

# 和漢薬のターゲットタンパク質のインシリコ探索

統括者	梅壽 雅人	国際共同研究分野（現 情報科学分野）	特命准教授
所内共同研究者	門脇 真	消化管生理学分野	教授
申請代表者	立川 仁典	横浜市立大学大学院生命ナノシステム科学研究科	教授
所外共同研究者	山西 芳裕	九州大学高等研究院生体防御医学研究所	准教授

## 【報告セミナー要旨】

和漢薬の化学的学術情報と臨床効果に関する情報は蓄積されてきたが、その作用機序の科学的根拠は未だ乏しい。本研究では、複雑系である和漢薬の作用機序解明を目指し、既知の和漢薬標準化合物に関するケモインフォマティクス分野の情報と、ヒトゲノム解析などのバイオインフォマティクス分野の情報から有益な情報を抽出し、和漢薬標準化合物の標的タンパク質探索を目的とするインシリコ解析を実施する。本課題が採択されている公募型研究において提供されている和漢薬標準化合物（探索研究プロジェクト：96種類）のヒト全タンパク質に対するインシリコでの結合シミュレーション解析や和漢薬成分の標的タンパク質予測のための機械学習法を開発し、和漢薬の標的タンパク質を解析するためのデータベースを確立する。

初年度は、化合物の化学構造や薬理情報を表すケミカル空間とヒトのタンパク質をコードする全遺伝子の様々な情報を表すゲノム空間の関係を探索するケモゲノミクスの枠組みで、和漢薬標準化合物の標的タンパク質の大規模探索を行った。まず、96種類の和漢薬標準化合物に関して、ヒト全タンパク質立体構造に対する結合シミュレーション（約400万パターン）を行った。次に、機械学習による予測のための予備解析として、上記96種類の和漢薬標準化合物や様々な公共データベースに蓄積されている膨大な数の化合物に対し、最新の化学構造記述子に基づく化学構造類似度を計算した。そして、タンパク質との相互作用情報を持つ約50万個の化合物に対して類似度検索を行い、和漢薬成分とタンパク質のペアの候補を6000件以上列挙した。これらの結合シミュレーションの結果や、化学構造類似度から予測された標的タンパク質のリストをまとめたデータベースを構築した。予測されたタンパク質のオンロジー情報、パスウェイ情報、そのタンパク質が関連する疾患情報なども付加して、結果を解釈しやすい環境を整備した。今後、更にインシリコ手法やデータベースを発展させていく予定であり、和漢薬成分の作用機序の解明や新規効能の発見などに応用することが期待できる。

## ■背景および目的

和漢薬の化学的学術情報と臨床効果に関する情報は蓄積されてきたが、その作用機序の科学的根拠は未だ乏しい。そこで、複雑系である和漢薬の作用機序解明を目指し、既知の和漢薬標準化合物に関するケモインフォマティクス分野の情報と、ヒトゲノム解析などのバイオインフォマティクス分野の情報から有益な情報を抽出し、和漢薬標準化合物の標的タンパク質探索を目的とするインシリコ解析を実施する。具体的には、本課題が採択されている公募型研究において提供されている和漢薬標準化合物（探索研究プロジェクト：96種類）のヒト全タンパク質に対するインシリコでの結合シミュレーション解析を実施すると共に、既知医薬品等の化合物と標的タンパク質間の相互作用情報から化学構造類似度に基づいた和漢薬成分の標的タンパク質の予測システムを構築し、和漢薬の標的タンパク質を解析する手法を確立する。

初年度の目的としては、膨大な化合物の化学構造や薬理情報を表すケミカル空間とヒトのタンパク質をコードする全遺伝子の様々な情報を表すゲノム空間の関係を探索するケモゲノミクスのインシリコ手法を適用し、和漢薬標準化合物の標的タンパク質の大規模探索を行う。そのため、既知の化合物・タンパク質間相互作用の情報を、様々なデータベースや文献情報から抽出し、ケモゲノミクスの予測モデルを学習するための大規模な化合物・タンパク質間相互作用の学習データを構築する。それに基づき学習した予測モデルを、本研究で対象とする和漢薬標準化合物に対して適用し、それらの標的タンパク質やオフ標的タンパク質（本来目標としていない標的タンパク質）をゲノムワイドに予測する。これらの予測結果や解析結果を系統的に整備し、データベースとして公開する。

上記の和漢薬標準化合物の標的タンパク質解析では、『「1成分」対「複合ターゲット」』という複雑系の解析となるが、最終的には『「多成分」対「複合ターゲット」』である和漢薬の複雑な作用メカニズムの解明を目指して、生薬に含まれる多数の化合物と、それらの標的となる多数の標的タンパク質の相互作用解明を目標とした複雑系解析手法の開発を目的とする。

## ■方 法

### 1. 和漢薬成分の標的タンパク質の予測システム開発：・ インシリコ結合シミュレーションによる予測

上記、96種類の和漢薬成分に関して、ヒト全タンパク質立体構造（約4万種類）に対する結合シミュレーション（約400万件）を実施し、標的タンパク質及びその作用部位を予測する。

