

RetroAgent: Harnessing LLMs to Search Over Structured Memory for Agentic Retrosynthesis Planning

Yanqiao Zhu^{*1}, Jingru Gan^{*1}, Xiaoqi Sun², Fang Sun¹, Yidan Shi¹, Md Mofijul Islam³, Chao Shang³, Wenhao Gao⁴, Connor W. Coley², Yizhou Sun¹, Wei Wang¹
¹UCLA ²MIT ³Amazon ⁴UPenn ^{*}Equal contribution

UCLA Samueli
Computer Science

Motivation

Multi-Step Retrosynthesis

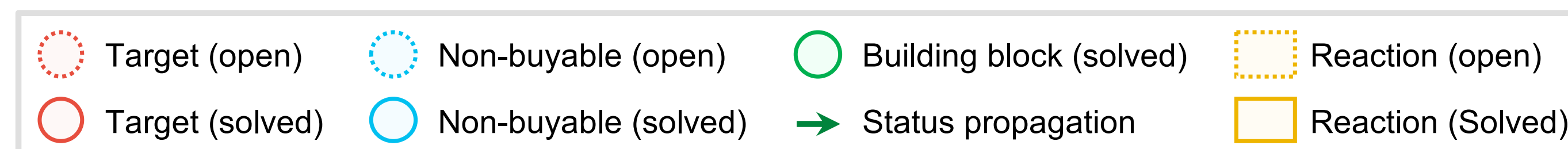
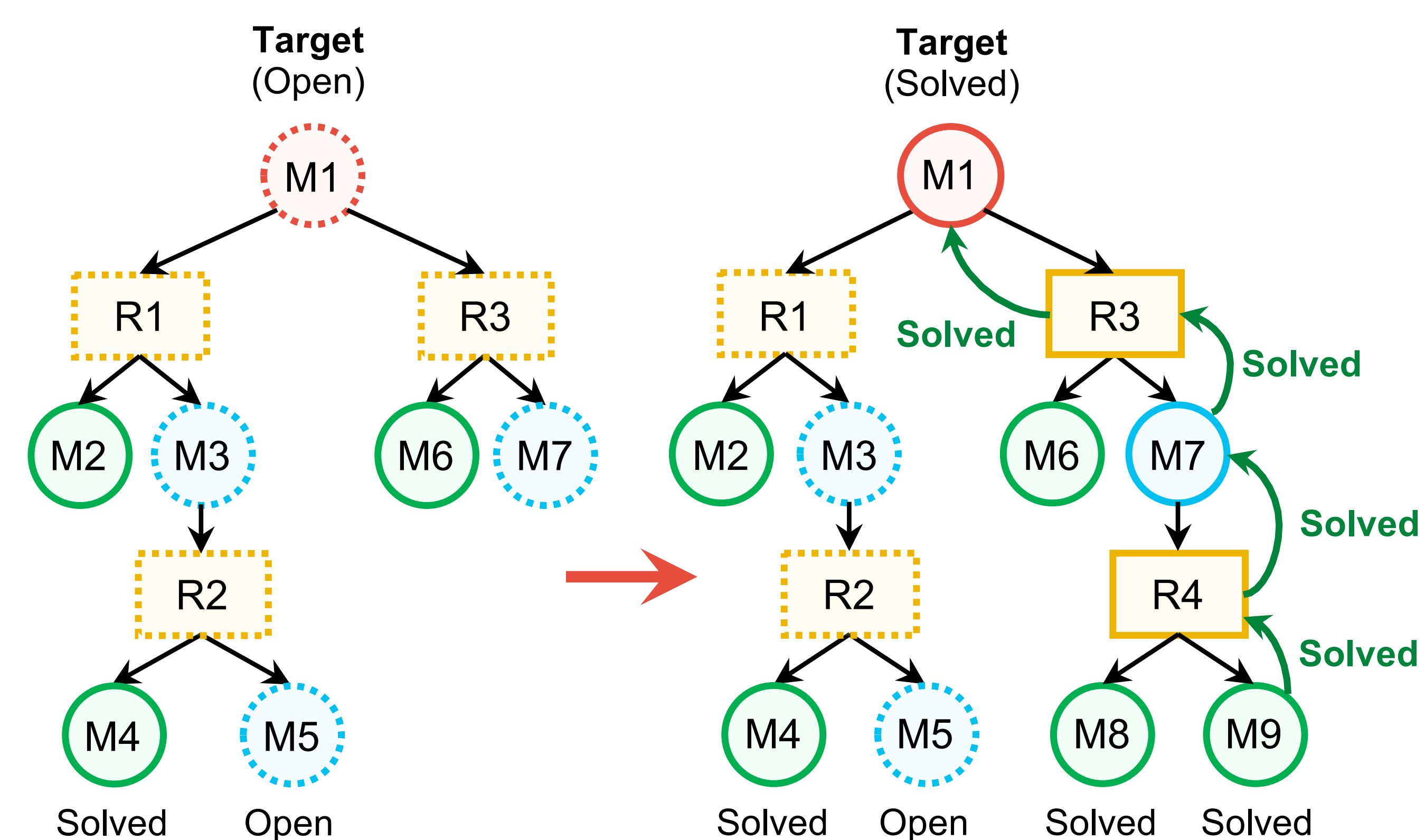
- Break a target molecule down into purchasable building blocks
- Long-horizon planning: The search space grows combinatorially with depth

