Multi-objective synthesis planning by means of Monte Carlo Tree search

Artificial Intelligence in the Life Sciences 7 (2025) 100130

Contents lists available at ScienceDirect

Artificial Intelligence in the Life Sciences

journal homepage: www.elsevier.com/locate/ailsci

Short communications

Multi-objective synthesis planning by means of Monte Carlo Tree search

Helen Lai [a,*], Christos Kannas [b], Alan Kai Hassen [c,d], Emma Granqvist [b,e], Annie M. Westerlund [b], Djork-Arné Clevert [d], Mike Preuss [c], Samuel Genheden [b]

[a] Molecular AI, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK
[b] Molecular AI, Discovery Sciences, R&D, AstraZeneca, Gothenburg, Sweden
[c] Leiden Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands
[d] Machine Learning Research, Pfizer Research and Development, Berlin, Germany
[e] Department of Computer Science and Engineering, Chalmers University of Technology, Gothenburg, Sweden

ARTICLE INFO

Keywords:
Monte Carlo Tree search
Pareto optimality
Multi-objective optimization
Reinforcement learning
Markov Decision Process
Policy network
Tree edit distance

ABSTRACT

We introduce a multi-objective search algorithm for retrosynthesis planning, based on a Monte Carlo Tree search formalism. The multi-objective search allows for combining diverse set of objectives without considering their scale or weighting factors. To benchmark this novel algorithm, we employ four objectives in a total of eight retrosynthesis experiments on a PaRoutes benchmark set. The objectives range from simple ones based on starting material and step count to complex ones based on synthesis complexity and route similarity. We show that with the careful employment of complex objectives, the multi-objective algorithm can outperform the single-objective search and provides a more diverse set of solutions. However, for many target compounds, the single- and multi-objective settings are equivalent. Nevertheless, our algorithm provides a framework for incorporating novel objectives for specific applications in synthesis planning.

1. Introduction

Retrosynthesis is a fundamental method in organic chemistry, enabling chemists to systematically deconstruct complex molecules into simpler precursors [1]. Computer-aided techniques have increasingly become powerful tools to aid chemists in this endeavour, largely coinciding with the rise of deep learning models [2,3]. In this field, this task is typically divided into single-step retrosynthesis, where a single compound is broken down, and multi-step retrosynthesis where this processes is repeated until some conditions are met [4]. Despite receiving significant attention, the problem remains challenging due to the vast chemical space and the complexity of reaction pathways. In this work, we build on recent progress in multi-objective search techniques developed within the evolutionary optimization community and adapt them to the problem of multi-step retrosynthesis.

Advancements in retrosynthesis research have demonstrated success in using search algorithms to navigate the vast chemical space. Based on the number of publications, the two most popular approaches are A* search [5,6] and Monte Carlo Tree Search (MCTS) [7,8], although other algorithms have been used as well [9–13] but to a lesser extent. A* and MCTS differ in how they estimate future costs or values when expanding nodes. In A* search, cost estimation is learned offline, typically through supervised training using routes from reaction databases like the United States Patent and Trademark Office (USPTO) [6]. This approach necessitates training a new model whenever a new search objective is introduced. In contrast, MCTS performs value estimation online during the roll-out stage (discussed further in Section 3), providing the flexibility to add objectives without needing separate model training [8,14]. In addition to its flexibility, a recent benchmark study showed that while MCTS and A* search (Retro*) are comparable in finding synthesis routes with all starting materials available in stock, MCTS outperforms A* in recovering reference routes and generating more diverse pathways [15]. Given that MCTS estimate the values of nodes in an online fashion and thus is more straightforward to adapt to novel objectives, this paper focuses on MCTS as the preferred algorithm.

Although several variations of MCTS and novel search algorithms have been presented for retrosynthesis [16–23], only a single scalar reward has been used to represent the desired properties of the synthesis route [8,14]. When multiple properties are required, existing methods assume these objectives can be linearly combined into one [24]. However, every single linear combination will only lead to one point on the Pareto front of optimal solutions in case of conflicting objectives. The weight assigned to each objective can be arbitrary, as can be the choice of a normalization method when objectives are measured on different scales. Any inappropriate decisions regarding these

* Corresponding author.
E-mail address: helen.lai1@astrazeneca.com (H. Lai).

https://doi.org/10.1016/j.ailsci.2025.100130
Received 7 October 2024; Received in revised form 16 January 2025; Accepted 9 February 2025
Available online 19 February 2025

issues may introduce undue bias during the search process. Recent multi-objective MCTS algorithms have been developed to address this issue [25–28], but their application has been largely confined to the computational game domain, and their applicability to retrosynthesis remains unknown.

Among the existing methods, one method focuses on modelling a posterior distribution over the expected future returns rather than its point estimate [25]. The authors argue that this approach is superior for preventing negative outcomes in stochastic environments. However, in our set up, the state transitions are deterministic, and since the routes are scored based on their physicochemical properties, the concept of risks is not applicable. The other line of work focuses on extending the Upper Confidence Bound (UCB) criterion used in single-objective MCTS to the multi-objective set up. The initial attempt [26] reduced the multi-objective problem into a single objective one. They achieved this by keeping track of a Pareto-front archive of rewards received at terminal nodes and iteratively calculating the hyper-volume (HV) of the Pareto-front at each selection stage [26]. While this algorithm demonstrated the state-of-the-art performance in two game domains, it did so at the expense of high computational cost. The later attempt by Perez et al. (2014) addresses this problem by maintaining the local pareto-front at each node, and only updating the pareto-front of parent node if the reward is received at its child node is dominating its current local pareto-front [27]. Subsequently, another multi-objective variant was introduced with a different formulation of UCB criterion. Instead of keeping track of a pareto front over the final solutions, each node maintains the pareto set of its child nodes. The method also established a logarithmic lower bound on the frequency of sub-optimal node selections and a polynomial convergence rate towards optimal solutions [28].

Here, we extend and tailor these developments to the retrosynthesis domain, and integrate multi-objective MCTS with AizynthFinder, a widely used open-source MCTS retrosynthesis library [29,30]. Similar to [28], our method aims to sample valid synthesis routes from solutions on the Pareto frontier [28]. In addition to the multi-objective extension, we also introduce two novel scoring objectives. The first one focuses on synthetic complexity, aiming to encourage a decrease in synthetic complexity from the intermediate reactants to the starting materials. The second one focuses on route similarity, aiming to steer the search towards some known reference routes. The impact of these two objectives are evaluated in both the single and multi-objective setting.

