

平成 22 年 4 月 30 日現在

研究種目：若手研究（スタートアップ）
 研究期間：2008～2010
 課題番号：20890261
 研究課題名（和文）
 バーチャルスクリーニングの高精度化へ向けた COMBINE スコア関数の開発
 研究課題名（英文）
 Development of COMBINE score function to improve accuracy of virtual screening
 研究代表者
 中村 真也（NAKAMURA SHINYA）
 近畿大学・薬学部・助手
 研究者番号：10510109

研究成果の概要（和文）：

本研究では、コンピュータを利用して高速に医薬品候補化合物を探索するバーチャルスクリーニングの高精度化のために、構造活性相関（QSAR）手法のひとつである COMBINE 法を評価関数として利用した。既知のターゲットを用いて検証した結果、一定の前処理が必要となるが、高活性の化合物を高速で見出すことが可能であり、同時にターゲットタンパク質への結合モードも精度よく予測できることを見出した。

研究成果の概要（英文）：

To improve accuracy of virtual screening in which candidate medical compounds were searched fast by computer, COMBINE method that is one of the Quantitative Structure-Activity Relationship (QSAR) was applied as the scoring function. As the results of several validations, active compounds were able to be extracted at high throughput. Furthermore, binding mode of the compounds to their target protein could be predicted at the same time.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008 年度	1,050,000	315,000	1,365,000
2009 年度	900,000	270,000	1,170,000
年度			
年度			
年度			
総計	1,950,000	585,000	2,535,000

研究分野：医歯薬学

科研費の分科・細目：薬学・物理系薬学

キーワード：構造活性相関・バーチャルスクリーニング・SBDD・IT創薬・計算化学

1. 研究開始当初の背景

近年の医薬品開発において、創薬のプロセスを効率的に行うために、コンピュータを利用した *in silico* 創薬が注目されている。その中で、コンピュータ上で医薬品候補化合物を探索するバーチャルスクリーニングという手法が存在する。この手法では、薬物の活性

値を高速に予測することが重要であるが、精度の面から確立された手法は存在しておらず、改良が求められていた。そこで、高速かつ高精度に予測を行うため、構造活性相関の手法を取り入れることで、従来にないターゲット特異的なスコア関数を作成し、問題点の

解決に臨んだ。これは汎用性に捉われず、予測に用いるトレーニング群の情報を増やせば増やすほど、その系に対しての予測精度を向上できるという大きな利点を持てると期待できた。

しかし多くの構造活性相関手法は、共通する構造を持っているものに利用が限定されるという特徴があった。そこで既存の構造活性相関手法のひとつとして知られていた COMBINE 法に着目した。この手法はターゲットタンパク質の3次元構造を利用し、それとの相互作用から活性値を見積もる手法である。本研究代表者は、COMBINE 法では化合物群の骨格に依存せず予測が可能であることをすでに示しており、これを利用することで予測能向上の達成を目指した。

さらにリードホッピングと呼ばれる過程が創薬において注目を集めている。これは創薬研究において、活性値は十分であるにも関わらず、物性が不利という点でドロップアウトしてしまった化合物群を有効利用するものであり、同じターゲットを志向しつつも化合物の骨格を変換することで、ターゲットへの活性値を保ったまま、物性を向上させる過程である。ここで COMBINE 法は骨格に依存せず高い精度で予測が可能であるため、既存情報が多いほど精度が向上できるという点において、リードホッピングに極めて親和性が高い。この点において、本手法は創薬の現場において重要な計算手法になることが期待された。

2. 研究の目的

本研究では、バーチャルスクリーニングにおける活性値の予測精度を、高速性を保ったまま向上させることを目的とし、可能であればターゲットタンパク質へ結合している状態のコンフォメーションや相互作用様式を予測することを目的とした。予測を行うためのスコア関数は従来汎用なものであるため、COMBINE 法を利用することで、従来にないターゲット特異的なスコア関数を開発することも目的であった。しかし手法としては一般的に汎用に使用できる計算条件を決定する必要があるため、これの決定も目的とした。

またこれらの達成により、本手法を実際の創薬の現場で使用される計算手法にすることが最終的な目的であった。

3. 研究の方法

前項の目的を達成するために、4つの項目を段階的に検討することにした。すなわち、1. ターゲットの選定、2. ドッキング条件の探索、3. 構造予測の可否、4. 大規模ドッキングの4項目である。本来は段階的に研究を進める予定であったが、選定が難航したため実際には順次並行して計算を行うこと

