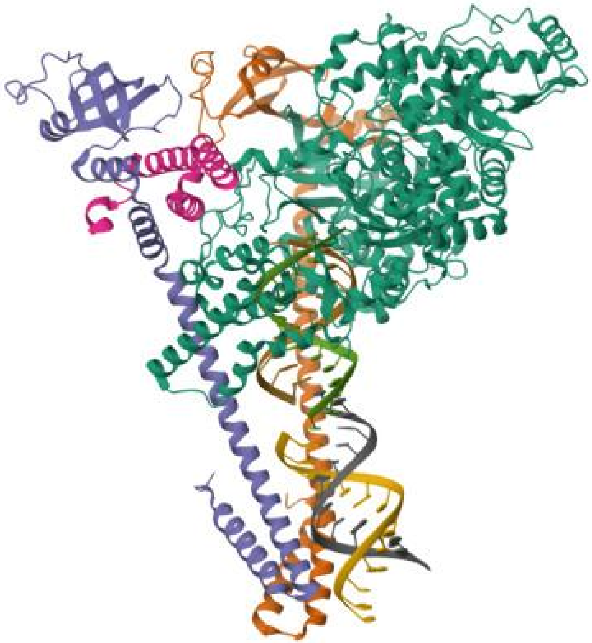
PLPro (papain-like protease)

7JIR [5]



RdRp (RNA-dependent RNA polymerase)

6YYT [6]



# Experiments: SARS-CoV-2

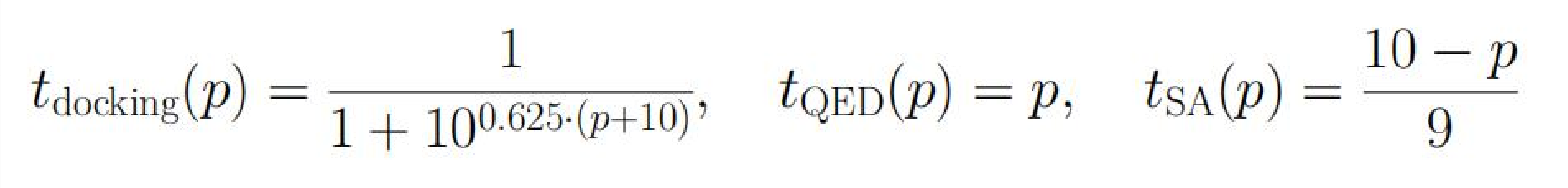
- Docking software: **Quick Vina 2**

For predicting binding modes and affinities (scores) between small molecules and protein targets

- Other oracles: **QED** (Quantitative Estimate of Drug-likeness), **SA** (Synthetic Accessibility)

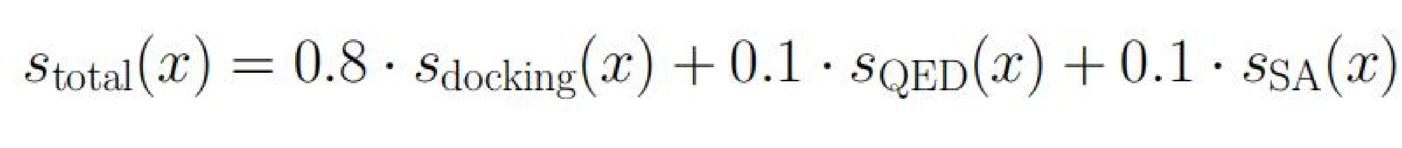
Commonly used in real-world drug design

- Transformation functions:



$$t_{\text{docking}}(p) = \frac{1}{1 + 10^{0.625\cdot(p+10)}}, \quad t_{\text{QED}}(p) = p, \quad t_{\text{SA}}(p) = \frac{10 - p}{9}$$

- Scoring function:



$$s_{\text{total}}(x) = 0.8 \cdot s_{\text{docking}}(x) + 0.1 \cdot s_{\text{QED}}(x) + 0.1 \cdot s_{\text{SA}}(x)$$

