上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

Retro-R1: LLM-based Agentic Retrosynthesis

Wei Liu, Jiangtao Feng, Hongli Yu, Yuxuan Song, Yuqiang Li,
Shufei Zhang, LEI BAI, Wei-Ying Ma, Hao Zhou*

Background

Retrosynthetic planning

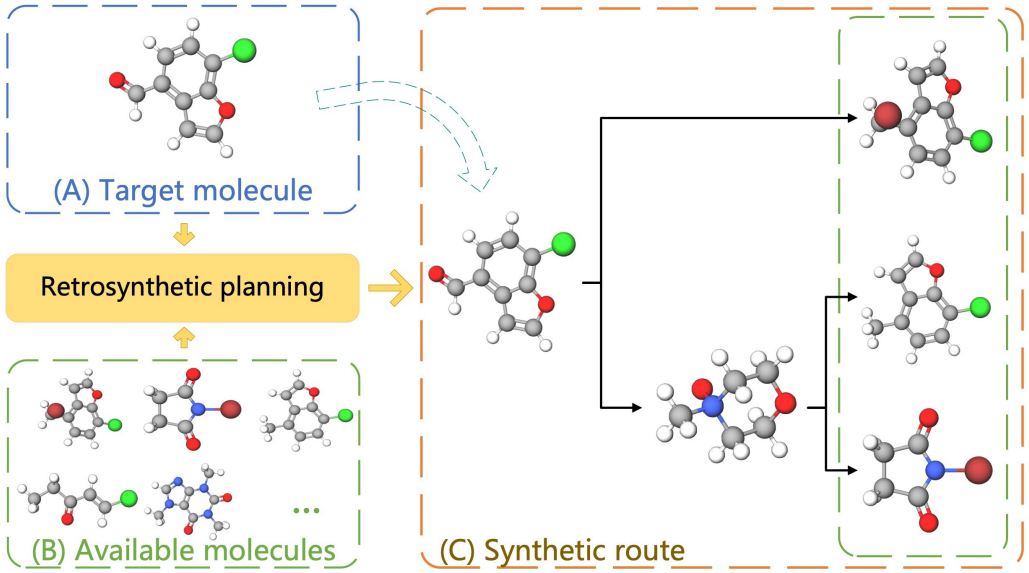
- A systematic process in organic chemistry for designing a synthesis pathway by working backward from a target molecule to simpler, commercially available starting molecules.

Building blocks

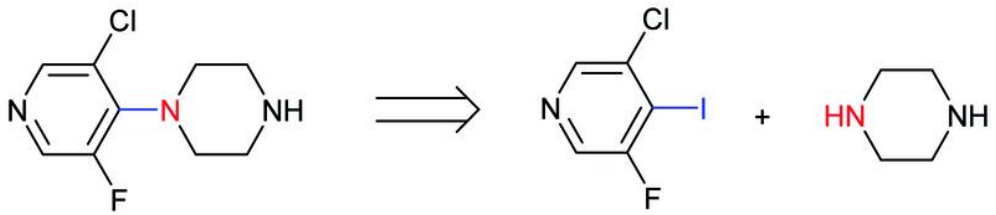
- A set of commercially available molecules. In this work, we use 23,081,633 commercially available molecules from eMolecules[1].

Single-step retrosynthesis

- Predict the reactants that can lead to a given product molecule through a single reaction step. In this work, we use a template based MLP from [2].



Example of retrosynthetic planning from [3]



Example of single-step retrosynthesis

[1] https://downloads.emolecules.com/
[2] Segler M H S, Waller M P. Neural-symbolic machine learning for retrosynthesis and reaction prediction[J]. Chemistry–A European Journal, 2017, 23(25): 5966-5971.
[3] Xie S, Yan R, Han P, et al. Retrograph: Retrosynthetic planning with graph search[C]//Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2022: 2120-2129.

