

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 6 月 3 日現在

機関番号：14401

研究種目：基盤研究（C）

研究期間：2010～2012

課題番号：22500274

研究課題名（和文） 複合体立体構造データを利用した標的蛋白質に結合する化合物群の探索法開発

研究課題名（英文） A searching method for interacting chemical compounds for target proteins using 3D structures of compound-protein complexes

研究代表者

川端 猛（KAWABATA TAKESHI）

大阪大学・蛋白質研究所・招へい研究員

研究者番号：60343274

研究成果の概要（和文）：結合化合物の2D構造の比較から原子の1対1対応を計算するプログラム *kcombu* の開発を行った。この方法は最大共通部分構造 (MCS) という古典的なアイデアに基づいているが、組み上げ法という高速な近似解法を採用したこと、連結・非連結 MCS のほか、より実際の対応に近い TD-MCS (トポロジー距離拘束付き非連結 MCS) の計算を実装したことに新規性がある。さらに、このプログラムを用いて、鋳型利用式の立体構造予測法の開発を行った。また、同族蛋白質の異なる化合物の複合体立体構造から、結合化合物の族に共通する特徴、各蛋白質に個別の特徴に関する調査も行った。

研究成果の概要（英文）：We developed the program *kcombu* to calculate one-to-one atomic correspondences from 2D structures of chemical compounds. The method is based on the classical concept, maximum common substructure (MCS). Our method employs the fast heuristic “build-up” algorithm, and calculates topologically-constrained disconnected MCS (TD-MCS) in addition to conventional connected and disconnected MCS. Based on this program, we developed a template-based structural prediction method of chemical compounds. We also researched common and specific 3D features of chemical compounds, which bind to proteins of the same family.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	1,500,000	450,000	1,950,000
2011年度	1,100,000	330,000	1,430,000
2012年度	700,000	210,000	910,000
年度			
年度			
総計	3,300,000	990,000	4,290,000

研究分野：総合領域

科研費の分科・細目：情報学・生体生命情報学

キーワード：バイオインフォマティクス、構造バイオインフォマティクス

## 1. 研究開始当初の背景

近年、分子生物学の潮流は生命体を構成する部品の列挙から、それらの相互作用の解析に重点が移ってきた。特に、低分子化合物と蛋白質との相互作用は、生体内の化学

反応の理解、実験用試薬や医療用薬剤の開発につながるため、その科学的・産業的な重要性は高い。しかし、人工的な化合物は膨大な種類が存在するため、包括的に数多くの化合物と蛋白質との結合実験を行うこ

とはコストが高い。よって、注目する蛋白質に選択的に結合する化合物を、そのアミノ酸配列や立体構造だけから予測できる方法が望まれている。この目的を達成するために、従来、標的蛋白質の立体構造データを利用したドッキング計算がよく研究されてきたが、その予測精度はそれほど高くない。一方、部分的ではあるものの、化合物群の結合性のデータ、および、化合物と蛋白質の複合体の立体構造データも入手可能になっており、これら既存のデータを利用したスクリーニング法の開発が望まれている。

## 2. 研究の目的

本研究は、標的蛋白質に結合する低分子化合物を、化合物データベースから検索（探索、スクリーニング）する手法を開発することを目的とする。検索のクエリとしては、既知の結合化合物の構造式(2D構造)、複合体立体構造が入手できる場合は結合化合物の立体構造も利用して、化合物データベース内を検索する。立体構造データが未知のケースが多いことを鑑み、類似性ベースのドッキング計算法の開発も行い、化合物の結合立体構造の予測も行う。本研究の成果は、酵素の基質同定や薬剤設計など分子生物学・薬学分野への広い応用が期待される。

## 3. 研究の方法

### (1) 化合物の2D構造の比較法の開発

化合物検索の基本となる化合物の結合情報だけを用いた2D化学構造の比較を行う。最大共通部分化学構造(MCS)を算出し原子の一対一対応を計算する手法の開発を行う。MCS問題はNP完全であるため、現実的な時間で計算を行うため、組み上げ(build-up)法という近似解法を採用する。