2. Theoretical background

2.1. Problem formulation

The synthesis planning problem involves breaking down a target molecule $m_0 \in \mathcal{M}$ into its reaction precursors until the reactants are in stock or some computational budget is reached. M is the set of possible reagents molecules that could be involved in a reaction. The synthesis route follows a tree-like structure that is built incrementally from a root node that represents the target compound m_0. The tree building process could be viewed as a sequential decision-making problem, which fits naturally in the Markov Decision Process (MDP) framework. Mathematically an MDP is defined by the following components:

1. The state space $S := 2^{\mathcal{M}}$ is formally defined as the powerset of the set of molecules $\mathcal{M}$; in other words, a state $s \in S$ represents some combination of molecules from $\mathcal{M}$. The starting state $s_0 := \{m_0\}$, represents the state corresponding to a single starting molecule m_0 before undergoing retrosynthesis.
2. A state dependent action function A

$$A : S \to U, \quad U := \bigcup_{s \in S} A(s),$$

where $A(s)$ is the set of possible actions reachable from s, and each action $a \in A(s)$ corresponds to a single step reaction for one of the molecules in state s. The single step reaction is defined by the reaction templates which is a mapping between reactants and products by specifying which bonds are formed or broken and how atoms are rearranged.
3. A transition probability function $P(s\prime|s,a) : S \times A \times S \to \mathbb{R}[0,1]$. Since this is commonly assumed to be deterministic in the context of synthesis planning, we use transition function $\Gamma(s,a)$ to represent the state transition from s to $s\prime$ by following action a.
4. a reward function $R : S \times A \to \mathbb{R}$.

At this point, it is important to clarify that the state, as defined above, is a mathematical concept representing the snapshot of the set of reagents involved at a specific stage of the retrosynthesis process. Moving forward, the term node will be used to refer to specific points within the search tree, a data structure used in the retrosynthesis framework. Each node in the search tree is associated with a state. Multiple nodes may correspond to the same state, as the tree structure allows different paths to converge on identical retrosynthesis scenarios. Furthermore, since nodes are sequentially added to the tree, each node allows tracing back the synthesis route leading to it. At the terminal node, this process reveals the complete synthesis route for the molecule m_0.

In the multi-objective setting, there exists more than one reward function, which extends the classical MDP framework to its more generalized form where a set of reward functions $\{R_i\}_{i=1}^{d}$ are considered. Consequently, instead of using scalar reward to represent the value of a given state–action pair, we have a $d-$dimensional vectorial reward $\vec{r}$. The MDP adapted for the synthesis planning problem also has finite horizon with non-discounted rewards that are only seen at the terminal node. A node is considered terminal if any of the following conditions is satisfied:

- no reaction template is available for any of the molecules in its state s
- all molecules in its state s are in stock
- the node is at a depth greater than a certain user-specified cut-off.

2.2. Single objective planning goal

Although the focus of the paper is the multi-objective setting, it is also worth briefly introducing the single objective setting as it is used as a benchmark in our experimental set up. Whereas in the multi-objective case we would have a vector based reward $\vec{r} := (r_1, \ldots, r_d)$, in the single objective case a scalar reward signal is used. In our case, we obtained this by mapping each multi-objective reward vector to a scale, by computing the average of the components, namely $c(\vec{r}) := \frac{1}{d}\sum_{i=1}^{d} r_i$. In this setting, the goal is to approximate the optimal value function $V^*(s), \forall s \in S$:

$$V^*(s) = \max_a V^*(\Gamma(s,a)), \quad V^*(s) = R(s)\ \forall s \in G$$

where G is the set of terminal nodes. An optimal policy π^* for any state can be obtained from $V^*(s)$ by defining

$$\pi^*(s) := \arg\max_{a \in A(s)} V^*(\Gamma(s,a)),$$

with ties broken arbitrarily. Note that treating the multi-objective setting as a weighted sum of single objectives can hide optimal solutions. As discussed in [24], section 4.1, there are instances where a multi-objective approach can achieve solutions that are inaccessible through any single-objective weighting. While the extremal points – solutions that optimize one criterion – can always be obtained, the more valuable compromise solutions in the middle often cannot be reached with single-objective methods.

2.3. Multi-objective planning goal

In the multi-objective setting, we need to introduce a different notion of ordering, *dominance*, to compare vector based outcomes, which will allow us to establish a Pareto front.

Pareto optimality. Assuming maximization, given two synthesis routes x_i, x_j, and a d - dimensional route evaluation criteria $R(.)$, x_i is said to dominate x_j, which we will denote by $x_i \prec x_j$, provided the following conditions hold jointly:

1. $R_k(x_i) \geq R_k(x_j) \;\; \forall k \in \{1, \ldots, d\}$;
2. $\exists q \in \{1, \ldots, d\}$ such that $R_q(x_i) < R_q(x_j)$.

Given a set of solutions $\mathcal{X}$, the Pareto optimal set of solutions is defined as

$$P(\mathcal{X}) := \{x \in \mathcal{X} : \nexists y \in \mathcal{X} \setminus \{x\} \text{ such that } y \prec x\}.$$

Any solution satisfying $x^* \in P(\mathcal{X})$ is said to be a *Pareto optimal solution*. The *Pareto Front* of a set of solutions can now be defined as

$$F(\mathcal{X}) := \{R(x)\}_{x \in P(\mathcal{X})}.$$

Following the above definition, our goal in the multi-objective setting is to provide a solution which we estimate to lie on the Pareto front. Ignoring computational restrictions, ideally we would sample a solution uniformly from the set of solutions on the Pareto front. However, to achieve this goal in a tractable computation manner, we follow the approach of Chen and Liu (2021). Instead of approximating the optimal value function $V^*(s)$, for each visited node s we recursively estimate a local Pareto front defined over its child nodes. The estimated optimal policy $\pi\hat{(s)}$ is then sampled uniformly at random from the local approximated Pareto optimal set, which in the limit of infinite compute guarantees that *some* solution on the global Pareto front will be returned [28].

3. The multi-objective Monte Carlo Tree Search algorithm

Having outlined the theoretical foundations of single- and multi-objective MCTS, this section delves into the implementation details of the algorithms being evaluated. Specifically, the MCTS implementation in AiZynthfinder [29,30] is chosen as the single-objective benchmark, while for the multi-objective case, we adapt the algorithm proposed by Chen and Liu (2021) [28] for the synthesis planning problem.

At a high level, single-objective MCTS repeatedly selects and samples various parts of the tree to approximate the expected return down the various parts of the tree. The multi-objective version follows a similar idea but instead, the iterative sampling and updating is aimed to approximate a Pareto front in the solution space from which we could sample synthesis routes from. To describe the algorithm in detail, we shall break it down into its four main key stages: selection, expansion, roll-out, backup.

3.1. Selection

The selection stage involves a guided traversal of the visited nodes of the tree. At each visited node, it faces a multi-armed bandit problem which involves choosing one of the many possible actions based on some chosen search heuristics. In the single objective setting, the Upper Confidence Bound criterion [31] is used to perform the selection by approximating $\pi^*(s)$ as follows:

$$\pi\hat{(s)} \in \underset{a \in A(s)}{\arg\max} \frac{1}{q} \sum_{i=1}^{q} r_{i,\Gamma(s,a)} + c \frac{\sqrt{2log(T_s^q)}}{T^q_{\Gamma(s,a)}}$$

where $T^q_{\Gamma(s,a)}$ is the number of times the state–action pair s, a is visited after q simulations and $T_s^q := \sum_a T_\Gamma^q(s,a)$. The first term in UCB criterion encourages exploiting the known promising routes, while the second term encourages exploration of the unknown and less visited routes. The balance between the two terms is controlled by constant $c \in [0, 1]$.