Background

Retrosynthetic planning

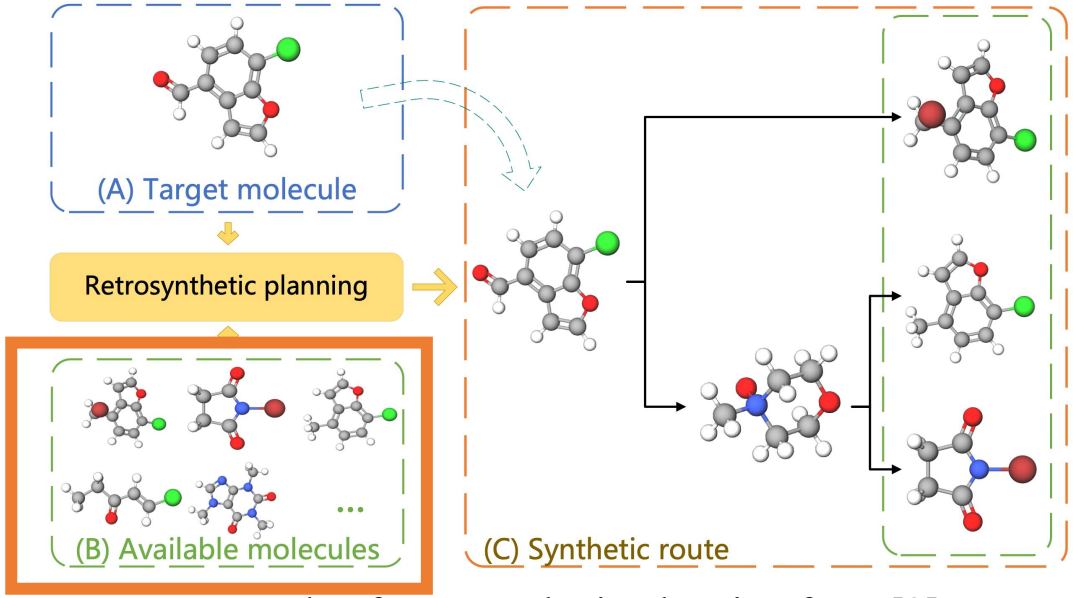
- A systematic process in organic chemistry for designing a synthesis pathway by working backward from a target molecule to simpler, commercially available starting molecules.

Building blocks

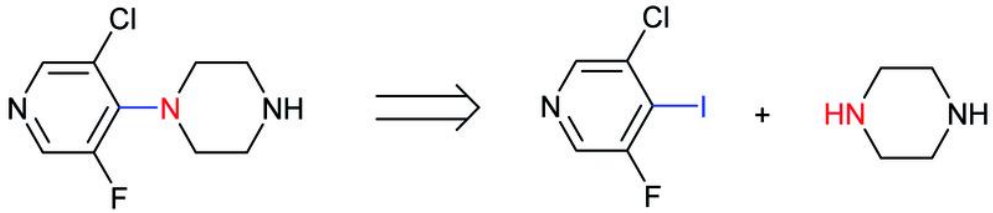
- A set of commercially available molecules. In this work, we use 23,081,633 commercially available molecules from eMolecules[1].

Single-step retrosynthesis

- Predict the reactants that can lead to a given product molecule through a single reaction step. In this work, we use a template based MLP from [2].



Example of retrosynthetic planning from [3]



Example of single-step retrosynthesis

[1] https://downloads.emolecules.com/
[2] Segler M H S, Waller M P. Neural-symbolic machine learning for retrosynthesis and reaction prediction[J]. Chemistry–A European Journal, 2017, 23(25): 5966-5971.
[3] Xie S, Yan R, Han P, et al. Retrograph: Retrosynthetic planning with graph search[C]//Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2022: 2120-2129.

Background

Retrosynthetic planning

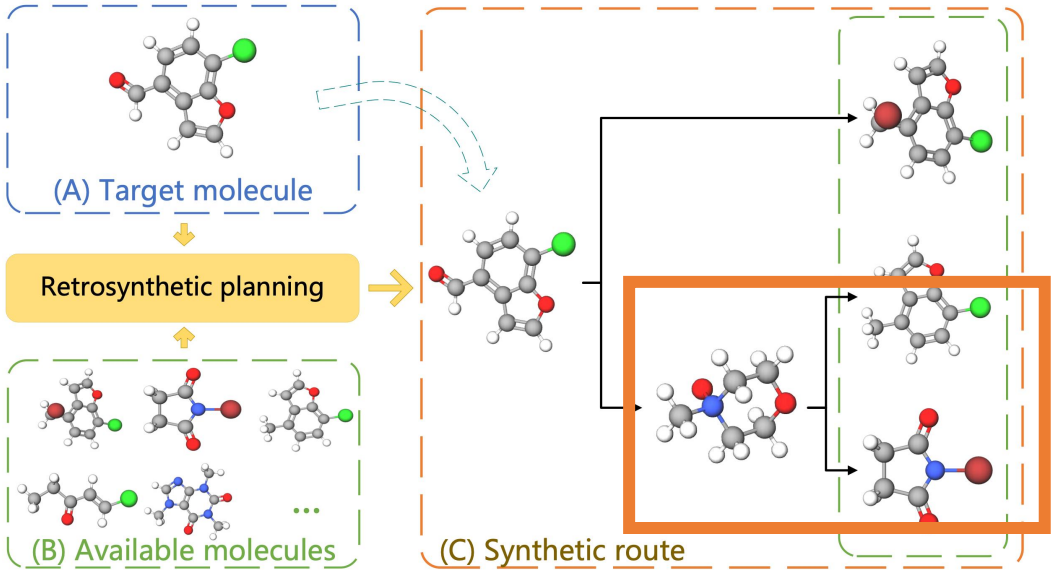
- A systematic process in organic chemistry for designing a synthesis pathway by working backward from a target molecule to simpler, commercially available starting molecules.

