

京都大学大学院 薬学研究科 准教授 奥野 恭史

[研究の目的]

医薬品開発には、医薬品と生体との作用点となるタンパク質（遺伝子）と化合物の相互作用を解明することが不可欠であり、本研究では、タンパク質と化合物の相互作用予測により医薬品候補化合物を探索する創薬支援システムの開発を行う。具体的には、遺伝子・タンパク質の配列と化学物質の化学構造を入力した際に、そのタンパク質と化学物質との相互作用（結合、薬理活性など）の有無を判別することを目指している。具体的には、遺伝子・タンパク質の配列と化学物質の化学構造を入力データとして、それらをプロファイル表現化するステップと、作成したプロファイルを機械学習し予測システムを構築するステップと、相互作用予測結果を実験により検証するステップの3ステップを計画している。

[研究の内容、成果]

1. サポートベクターマシンを用いた相互作用予測モデルの開発

活性化合物を効率よく迅速に探し出すために、計算機を用いた候補化合物の絞り込み手法「バーチャルスクリーニング (VS)」が開発されてきた。現在よく用いられているVSとして、既知リガンドとの構造類似性に基づく「Ligand-based virtual screening (LBVS)」と標的タンパク質の立体構造に基づく「Structure-based virtual screening (SBVS)」がある。この二つの手法は、近年の情報技

術の進歩と相まって、この10-20年で著しい発展を遂げ、ゆるぎない地位を確立した。しかしながら現在、VSのヒット確率は1% - 10%もあればよしとされており（例えば、LBVSでは、既知活性化合物の骨格構造に強く影響される嫌いがあるし、SBVSでは、パラメータの恣意性、予測的中率の低さなどが指摘されている）、さらなる技術的な改良や革新的技術の開発が切望されていることは間違いない。

そこで本研究では、LBVSやSBVSとは概念の異なる第三のVSとして、「ケミカルゲノミクスに基づくVS手法 (Chemical Genomics-based virtual screening: CGBVS)」に着手した。「ケミカルゲノミクス」は、興味を持つ化合物が生物に与える影響についてゲノム規模で研究する学問であり、マイクロレイやハイスループットスクリーニングなどの同時大量解析技術の革新に後押しされ、近年、注目を浴び始めている。それに伴い、化合物と遺伝子の関連性について、膨大な実験データが蓄積されている。そこで我々は、情報科学技術の一つであるパターン認識技術を用いて、タンパク質と化合物との結合情報（ケミカルゲノミクス情報）から抽出したタンパク質のリガンド認識パターンに基づいて活性化合物を効率的に発見することを新手法「CGBVS」のコンセプトとした。

タンパク質とリガンドとの相互作用パターンの認識とそのリガンド予測を開発するために、我々は学習アルゴリズムの一種であるサポートベクターマシン (SVM) を用いた。SVMは、2クラス分類器

の一種であり、与えられた2つのグループに属する特徴ベクトルを最大マージンで分離するような超平面を構築する。ここで、最大マージンとは、分離した超平面から各サンプル間までの最短距離を指す。我々は、このSVMを用いて、化合物-タンパク質相互作用の有/無を判別する手法を開発した。その手法の流れを図1に示す。まず、収集した相互作用をベクトルとして表現するために、各化合物の化学構造、各タンパク質のアミノ酸配列について、様々な属性（記述子と呼ぶ）を計算する。次に、正例（相互作用する化合物-タンパク質ペア）および負例（相互作用しない化合物-タンパク質ペア）に対応する記述子をそれぞれ組み合わせて特徴ベクトルを構成し、SVMを用いて学習モデルを構築する。このモデルが得られると、（未知の化合物-タンパク質ペアに相当する）新しいベクトルが相互作用 有/無 のどちらのクラスに属するか、を予測することができる。

