

先端科学技術研究科 修士論文要旨

所属研究室 (主指導教員)	計算システムズ生物学 (金谷 重彦 (教授))		
学籍番号	2111262	提出日	令和 5年 1月 20日
学生氏名	村田 友真		
論文題目	SARS-CoV-2のメインプロテアーゼをターゲットとした深層学習を用いたリガンド- プロテイン相互作用の回帰モデルの構築		
要旨			
<p>現在、COVID-19(新型コロナウイルス感染症)の原因ウイルスであるSARS-CoV-2の感染拡大を抑制できる薬の開発が一刻も早く望まれている。SARS-CoV-2の抑制のためには、感染から増殖に至るプロセスの中で必要とされているタンパク質をターゲットとして、それに結合する化合物を投与することでその機能を阻害することが有力な手法とされている。ターゲットとなるタンパク質の候補はいくつかあるが、本研究ではSARS-CoV-2の必須タンパク質の合成に関与するメインプロテアーゼ (Mpro) をターゲットとして薬剤候補化合物を探すことを考える。既存の実験データから Mpro との結合強度を予測する深層学習モデルを構築し、転移学習を用いる。天然化合物のような多様性の高い分子をモデル化するためには、適切な分子記述子や特徴量を選択する必要がある。そのため、深層学習モデルとして、分子構造から分子グラフ畳み込みニューラルネットワーク (GCNN) を用いて特徴量の選択を最適化する回帰モデルを構築した。GCNN は画像分野で大きな成功を収めた畳み込みニューラルネットワーク (CNN) を、グラフを入力とした場合でも扱えるように設計された手法である。GCNN では、CNN と同様に畳み込み処理によって各ノードだけでなく周囲のノードの特徴も考慮することができるため、分子構造の場合、化合物に含まれるそれぞれの原子の特徴を考慮した結果を出力できる。</p> <p>このモデルを用いてまず大規模データベースを用いて Mpro との結合強度 (pAC50) を分類するモデルを学習し、分子構造からの特徴抽出を学習した。さらに、より精度の高いデータで転移学習を行うため、Mpro との結合強度 (%Inh) を回帰するモデルとして学習させた。最終的には、天然物データベース KNApSAcK をもとに、57,765件の天然化合物から候補分子をスクリーニングし、有意に結合強度が高いと予想される分子を86個検出した。</p>			