Limitations of Prior Approaches

- Search planners:** Value functions score molecules in isolation rather than the whole route
- Prior LLM agents:** A two-action interface cannot explore alternative decompositions

Structured Memory

- The LLM agent edits an AND-OR graph $\mathcal{G} = (\mathcal{V}_M \cup \mathcal{V}_R, \mathcal{E})$ through tool calls
- Molecules are OR nodes:** Solved if *any* child reaction is solved; Open leaves form the frontier
- Reactions are AND nodes:** Solved only if *all* children are solved
- The environment propagates status to ancestors; the agent does not need to recompute it

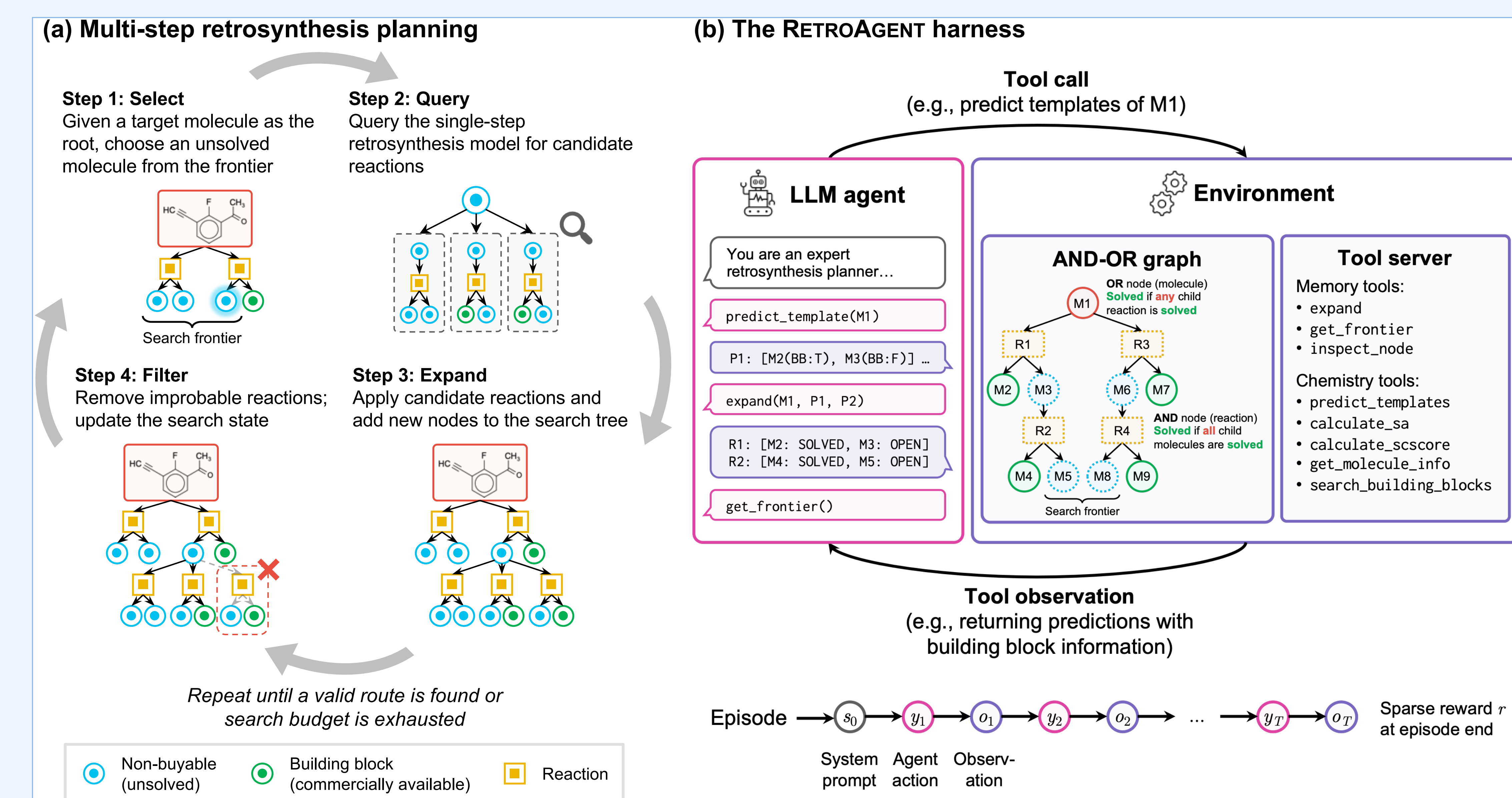


The RetroAgent Framework

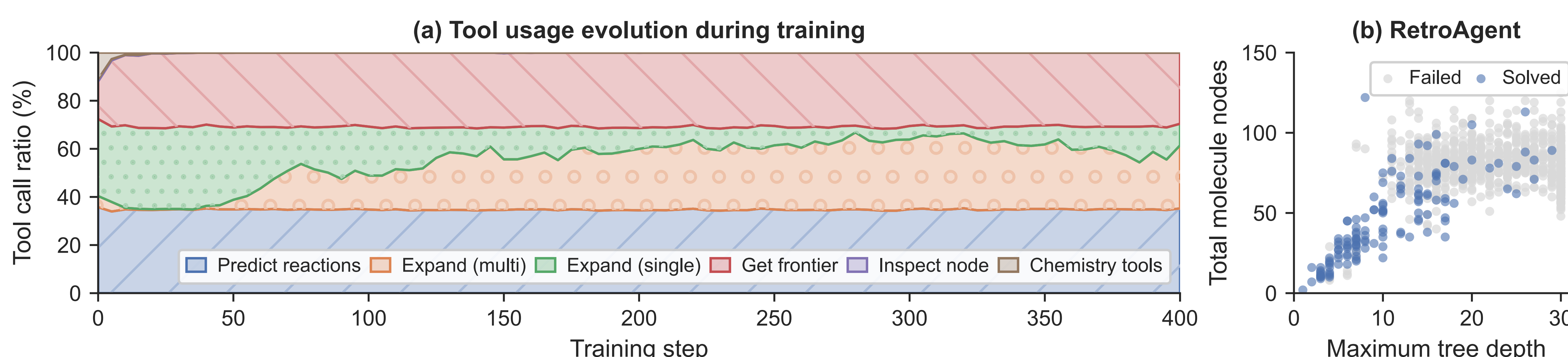
TL;DR: RetroAgent gives an LLM direct control of an AND-OR graph as structured memory; the environment tracks node status and reaction validity, and RL trains the search strategy

