（ 続紙 1 ）

京都大学	博士（薬科学）	氏名	白石 慧
論文題目	ケミカルゲノミクス情報に基づくタンパク質と化合物の相互作用予測法の開発		

（論文内容の要旨）

ケミカルバイオロジーやケミカルゲノミクスの進展により、化合物とタンパク質の相互作用は１分子対１分子の単純な関係ではなく、多対多の複雑な関係にあることが明らかになってきた。実際に、分子標的薬として開発された医薬品でさえ、複数のタンパク質を標的としていることが報告され、それら潜在的な相互作用が化合物の生理活性や薬物の副作用と大きく関わっていることが注目されている。しかし、膨大な化合物とタンパク質全ての組み合わせの相互作用を実験的手法により明らかにするのは極めて困難であるため、高速かつ網羅的に相互作用を予測できる計算手法に対する期待が高まっている。このため、本研究では、化合物とタンパク質の膨大な相互作用の包括的な予測ならびに機序の解明を目的として、ケミカルゲノミクス情報を用いた新規計算手法の開発および評価を行った。

第一章　キナーゼファミリーと化合物の網羅的相互作用解析と予測

キナーゼファミリーは活性部位である ATP 結合ドメインの保存度が高いことから、特定のキナーゼに選択的な分子の設計が困難とされている。そこで本研究では、網羅的なアッセイデータ、すなわちケミカルゲノミクスデータをもとに、各種キナーゼと阻害剤との相互作用を解析するための新たな計算手法を開発し、キナーゼファミリーに対する化合物の選択性の予測精度向上と選択性を担う機序の解明を行った。具体的には、DCNB (dual-component naive Bayes)モデルと DCSVMs (dual-component support vector machines)の二種類の相互作用予測法を開発した。これらの手法は、タンパク質をアミノ酸残基フラグメントにより、また化合物を部分構造フラグメントにより表現することで、キナーゼプロファイルデータを分析、学習するという特色を有している。特に DCSVMs は、内部評価、外部評価のいずれにおいても非常に高い予測性能を有していることが示された。また、DCNB は化合物の選択性に関わる特徴的な化合物部分構造とアミノ酸残基を簡便に抽出することが可能であり、そのようにして抽出された特徴は、創薬のための化合物ライブラリの構築や構造展開に役立つことが期待できる。

第二章　GPCR ファミリーと化合物の網羅的相互作用解析と予測

GPCR(G タンパク共役型受容体)は結晶構造の取得が非常に難しいことから、三次元構造を基にした相互作用予測や解析の非常に難しいタンパク質ファミリーとして知られている。このことから、タンパク質配列のみを用いて化合物相互作用の予測を可能とするケミカルゲノミクス情報に基づく計算手法が注目されてきた。本研究では、GPCR に特化した化合物－タンパク質間相互作用の記述子・カーネル表現を新たに開発し、予測精度の向上と相互作用機序の解明を行った。具体的には、GPCR 配列の膜貫通領域のマルチプルアライメントをと

り、それに基づくカーネル表現を開発した。また、GPCR のアミノ酸残基と化合物部分構造との組み合わせから相互作用への統計的寄与率を算出することにより、予測器の構築に必要な負例の合理的な選択法も開発した。この結果、内部および外部データの活性を高い精度で予測可能な GPCR－化合物間相互作用予測器の開発に成功した。また、本計算手法によって抽出された相互作用に関わるアミノ酸残基と化合物部分構造の組み合わせについて、既報の結晶構造と比較したところ、結晶構造上でも同様の化合物部分構造-タンパク質残基間の相互作用が確認されたことから、本計算手法が妥当であることが示された。