### (2) 鋳型利用式の結合3D構造予測プログラムの開発

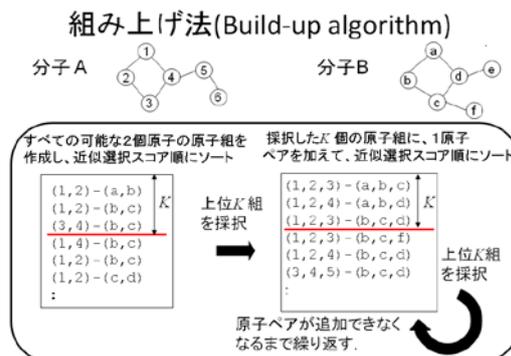
複合体形成時の結合立体構造がわかっている化合物を鋳型として用いて、標的化合物の結合立体構造を予測する、鋳型利用式の低分子構造予測法を開発を行う。鋳型利用式では、鋳型化合物とよく重なるように標的化合物構造を変換・変形する必要がある。従来法の多くは、非特異的なガウス関数を最大化する方法を採用している。本研究では、(1)で開発したMCS(最大共通部分構造)による原子対応を用い、対応する原子間の距離が最小となるよう変換・変形操作を行う方式を採用する。

### (3) 同族蛋白質の異なる化合物の複合体立体構造の調査

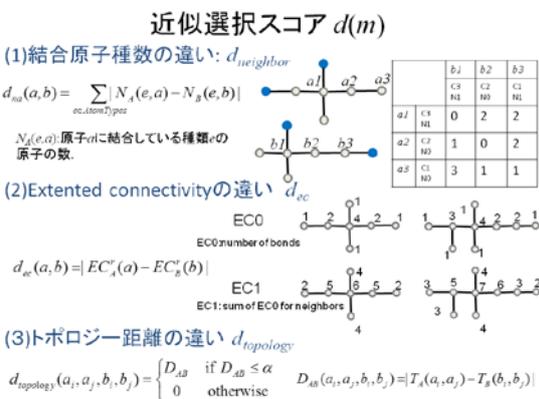
PDBデータベースの中には、同族の蛋白質について多数の異なる化合物が結合した複合体の立体構造が多数登録されている場合がある。これらの化合物の結合構造を蛋白質構造を基準に重ね合わせることで、異なる化合物の立体構造を重ね合わせることができる。これらの重ね合わせ構造は、前述の2Dや3Dの化合物比較の正解として用いることができるほか、その族に結合する化合物の族に共通の特徴や各蛋白質に個別の特徴を解析できるはずである。本研究では、主にタンパク質キナーゼ族について解析を行う。

## 4. 研究成果

(1) 化合物の2D構造の比較法 *kcombu* の開発と性能評価  
組み上げ法(build-up法)という近似アルゴリズムを用いて最大共通部分構造(MCS)を算出するプログラム *kcombu* (Kemical COMparison using Build-Up algorithm) の開発を行った。組み上げ法は、1原子の原子対応から開始し、逐次的に原子対応数を追加していく方法である。

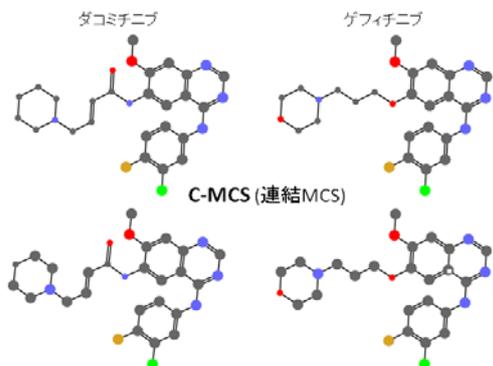


候補の選択には、下図のように、結合原子種、Extended Connectivity, トポロジー距離の違いの和による近似選択スコアを用いた。



また、MCSそのものの定義についても検討を加えた。これまで、MCSにはいくつかの異なる

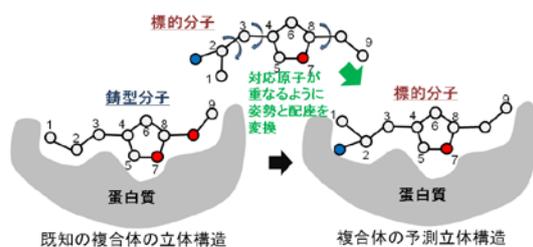
る定義が用いられてきた。本プログラムでは、これまで多く用いられてきた連結MCS (C-MCS)、非連結MCS (D-MCS)に加えて、より現実的な対応を目指し、トポロジー距離の拘束を加えたTD-MCS (Topologically constrained D-MCS) も実装した。下図にC-MCSとTD-MCS1の例を示す。



