

【目的】 時間と費用を削減して効率的に創薬を進めるため、薬の化学構造と生理活性の関係性を見出す研究への期待は大きく、近年は特に人工知能を始めとした計算科学手法が注目されている。昨年末から世界中で脅威をふるっている新型コロナウイルスは治療薬の早急な開発が求められており、承認までの早さから既存薬の転用に向けた検討が行われているが、順調とは言えない。問題として「患者数が少なく、治療薬の十分な効果と安全性を確認できない」ことがあるため、医薬品候補を探す計算科学手法の重要性が改めて認識されてきている。薬の性質はタンパク質との立体的な相互作用が大きく関わっているが、機械学習手法に使用される記述子は、ほとんどが2次元の情報である。そこで、従来の構造記述子と3次元の位相的な距離を特徴量として使用し、Mproの阻害性を予測する手法の検討を行った。

【方法】 COVID19のメインプロテアーゼであるMproの阻害性試験のデータを収集し、50%阻害濃度の実験値が存在する659化合物のデータを分析対象とした。部分構造を表す構造特徴量として、4つのFingerprintを算出し、Mproの阻害予測に最も適した特徴量を選択した。また、構造特徴量だけでは見つけることができない情報を取得するため、Mproの立体構造を取得してドッキングシミュレーションを行い、化合物の3次元結合座標を取得した。発生させた化合物の3次元構造から位相的な特徴量を算出するPersistent Homologyを実装した。選択した構造特徴量と位相的特徴量を合わせて、Random Forestによって判別モデルを構築した。

【結果】 構造特徴量の検討によりMACCSとMorganが適した構造特徴量であると決定し、位相的特徴量と合わせてRandom Forestによる判別モデルを構築したところ、多くの特徴量で80%を超える正解率を出した。著しい精度の違いはなかったが、Morganが同じであるにも関わらず、阻害性が大きく異なる2つの化合物を比べたところ、Persistent Homologyによって立体的な特徴量を得ることができたことが確認できた。タンパク質との立体的な結合を表すためには、従来使用していた構造特徴量だけでなく、位相的特徴量も組み合わせて検討を行うことが望ましいことが示唆された。

位相的距離を用いた化合物の立体構造の特徴量化による重要構造の探索