### 2. 和漢薬成分の標的タンパク質の予測システム開発：機械学習手法による予測

既存医薬品・薬理活性物質について、文献情報などより相互作用するタンパク質の情報を抽出する。また、ヒト全タンパク質情報と既存医薬品・薬理活性物質の薬理作用情報を整備して、それらのデータを基に和漢薬成分の標的タンパク質を予測するシステムを開発する。具体的には、ChEMBL、BindingDBのような、「化合物（既存医薬品・薬理活性物質）－タンパク質相互作用情報」を有している公共データベースに収載されている化合物と、上記96種類の和漢薬成分の化学構造類似度に基づいて標的タンパク質を予測する機械学習の手法を開発する。

### 3. 和漢薬成分の標的タンパク質の予測システム開発：漢方薬成分の作用機序解析

次に、上記方法で漢方薬成分が相互作用すると予測された標的タンパク質について、タンパク質の生物学的機能で分類するオントロジー解析（Gene Ontologyなど）を行う。さらに予測された標的タ

ンパク質を既知のパスウェイ（KEGGやReactomeなどのシグナリング経路や代謝経路）にマッピングし、影響を及ぼしているパスウェイなどを解析する。

これらの結果を比較検討し、両方の手法で高い信頼度で予測される標的タンパク質の候補を選定する。

## ■結果および考察

96種類の和漢薬標準化合物に関して、ヒト全タンパク質立体構造に対する結合シミュレーション（約400万パターン）は、ほぼ計算が終了しその結果（結合エネルギー、結合状態の分子イメージなど）を表示するデータベースを構築している（試作公開版：<http://160.26.6.205/ligand/wakan/koba3.jsp>、図1）。

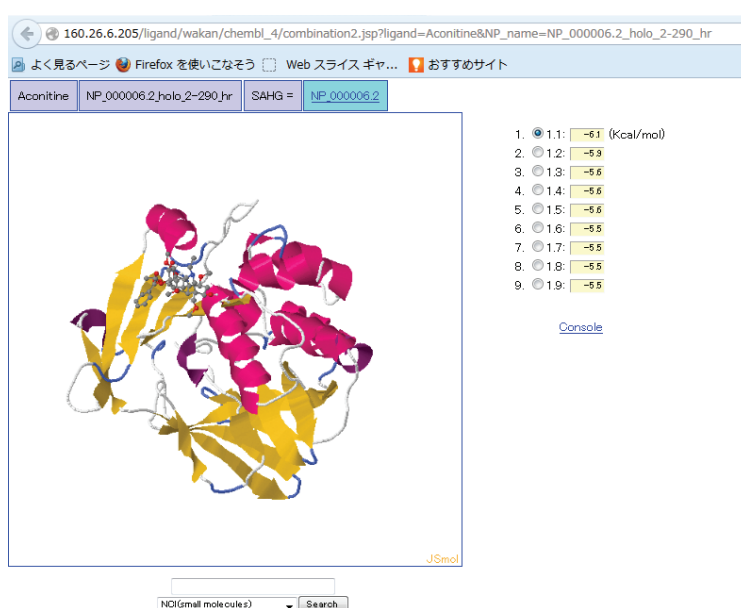


図1 和漢薬標準化合物とヒト全タンパク質との結合シミュレーション結果表示。

この結合シミュレーション結果の表示には「JsMol」を利用しているため、従来の「Jmol」を使った場合のセキュリティ懸念による非表示という問題はない。タンパク質は分子表面で表示され、化合物はボール&スティックで表示される。結合状態はエネルギー順位のラジオボタンで選択することができ、全ての立体構造は「JsMol」による3次元表示や拡大・縮小などの操作を利用できる。

機械学習手法による予測手法確立のために、上記96種類の和漢薬標準化合物に関して、最新の化学構造記述子に基づく化学構造類似度を計算した。少なくとも一つ以上のタンパク質との相互作用情報を持つ約50万個の化合物に対して類似度検索を行い、和漢薬成分とタンパク質のペアの候補を6000件以上列挙した。その結果をデータベースとしてまとめている（試作公開版：[http://160.26.6.205/ligand/wakan/chembl\\_4/](http://160.26.6.205/ligand/wakan/chembl_4/)、図2）。また、試作公開内部版（富山大学杉谷キャンパス内のコンピュータからのみ閲覧でき、IDとパスワードで保護している）として、予測された標的タンパク質と類似化合物の「結合強度情報」や、そのタンパク質が関連する「疾病情報」に関する情報を閲覧できるようにしている（これらの解析手法には特許に絡む事項があるため公開を内部に限定している）。6,000件以上の予測標的タンパク質に関する情報のうち、「結合強度情報」が提示され

