

機械学習を使って薬になる分子構造を探すには

部分グラフマイニングと強化学習を使った分子構造生成

○山田 正嗣, 杉山 磨人
(国立情報学研究所)

どんな研究？

薬となる分子構造を探索・設計することは非常に難しいです。

機械学習を組み合わせることで効率よく分子設計する方法を提案しています。

何がわかる？

分子の「**部分構造**」に着目することで、解釈性のある分子構造生成が行えます。

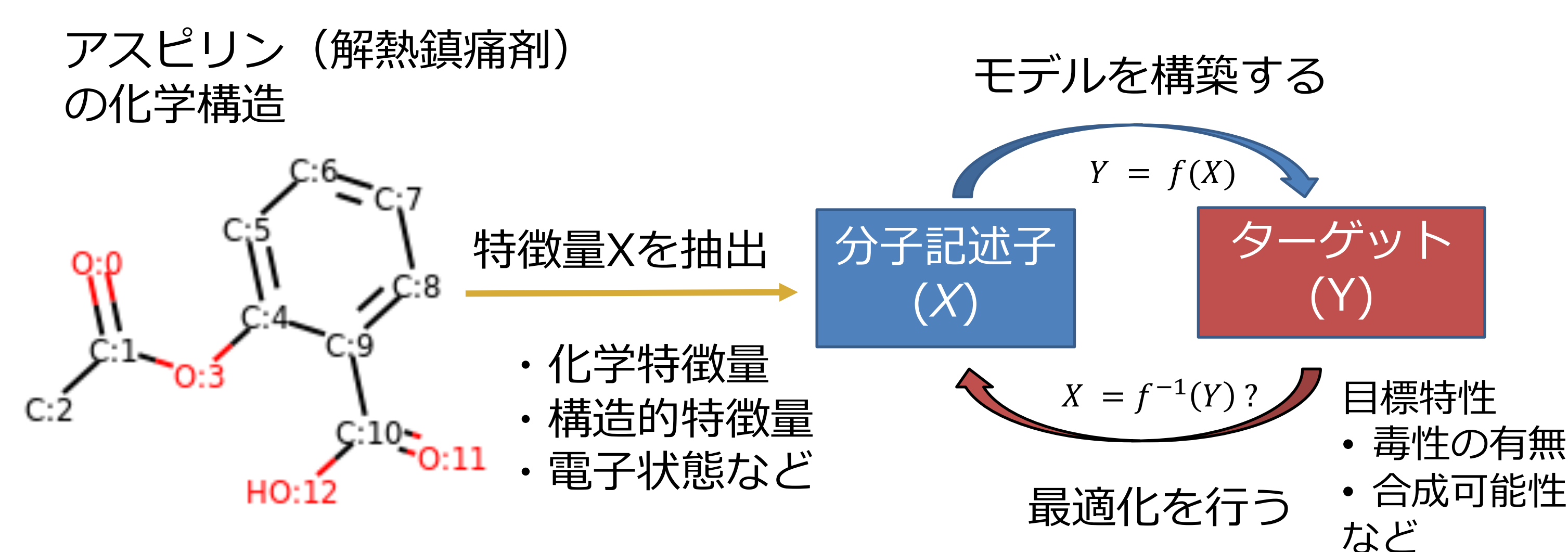
「**強化学習**」と組み合わせることで最適な分子の探索・設計ができます。

背景・目的

分子構造生成の難しさ

- **組合せ爆発**：薬になる分子の組合わせ数は、 10^{60} にもなると言われています。最適な分子を探すことは非常に難しいです。
- **分子構造の表現**：どのような分子構造を持つと、薬もしくは毒になってしまうのでしょうか。これらを予測することも創薬における課題の1つです。