TD-MCS1 (距離拘束付き非連結MCS, 距離のずれの許容=1)

同一蛋白質に結合する化合物の複合体立体構造を収集し、それらを蛋白質構造に従って重ね合わせた構造を正しい重ね合わせとみなし、そこから得られる原子対応を正解として本プログラムの評価を行った。その結果、D-MCSやC-MCSに比べて、TD-MCSの一致度が最も高いことがわかった。本手法は、厳密解を求める方法に比べ、はるかに高速であるため、大規模な化合物ライブラリに対する検索にも十分用いることができる。この手法については、既に論文にまとめ、2011年に、J. Chem. Info. Model. 誌に出版している。また、ソースコードもWEBサーバにて公開を行っている。

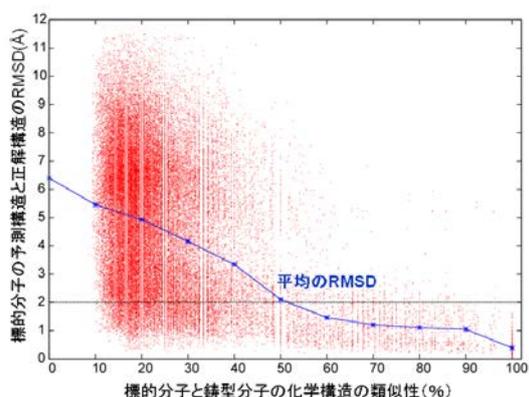
(2) 鋳型利用式の結合3D構造予測プログラム *fkcombu* の開発と性能評価  
*kcombu* で計算したMCSによる原子の1対1対応を基として、鋳型利用式の低分子の結合3D構造予測法の開発を行った。鋳型分子に重なるように標的分子の姿勢と配座を変換するプログラム *fkcombu* (flexible transforming *kcombu*) を開発した。下図にその概略を示す。



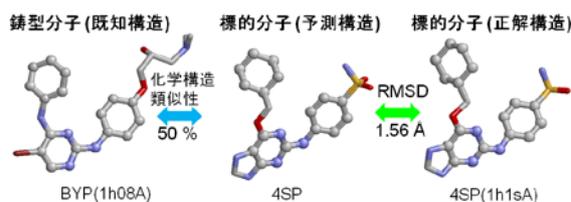
姿勢と配座の変換は、鋳型分子の回転角をそのまま捺印する操作、及び最急降下法によるエネルギー最小化操作を組合せて行う

よう実装した。エネルギー関数は、MCS に対応した原子間のバネ関数と標的分子内の衝突ペナルティ関数を主に用い、最適化は分子全体の並進・回転および結合回転角について行った。

2D構造の試験と同様に、同一蛋白質に結合する化合物の複合体立体構造を収集し、それらを蛋白質構造に従って重ね合わせた構造を正しい重ね合わせとみなして *fkcombu* プログラムの性能評価を行った。下図に CDK2 に結合する化合物を用いた場合の、標的分子と鋳型分子の化学構造の類似性 (タニモト係数) と標的分子の予測構造と正解構造のずれ (Å) を示す。



この図からわかるように、予測立体構造の精度は、鋳型分子と標的分子との類似性に依存し、類似度が50%を超えるなら、平均で2Å以下のずれで予測できることがわかった。下図に、鋳型分子、標的分子の予測構造、標的分子の正解構造の例を示す。これは鋳型一標的間の化学構造の類似性が50%で、1.56ÅのRMSDを得た例である。



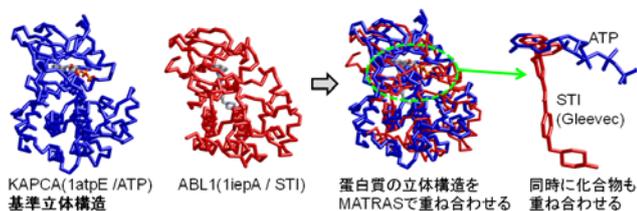
これらは、十分類似した鋳型分子の結合立体構造が得られる場合、本手法は、高精度に結合立体構造を予測できることを示す。計算結果は1ペアあたり数秒しかかからないので、大規模なライブラリに対する検索計算に応用できる可能性もある。開発した鋳型利用式の低分子構造予測プログラム *fkcombu* はWEBサーバを通じて実行可能である。その成果をまとめた論文を投稿準備中であり、ソースコードの配布準備も進めている。

