

平成 21年 6月 11日現在

研究種目：基盤研究（C）

研究期間：2007～2008

課題番号：19500253

研究課題名（和文） 分子間相互作用予測を目指したタンパク質複合体結合部位の幾何形状解析

研究課題名（英文） Analysis of geometric shapes of binding sites on protein-ligand complex for prediction of molecular interactions

研究代表者

川端 猛 (KAWABATA TAKESHI)

奈良先端科学技術大学院大学・情報科学研究科・准教授

研究者番号：60343274

研究成果の概要：単体の蛋白質の立体構造データから結合する分子の種類、結合する場所を予測することを最終目標として研究を進め、蛋白質表面の結合部位の候補を認識するプログラム GHECOM を開発した。このプログラムはモルフォロジーという画像処理技術を使用し、様々な深さのマルチスケールなポケットを同時に認識することができ、その深さから結合する分子の種類などの情報を推定できる。プログラムはフリーウエアとして一般公開している。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	2,100,000	630,000	2,730,000
2008年度	1,400,000	420,000	1,820,000
年度			
年度			
年度			
総計	3,500,000	1,050,000	4,550,000

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、タンパク質、分子認識、分子形状、モルフォロジー

## 1. 研究開始当初の背景

ゲノム配列決定の急速な進展により、分子生物学の研究の焦点は、各遺伝子とその産物であるタンパク質の機能的な特徴づけに移行した。マイクロアレイによる発現解析、酵母2ハイブリッド法によるタンパク質間相互作用解析などの、ゲノム規模での機能解析が次々に行われている。また、構造ゲノム科学プロジェクトの進展により、生体高分子の立体構造のデータも急速に集積しており、特にタンパク質が他の分子と結合した複合体の立体構造データが得られた場合、その分子の機能メカニズムに関して深い知見が得られ

る。しかし、単量体の蛋白質の立体構造データは得られているものの、他分子との複合体の構造については未知である場合は多い。これは一般に分子量が大きくなると構造解析は困難になっていくこと、また、シグナル伝達などその機能の必然として、一時的な不安定な複合体しか形成しない蛋白質が多いことによる。また、創薬の現場においては、標的のタンパク質構造に特異的に結合する分子を新規に設計・探索することが求められている(ドラッグデザイン)。

## 2. 研究の目的

本研究では、蛋白質の立体構造データを利用して、幅広いタンパク質の機能の中で、主に分子間の結合に焦点をあて、単体のタンパク質の立体構造データから、そのタンパク質に結合する分子の種類、結合する場所、その強さを予測することを最終的な目標とする。こうした目的の理論的な研究は、「ドッキング計算」を行うのが正攻法である。しかし、ドッキング計算は、タンパク質の立体構造予測と似た二つの困難をかかえている。一つは、会合に伴う結合自由エネルギーを精確かつ効率的に計算することが困難であること、もう一つは、複合体の構造について、様々な配置を探索することが必要であり、その探索空間が膨大であることである。また、ドッキング計算を行うためには少なくとも、結合する分子種が既知であることが前提となる。本研究では、これらの困難を次の二つのアプローチで軽減することを目指す。一つは、タンパク質構造の形状的・幾何学的な情報を使って、結合サイト・結合分子の候補を効率的に絞り込むことである。結合部位はその表面に特徴的な形状を持つことが知られており、特に低分子の結合サイトはポケット型の凹形状をしており、また、蛋白質間相互作用面は、その形状の凹凸が相補的に組み合わさっていることが知られている。申請者は、これまでに大小二つのプローブ球を用いたポケット発見プログラム **PHECOM**(Probe-based HECOMi finder)を既に発表しており、格子版のポケット発見プログラムの開発も開始している。これらのプログラムにより、ポケット部の大きさ・深さなど、様々な幾何学量を高速に計算できる。もう一つは、既知の分子間相互作用データベースの効果的な利用である。数は限定されているとはいえ、PDBには相当数の複合体構造のデータが登録されており、既知の結合部位について、前述の幾何学形状による特徴化、さらに、疎水性・極性などの物性的な特徴量、結合原子の空間配置などを比較することで、ある分子が結合するポケットの性質を抽出することができ、その情報を用いて、結合部位、結合分子予測につながる知見が得られると考えられる。

### 3. 研究の方法

本研究は以下のような手順で進めた。

#### (1) マルチスケールなポケット認識アルゴリズムの開発とソフトウェアの整備

蛋白質の立体構造から、結合ポケットを推定することは本研究の要である。申請者は以前から行っていた大小二種のプローブ球を用いるアルゴリズムを、3次元格子を用いたグリッドベースに拡張することで、多数の大きさのプローブ(マルチスケール球)を用いて、より高速で、拡張性の高いプログラムの開発を進めた。

(2) 同じ分子が結合するポケットの比較解析  
結合する分子種によって、共通した性質がポケット部に見られるか解析を進めた。

#### (3) 非球体プローブを用いた結合部位予測とポケットの特徴量を用いた結合分子種の予測

非球体のプローブを採用することで結合部位予測の精度を向上を試みた。またそこから得られた様々なポケットの特徴量を用いて、それに結合する分子種が予測できるかどうか、予測システムを開発して検討した。

#### (4) 病因性の nsSNP と蛋白質のポケット部との関連の解析

発展的な研究として、病因性の nsSNP と中立な nsSNP が、その蛋白質の構造の特徴においてどのような違いがあるか、検討を行った。どの程度のポケット部に集中しているかについても検討した。