ているデータが約1,700件以上あり、「疾病情報」が提示されているデータが約1,300件以上存在している。「結合強度情報」や「疾病情報」がない類似化合物データについても、全ての化合物に「相互作用する標的タンパク質の情報や、そのタンパク質のオントロジー情報、パスウェイ情報」などが付加されており、今後の解析で和漢薬成分の作用機序の解明に有用なデータが抽出されることが期待できる。

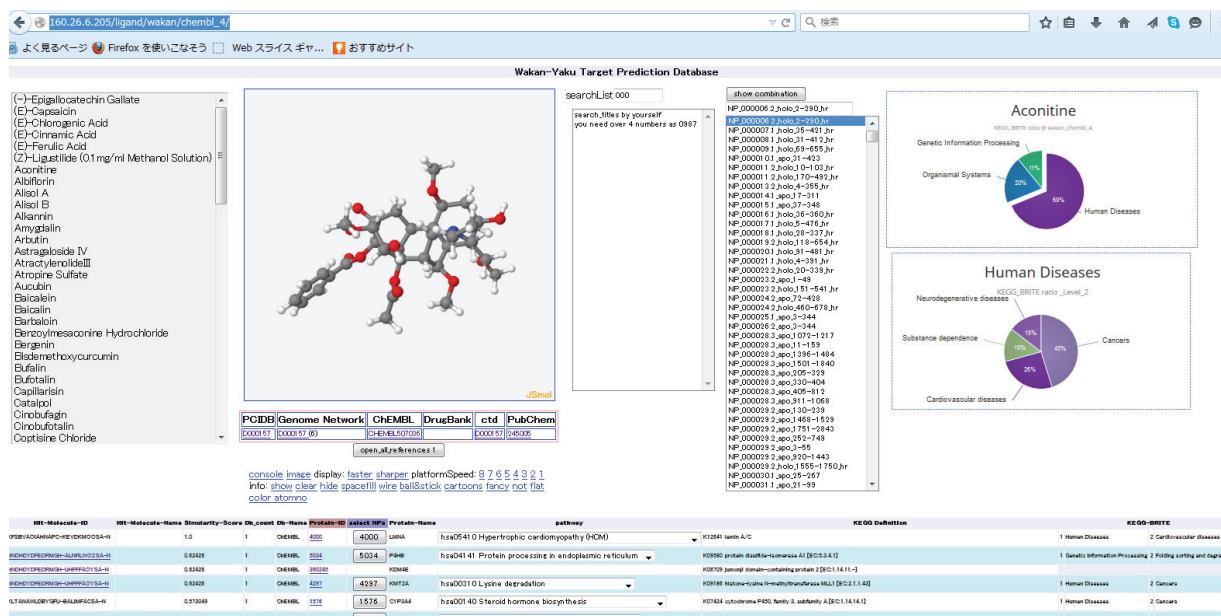


図2 化学構造類似度に基づいた標的タンパク質の情報表示結果。

96種類の和漢薬成分の化学構造類似度に基づいて、標的タンパク質を予測する機械学習の解析結果を表示したものである。左上のカラムから化合物名を選択すると、その横のボックスに分子構造が表示される。この分子構造表示の下にあるボタンによって、選択している和漢薬標準化合物に関する外部データベース（PubChem、CTDなど）のリンクを開くことができる。右下の一覧表には、化学構造の類似度を基にした標的タンパク質の予測結果が表示されており、そのタンパク質の情報へのリンクなどを利用できる。

現在、これらの結果を元にオントロジー解析や、パスウェイ解析の方法を開発しており、結合シミュレーションと機械学習法の結果をリンクさせる機能も構築している。今後、オントロジー解析やパスウェイ解析などの可視化及び、結合シミュレーションとの連携を進めて、和漢薬標準化合物の標的タンパク質候補を予測する。また、複合薬物である和漢薬の作用機序の解明を目指して、構成生薬の主な含有成分に関するデータベースを構築している研究者と連携し、生薬、和漢薬の標的タンパク質予測のための手法を開発する予定である。

## ■結論

特定研究課題「和漢薬のターゲットタンパク質のインシリコ探索」の初年度の研究として、探索研究プロジェクトで提供している96種類の和漢薬標準化合物について、ヒト全タンパク質との結合シミュレーション（約400万件）を実施し、その結果を表示するデータベースを構築した。また、「化合物（既存医薬品・薬理活性物質）-タンパク質相互作用情報」と、上記96種類の和漢薬成分の化学構造類似度に基づいて標的タンパク質を予測する機械学習の手法を開発した。この結果についても、

外部データベースへのリンクなどの情報を付加したデータベースを構築した。このデータベースと、上記結合シミュレーションのデータを関連付けて表示する機能も構築している。

今年度を実施した和漢薬標準化合物の標的タンパク質解析では、『「1成分」対「複合ターゲット」』という複雑系の解析手法を開発したが、来年度は『「多成分」対「複合ターゲット」』である和漢薬の複雑なメカニズムの解明を目指して、生薬に含まれる多数の化合物と、それらの標的となる多数の標的タンパク質の相互作用の解明を目標とした複雑系解析手法の開発を実施する。