

創薬支援推進事業・産学連携による次世代創薬 AI 開発
中間評価報告書

研究課題名	最先端の AI 技術を用いたマルチターゲット予測と構造発生を組み合わせた包括的な創薬 AI プラットフォームの開発	
機関名	国立研究開発法人理化学研究所	
研究担当者名	所属 役職	生命機能科学研究センター 制御分子設計研究チーム チームリーダー
	氏名	本間 光貴

1. 研究概要

創薬の現場で活用できるインターフェースを備えた包括的に創薬の過程で必要となる評価項目を予測しながら有望な新規化合物を提案できる創薬 AI プラットフォームを構築し、日本のアカデミア及び製薬企業に提供して、創薬プロセスを飛躍的に効率化する。創薬 AI プラットフォームは、新規化合物提案 AI、化合物プロファイル予測 AI、オミクス情報を含む多階層データを考慮した創薬 AI から構成され、ユーザーインターフェースによって創薬研究者が創薬の早期から後期まで包括的に設計支援を受けられるものとする。

2. 研究成果

新規化合物提案 AI については、広大なケミカル空間の中で創薬の様々な目的に対応できる構造発生 AI として、ChemTS、RLS 法、XGG 法の 3 種類を実装した。それぞれ新規性、医薬品らしさ、医薬品の骨格変換を重視した手法である。モンテカルロ木探索による強化学習を特徴とした ChemTS については、10 種類の項目の同時最適化を行う能力があることを検証した。予測 AI について半教師あり及び自己教師あり学習による精度改善手法を開発した。医薬品らしさ、合成難易度、特許取得のための新規性の評価手法を開発した。創薬 AI プラットフォームについては、構造発生 AI、化合物プロファイル予測 AI を登録するシステム及びユーザーインターフェースを開発した。アカデミア創薬の 4 個のテーマで創薬 AI プラットフォームを利用した創薬支援を開始した。

化合物プロファイル予測 AI については、オン・オフターゲットと ADMET の項目について公共データベースの整備を行い、予測 AI を構築した。GCN 記述子によるマルチタスク及びマルチモーダル学習によって、平均で回帰モデルの R^2 が 0.767、2 値分類モデルで 100 nM、1 μ M、10 μ M の閾値での ROC-AUC がそれぞれ 0.946、0.935、0.931 を示した。オン・オフターゲットについては、連携に参加した製薬企業から 400 種類以上の 1000 万データポイント以上のアッセイデータの提供を受けプロトタイプの予測 AI を構築したところ、予測精度の向上が見られた。Federated Learning について、PyTorch によって開発した予測手法が良好な結果を示し、クラウドのファイル共有サービスを中継して、モデルパラメータを受け渡す自動型のシステムの構築を行い、製薬企業への試験運用を準備している。

オミクス情報を含む多階層データを考慮した創薬 AI については、化合物応答遺伝子発現プロファイルなどオミクス情報から化合物の潜在的な標的タンパク質を予測する手法を開発した。欠損値や未観測値が多いため、テンソル分解を適用してデータ補完を行う手法を開発した。遺伝子発現プロファイルなどのオミクス情報とケミカル情報を融合し、変分オートエンコーダや類似度検索などを組み合わせて構造発生 AI を開発した。また、新規な構造発生アルゴリズムとして、casVAE を開発し、合成反応を考慮した構造発生ができるようになった。ケミカル空間を拡充するために、三次元ビシクロ化合物を中心とした新奇性の高い化合物の合成法を開発した。