Building blocks

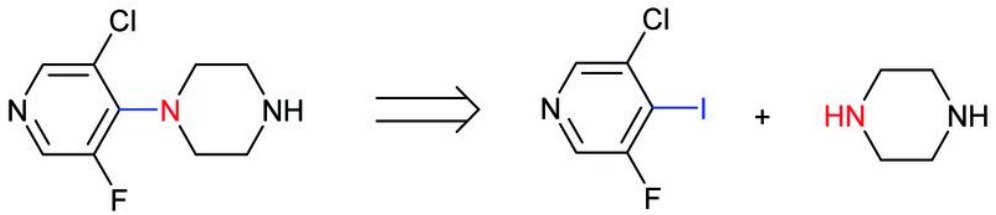
- A set of commercially available molecules. In this work, we use 23,081,633 commercially available molecules from eMolecules[1].

Single-step retrosynthesis

- Predict the reactants that can lead to a given product molecule through a single reaction step. In this work, we use a template based MLP from [2].



Example of retrosynthetic planning from [3]



Example of single-step retrosynthesis

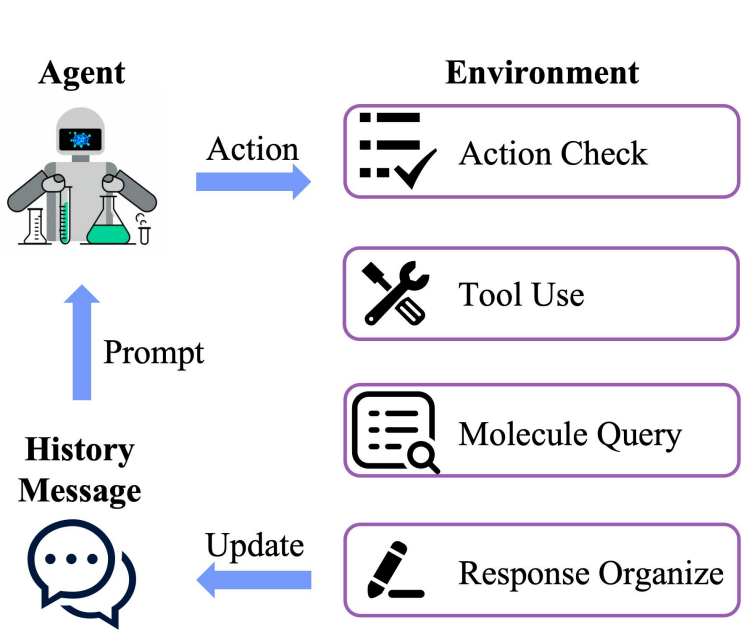
[1] https://downloads.emolecules.com/
[2] Segler M H S, Waller M P. Neural-symbolic machine learning for retrosynthesis and reaction prediction[J]. Chemistry–A European Journal, 2017, 23(25): 5966-5971.
[3] Xie S, Yan R, Han P, et al. Retrograph: Retrosynthetic planning with graph search[C]//Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2022: 2120-2129.