In the multi-objective MCTS proposed by Chen [28], for each node s, the Pareto Upper Confidence Bound (Pareto UCB) is used to build its approximate Pareto optimal set that is defined over the possible child nodes reachable from the set of molecules in s, $P(C_s)$, where $C_s := \{\Gamma(s,a)\}_{\forall a \in A(s)}$. $P(C_s)$ is constructed by first evaluating the Pareto-UCB for each node in C_s as shown in equation below, and then compute the set of non-dominated nodes according to the definition of dominance in 2.3 by using Pareto-UCB as the evaluation criteria.

$$Pareto - UCB\,(\Gamma(s,a)) = \frac{1}{q} \sum_{i=1}^{q} \vec{r}_{i,\Gamma(s,a)} + c \frac{\sqrt{2log(T_s^q)}}{T^q_\Gamma(s,a)}, \quad \forall \Gamma(s,a) \in C_s$$

$$\hat{\pi}^{multi}(s) : s \sim U(P(C_s))$$

Given the approximate Pareto optimal set $P(C_s)$, the selection problem at each node s is then reduced to sampling one of the nodes from $P(C_s)$ uniformly at random.

3.2. Expansion

As described previously, the selection stage only provides guidance for navigating the explored sections of the search tree, once we reach a node s to which no further child nodes are attached, an expansion policy is executed to add new child nodes to the tree. In the current paper, both single- and multi-objective setting follow the same strategy where a policy network $f_\theta(m)$ is trained to produce a categorical distribution over the possible reactions for a given molecule m. Consequently, the action for each m is sampled as follows:

$$a \mid m \sim Categorical(f_1(m,\theta), \ldots, f_K(m,\theta)), \quad \forall m \in s$$

$$P(a_k \mid m) = f_k(m,\theta), \quad \forall k = 1, 2..., K$$

where $f_k(m,\theta)$ is the output of the soft-max operator of the policy network for reaction template k and K is the total number of reaction templates. For each $m \in s$, the top N reactions with the highest $P(a_k \mid m)$ are executed to produce the corresponding reactant precursors. The probability of the top N actions are re-normalized such that $\sum_{k=1}^{N} f_k(m,\theta) = 1$.

3.3. Roll-out

After the new child nodes are added to the search tree, a stochastic simulation is initiated, traversing the tree from the leaf node to a terminal node. Since the nodes to be visited during the simulation has not accumulated any reward statistics, actions are chosen according to a default policy, where the output from $f_\theta(m)$ and its vectorized form is used in place of $r_{i,\Gamma(s,a)}$ and $\vec{r}_{i,\Gamma(s,a)}$ as follows:

$$\pi(s)^{\text{single}}_{\text{default}} \in \underset{a \in A(s)}{\arg\max}\, P(a \mid m)$$

$$Pareto - UCB_{\text{default}}(\Gamma(s,a)) = \vec{p}_a, \quad \forall \Gamma(s,a) \in C_s,$$

$$\vec{p}_a \in \mathbb{R}^d = \begin{pmatrix} P(a|m) & P(a \mid m) & \ldots & P(a|m) \end{pmatrix}^{\mathrm{T}}$$

3.4. Backup

When the roll-out episode is complete, the r_i and $\vec{r}_i$ are backpropagated through the path visited during selection. For each of the visited node s, in the single-objective case the back-up operation is simply an accumulated sum over the r_i seen so far at s, i.e $\sum_i r_{i,s}$, for the multi-objective case, it becomes $\sum_i \vec{r}_{i,s}$. The approximate local pareto-front $P(C_s)$ is then recomputed based on these updated values. In addition to the reward, also the visiting counts are updated among the nodes.

Table 1
The objectives used in this study.

Symbol	Functional form	Description
O_{stock}	$N_{\text{in stock}}/N_{\text{total}}$	The fraction of starting material in the stock
O_{steps}	$\frac{6-N_{\text{reactions}}}{5}$	The total number of reactions, scaled
O_{SC}	$\frac{\min(SC_{\text{intermediate}}-SC_{\text{starting material}})+1.5}{5.5}$	The difference in SCScore between starting material and preceding intermediate
O_{ref}	$1-\frac{1}{1+\exp(-0.5\text{TED}+5)}$	The tree edit distance (TED) between the route and the reference route

4. Method

4.1. Datasets

We selected the set-n1 from PaRoutes as our benchmarking set [15]. This consists of 10,000 target compounds with associated reference routes extracted from the US patent and trademark office (USPTO) reaction set. The stock collection used as stopping criteria for the search was a public dataset from eMolecules consisting of 25M compounds, downloaded in January of 2023.

4.2. Objectives

We considered four objectives or scoring functions for the routes in this study, and they are summarized in Table 1. All scoring functions are scaled such that they should be maximized and ranges between 0 and 1. To score the starting material, we used the fraction of starting material that can be found in the eMolecules stock (O_{stock}). To score the length of the route, we use the total number of reactions scaled so that the score is maximum at one reaction, and minimum at six reactions (O_{steps}). Objectives similar to these is what usually is used in the single-objective reward function employed in Monte Carlo tree search (MCTS) [14] because they are the simplest representation of the goal of synthesis planning, i.e. finding a plan to commercial material with as few steps as possible. We developed two additional objectives: the first one is based on the minimum difference between the SC-Score [32] of a starting material and the intermediate in the synthesis plan preceding the starting material with three steps (O_{SC}). This is to enforce a reasonable decrease in synthetic complexity between the intermediates and the starting materials, and we scale this to have a minimum at a difference of −1.5 and maximum of 4.0 (the SCScore goes between 0 and 4). The second additional objective is based on the route distance between a route and the PaRoutes reference route (O_{ref}), calculated with the fast LSTM model previously published [33]. This enforces the generation of routes similar to the reference route, and we scale this with a sigmoid-like function (see Table 1). It should be noted that except for the O_{stock}, all objective functions can in principle be normalized with different functions, herein we have chosen scaling functions that are reasonable based on our experience — but we have made no effort in optimizing them. We will leave that work for future studies.

4.3. Retrosynthesis experiments

For each target compound in the PaRoutes set, we carried out a number of single- and multi-objective experiments with AiZynthFinder [29,30]. In all experiments, we employed a template-based retrosynthesis model and a filter model trained on the USPTO dataset as detailed previously [34,35]. The same template-based model was used for both expansion and rollout. We set the maximum depth of the search to six, let the MCTS algorithm perform 100 iterations (one iteration consisting of selection, expansion, roll-out and backpropagation), and extracted between 10 and 30 routes. For the single-objective experiments, we employed the same scoring function as used in the MCTS objective and selected first ten routes, and then filled up to a maximum 30 routes if they had the same score as the previously selected routes. For the multi-objective experiments, we extracted first the routes on the Pareto front using the same objectives as in the MCTS. If they were less than 10, we continued adding routes at lower Pareto ranks until we had extracted at least 10 but no more than 30.