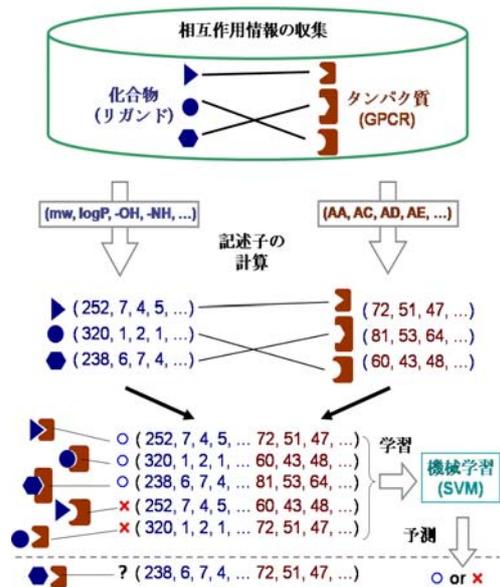


図1 CGBVSの流れ

2. 交差検定法による予測モデルの評価

既存のVS手法との比較検討を行うため、今回開発したCGBVSとLBVSとの予測性能を比較した。収集した化合物-GPCR 相互作用を用いて、負例を交換しながら5分割交差検定法 (5 fold cross-validation) を試行した。交差検定の結果、最近傍法を用いたLBVSでは84.4±0.3%、CGBVSでは91.6±0.2%の相互作用を正しく予測した。また、ROC曲線からも、CGBVSの予測性能の高さが確認された(図2)。したがって、ケミカルゲノミクス情報の活用がリガンド予測性能の向上につながったといえる。なお、予測に要する計算的負荷が一般的なLBVSによるものとほとんど変わらないことから、CGBVSは大規模スクリーニングに十分適用可能と考えられる。

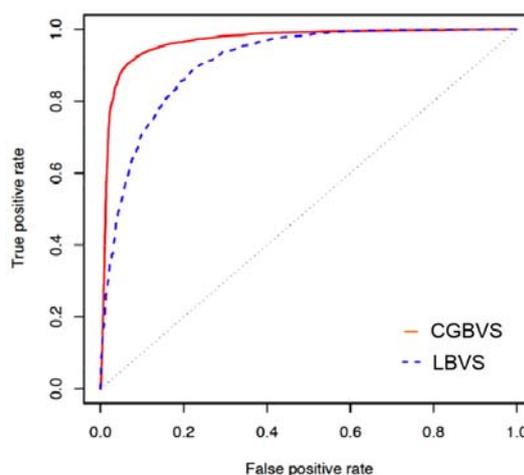


図2 ROC曲線による相互作用予測性能の比較
赤い実線がCGBVS、青い破線が最近傍法を用いたLBVSを示す。

3. ヒトβ2アドレナリン受容体リガンド予測への適用

ヒトβ2アドレナリン受容体(β2AR)を標的GPCRとし、構築した学習モデルを用いてリガンド予測を行った。リガンド予測の対象化合物は、他GPCRのリガンドとして知られている826化合物とした。

そして、CGBVSにより予測された β 2ARリガンド候補上位50の化合物のうち、文献・特許調査により14種の化合物について β 2ARとの相互作用に関する報告を確認した(図3A左)。さらに、残りの相互作用不明な化合物のうち、入手可能な21種類についてin vitro結合阻害実験を行ったところ、17種類の化合物が相互作用($10^{-5}\text{M} < \text{IC}_{50} < 10^{-3}\text{M}$)を示した(図3A右)。結合阻害実験のヒット率は81%(17/21)にのぼり、ここにおいても高い予測的中率が示された。

一方、下位50化合物については、 β 2ARリガンドとして報告されているものは文献および特許調査では確認されなかった(図3B左)。さらに、残りのうち入手可能な9化合物についても結合阻害実験を行ったが、2個の化合物が同程度の強さの相互作用を示したものの、残りの7化合物は相互作用を示さなかった(図3B右)。

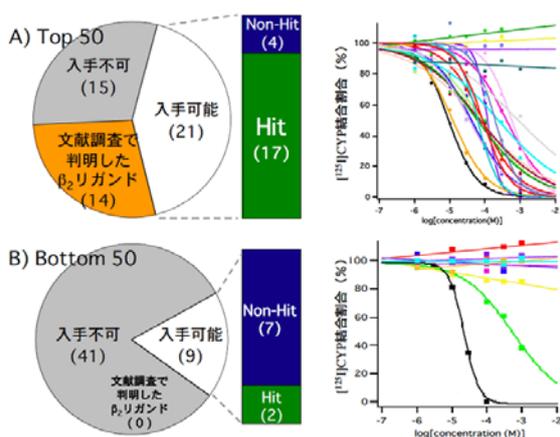


図3 β 2ARリガンド予測結果の検証

予測スコア上位50 (A) と下位50 (B) の化合物に対する調査・実験結果を示す。(左: 文献調査および実験検証で判明した化合物の内訳。右: $[^{125}\text{I}]$ -シアノピンドロールに対する結合阻害曲線。縦軸が阻害の割合、横軸が各化合物の濃度を示す。)

これらの予測結果をLBVSによるものと比較したのが図4である。結合阻害実験で相互作用を確認した化合物の半数近くは、LBVSではスコアが低かった。実際に、これらの化合物は、典型的な β 2AR作動薬の構造(カテコラミン骨格、イソプレナリン誘導体)および β 2AR拮抗薬の構造(アリルアルキルアミン誘導体)とは異なる多様な骨格を持っており、化合物の構造類似性に基づく従来の方法では発見が困難なリガンド群である。また、これらの化合物の中には、ニューロペプチド受容体アンタゴニストなど、従来はペプチド受容体に作用する化合物として知られていたものも含まれており、遠縁にあたる β 2ARとも相互作用することが今回の結合阻害実験により確認された。以上から、ケミカルゲノミクス情報が、リガンド予測精度の向上のみならず、新規骨格を持つリガンドの検出にも有用であることが示唆された。

