

先端科学技術研究科 博士論文要旨

所属研究室 (主指導教員)	データ駆動型化学研究室 (浦岡行治教授)		
学籍番号	2221023	提出	令和6年12月17日
氏名	前田樹		
題目	リガンドベース手法を用いたスキュフォールド・ホップ化合物同定手法の開発		

低分子医薬品開発の初期段階において、標的マクロ分子に対して活性を示す化合物を同定するため、機械学習等を用いて化合物ライブラリを探索する仮想スクリーニング (VS) が行われている。リガンドベース手法では訓練データに含まれる化合物との構造類似性に基づいて生物学的活性の有無を判断するため、既知の活性化合物と構造の異なる化合物の検出は困難である。一方、実用的には活性を維持しながら既知の活性化合物と骨格の異なる化合物の同定が求められる。このような化合物をスキュフォールド・ホップ (SH) 化合物[1]と呼ぶ。SH 化合物の同定は、薬物動態特性の向上や合成可能性の改善、特許侵害の回避につながることから、創薬化学の中心的な課題である。本研究では、SH 化合物を検出できる VS モデルの開発を目的とした。

第1章では、既知の活性化合物 (リガンド) の情報に基づく VS (LBVS) による SH 化合物探索の現状を紹介する。また、既存研究の問題点を示しながら、本論文の目的を述べる。

第2章では、リガンドベース手法における SH 化合物同定に適した分子表現 (記述子) を評価した。LBVS において、SH 化合物が標的マクロ分子との分子間相互作用の点で既知活性化合物と類似していることに基づく様々な記述子が開発されてきたが、これらの手法間の比較検証は不十分である。本研究では、いくつかの記述子の SH 化合物同定能力を、後方視的検証と前方視的検証の両方で公平に評価した。2つのスクリーニングアルゴリズム (類似性検索 (SS)、support vector machine (SVM)) と4つの記述子 (ECFP4 [2]、CATS [1]、WHALES [3]、ROCS [4, 5]) の組み合わせを、4つの標的マクロ分子に対する制御されたデータセットを用いて公平に評価した。全体として、SVM-ECFP4 と SVM-ROCS が高い性能を示した (図1)。さらに、SVM-ROCS で上位に順位付けされた化合物は訓練活性化合物と異なる構造を有していた。前方視的検証では、後方視的検証で最も高精度であった SVM-ROCS を用いて VS を行った。VS により提案された93個のSH候補化合物について相互作用解析を行ったところ、5つのヒット化合物を得たが、いずれも特異的な結合を確認できなかった。

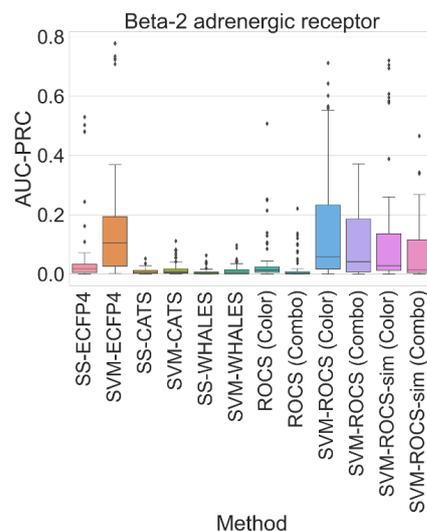


図1 スクリーニングアルゴリズムと記述子間の VS 性能の比較[6]