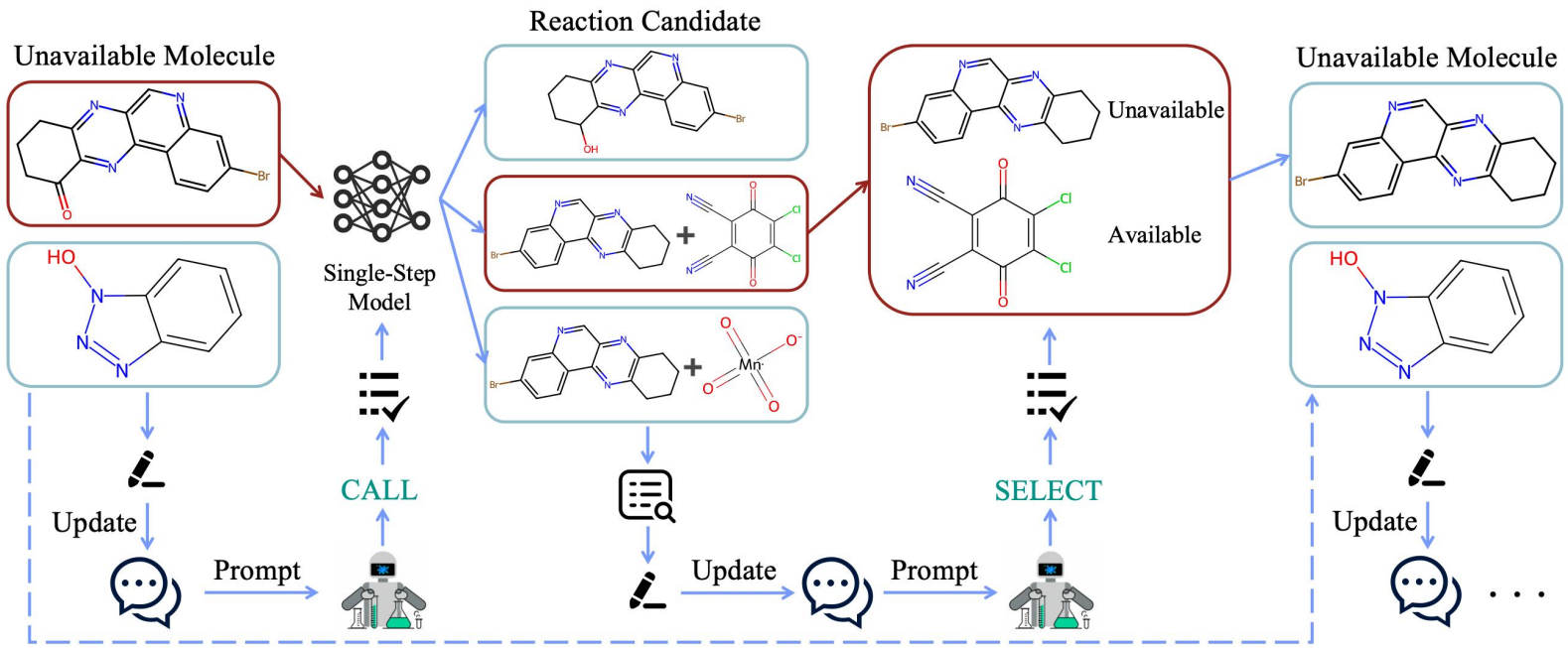
Method

Agent workflow

- The LLM-based agent interacts with the environment with structured text.
- With the increase of interactions, complex molecules are broken down gradually.

```
<think>
...
</think>
<tool_call>
{"name": "CALL", "arguments":{"molecule": "2-1"}}
</tool_call>
```