Agent: Picks the molecule to expand, the reactions to apply, and how many branches to open

Environment: Maintains the graph, rejects cyclic, duplicate and invalid reactions, propagates status, and reports the frontier



Tool Usage During Training



- Multi-expand:** Becomes the dominant expansion mode after step 350
- Chemistry tools:** Calls drop to near zero by step 10, since the building block annotations already carry the same information

Agent Interface

Tools Memory tools edit the graph; chemistry tools give heuristic guidance

- Memory:** predict_templates, expand, get_frontier, inspect_node
- Chemistry:** get_molecule_info, search_building_blocks, calculate_scscore, calculate_synthetic_accessibility

Design Choices

- Multi-expand:** Apply several predictions per call, opening many OR branches at once
- Building block annotations:** Candidates report per-reactant purchasability; no extra call is needed
- Symbolic IDs:** Molecules M#, reactions R#, predictions P#, never raw SMILES, which removes copying errors

RL Training

- Qwen3-4B-Instruct-2507 (non-thinking) trained on 8xH200 with GSPO, a critic-free sequence-level policy gradient
- One episode is one search trajectory with an outcome-only reward, and tool output is masked from the loss
- Reward design:** Budget and depth penalties

$$r = \begin{cases} 1.0 - \delta_N - \delta_D + p, & \text{route found} \\ -1.5, & \text{no valid call} \\ -1.0, & \text{otherwise} \end{cases}$$

where δ_N (budget) and δ_D (depth) grow by 0.02 per unit past 20, capped at 0.3, and p accumulates -0.05 per turn with a tool error

- Adaptive width:** Five branches at the root, one when only a single prediction has purchasable reactants
- Candidates:** The top 50 predictions are shuffled and truncated to 20, so rank carries no signal

Main Results

	Method	Pass@1	N=50	N=200	N=500
USPTO-190	Greedy DFS	19.47	19.47	19.47	19.47
	Retro*	20.53	38.95	68.95	79.47
	Retro*-0	19.47	27.37	55.79	74.21
	Retro*+	30.00	55.79	79.47	84.21
	Retro*+-0	25.26	47.89	76.84	86.32
	EG-MCTS	46.84	58.42	69.47	81.05
	PDVN	42.63	64.74	82.63	92.11
ChEMBL-1000	Retro-R1	50.00	64.63	78.11	84.32
	RetroAgent	53.26	58.42	78.63	85.05
	Greedy DFS	38.10	38.10	38.10	38.10
	Retro*	47.70	65.60	71.60	74.40
	Retro*-0	38.10	64.60	70.30	73.40
	Retro*+	53.50	70.50	76.20	79.40
	Retro*+-0	49.00	69.40	75.60	78.50
	EG-MCTS	59.70	71.20	76.90	79.30
	PDVN	60.00	73.90	78.10	80.60
	Retro-R1	68.52	73.92	80.36	82.42
	RetroAgent	73.82	76.36	81.96	84.40

- Highest Pass@1** with a 4B model, in-distribution (USPTO-190) and out-of-distribution (ChEMBL-1000)
- ChEMBL-1000:** +5.30 points over the best baseline
- Mean expansion width:** 1.71 per expand call

Ablations

Configuration	Pass@1	N=500	Short routes
RetroAgent	53.26	85.05	38.3
Zero-shot, no RL	4.42	16.63	16.4
w/o depth penalty	51.16	81.26	31.4
w/o multi-expand	40.63	79.05	32.8
w/o both	39.58	77.37	26.6

- Zero-shot:** No RL reaches 4.42%, ruling out pretraining leakage
- Multi-expand:** -12.63 Pass@1 without it
- Depth penalty:** Raises shorter routes from 31.4 to 38.3 of 190 targets