# Experiments: SARS-CoV-2

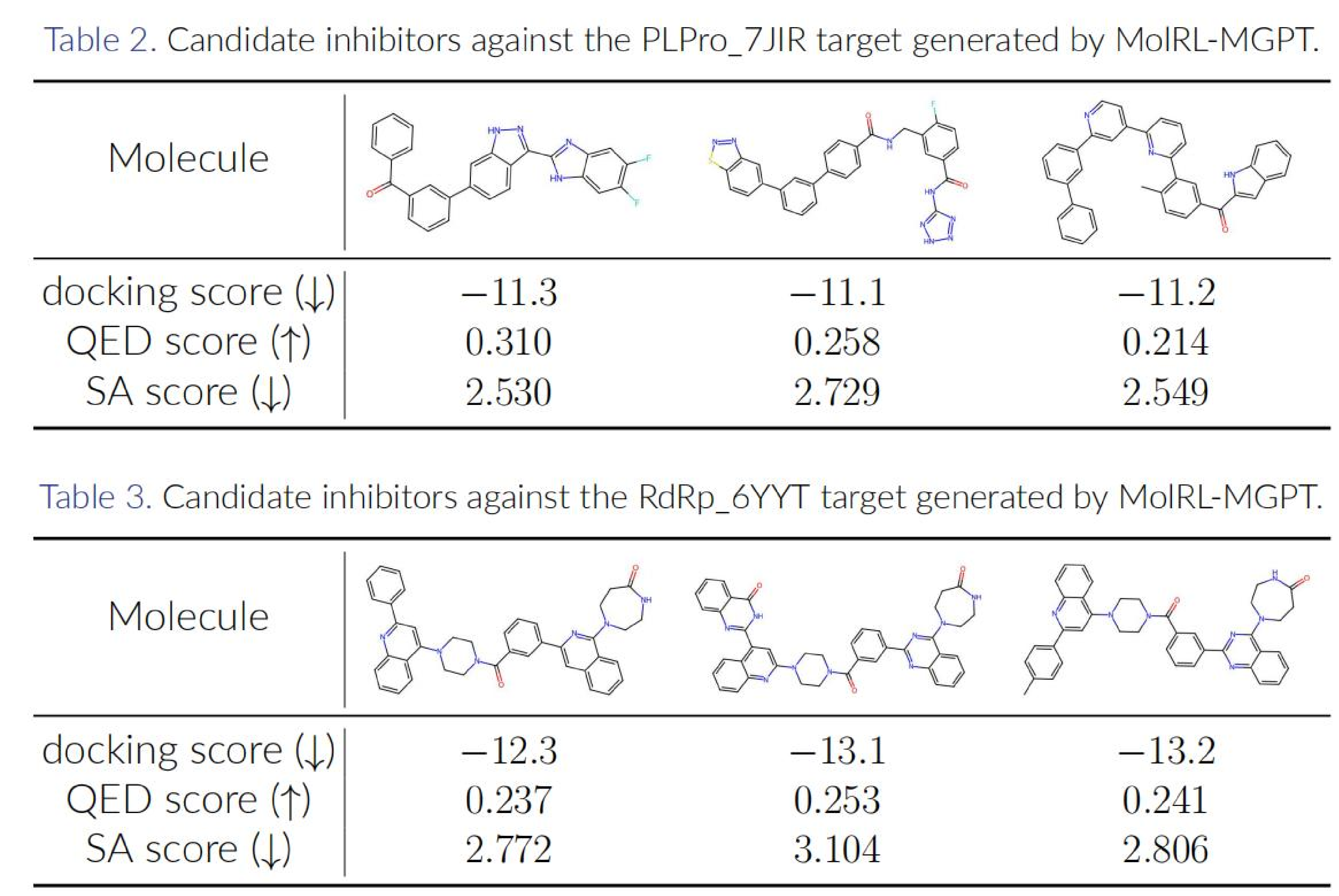


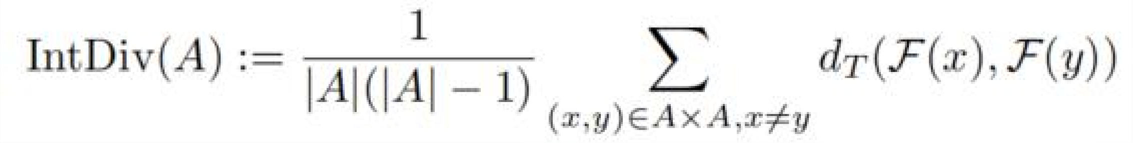
Table 2. Candidate inhibitors against the PLPro_7JIR target generated by MolRL-MGPT.

| Molecule | | | |
|---|---|---|---|
| docking score (↓) | −11.3 | −11.1 | −11.2 |
| QED score (↑) | 0.310 | 0.258 | 0.214 |
| SA score (↓) | 2.530 | 2.729 | 2.549 |

Table 3. Candidate inhibitors against the RdRp_6YYT target generated by MolRL-MGPT.

| Molecule | | | |
|---|---|---|---|
| docking score (↓) | −12.3 | −13.1 | −13.2 |
| QED score (↑) | 0.237 | 0.253 | 0.241 |
| SA score (↓) | 2.772 | 3.104 | 2.806 |

# Experiments: Ablation and Comparison



$$\mathrm{IntDiv}(A) := \frac{1}{|A|(|A|-1)} \sum_{(x,y)\in A\times A, x\neq y} d_T(\mathcal{F}(x), \mathcal{F}(y))$$