分子構造探索の流れ



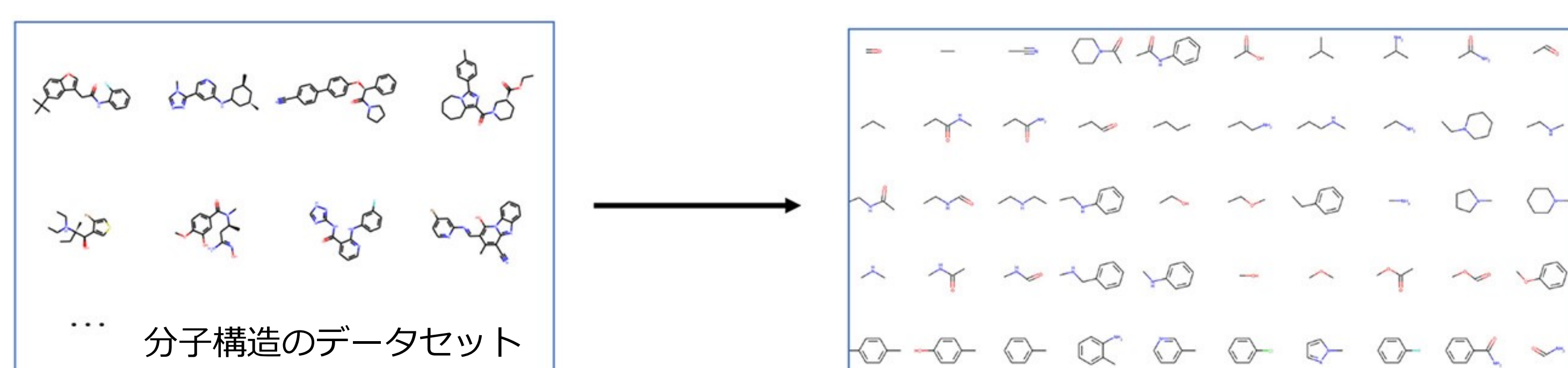
研究内容（方法・結果・結論）

方法

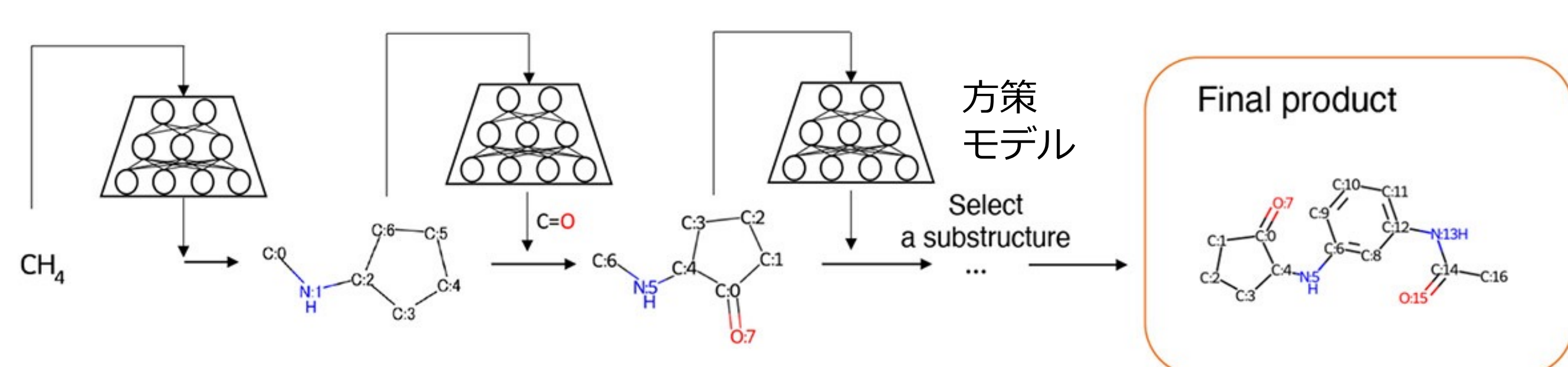
部分構造に着目した分子構造生成手法MOLDRを提案しました。これらは2つのプロセスから成り立っています。

- **分解**：分子のデータベースから共通する部分構造を抽出し、これを分子構造の部品とします。
- **再構築**：強化学習を利用し、目的の特性を見やすく、最適な部分構造を選択していきながら、最終的なターゲットの分子を生成していきます。