(3) タンパク質キナーゼ族蛋白質の異なる

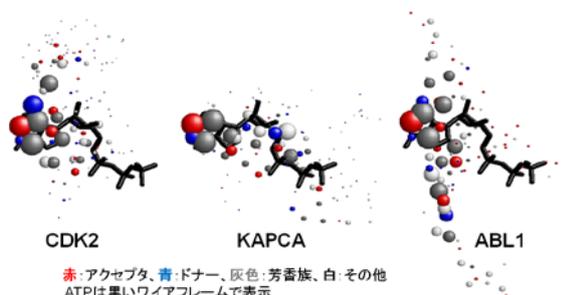
## 化合物の複合体立体構造の調査

タンパク質キナーゼは、ヒトのゲノム内に500以上コードされており、細胞内のシグナル伝達系の中核として、様々な細胞活動の制御に関わっている。また、癌など多くの疾病の標的蛋白質としても知られている。しかし、多くの類縁キナーゼが存在することから、選択的に一つのキナーゼだけに結合する化合物の設計は困難であり、その設計指針の確立が望まれている。PDBには、100以上のキナーゼについて、多種の化合物との複合体結晶データが1000以上登録されている。本研究では、それらの結合3D構造を解析することで、結合化合物のキナーゼ全体に共通する特徴、さらに、各キナーゼの結合化合物の個別的特徴を明らかにすることを目的とする。

以下のような手続きでキナーゼの結合化合物の結合状態の立体構造を入手した。まず、PDBからキナーゼ族に属するヒトの蛋白質のデータを入手し、その中から異なる化合物との複合体のデータを得る。これから、さらに10種以上の化合物の複合体構造が得られているキナーゼだけに限定すると700程度のデータが得られる。これらの構造データを下図のように、一つの基準立体構造にプログラムMATRASを用いて蛋白質の立体構造を重ね合わせ、同時に化合物も重ね合わせる。この操作を繰り返すことで、約700個の化合物を全て重ね合わせた。



こうして、重ね合わせた化合物の原子を、水素結合のアクセプタ原子、ドナー原子、芳香族原子、その他の原子に分類し、各キナーゼごとにクラスタリングを行うことで、その原子分布を観察した。下図にCDK2, KAPCA, ABL1についてその分布を示す。



この図から、全てのキナーゼの結合化合物に共通する特徴として、ATPのアデニン環に対応する部分に、アクセプタと芳香族の原子が結合する傾向があることがわかる。また、それぞれのキナーゼごとに、結合化合物の原子

分布に個別の特徴があることがわかった。これらの共通性、個別性は、あるキナーゼのみに選択的に結合する化合物を設計する上で重要な指針を与えると期待している。この結果は、2011年に日本蛋白質科学会年会で発表し、現在投稿準備中である。

## 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 1 件)

- ① Takeshi Kawabata, Build-up algorithm for atomic correspondence between chemical structures. J.Chem. Info. Model., 査読有、Vol.51、2011年、1775-1787.

[学会発表] (計 5 件)

- ① 川端 猛、中村春木、リガンド結合構造の比較モデリング—リガンド類似性・蛋白質相同性の予測精度への影響—、第12回日本蛋白質科学会年会、2012.6.21、名古屋国際会議場
- ② 宮田雄輔、川端 猛、複合体立体構造群を用いたタンパク質キナーゼ阻害剤の構造特徴解析、第11回日本蛋白質科学会年会2011.6.9、ホテル阪急エキスポパーク
- ③ 川端 猛、The build-up algorithm for matching chemical substructures、2010年日本バイオインフォマティクス学会年会、2010.12.14、九州大学医学部百年講堂
- ④ 川端 猛、KCOMBU: a program for matching chemical structures using the build-up algorithm、2010.9.15、学術総合センター
- ⑤ 加藤文彦、川端 猛、結合部位予測とドッキング計算の組み合わせによるリガンド・蛋白質複合体予測、第10回日本蛋白質科学会年会、2010.6.16、札幌コンベンションセンター

[その他]

プログラム *kcombu* の計算サービス・ソフトウェア配布用 WEB サーバ

<http://strcomp.protein.osaka-u.ac.jp/kcombu>

## 6. 研究組織

(1) 研究代表者

川端 猛 (KAWABATA TAKESHI)

大阪大学・蛋白質研究所・招へい研究員

研究者番号：60343274