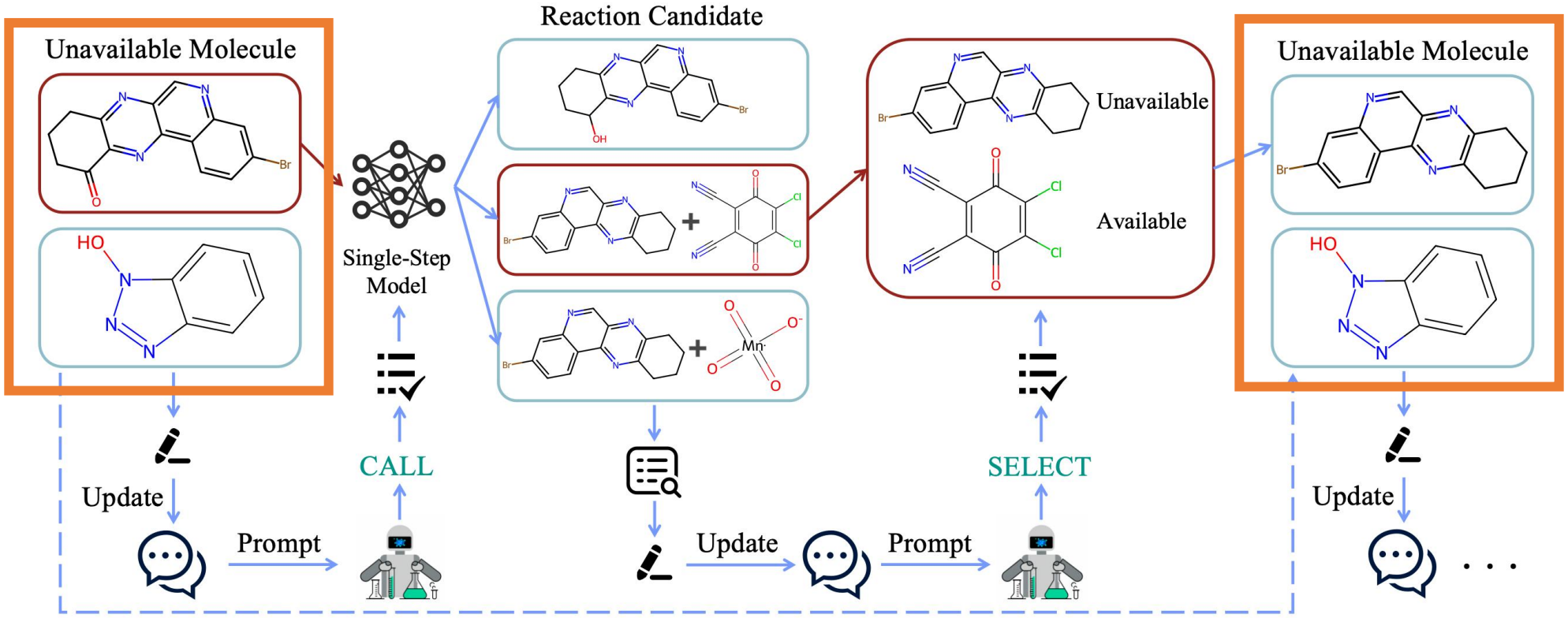




Method

Molecular state

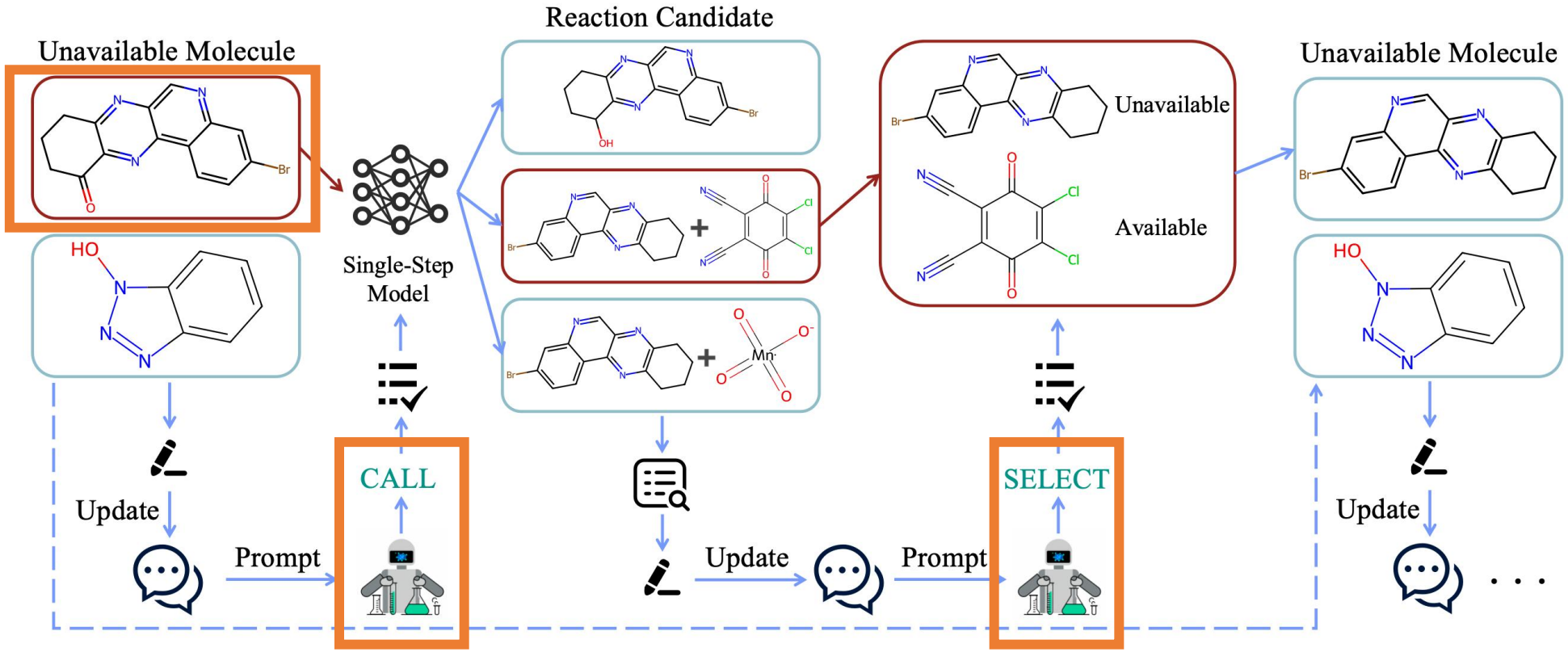
- Containing all molecules that have to be synthesized (unavailable).
- The first molecular state contains only the target molecule. The successful molecular state contains no molecules.



