Focus! Don't Spread: Prioritized Recursive Sub-Search for Retrosynthesis Planning

Guillaume Schmitz[1,2], Almaz Gilmullin [2], Alexandre Varnek [2]

[1]*Complex Systems Chemistry (CSC) Graduate School* , [2]*Laboratoire de Chémoinformatique, Université de Strasbourg, 4 rue Blaise Pascal, Strasbourg, France.*

The concept of retrosynthesis — the systematic disconnection of a target molecule into simpler precursors — was formalised by E.J. Corey [1] and underpins modern computer-aided synthesis planning (CASP). Over the past decade, Monte Carlo Tree Search (MCTS) has emerged as the dominant search framework [2], coupling neural network-guided expansion policies with tree-based exploration to navigate the combinatorial space of retrosynthetic routes. SynPlanner [3] exemplifies this approach, pairing a graph neural network trained on USPTO reactions with MCTS to generate multi-step routes toward commercially available building blocks.

To evaluate planners across difficulty tiers, we use the synthetic accessibility (SA) score [4] to partition a pre-curated, high-diversity set of 700 drug-like molecules into seven bins (SA 1.5-8.5). We first benchmarked a range of MCTS evaluation heuristics across all bins; heavy atom count consistently outperformed SA score, molecular weight, and composite functions, and was adopted as the default evaluation function throughout.

Despite strong performance on easy targets, MCTS-based planners face a systematic limitation as complexity increases: the tree grows wide before growing deep. Each expansion produces multiple co-precursors that must all be independently solvable. At nodes where all but one co-precursor is a known building block, the route depends entirely on that single fragment, which we term « **sole blocker** ». Standard MCTS dilutes search effort across the global tree rather than concentrating it on the one fragment whose resolution would immediately unlock one or more complete routes.

We address this with a **recursive solver** integrated into SynPlanner. After a main MCTS phase, the solver identifies sole blockers in the tree, ranks them by size and occurrence across partial routes, and launches a focused MCTS sub-search on the top candidate using the same policy and building block stock. If solved, its route is merged to the parent route; otherwise, the next sole blocker is attempted, up to a configurable limit.

On our SA-score benchmark, the recursive solver achieves an overall 57% improvement in solvability over the rollout baseline and 18% over the heavy-atom heuristic, with the most pronounced gains in mid-to-hard bins (SA 3.5-6.5). Such results were obtained using the same total iteration budget, which demonstrates that targeted sub-search on sole blockers is a computationally efficient complement to improved evaluation functions.

These results establish prioritized recursive sub-search as a simple mechanism to increase solvability, orthogonal to policy improvements and broadly applicable within MCTS-based planners.

Bibliography :

[1] E. J. Corey et al., Computer-Assisted Analysis in Organic Synthesis. Science **228**, 408-418 (1985).
[2] M. Segler, M. Preuss, & M. Waller. Planning chemical syntheses with deep neural networks and symbolic AI. Nature **555**, 604-610 (2018).
[3] Tagir Akhmetshin, et al., SynPlanner: An End-to-End Tool for Synthesis Planning. Journal of Chemical Information and Modeling. **65** (1), 15-21 (2025).
[4] P. Ertl and A. Schuffenhauer. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. J Cheminform **1**, 8 (2009).