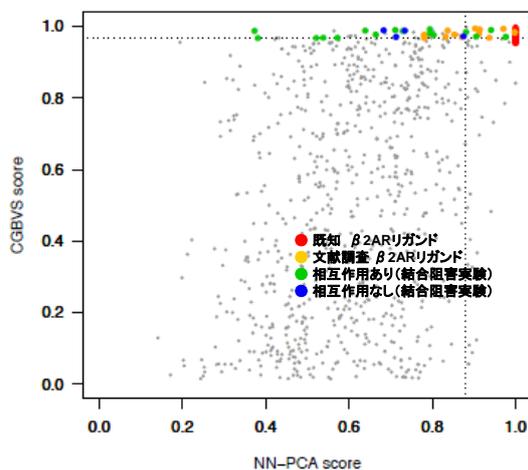


図4 CGBVSとLBVSによる β 2ARリガンド予測結果の比較
各点が化合物であり、縦軸はCGBVS、横軸はLBVS(主成分空間における最近傍法)による相互作用予測スコアを示している。文献調査および結合阻害実験の結果を色分けにて表示した。なお、図中の点線は、各手法において上位50位のスコアを示す。

[今後の研究の方向、課題]

計算機によって創薬を行うインシリコ創薬システムの理想形は、「新規性・高活性・選択性・安全性・リード展開性」の全ての点において優れた化合物、すなわち医薬品候補化合物として総合的に良質な化合物を「高精度」に計算予測するシステムと言える。本研究では、「高精度・新規性」の点において優れた性能を示す方法論の開発に成功した。今後の研究展開としては、残りの課題である「選択性・高活性・安全性・リード展開性」における技術開発を行い、これらを統合化したインシリコ創薬システムの実用化へと繋げたい。

[成果の発表、論文等]

論文発表

1. Yabuuchi, H., Nijima, S., Hirokawa, T., Hara, T., Minowa, Y., Tsujimoto, G. and Okuno, Y. “Chemical genomics-based virtual screening identifies novel ligands and cross-reactivities.” *Science*, in submission
2. Nijima, S. and Okuno, Y. “Laplacian Linear Discriminant Analysis Approach to Unsupervised Feature Selection.” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, in press
3. Okuno, Y., Tamon, A., Yabuuchi, H., Nijima, S., Minowa, Y., Tonomura, K., Kunimoto, R. and Feng, C. “GLIDA: GPCR-Ligand Database for Chemical Genomics Drug Discovery – Database and Tools Update.” *Nucleic Acids Research*, **36**, D907-12, 2008
4. Kitajima, M., Minowa, Y., Matsuda, H. and Okuno, Y. “Compound-Transporter Interaction Studies using Canonical

Correlation Analysis.” *Chem-Bio Informatics J.*, **7**, 24-34, 2007

学会発表

1. 日本薬学会 128 年会 日本薬学会奨励賞受賞講演「バイオ空間とケミカル空間の包括的相関解析とそのインシリコ創薬への研究展開」 (2008. 3. 28)
2. 日本薬学会 128 年会 日本薬学会・日本学術会議薬学委員会共催シンポジウム「バイオインフォマティクスの薬学研究・薬学教育への応用と展開」 Bio-Informatics for Pharmaceutical Sciences and Pharmacy Education 「バイオ空間とケミカル空間の包括的相関解析とそのインシリコ創薬への研究展開」 (2008. 3. 26)
3. 第 3 回三重ゲノム創薬フォーラム「薬学研究におけるアレイインフォマティクス」 (2008. 2. 15)
4. 平成 19 年度 第 2 回産業情報交流会「ケミカルゲノミクスに基づくインシリコ化合物探索他」 (2007. 10. 22)
5. 新産業を創る先端科学技術フォーラム 2007 「ポストゲノム創薬のための新技術」セッション「ケミカルゲノミクスに基づく創薬インフォマティクス」 (2007. 10. 18)
6. 第 66 回日本癌学会学術総会 International Session 「Chemical Genomics for Cancer Research」 「Knowledge Discovery and Data Mining in Chemical Genomics」 (2007. 10. 3)
7. 日本ケミカルバイオロジー研究会 第 2 回年会「ケミカルゲノミクス情報を利用した GPCR の in silico リガンド探索手法の開発」 (2007. 5. 10)