Method

Molecular state transition

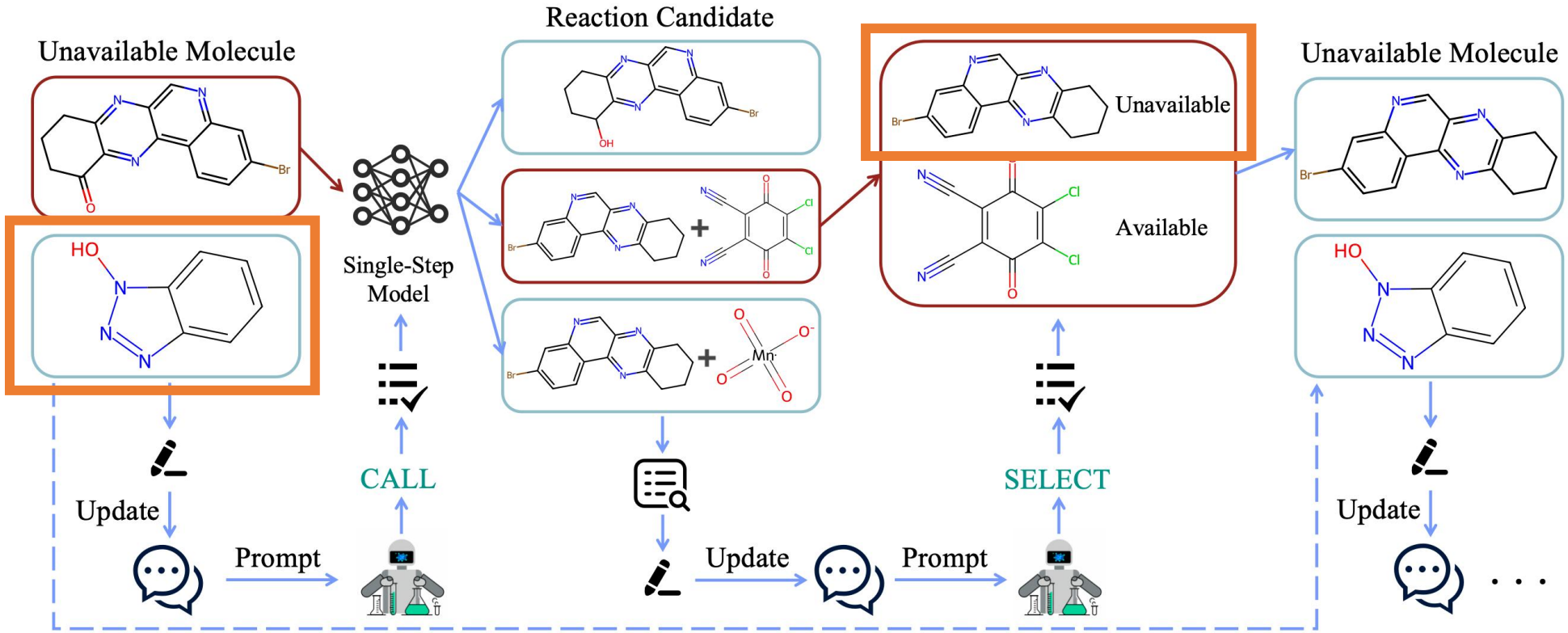
- In a molecular state transition, agent Retro-R1 process one molecule with two actions.
- ‘CALL’ uses the single-step model to propose multiple reaction candidates. ‘SELECT’ selects one from them.



Method

Molecular state transition

- In a molecular state transition, agent Retro-R1 process one molecule with two actions.
- ‘CALL’ uses the single-step model to propose multiple reaction candidates for a molecule. ‘SELECT’ selects one from them.



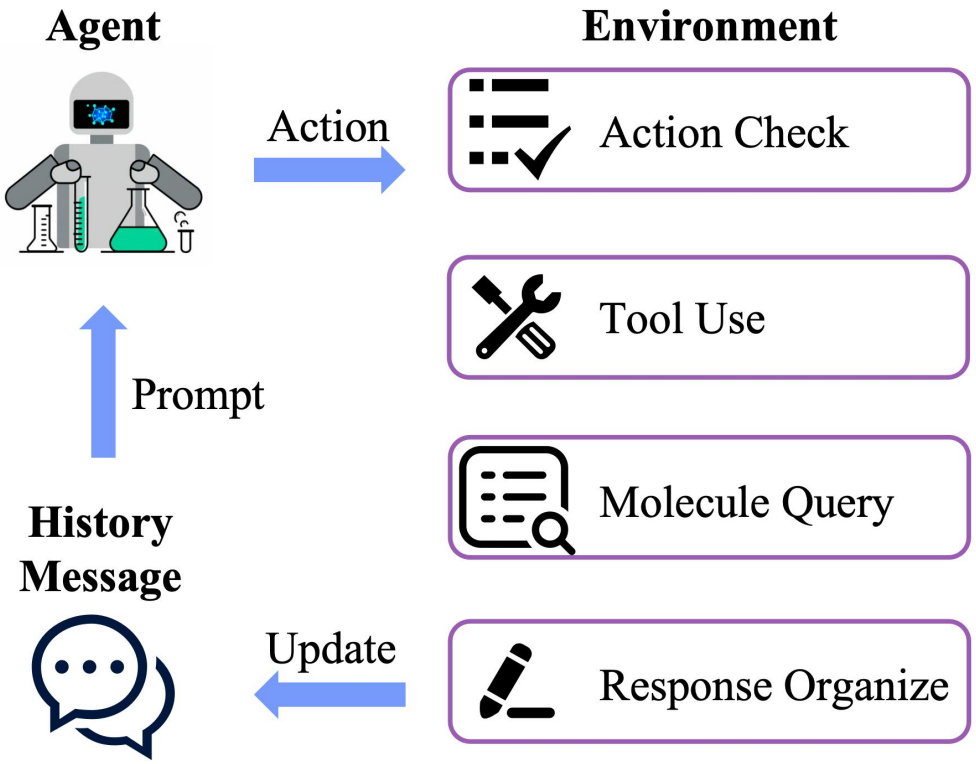
Method

Environment

- Check if the action is legal.
- Use tools according to the actions.
- Query molecules in building blocks.
- Organize responses based on the results of the other operations.
- Checks for termination conditions.
- Calculate the reward after termination.