1. グラフマイニングによる分解

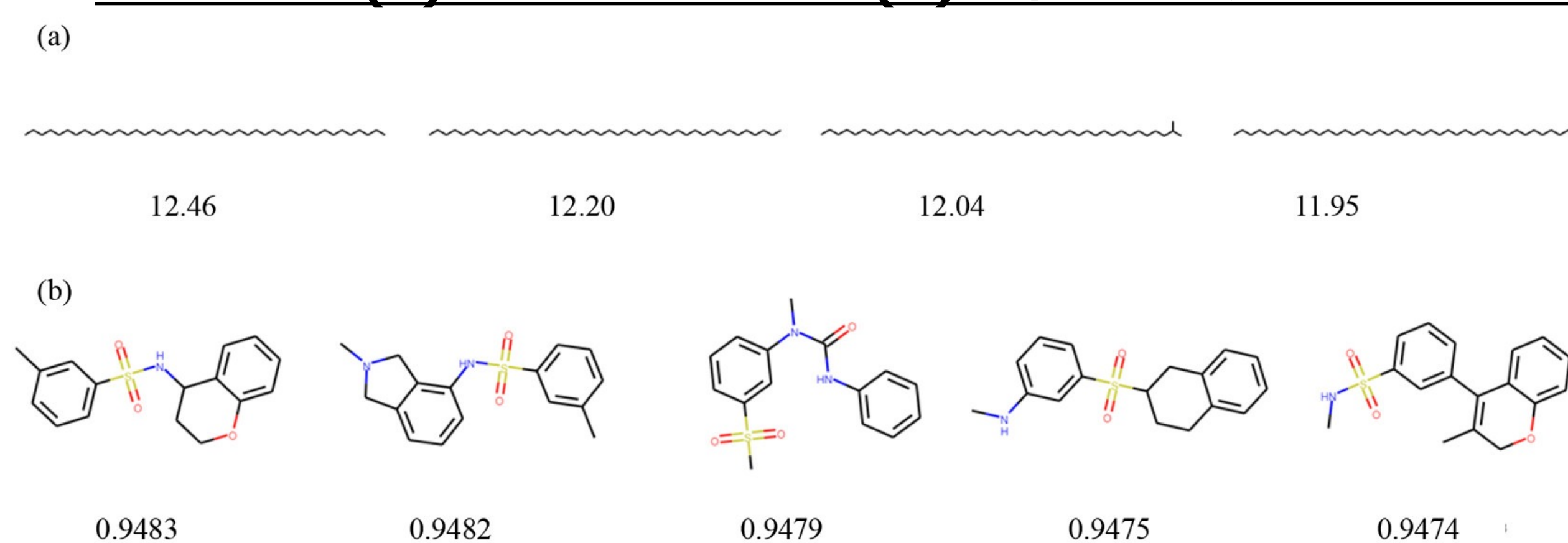


2. 強化学習を用いた再構築



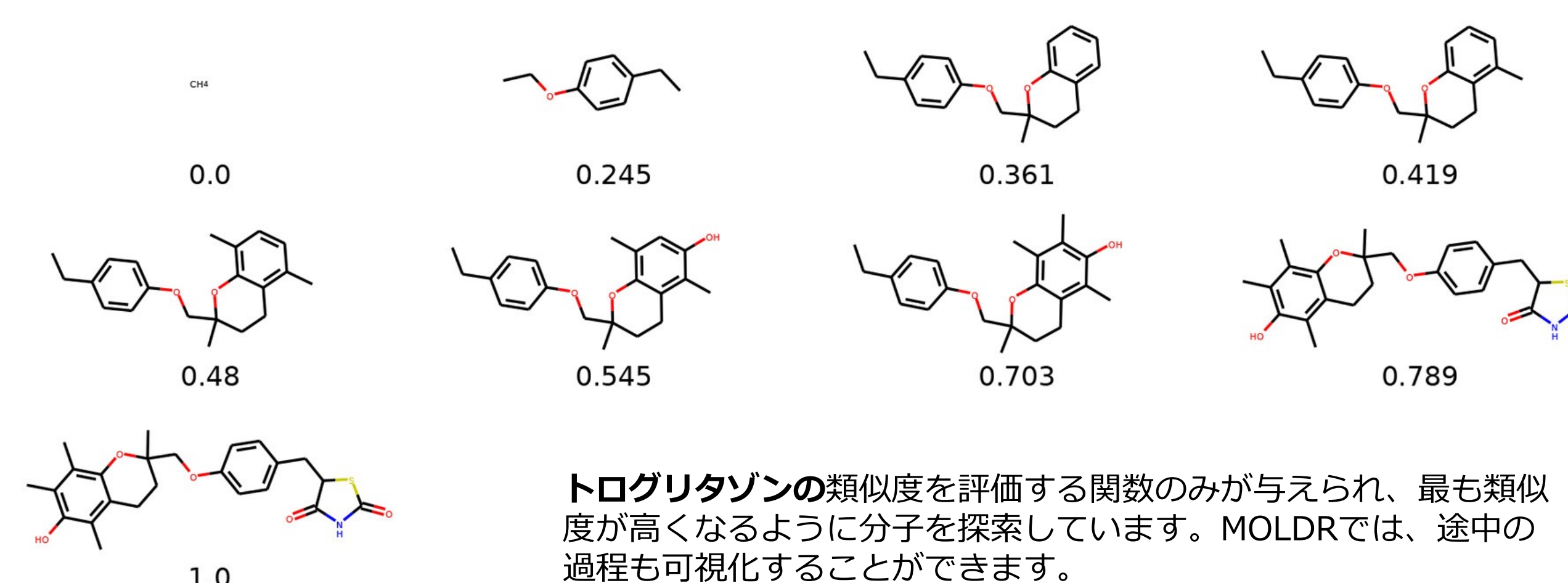
結果

疎水性指標(a)と薬らしさの指標(b)を最適化した時の分子構造



(a) 疎水性指標のため、最適化後、炭素(C)を含む構造のみが生成されるようになりました。(b)では、薬らしさの指標のため、窒素(N), 硫黄(S), 酸素(O)も含む構造も生成されています。

トログリタゾン（チアゾリジン系抗糖尿病薬）の再発見



結論

「グラフマイニング」と「強化学習」を組合わせた分子生成アルゴリズムMOLDRを提案しました。

特定の指標を最大化させた分子を生成するだけでなく「薬剤の再発見」に対しても途中結果を可視化しつつ生成が行えることを示しました。