4.4. Evaluation

The success rate of the retrosynthesis experiments was measured as the fraction of targets for which we find a synthesis plan to commercial material. The packing number of the extracted routes was measured as the fraction of the routes that has no overlapping reactions [36]. This metric was suggested as a metric of diversity of the routes [36], but is rather restrictive as it only considered routes that are completely non-overlapping. The volume of the route space was measured by first calculating the latent space encoding of the routes for a particular target using the LSTM model trained to estimate route distances [33], and then reducing the latent space with principal component analysis. The cubic volume spanned by the first three principal components were taken as the volume of the route space as they explain most of the variance. This is also a diversity metric, but takes all routes into account, even if they just show a small variation, e.g. a change in step order. We also extracted the average number of molecules and reactions in the routes, and computed the percentage of convergent (branched) routes. To measure the size of the space enclosed by the routes in objective space, we utilize the concept of hypervolume [37]. For experiments with a single objective, we turned the solutions into a multi objective-problem and construct a Pareto front of the objectives that are linearly combined into the single objective. For MO-MCTS experiments, where one objective was a combination of two objectives, we also create such a Pareto front in three-dimensions. The hypervolume of the routes for a particular target were computed by the pygmo library [38] using a constructed or true Pareto front in two or three dimensions and a reference point at origin. The hypervolume is naturally a much coarse grained featurization of route space than the route space volume, because routes that are rather chemically diverse could be evaluated to the same values of the objectives.

5. Results and discussion

We performed eight different retrosynthesis experiments on the PaRoutes targets with either single- or multi-objective MCTS and different combinations of the objectives in Table 1. In Table 2, we collect general statistics from the experiments. The setup of the experiment, i.e. the employed rewards and if we use single- or multi-objective MCTS, impacted neither success rate nor search time in any noticeable way; in all setups, we can find synthesis plans to commercial material for 77%–80% of the targets in about half a minute. The difference in success rate is likely not significant; in our experience the uncertainty of these experiments is around 2%. Although there are differences between related experiments of up to 10 s, this is not considered to be a difference that holds any practical significance. The diversity, i.e. the fraction of routes with no overlapping reactions is somewhat affected by the setup. In general, we can increase the packing, i.e. fraction of non-overlapping routes, by adding the O_{ref} objective, showing that forcing the search to explore an additional objective might increase the diversity in this regard. Furthermore, the packing is not correlated with the route space explored by the different setups, highlighting that different diversity metrics can indicate different salient features of the solutions. In general, we explore a larger route space with the MO-MCTS experiments, and it appears that going from two to three objectives in the search also leads to an increase. It is harder to interpret how O_{SC} and O_{ref} is affecting the route space exploration, although

Table 2
Statistics from the retrosynthesis experiments.

Algorithm	Objective(s)	Success rate	Mean search time	Mean packing	Mean route space volume	Mean # mols.	Mean # reactions	% of convergence
SO-MCTS	$0.5O_{stock} + 0.5O_{step}$	0.79	25.2	0.44	295.1	6.2	3.0	11.1
MO-MCTS	$\{O_{stock}, O_{step}\}$	0.80	36.6	0.43	436.6	6.4	3.1	12.1
SO-MCTS	$1/3O_{stock} + 1/3O_{step} + 1/3O_{SC}$	0.78	24.5	0.47	269.8	6.3	3.0	12.6
MO-MCTS	$\{0.5O_{stock} + 0.5O_{step}, O_{SC}\}$	0.79	32.3	0.44	609.0	7.2	3.5	17.5
MO-MCTS	$\{O_{stock}, O_{step}, O_{SC}\}$	0.80	27.1	0.43	838.1	7.6	3.8	20.0
SO-MCTS	$1/3O_{stock} + 1/3O_{step} + 1/3O_{ref}$	0.77	29.9	0.50	149.0	5.8	2.8	8.4
MO-MCTS	$\{0.5O_{stock} + 0.5O_{step}, O_{ref}\}$	0.78	34.5	0.48	228.5	6.2	3.0	9.8
MO-MCTS	$\{O_{stock}, O_{step}, O_{ref}\}$	0.80	32.5	0.49	372.4	6.4	3.2	11.8

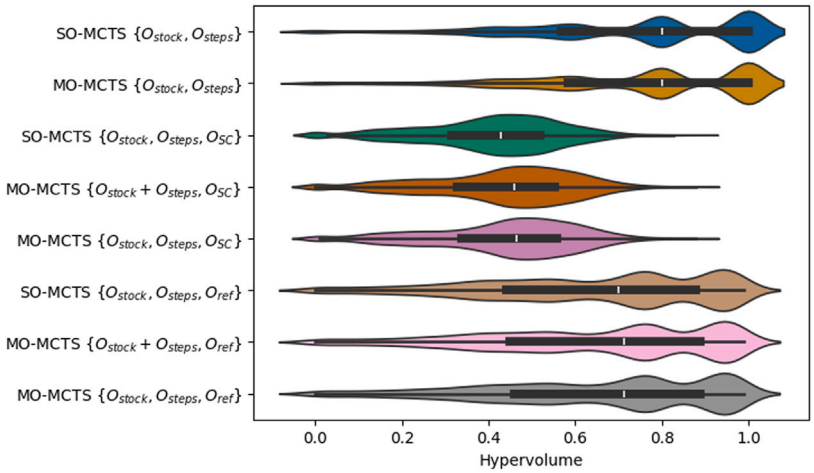

Fig. 1. Hypervolume distributions for the retrosynthesis experiments.

it appears to adding O_{SC} leads to a greater increase in the explored route space than adding O_{ref}. The average number of molecules and reactions in the routes are increased with MO-MCTS experiments, compared to SO-MCTS corroborating the effect of large route space exploration. The MO-MCTS experiment also results in more convergent routes on average, and this is especially true for the experiments including O_{SC}. When comparing the MO-MCTS with O_{SC} it appears that using three objectives leads to increase molecules, reactions and convergent routes compared to using two objectives — in line with the increased route space. The same observations can be seen when comparing the MO-MCTS experiments with O_{ref}.