History message

- Organized into multi-turn dialogues.
- Containing all the action-response pairs.



Method

Reinforcement Learning Algorithm

- We generalize the Proximal Policy Optimization (PPO) to multi-turn dialogues.

Reward

- A rule-based reward.
- If terminating successsfully, +0.9.
- If terminating in failure, -1.0.
- When success, if the number of interactions l is larger than 40, $(40 - l) \times 0.02$.

Method

Policy model loss function

- $\mathcal{D}$ is the training set, x is a target molecule, E is the environment.
- π_{old} is the old policy model, s is the token sequence of history message after termination.
- Define a token mask function $I(s_t)$, which is 1 if s_t is a token generated by the LLM, also 0.
- $r_t(\theta) = \frac{\pi_\theta(s_t|s_{<t})}{\pi_{\text{old}}(s_t|s_{<t})}$

$$\mathcal{L}^{\text{CLIP}}(\theta) = -\mathbb{E}_{x\sim\mathcal{D},s\sim\pi_{\text{old}}(\cdot|x;E)}$$
$$\left[\frac{1}{\sum_{t=1}^{|s|} I(s_t)} \sum_{t=1}^{|s|} I(s_t) \min\left(r_t(\theta)\hat{A}_t, \text{clip}\left(r_t(\theta), 1-\epsilon, 1+\epsilon\right)\hat{A}_t\right)\right]$$

Method

General Advantage Estimation（GAE）

- The original formula of GAE is

$$Adv_t = (R_t + \gamma * V_{t+1} - V_t) + \gamma * \lambda * Adv_{t+1}$$
$$Adv_T = R_T - V_T$$

- Set $\lambda = 1$, $\gamma = 1$, $V_T = 0$

$$\hat{A}_t = \sum_{k=t}^{T-1} R_k - V_\phi(s_t)$$

- In R_k, the KL term for environment tokens is set to 0.

Method

Value model loss function

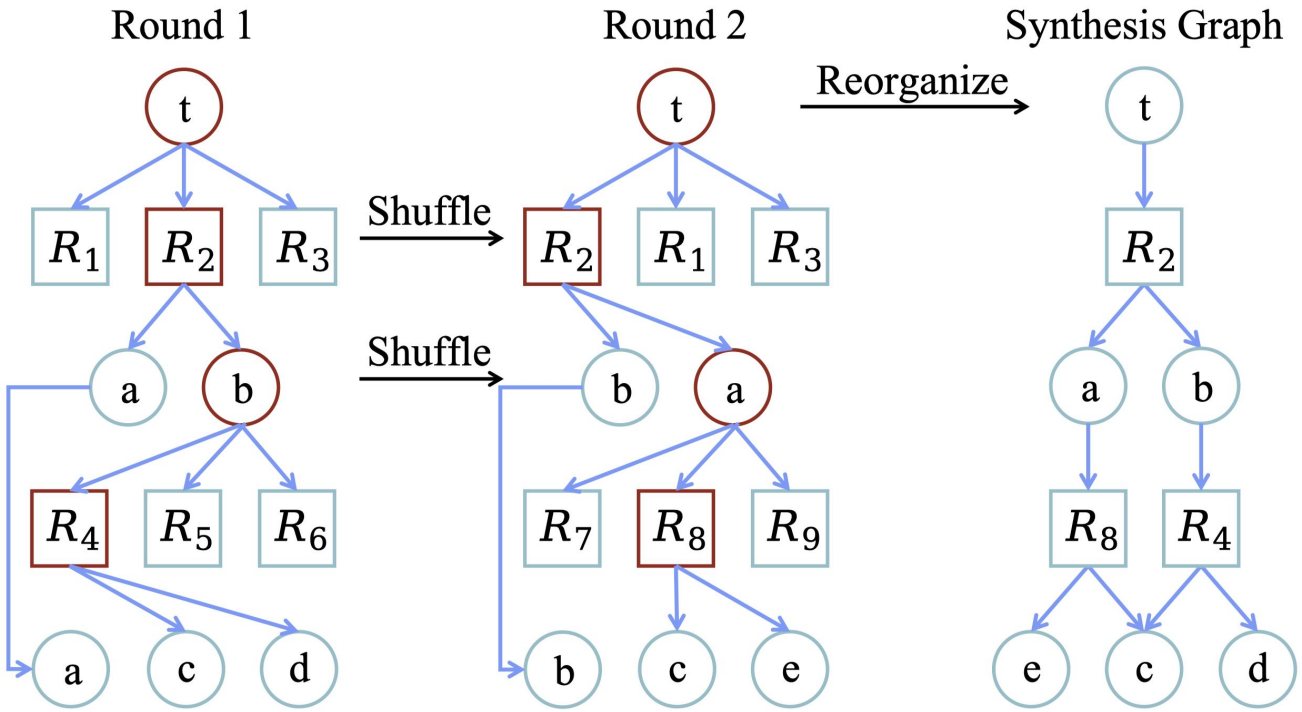
- For LLM tokens, it's the same as the original function.
- For environment tokens, the target is set to 0.