第3章では、活性予測のための分子表現提案手法を検討した。第2章の検討で、高い予測精度を示す記述子は標的マクロ分子によって異なることが明らかになった。そこで、少数かつ類似した訓練データから適切な記述子を構築する枠組みを考案した。提案手法は重み付き組み合わせカーネル関数で、これは複数の記述子を扱う柔軟な方法である。骨格に依存しない表現を抽出するため、ハイパーパラメータは類縁体ベースの交差検証 (CV) で最適化した。類縁体分子群として構造活性相関行列 (SARM) を使用し leave-one-SARM-out (LOSO) CV を行った。15 個の多様な標的マクロ分子について、4 つの記述子を用いて提案手法の効果を検証した。スクリーニングにおいて、組み合わせ記述子はより良い性能を示す単一の記述子に匹敵する性能を示した (図2)。2 つの記述子の性能の差と重みの関係を調査したところ、通常の無作為な分割を用いるとほとんどの重みが 0 に最適化される状況で、LOSO は 0 以外の重みを選択した (図3)。

第4章では、スクリーニング失敗のリスクを最小化するためのスクリーニング手法を提案した。第2章の前方視的検証では非特異的な相互作用を示す化合物は発見できなかった。そこで、VS において少なくとも一つの活性化合物を提案する手法を開発した。本手法は、最も活性である可能性が高い化合物の予測と、その化合物を不活性であると仮定してのモデルの更新を繰り返す (図4)。

各モデルで予測された上位化合物を提案化合物セットとすることで、多様な構造を持つ化合物の獲得と、活性化合物を一つも発見できないリスクの回避を目指す。二次元フィンガープリントのスクリーニングのために用意された activity class [7] に対して、SVM と ECFP4 を用いてモデルを構築した。一般的な VS 手法 (図4の一段目のみに相当) で 100 位以内に活性化合物を検出できなかったデータセットに対して、1-2 回のモデルの更新で 10 位以内に活性化合物を検出できた。