とした。

まずターゲットの選定に関しては、HIV protease, Adenosine Deaminase, p38 MAPK など、PDB 結晶構造データベースから既存の阻害薬との複合体結晶構造が登録されているものを複数のターゲットとして選出した。これらを元に各ターゲットについて、ドッキング計算を行い COMBINE 法により予測スコアを得た。特に HIV protease と Adenosine Deaminase については構造予測ができるかどうかも考慮した。また予測の有用性の比較として、デコイとして化合物データベースの化合物群から既存阻害剤が有意に抽出されてくるか比較を行った。

4. 研究成果

本研究の成果として、COMBINE スコアによる構造と活性の同時予測の実証と、COMBINE スコアの有意性の実証の2点が大きく分けて挙げられる。

まず構造予測の可否について、HIV protease について 検証した結果を示す。32個の化合物をトレーニング群とし、16個の化合物の結合構造および活性値を同時に予測した。その結果対照に用いた DOCK スコアが14個の結晶構造を最上位にランクしたのに対し、COMBINE スコアも14個を最上位としてランクした。この点においては COMBINE スコアの有意性を示すことはできなかったが、活性予測を DOCK スコアに加えて他の既存スコア (LigScore2, PLP2, JAIN, PMF) と比較したところ、表1に示す通り COMBINE スコアの予測能が最も高かった。すなわち、構造に対する予測精度は既存のスコアと同等の精度を保ちながら、活性値の予測精度は既存のスコアを上回っており、同時予測を行うことができるという点において COMBINE 法は優れていると考えられる。

表1. HIV protease の活性予測

スコア	R ²
COMBINE	0.74
DOCK	0.63
LigScore2	0.30
PLP2	0.60
JAIN	0.20
PMF	0.66

また同様に Adenosine Deaminase においても構造予測を行った。このときは27個のトレーニング群を用いて、11個の化合物の結

晶構造を予測した。11個のうち1個はドッキングツールを用いても、得られた候補構造群の中に結晶構造を見出すことができなかつたため、解析から除外した。残り10個のうち、ドッキングに用いた U_{dock} スコアは10個全て予測可能であったのに対し、COMBINE スコアでは9個が予測可能であった。構造の予測精度という点においては、既存のそれに劣るかほぼ同等であるという結論が得られた。

この予測できなかった1個を精査することで、COMBINE スコアの改良すべき点が示唆された。図1に結晶構造と COMBINE 法による最上位構造との比較を示した。結晶構造では、分子内水素結合を形成することにより、化合物自身が分子として安定化すると同時にターゲットタンパク質に結合することで安定化を図っている。一方 COMBINE 法で予測された構造はタンパク質との相互作用の身を考えると結晶構造と全く変わらない相互作用様式が得られている。実際に、実験値の pK_i は 8.12 であるのに対し、結晶構造のスコアは 7.96、最上位の構造は 8.32 と非常に良い精度で実験値の予測が可能であった。それにも関わらず実験構造を上回るスコアを算出した原因としては、分子間相互作用を考慮に入れていないためであると推測できる。逆にドッキングの条件を工夫し、一定以上の分子内ひずみをもつコンフォメーションを除外するような条件設定でドッキングを行えば十分に最良に比肩することが推測できると期待できる。

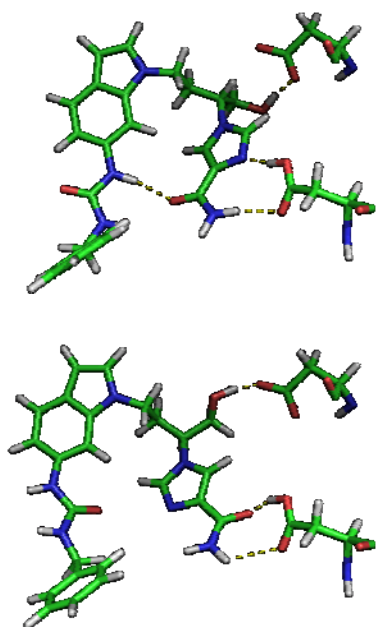


図1 (上) 結晶構造 (下) COMBINE スコアで最上位の構造

さらに、今回用いたドッキングツールの他のスコアと活性値の予測精度を比較したところ、僅かではあるが COMBINE スコアが最も良い予測精度を示した。すなわちこの系においても、構造に関する予測精度は既存のスコア関数とほぼ同等であり、活性値の予測精度は最も優れていたと言える。ここで最も特筆すべき点は、 U_{dock} スコアの算出には1化合物あたり数秒程度というドッキングにおいては比較的時間のかかる計算が必要であり、バーチャルスクリーニングのような数百万化合物にもおよぶ計算には適さないのに対し、COMBINE スコアの算出は1秒未満という高速な予測が可能となっている。これらの点については国際学会「The 8th China-Japan Joint Symposium On Drug Design and Development」において発表を行った。