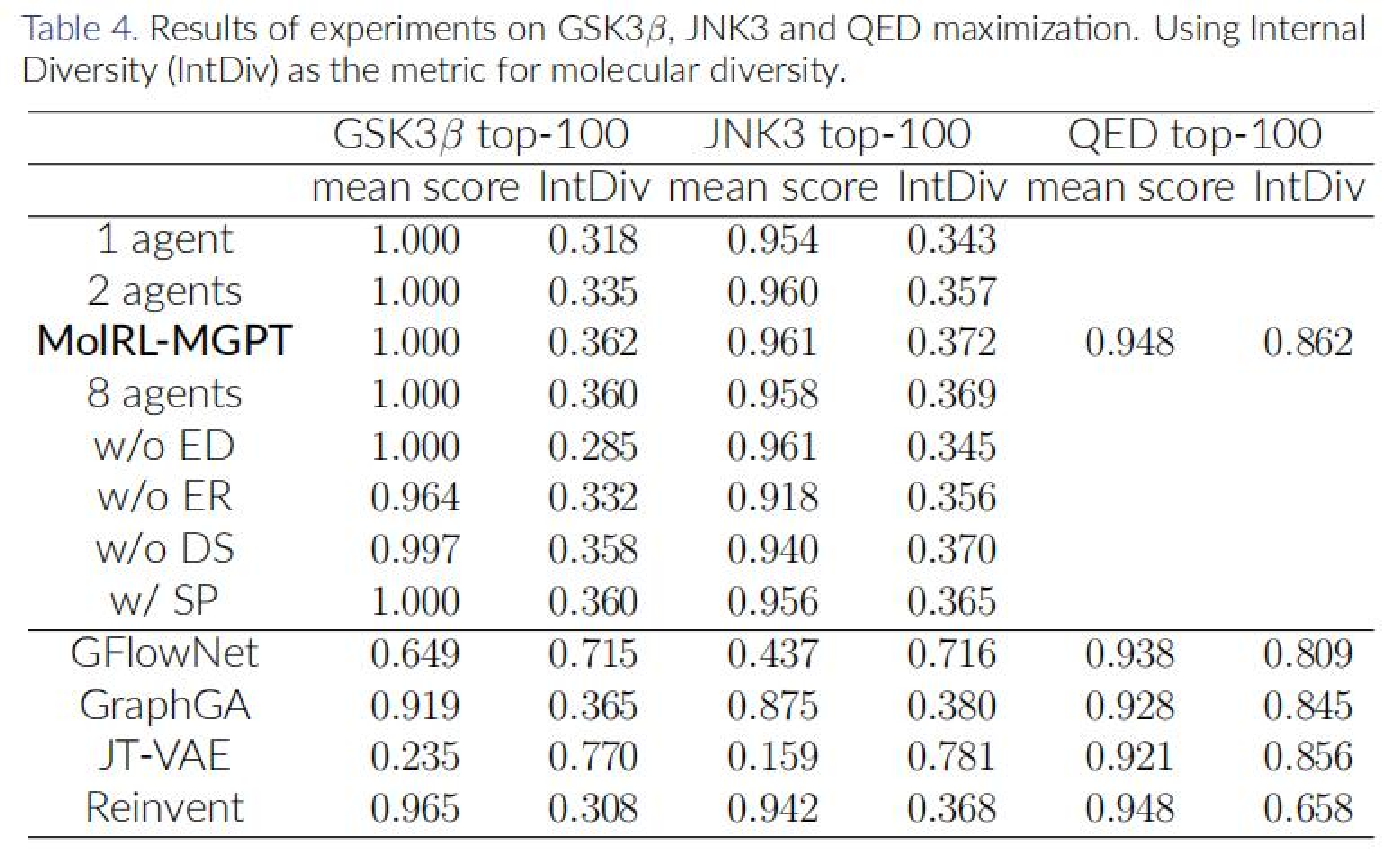


Table 4. Results of experiments on GSK3$\beta$, JNK3 and QED maximization. Using Internal Diversity (IntDiv) as the metric for molecular diversity.

| | GSK3$\beta$ top-100 | | JNK3 top-100 | | QED top-100 | |
|---|---|---|---|---|---|---|
| | mean score | IntDiv | mean score | IntDiv | mean score | IntDiv |
| 1 agent | 1.000 | 0.318 | 0.954 | 0.343 | | |
| 2 agents | 1.000 | 0.335 | 0.960 | 0.357 | | |
| **MolRL-MGPT** | 1.000 | 0.362 | 0.961 | 0.372 | 0.948 | 0.862 |
| 8 agents | 1.000 | 0.360 | 0.958 | 0.369 | | |
| w/o ED | 1.000 | 0.285 | 0.961 | 0.345 | | |
| w/o ER | 0.964 | 0.332 | 0.918 | 0.356 | | |
| w/o DS | 0.997 | 0.358 | 0.940 | 0.370 | | |
| w/ SP | 1.000 | 0.360 | 0.956 | 0.365 | | |
| GFlowNet | 0.649 | 0.715 | 0.437 | 0.716 | 0.938 | 0.809 |
| GraphGA | 0.919 | 0.365 | 0.875 | 0.380 | 0.928 | 0.845 |
| JT-VAE | 0.235 | 0.770 | 0.159 | 0.781 | 0.921 | 0.856 |
| Reinvent | 0.965 | 0.308 | 0.942 | 0.368 | 0.948 | 0.658 |

# Thanks!