最後に第5章では、本研究で提案した SH 化合物同定手法をまとめ、今後の展望を述べる。

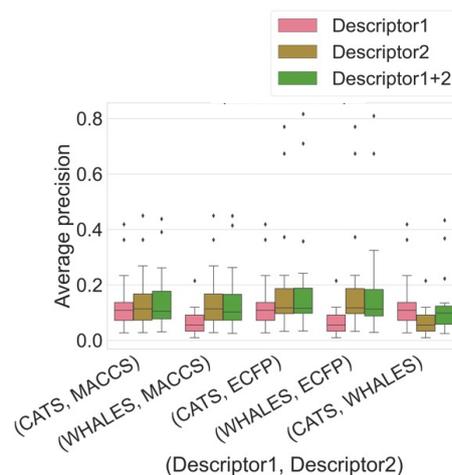


図2 組み合わせ記述子の性能

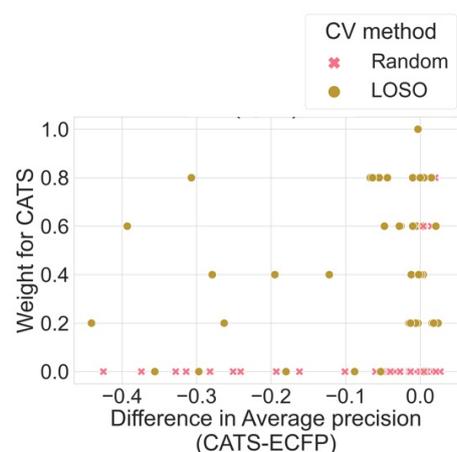


図3 性能と重みの関係

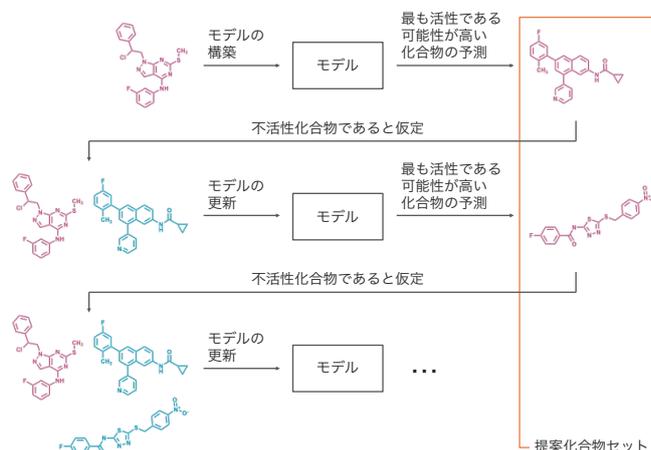


図4 提案スクリーニング手法の流れ

参考文献 : [1] G. Schneider et al. *Angew. Chem. Int. Ed.* **38**, 2894–2896 (1999). [2] D. Rogers & M. Hahn, *J. Chem. Inf. Model.* **50**, 742–754 (2010). [3] F. Grisoni et al. *Commun. Chem.* **1**, 1–9 (2018). [4] P. C. D. Hawkins et al. *J. Med. Chem.* **50**, 74–82 (2007). [5] T. Sato et al. *J. Chem. Inf. Model.* **52**, 1015–1026 (2012). [6] I. Maeda et al. *J. Chem. Inf. Model.* **64**, 5557–5569 (2024). [7] K. Heikamp & J. Bajorath, *J. Chem. Inf. Model.* **51**, 1831–1839 (2011).

(論文審査結果の要旨)

低分子創薬において、コンピュータを利用した化合物仮想スクリーニングが行われる。本論文は、スクリーニング手法として、標的マクロ分子の情報を利用せず、低分子化合物情報のみを利用する「リガンドベース」手法に焦点を当てる。「リガンドベース」のスクリーニングでは、分子構造がコンピュータ上で数値表現され、活性化合物を識別する機械学習モデルが構築され、利用される。この分野で重要な課題の一つとして、「既存の活性化合物から分子骨格が異なる新規活性化合物を同定すること」が挙げられる。この課題を解決するために、様々な分子表現が提案されてきたが、それらの表現がもつ情報や性能をシステムティックに評価した研究はなく、分子表現の優劣は定まっていなかった。また、化合物スクリーニングを実施する際には、「少なくとも一つの活性化合物を提案する」ことが、プロジェクトの前進に必要であるが、そのための手法は存在しなかった。これら2つの課題に本論文は取り組む。論文の主要な結果は以下の通りである。

1. 「既存の活性化合物から分子骨格が異なる新規活性化合物」を同定するための様々な分子表現の特徴を明らかにした。頻繁に利用される **extended connectivity fingerprint** を用いた場合のスクリーニングでは「活性化合物における部分構造の組み合わせによる新規骨格を有する化合物」を選択するケースが多く、分子形状とファーマコフォア情報に基づく表現は、「部分構造に依存しない共通する特徴を持つ化合物」を選択する特徴が複数のデータセットによる検証から示された。これらの表現の違いは、距離相関プロットにより明示された。加えて、提案手法の「分子形状とファーマコフォア」情報に基づいた表現を利用して仮想スクリーニングを実施し、一次スクリーニング結果としては、分子骨格の異なる活性化合物を同定することに成功した。

2. 「少なくとも一つの活性化合物を提案する」ための新規スクリーニング手法を考案した。提案手法は、機械学習モデルにより活性化合物とランクづけされた上位化合物を不活性化合物と仮定し、モデルを再構築し、再構築したモデルにより再度スクリーニングを実施、という過程を繰り返す方法である。この手法は確率に基づく理論的な裏付けから提案されている。レトロスペクティブな検証において、既往手法である機械学習モデルによるランキングに基づいた化合物選択と比較して、本手法は優位に高い確率で活性化合物を選択できていることが示された。

以上、本論文は、「リガンドベース」の仮想スクリーニング手法として分子骨格の異なる活性化合物を選択する手法開発と、新しいスクリーニングの評価関数(基準)を提案する。これらの結果は、インシリコ低分子創薬に新しい知見を与えるものであり、学術的にも大きな意義がある。よって、審査委員一同は本論文が博士(工学)の学位論文として価値あるものと認めた。