第三章　ケミカルゲノミクス情報に基づく相互作用予測法の異なるタンパク質ファミリーにおける評価

化合物やタンパク質の特徴や類似度を数値で表現する記述子やカーネルといったデータ表現は、ケミカルゲノミクス情報を用いて相互作用予測を行う上で必要不可欠であり、かつ予測器の精度を決定付ける重要な因子の一つである。本研究では、異なるタンパク質ファミリーにおける網羅的な相互作用予測に対する種々のカーネル表現の効果を明らかにすることを目的として、GPCR、キナーゼ、核内受容体、イオンチャネル、トランスポーターそれぞれのファミリーの大規模データに対して、カーネルに基づく予測器の予測性能を比較評価した。本研究により、化合物カーネルとタンパク質カーネルの組み合わせが予測精度に大きく関わること、更にその影響の度合いは対象とするタンパク質ファミリーによって異なることが明らかとなった。

以上、ケミカルゲノミクス情報を用いた新たな計算手法を開発し、タンパク質の結晶構造の有無や非相互作用データの有無に関わらず高い予測精度を示す相互作用予測器の開発ならびにアミノ酸残基と化合物部分構造からなる特徴的な相互作用ペアの抽出法の開発に成功した。本研究の成果は、医薬品開発におけるリード化合物の探索や最適化に資するものと期待できる。

（続紙　２　）

（論文審査の結果の要旨）

近年、特定のタンパク質を標的として設計された分子標的薬が、標的以外のタンパク質と作用し、意図せぬ副作用や薬効の原因になる「ポリファーマコロジー」という観点が注目されつつある。しかしながら、ポリファーマコロジーは、薬物と複数のタンパク質との相互作用を評価しなければならないことから、薬剤の分子設計においてこれらを加味することが非常に困難である。

申請者の研究は、この問題に対応するための新しい情報科学的アプローチであるといえる。すなわち、過去の知見から得られるタンパク質と化合物との膨大な相互作用データ（ケミカルゲノミクス情報）の機械学習や統計解析を通じて、ポリファーマコロジーを考慮した化合物探索や相互作用因子の抽出に取り組んでいる。

申請者は、キナーゼタンパク質ファミリーと GPCR タンパク質ファミリーを対象に、大規模な化合物との相互作用を解析する数理的枠組みの開発に着手した。ここで、タンパク質ファミリーのマルチプルアライメントによって独自のタンパク質記述子を作成し、化合物に対しては部分構造を断片化したフラグメント記述子を独自に考案している。また、これらの独自の記述子を用いて、新しい予測器を開発し、その予測性能を評価したところ、従来型の予測法よりも、いずれも高い予測性能を示すことが証明された。さらに、タンパク質記述子と化合物記述子を用いて、タンパク質の構成アミノ酸と化合物の部分構造の共起性を統計的に評価することで、大規模なタンパク質と化合物との相互作用データから、分子間相互作用の要因となる構成アミノ酸と化合物部分構造のペアの統計的抽出にも成功している。

また、申請者は、これらのケミカルゲノミクス情報に基づく相互作用予測法を、キナーゼと GPCR 以外の標的タンパク質である核内受容体、イオンチャネル、トランスポーターにも適用し、複数のタンパク質ファミリーへの方法論の汎化性を評価するとともに、異なるタンパク質記述子や化合物記述子の予測性能に与える影響を評価した。その結果、標的とするタンパク質ファミリーの種類によって、最適なタンパク質記述子と化合物記述子の組合せが異なることが明らかとなり、対象とする標的タンパク質によって最適な記述子を選択しなければならないことを明らかにした。

申請者が開発した相互作用予測法や相互作用因子の抽出法は、医薬品開発過程において、ポリファーマコロジーを考慮したリード化合物探索やリード最適化に適用されることが期待される。

よって本論文は博士（薬科学）の学位論文として価値あるものと認める。

さらに、平成２５年２月２２日、論文内容とそれに関連した口頭試問を行った結果、合格と認めた。

論文内容の要旨及び審査の結果の要旨は、本学学術情報リポジトリに掲載し、公表とする。特許申請、雑誌掲載等の関係により、学位授与後即日公表することに支障がある場合は、以下に公表可能とする日付を記入すること。

要旨公開可能日：　平成２６年３月３１日以降