The distributions of hypervolume of the extracted routes for all experiments are shown in Fig. 1. It should be noted that we are analysing hypervolumes for objectives that technically might not have been explored in the retrosynthesis search, because we linearly combine them. However, it is still of interest to perform some comparisons. Using the two basic objectives O_{stock} and O_{steps}, we do not observe any difference between the single- and multi-objectives algorithms. There are small individual differences for the different target compounds; for 12% and 4% of the targets, the hypervolume is larger with MO-MCTS and SO-MCTS, respectively. This is expected, because these objectives can take discrete and limited number of values. The advantage of the MO-MCTS setup is rather that one can combine objectives without having to construct an arbitrary weighting. We constructed two such objectives and included them in both single- and multi-objective MCTS to investigate if we can bias the search towards the objectives, increased difference in synthetic complexity and similarity to a reference route, respectively. Using the O_{SC}, we observe an increase in the hypervolume explored with the MO-MCTS using three objectives compared to SO-MCTS on average; for 76% and 8% of the targets, the hypervolume explored is larger with MO-MCTS and SO-MCTS, respectively. However, comparing the two MO-MCTS experiments with O_{SC}, it appears advantageous to use three objectives, as for 52% of the targets, the hypervolume explored is larger when using three objectives instead of two. The relative increase in hypervolume is also reflected in an increase in route space volume, for instance for 93% of the targets, we explore a larger route space volume when using MO-MCTS with three objectives than in the SO-MCTS. Therefore, we will continue with the analysis of the MO-MCTS experiments using three objectives. The shift in hypervolume and route space volume between SO-MCTS and MO-MCTS can likely be attributed to a shift in the distribution of ΔSC as can be seen in the top row of Fig. 2. For the MO-MCTS experiment including O_{SC} we see a shift towards higher ΔSC compared to both the SO-MCTS experiment the MO-MCTS without O_{SC}.

Next, if we include the O_{ref} we see a much smaller difference between SO-MCTS and MO-MCTS on average; for 50% and 12% the hypervolume explored is larger with MO-MCTS and SO-MCTS, respectively. Furthermore, the distribution of the tree edit distance (TED) is wider in the MO-MCTS as can be seen in the bottom row of Fig. 2. For both SO-MCTS and MO-MCTS, we can see a shift towards smaller TED values than in the MO-MCTS experiment without O_{ref}. The difference between the two MO-MCTS experiments with O_{ref}, we see that for 30% of the targets, the hyperspace volume explored is larger when using three objectives than when using two. This is also corroborated with that for 53% of the targets the route space volume is increased when using three objectives instead of two. Therefore, we will continue the analysis with the MO-MCTS using three objectives.

5.1. Case study 1 - US20130137689A1

We provide one example where the MO-MCTS outperforms SO-MCTS when using two objectives, O_{step} and O_{stock} to illustrate how we can analyse the experiments. This is a target from the US20130137689A1 patent. In 3A we plot the extracted routes in the two-dimensional space of the two objectives to illustrate the hypervolume (area) spanned by these solutions. Notice that many routes have the same values for the two objectives, and thus the number of points in the plot is typically less than the number of extracted routes.

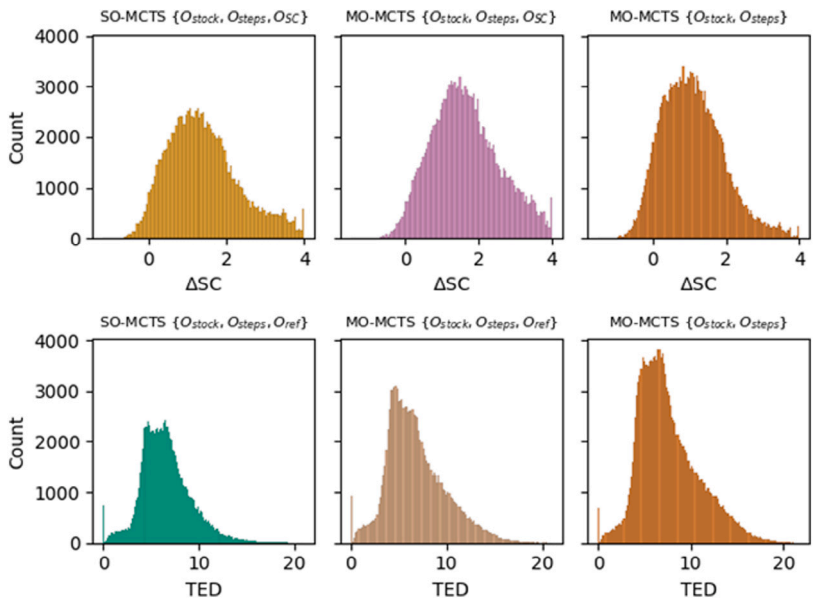

Fig. 2. Difference in SCscore (ΔSC) and tree edit distance (TED) in selected retrosynthesis experiments.

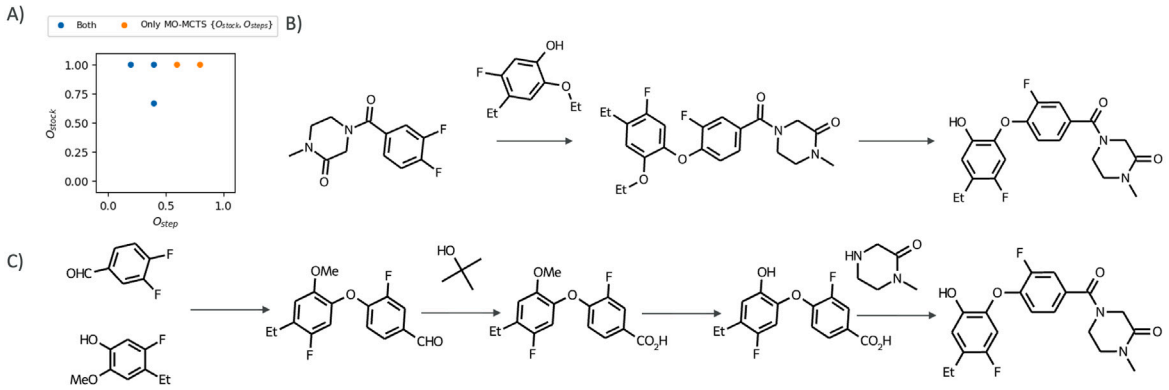

Fig. 3. Case study 1 (A) The objective space for SO-MCTS and MO-MCTS when using two objectives. "Both" in the legend, indicate that the particular solution was found by both SO-MCTS and MO-MCTS. (B) A top-ranked route from the MO-MCTS experiment (C) A top-ranked route from the SO-MCTS experiment.

In this specific example, the SO-MCTS experiment did not yield any solutions beyond those already identified by MO-MCTS, and there are three solutions found by both experiments. Two of them are on the first Pareto front and corresponds to routes where all starting material can be found in stock ($O_{stock} = 1$), but with different lengths. The third solution found by both experiments is on the second Pareto front, and was extracted because we chose to extract a minimum number of routes. This route has some starting material that is not in stock ($O_{stock} < 1$). Finally, for this example, the MO-MCTS produced two routes that was not found by the SO-MCTS. These two routes have all the starting material in stock, and are shorter in length than the other routes. Thus this example shows that MO-MCTS leads to a greater exploration of the two objectives with a hypervolume of 0.8, compared to 0.4 for SO-MCTS. Incidentally, this also lead MO-MCTS to produce a route that was shorter (see 3B) than the route found by SO-MCTS (see 3C). However, it should be noticed that this is not a guarantee and this is just one example out of all the targets used to benchmark the algorithm. For this example, we also find that the MO-MCTS also produced a more diverse set of routes with a route volume of 350.5 compared to 191.4 for SO-MCTS.

