

パレート最適化 MCTS を用いた *de novo* 分子生成手法の開発鈴木 敬将<sup>†</sup>安尾 信明<sup>‡</sup>東京工業大学情報理工学院<sup>†</sup>東京工業大学物質・情報卓越教育院<sup>‡</sup>関嶋 政和<sup>§</sup>東京工業大学情報理工学院<sup>§</sup>

## 1 背景

最近では創薬研究に 13.5 年以上、26 億ドルのコストがかかっている [1]。コストを低減するため、近年では計算機を用いた創薬手法の研究が進められている。機械学習による分子生成モデルの研究が盛んになっている。多くの生成手法は ChemTS[2] など一つの評価関数、あるいは ChemTSv2[3] などの重み付き和や重み付き積によって複数の評価指標を一つの評価関数に集約して用いている。実際の創薬では複数の評価指標を同時に最適化して化合物を探索するため、一つの評価指標では不足である。また、重み付きの総和や総乗は重みを設定する問題が残されており、設定も属人的である。本研究ではパレート最適化 MCTS を用いた分子生成手法の開発が目的である。

## 2 手法

本研究では ChemTS を拡張し、パレート最適化 MCTS を用いたモンテカルロ木探索 [4] による SMILES ベースの *de novo* 分子生成手法を提案する。モンテカルロ木探索は 4 ステップに

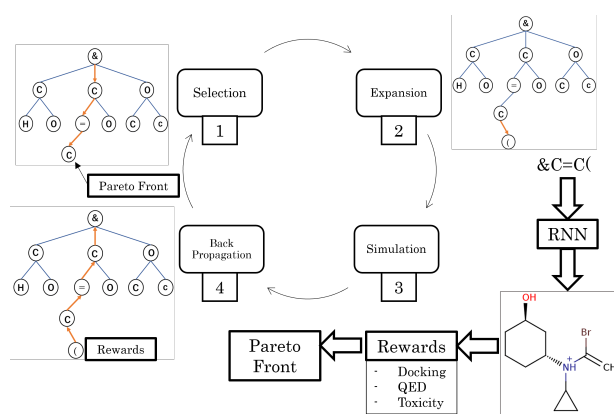


図1 提案手法概要

分かれている。探索木のノードは SMILES 文法の記号 1 つに対応している。図 1 は提案手法の概要を表す。

1. 選択：探索木から葉ノードをパレートフロントから一つ選ぶ。
2. 展開：選択した葉ノードに一つノードを追加する。
3. 評価：選択したノードを順にたどることで化合物を SMILES で表現したときの先頭の文字列を得られる。この部分文字列を RNN に入力することで化合物を得る。生成された化合物の複数の評価指標で評価し、報酬の値を決定する。またパレートフロントを計算する。
4. 反映：報酬の値を葉ノードから根まで経路上の各ノードに足し込む。

評価関数は任意の関数を設定できるが、本実験では AutoDock Vina のドッキングスコア、

Development of *de novo* molecular generation method with Pareto-Optimizing MCTS

<sup>†</sup> School of Computing, Tokyo Institute of Technology

<sup>‡</sup> Tokyo Tech Academy for Convergence of Materials and Informatics (TAC-MI), Tokyo Institute of Technology

<sup>§</sup> School of Computing, Tokyo Institute of Technology

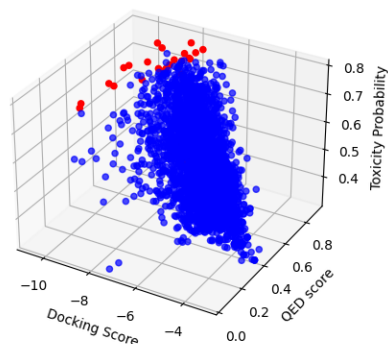


図2 DDR1 キナーゼでの探索空間

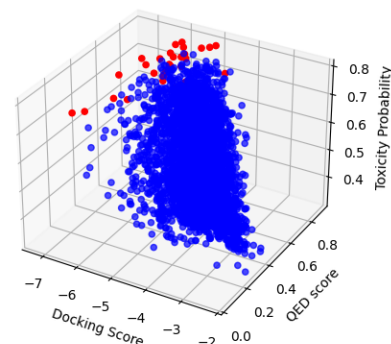


図3 SARS-CoV-2 メインプロテアーゼでの探索空間

QED, 毒性の3指標とした。

### 3 実験

実装はPythonとTensorflowを用いた。RNNの学習はChemTS[2]と同じデータセット、同じモデルを用いた。CPUにIntel Xeon E5-2680 V4を2機とGPUにTesla P100を4機使用した計算機で分子生成実験を14日間行った。標的タンパク質はDDR1キナーゼ(PDBID:3ZOS)とSARS-CoV-2メインプロテアーゼ(PDBID:6LU7)とした。

### 4 結果

図2は標的タンパク質をDDR1キナーゼに、図3はSARS-CoV-2メインプロテアーゼに設定して実験したときに生成された化合物の散布図である。赤色の点はパレートフロントに属する化合物、青色の点はそうでない化合物に対応し、一つの点は一つの化合物に対応している。

### 5 結論

本研究によりパレート最適化MCTSを用いた分子生成手法を提案した。従来では重み付けでないと多目的最適な分子生成が行えなかったが、本研究により重み付けなしで分子を生成できた。従来の生成手法に比べて性質の良い分子

を効率よく生成することに成功した。

### 参考文献

- [1] Steven M Paul, Daniel S Mytelka, Christopher T Dunwiddie, Charles C Persinger, Bernard H Munos, Stacy R Lindborg, and Aaron L Schacht. How to improve r&d productivity: the pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery*, 9:203–214, 2010.
- [2] Xiufeng Yang, Jinzhe Zhang, Kazuki Yoshizoe, Kei Terayama, and Koji Tsuda. Chemts: an efficient python library for de novo molecular generation. *Science and Technology of Advanced Materials*, 18:972–976, 12 2017.
- [3] Tatsuya Yoshizawa, Shoichi Ishida, Tomohiro Sato, Masateru Ohta, Teruki Honma, and Kei Terayama. Selective inhibitor design for kinase homologs using multiobjective monte carlo tree search. *Journal of Chemical Information and Modeling*, 0, 11 2022.
- [4] Weijia Wang, Michèle Sebag, Steven C H Hoi, and Wray Buntine. Multi-objective monte-carlo tree search. volume 25, pages 507–522. PMLR, 11 2012.