表2. Adenosine Deaminase の活性予測

	R^2
COMBINE	0.62
U_{dock}	0.57
ASE	0.40

上記のような有用性が期待されたため、十万規模の小規模な化合物データベースから既存化合物の抽出が可能であるかを検証した。ただし、計算機の納入の遅れに伴い計算機資源がかなり乏しくなってしまったため、真に大規模な検討が行えたとは言い難い。またそれに伴い、計算効率のより一層の向上を図るべく、計算のカットオフ距離の工夫による省力化を行い速やかな結果の算出にも注力した。

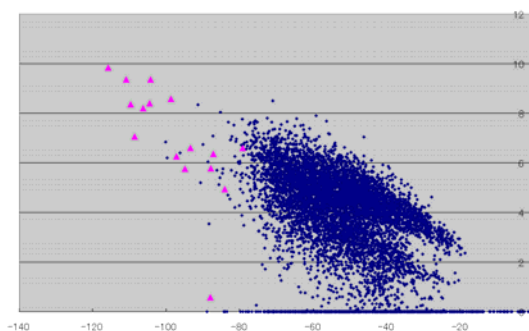


図2. スコアの比較
横軸：DOCK スコア
縦軸：COMBINE スコア
赤▲が既存薬を示す

Adenosine Deaminase によるスクリーニング検討結果を示すと、まず図2に既存スコアである DOCK スコアと COMBINE スコアの比較を行った場合明らかに DOCK スコアの方が、判別能が高いことがうかがえる。これは図3のようにスコアの分布図と照らし合わせた結果、不活性の化合物が活性として評価されてしまう、False Positive の傾向が高いためであると判断できた。これは COMBINE スコアに含まれる定数項に含まれるものであると考えた。つまりトレーニング群にある程度活性があるものが揃っていると、そのほぼ平均値が定数項として得られ、活性への寄与が少ない化合物であってもある程度活性があるとみなされてしまうためである。

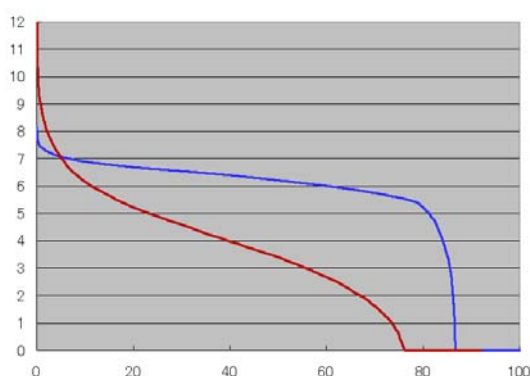


図3. COMBINE スコアの分布図

横軸：化合物数の割合
縦軸：COMBINE スコア
青線：定数項あり
赤線：定数項なし

そこで定数項を含まないように計算ツールを編集し予測を行ったところ分布が変化し、False Positive は大幅に低減し、判別能も向上した。しかしドッキング精度と回帰された係数の兼ね合いから、反発を生じた構造部分が好ましいと評価される場合が見受けられた。それにより物理的に妥当でないにもかかわらず高活性であると評価されるものも依然残っており、より洗練された前処理の方法が必要であった。

以上より、COMBINE スコアは手法としては有用であるが、ドッキングの結果の中から反発した物理的に妥当でない構造をとる化合物を前処理として合理的に除く必要があると考えられる。本研究の成果として、手法の完全な確立までは至らなかったが、本手法は一層の洗練により精度向上が得られるという期待が得られた。今後の展望としては、前処理の方法をより妥当かつ COMBINE スコアに馴染むように変更し、より多くの

系で検討を続け、本研究の有用性を示したい。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計0件)

〔学会発表〕(計3件)

①中村真也, 高平和典, 仲西功, サラシア属植物含有成分 salacinol の α グルコシダーゼへの結合シミュレーション、第3回食品薬学シンポジウム、2009年11月13日、近畿大学

②中村真也, 高平和典, 仲西功, 新規 α グルコシダーゼ阻害剤探索へ向けた salacinol の結合様式推定、第59回日本薬学会近畿支部総会・大会、2009年10月24日、近畿大学

③中村真也, 北浦和夫, 仲西功, Docking-pose prediction by a receptor-based tailor-made scoring function、The 8th China-Japan Joint Symposium On Drug Design and Development、2008年11月3日、神戸国際会議場

6. 研究組織

(1) 研究代表者

中村 真也 (NAKAMURA SHINYA)

研究者番号：10510109

(2) 研究分担者

なし

(3) 連携研究者

なし