5.2. Case study 2 – US20140162984A1

To exemplify the effect of the O_{SC} objective, we investigated a compound from the US20140162984A1 patent, a quinazoline derivative. In our experience, when the search struggles to make key disconnections in the target compound is resorts to repeating making simpler transformations like protections or functional group interconversions until the maximum depth is reached. The O_{SC} objective was created to address this issue, by forcing the MCTS to break down the target compound more aggressively. We plot the objectives in the extracted routes from the MO-MCTS experiments with or without O_{SC} in Fig. 4A. As we now have three objectives, we show the objectives in three 2D-plots because we find it clearer than a 3D plot. We can see that for this particular target compound, the MO-MCTS excluding O_{SC} does not produce any routes that are also not found when including O_{SC} in the search. We can also see that for all combination of objectives, the run with O_{SC} span a wider range of both O_{step} and O_{stock}. The computed hypervolume is 0.71 when including O_{SC} as an objective in the search. We can construct an Pareto front with three dimensions for the MO-MCTS experiments without O_{SC}, and the hypervolume enclosed by that front is 0.29. This analysis clearly shows that including this objective explicitly in the search, we obtain an algorithm that explores the objectives more widely. In 4B, we instead compare SO-MCTS and MO-MCTS when including O_{SC}. From these plots, we observe that both SO-MCTS and MO-MCTS generate routes not identified by the other method. However, MO-MCTS discovers more routes missed by SO-MCTS than vice versa. Notably, the highest scoring routes are found by both methods. This is also reflected in the hypervolume of

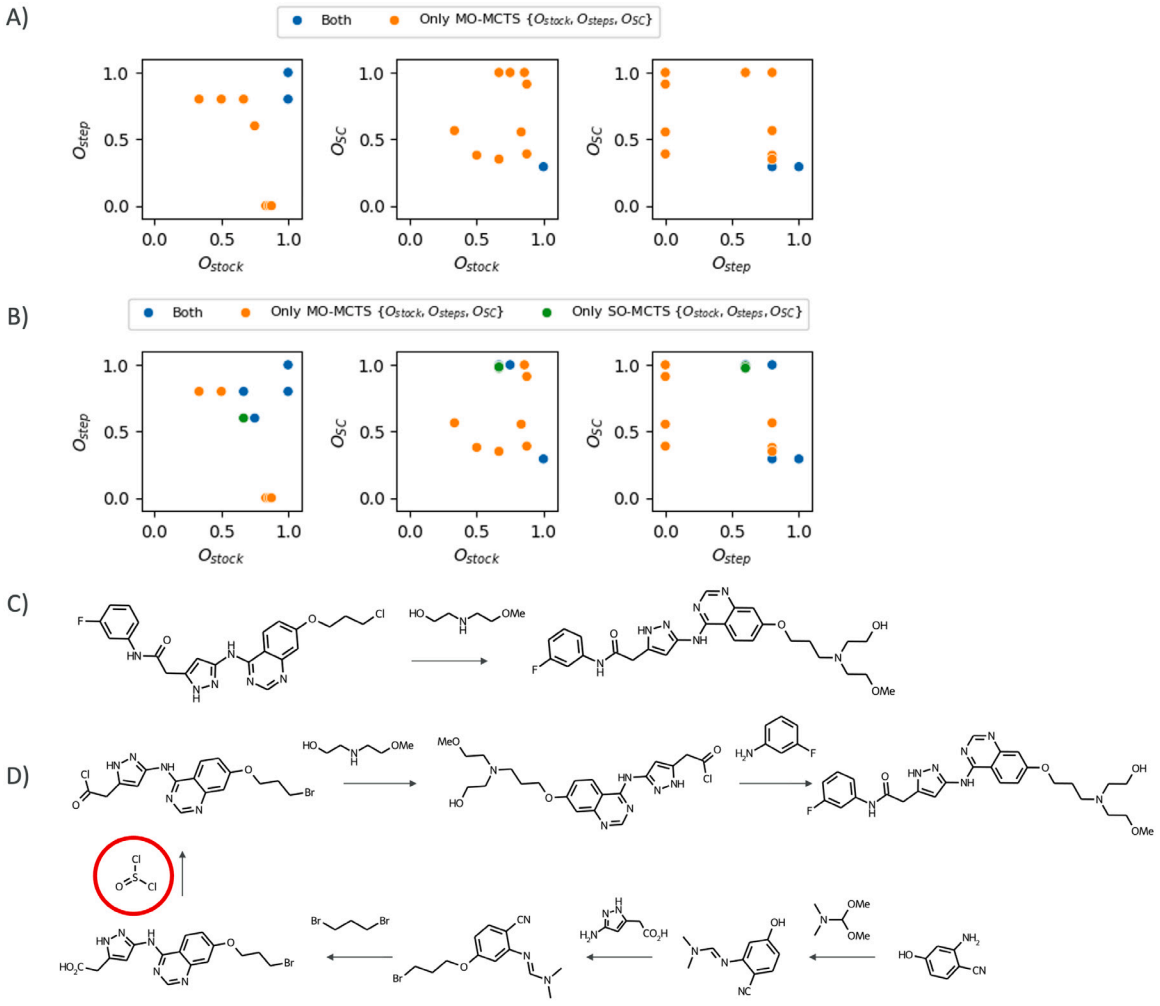

Fig. 4. Case study 2 (A) The objective space for MO-MCTS with or without the O_{SC} objective. "Both" in the legend indicates that the solution was found in both experiments that are being compared. (B) The objective space for MO-MCTS and SO-MCTS when including the O_{SC} objective. "Both" in the legend indicates that the solution was found in both experiments that are being compared. (C) A top-ranked route from the MO-MCTS run excluding O_{SC} (D) The route from the MO-MCTS run including O_{SC} that has the largest ΔSC for this target. A starting material not in stock is encircled in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

SO-MCTS, which is 0.71 and identical to MO-MCTS. However, it is also clear that within the hypervolume of the objectives, MO-MCTS produces qualitative more diverse solutions. This is reflected in the route volume that is 1349.3 for MO-MCTS, but only 34.4 for SO-MCTS and 6.8 for MO-MCTS when O_{SC} is excluded. Furthermore, for this target, we see a significant effect of O_{SC} on the produced routes. A top-ranked route when O_{SC} is not included is shown in Fig. 4C, and shows that the search identifies a near-analogue to the target in the stock, and simply disconnect a small substituent. This means that there is very little difference in SCScore between the target and the starting material. Contrary, when O_{SC} is included, we see that the search forces further disconnections (see Fig. 4B), and the result is a rather lengthy, convergent route. Unfortunately, a small starting material is used that is not found in the stock – although it is reasonable to believe that this agent can be replaced with something that is available.