$$\mathcal{L}^{\mathrm{VF}}(\phi) = I(s_t)(V_\phi(s_t|s_{<t}) - V_t^{target})^2 + (1 - I(s_t))V_\phi^2(s_t|s_{<t})$$

Method

Iterative planning strategy

- A memory mechanism is implemented to record all the intermediate information.
- The orders of molecules and reactions are shuffled to encourage exploration.
- Multiple rounds of planning are reorganized into an AND-OR graph and the best routes are searched.



Result

Dataset

- Retro*-190
- ChEMBL-1000

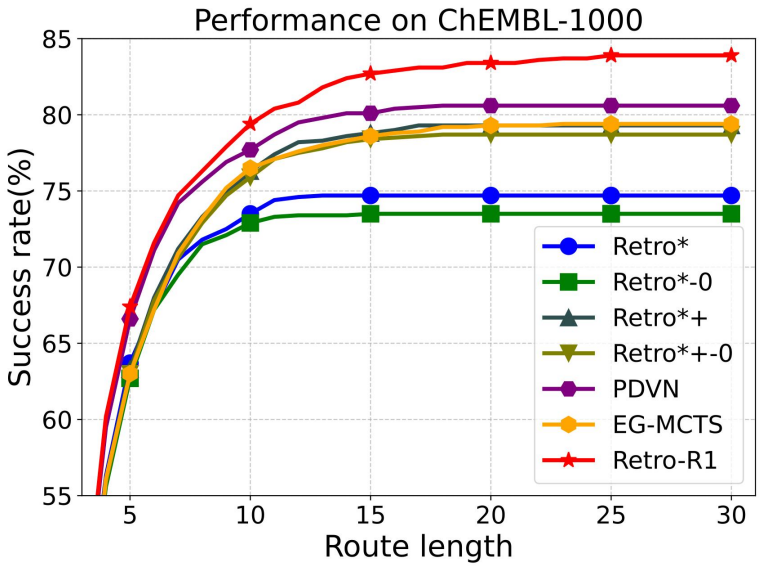


Table 1: Results on retro*-190 and ChEMBL-1000. Pass@1 is short for Pass@1 Success Rate. RETRO-R1 is tested in ten random runs, and the means and standard deviations are reported. All baselines are deterministic, meaning they produce the same results on repeated runs; therefore, standard deviations are not reported. Shorter routes refer to the number of molecules whose routes found at the iteration limit of 500 are shorter than the reference routes. This metric is not applicable to ChEMBL-1000 because it lacks reference routes. The best results are marked in bold, and the second-best results are marked with underlines.

		Success Rate (%) at Iteration Limit N						
Method	Pass@1 (%)	$N = 50$	$N = 100$	$N = 200$	$N = 300$	$N = 400$	$N = 500$	Shorter routes
Performance on Retro*-190								
Greedy DFS	19.47	19.47	19.47	19.47	19.47	19.47	19.47	8
retro*	20.53	38.95	50.00	68.95	74.74	77.89	79.47	64
retro*-0	19.47	27.37	37.89	55.79	62.63	72.11	74.21	57
retro*+	30.00	55.79	68.42	79.47	83.16	84.21	84.21	81
retro*+-0	25.26	47.89	61.05	76.84	82.11	85.79	86.32	79
EG-MCTS	46.84	58.42	64.74	69.47	75.26	78.42	81.05	68
PDVN	42.63	64.74	72.11	**82.63**	**87.89**	**91.58**	**92.11**	**90**
RETRO-R1	**55.79**	**73.21±0.80**	**77.21±1.47**	82.58±1.46	84.47±0.82	85.89±0.57	86.95±0.52	87
Performance on ChEMBL-1000								
Greedy DFS	38.10	38.10	38.10	38.10	38.10	38.10	38.10	–
retro*	47.70	65.60	69.00	71.60	73.20	74.00	74.40	–
retro*-0	38.10	64.60	67.20	70.30	71.80	72.70	73.40	–
retro*+	53.50	70.50	73.90	76.20	78.00	78.80	79.40	–
retro*+-0	49.00	69.40	73.60	75.60	76.70	77.60	78.50	–
EG-MCTS	59.70	71.20	74.20	76.90	78.30	79.00	79.30	–
PDVN	60.00	73.90	76.40	78.10	79.60	80.30	80.60	–
RETRO-R1	**73.30**	**77.88±0.35**	**80.21±0.18**	**81.79±0.23**	**82.62±0.29**	**83.28±0.25**	**83.72±0.17**	–

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

Thank you for listening.