5.3. Case study 3 – US20150336908A

To exemplify the effect of the O_{ref} objective, we investigated a compound from the US20150336908A patent, a tetrazolinone derivative. We plot the objectives in the extracted routes from the MO-MCTS experiments with or without Oref in Fig. 5A. O_{ref} is consistently lower in the MO-MCTS run that did excluded this objective, and we can also see that including this objective is essential for finding a plan to commercial starting material. The hypervolume when including O_{ref} is 0.73, and the corresponding volume for the experiment without O_{ref} is 0.23. The wider exploration of the objectives are also shown in the route space volume that is 344.9 in the experiments with O_{ref} and only 66.9 in the experiments without this objective. In 5B we compare MO-MCTS and SO-MCTS when including O_{ref}. As with the case study 2, we see that both MO-MCTS and SO-MCTS produce solutions not found by the other method. However, the solutions with the highest values of the objectives are often found by both and the hypervolume with SO-MCTS is identical to MO-MCTS. The route volume if 153.0 for the SO-MCTS, also showing the MO-MCTS produce a more diverse set of routes. The experimental reference route shown in Fig. 5C has two starting material that cannot be found in the stock used in the experiments. The MO-MCTS experiment excluding O_{ref} follows a similar disconnection strategy to the experimental route see (Fig. 5D) but has to introduce an additional step that leads to more starting material not in the stock. However, MO-MCTS with O_{ref} produces a route that is similar to the reference route (Fig. 5E), sharing the final step, although the retrosynthesis leads to starting material that is in stock.

6. Conclusion

We introduced a multi-objective Monte Carlo Tree Search (MCTS) algorithm to solve the retrosynthesis task. The algorithm, which is

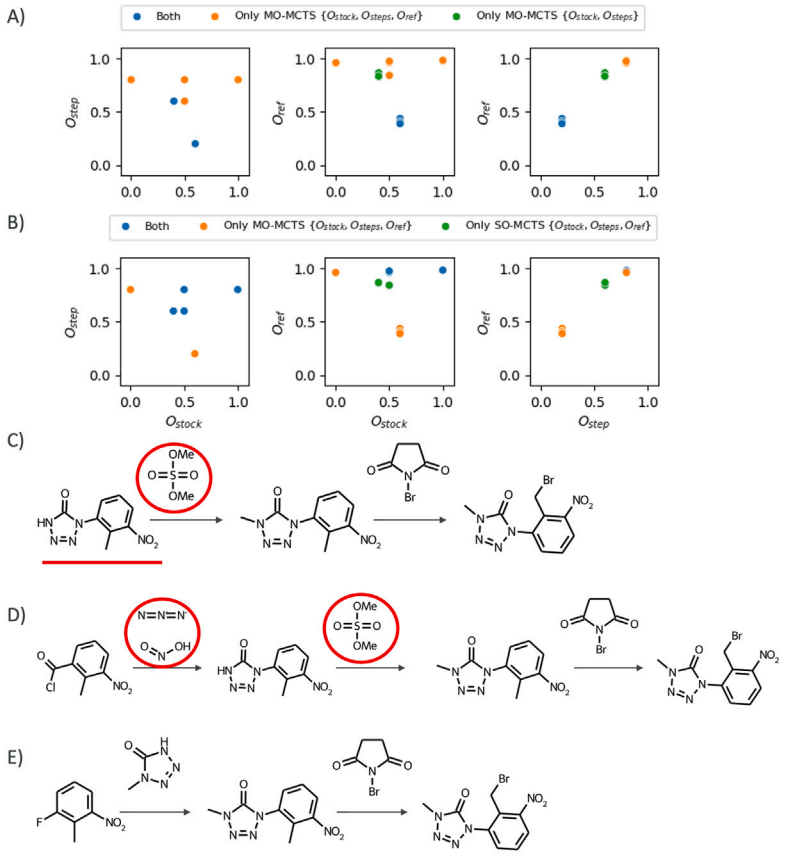

Fig. 5. Case study 3 (A) The objective space for MO-MCTS with or without the O_{ref} objective. Both" in the legend indicates that the solution was found in both experiments that are being compared. (B) The objective space for MO-MCTS and SO-MCTS when including the O_{ref} objective. "Both" in the legend indicates that the solution was found in both experiments that are being compared. (C) The reference route for this compound, The starting material encircled or underscored with red is not in stock. (D) The route from the MO-MCTS without O_{ref} that has the lowest TED. (E) The route from the MO-MCTS with O_{ref} that has the lowest TED.

an adaption of [28] modifies the selection and backpropagation steps of the MCTS algorithm to keep track of the local Pareto-fronts of the child nodes. A multi-objective search algorithm allows one to readily combine different objectives into the search without having to consider the scale of the objectives and how to weight them. To benchmark this algorithm we used four objectives in a total of eight retrosynthesis experiments on a dataset comprising of 10,000 targets from the PaRoutes set. We can conclude that the multi-objective search performs comparable to the standard, single-objective search when employing simple finite-range objectives based on the starting material and the number of steps. However, for more complex objectives based on synthetic complexity and route similarity, we can conclude that the multi-objective search is preferable, because it provides more diverse set of solutions (in terms of route space) that show a higher degree of exploration of the desired objective than the single-objective search. For most of the targets, this also means that the objective is closer to optimality. Finally, we can conclude that the careful design of objectives for retrosynthesis continues and depends greatly on the application. However, the framework developed here allows the user to seamlessly incorporate their specific objectives in the synthesis planning process.

CRediT authorship contribution statement

Helen Lai: Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Christos Kannas:** Writing – original draft, Validation, Investigation. **Alan Kai Hassen:** Writing – review & editing, Conceptualization. **Emma Granqvist:** Methodology. **Annie M. Westerlund:** Writing – review & editing, Methodology. **Djork-Arné Clevert:** Supervision. **Mike Preuss:** Writing – review & editing, Supervision. **Samuel Genheden:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Alan Kai Hassen was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Innovative Training Network European Industrial Doctorate grant agreement No. 956832 "Advanced machine learning for Innovative Drug Discovery". Emma Rydholm was partially supported by the Wallenberg Artificial Intelligence, Autonomous Systems, and Software Program (WASP), funded by the Knut and Alice Wallenberg Foundation, Sweden.

During the preparation of this work, the author(s) used ChatGPT to identify alternative phrasing for smoother transitions between paragraphs. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Data availability

Data will be made available on request.

References

[1] Corey EJ, Wipke WT. Computer-assisted design of complex organic syntheses. Science 1969;166(3902):178–92.

[2] Coley CW, Green WH, Jensen KF. Machine learning in computer-aided synthesis planning. Acc Chem Res 2018;51(5):1281–9, PMID: 29715002.

[3] Schwaller P, Vaucher AC, Laplaza R, Bunne C, Krause A, Corminboeuf C, et al. Machine intelligence for chemical reaction space. WIREs Comput Mol Sci 2022;12(5):e1604.

[4] Zhong Z, Song J, Feng Z, Liu T, Jia L, Yao S, et al. Recent advances in deep learning for retrosynthesis. WIREs Comput Mol Sci 2024;14(1):e1694.

[5] Hart PE, Nilsson NJ, Raphael B. A formal basis for the heuristic determination of minimum cost paths. IEEE Trans Syst Sci Cybern 1968;4:100–7.

[6] Chen B, Li C, Dai H, Song L. Retro*: Learning retrosynthetic planning with neural guided A* search. In: The 37th international conference on machine learning. 2020.

[7] Świechowski M, Godlewski K, Sawicki B, Ma'ndziuk J. Monte Carlo tree search: a review of recent modifications and applications. Artif Intell Rev 2021;56:2497–562.

[8] Segler MHS, Preuss M, Waller MP. Planning chemical syntheses with deep neural networks and symbolic AI. Nature 2017;555:604–10.

[9] Kishimoto A, Buesser B, Chen B, Botea A. Depth-first proof-number search with heuristic edge cost and application to chemical synthesis planning. In: Neural information processing systems. 2019.

[10] Shibukawa R, Ishida S, Yoshizoe K, Wasa K, Takasu K, Okuno Y, et al. CompRet: a comprehensive recommendation framework for chemical synthesis planning with algorithmic enumeration. J Cheminformatics 2020;12.

[11] Schwaller P, Petraglia R, Zullo V, Nair VH, Häuselmann R, Pisoni R, et al. Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. Chem Sci 2020;11:3316–25.

[12] Xie S, Yan R, Han P, Xia Y, Wu L, Guo C, et al. RetroGraph: Retrosynthetic planning with graph search. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2022.

[13] Franz C, Mogk G, Mrziglod T, Schewior K. Completeness and diversity in depth-first proof-number search with applications to retrosynthesis. In: International joint conference on artificial intelligence. 2022.

[14] Thakkar A, Kogej T, Reymond J-L, Engkvist O, Bjerrum EJ. Datasets and their influence on the development of computer assisted synthesis planning tools in the pharmaceutical domain. Chem Sci 2020;11(1):154–68.

[15] Genheden S, Bjerrum E. PaRoutes: towards a framework for benchmarking retrosynthesis route predictions. Digit Discov 2022;1(4):527–39.

[16] Wang X, Qian Y, Gao H, Coley CW, Mo Y, Barzilay R, et al. Towards efficient discovery of green synthetic pathways with Monte Carlo tree search and reinforcement learning. Chem Sci 2020;11:10959–72.

[17] Lin K, Xu Y, Pei J, Lai L. Automatic retrosynthetic route planning using template-free models. Chem Sci 2020;11:3355–64.

[18] Yu Y, Wei Y, Kuang K, Huang Z, Yao H, Wu F. GRASP: Navigating retrosynthetic planning with goal-driven policy. In: Neural information processing systems. 2022.

[19] Ishida S, Terayama K, Kojima R, Takasu K, Okuno Y. AI-driven synthetic route design incorporated with retrosynthesis knowledge. J Chem Inf Model 2022;62:1357–67.

[20] Tripp A, Maziarz K, Lewis S, Segler MHS, Hern'andez-Lobato JM. Retro-fallback: retrosynthetic planning in an uncertain world. 2023, ArXiv, abs/2310.09270.

[21] Liu G, Xue D, Xie S, Xia Y, Tripp A, Maziarz K, et al. Retrosynthetic planning with dual value networks. In: International conference on machine learning. 2023.

[22] Kreutter D, Reymond J-L. Multistep retrosynthesis combining a disconnection aware triple transformer loop with a route penalty score guided tree search. Chem Sci 2023;14:9959–69.

[23] Roucairol M, Cazenave T. Comparing search algorithms on the retrosynthesis problem. Mol Informatics 2024;e202300259.

[24] Emmerich MTM, Deutz AH. A tutorial on multiobjective optimization: fundamentals and evolutionary methods. Nat Comput 2018;17(3):585–609.

[25] Hayes CF, Reymond M, Roijers DM, Howley E, Mannion P. Distributional monte carlo tree search for risk-aware and multi-objective reinforcement learning. In: Proceedings of the 20th international conference on autonomous agents and multiagent systems. 2021, p. 1530–2.

[26] Wang W, Sebag M. Multi-objective monte-carlo tree search. In: Asian conference on machine learning. PMLR; 2012, p. 507–22.

[27] Perez D, Mostaghim S, Samothrakis S, Lucas SM. Multiobjective monte carlo tree search for real-time games. IEEE Trans Comput Intell AI Games 2014;7(4):347–60.

[28] Chen W, Liu L. Pareto Monte Carlo tree search for multi-objective informative planning. In: Proceedings of robotics: science and systems. 2019.

[29] Genheden S, Thakkar A, Chadimová V, Reymond J-L, Engkvist O, Bjerrum E. AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning. J Cheminformatics 2020;12(1):70.

[30] Saigiridharan L, Hassen AK, Lai H, Torren-Peraire P, Engkvist O, Genheden S. AiZynthFinder 4.0: developments based on learnings from 3 years of industrial application. J Cheminformatics 2024;16(57).

[31] Auer P. Using confidence bounds for exploitation-exploration trade-offs. J Mach Learn Res 2002;3(Nov):397–422.

[32] Coley CW, Rogers L, Green WH, Jensen KF. SCScore: Synthetic complexity learned from a reaction corpus. J Chem Inf Model 2018;58(2):252–61.

[33] Genheden S, Engkvist O, Bjerrum EJ. Fast prediction of distances between synthetic routes with deep learning. Mach Learning: Sci Technol 2021;3.

[34] Genheden S, Engkvist O, Bjerrum E. A quick policy to filter reactions based on feasibility in AI-guided retrosynthetic planning. ChemRxiv 2020;1–19.

[35] Genheden S, Norrby PO, Engkvist O. AiZynthTrain: Robust, Reproducible, and Extensible Pipelines for Training Synthesis Prediction Models. J Chem Inf Model 2022.

[36] Maziarz K, Tripp A, Liu G, Stanley M, Xie S, Gainski P, et al. Re-evaluating retrosynthesis algorithms with syntheseus. Faraday Discuss 2024.

[37] Zitzler E, Thiele L. Multiobjective optimization using evolutionary algorithms — A comparative case study. In: Eiben AE, Bäck T, Schoenauer M, Schwefel H-P, editors. Parallel problem solving from nature — PPSN v. Berlin, Heidelberg: Springer Berlin Heidelberg; 1998, p. 292–301.

[38] Biscani F, Izzo D. A parallel global multiobjective framework for optimization: pagmo. J Open Source Softw 2020;5(53):2338.