## 4. 研究成果

### (1) マルチスケールなポケット認識アルゴリズムの開発とソフトウェアの整備

開発を続けてきたポケット認識アルゴリズムを、格子表現を採用し、画像処理の分野で開発されたモルフォロジー理論 (mathematical morphology) を利用することで、厳密かつ高速なアルゴリズムの開発に成功した。開発されたプログラムを **GHECOM**(Grid-based HECOMi finder) と名付けた。蛋白質上には様々な深さのポケットがあり、深さに応じて結合する分子に違いがあることが知られている。このプログラムは、様々な深さのポケットを同時に高速に計算することができる。

本アルゴリズムでは、「小球  $S$  は入れるが大球  $P$  は入れない蛋白質  $X$  外の空間」をポケットとして定義している。この関係はモルフォロジーの表記を用い

#### モルフォロジーによるポケットの定義

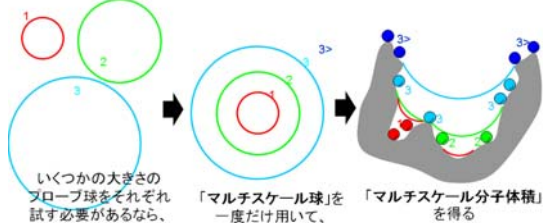


$$X.pocket(P,S) = [(X \cdot P) \cap X^c] \circ S$$

ると簡潔に表記できることが Masuya & Doi(1995)によって提案されている。ここで、大球  $P$  の大きさはポケットの深さ(狭さ)を規定するパラメータであり、いくつかの大きさを試して比較検討することが望ましい。ここで本研究では  $K$

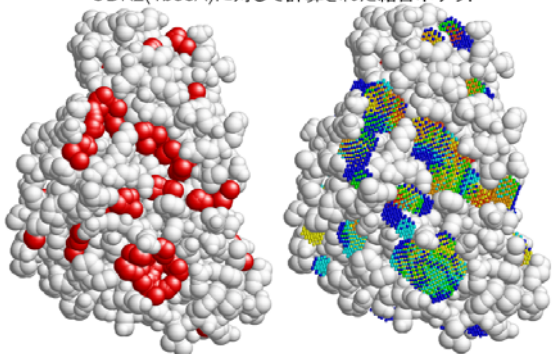
種の異なる大きさの地球 $\{P_1, P_2, \dots, P_K\}$ を用意し、それぞれの地球によるポケットを同時に高速計算するアルゴリズムを開発した。多くの異なる大きさの球のことをマルチスケール球と呼び、それによって計算された分子体積をマルチスケール分子体積、マルチスケールポケットと呼ぶこととする。このアルゴリズムを用いることによって、ポケットの深さ、最小非接触半径  $Rin_{access}$  を定量化することが可能である。

### マルチスケール球を用いた深さ(狭さ)の指標



このアルゴリズムは「大きな地球で定義されたポケットは小さな地球で定義されたポケットを含む」という性質に基づいている。この性質は一見自明であるが、これまできちんと証明されたことはなかった。本研究では、モルフォロジー理論に基づくシンボル操作によりこの性質を厳密に証明することに成功した。

### CDK2(1b38A)に対して計算された結合ポケット

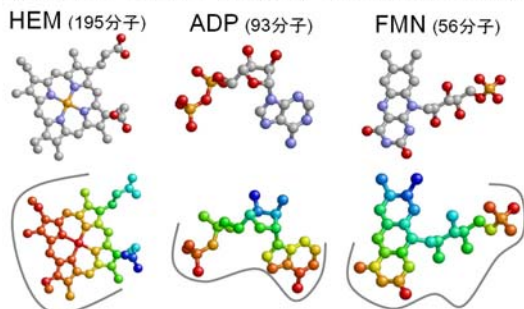


PHECOMで計算されたポケット。プローブ球の集合でポケットを表現。

本研究で新たに開発されたGHECOMで計算されたポケット。格子点の集合でポケットを表現。浅いポケットは青、深いポケットは赤で表示。

プログラム GHECOM を用いて既知の

### 結合分子内のポケット深さ(平均 $Rin_{access}$ ) の偏り



平均  $Rin_{access}$  による彩色 (赤: 深い、青: 浅い)

複合体立体構造データのポケット統計

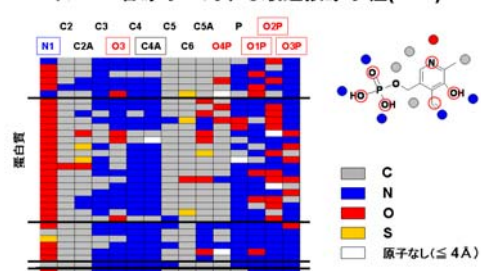
解析を行ったところ、分子種により固有の深さの偏好があること、同一分子種の各原子にも深さの偏好が見られ、分子の結合姿勢に偏りがあることが観察された。これらの知見は、結合ポケットの同定だけではなく、その「深さ」から、結合分子の絞り込みや、結合分子の結合姿勢の推定に寄与できることを示唆した。

本成果で開発したアルゴリズムについては、国内学会①において発表を行ったほか、既に *Proteins* 誌に投稿済みであり、現在審査中である。ソフトウェア **GHECOM** のプログラムソースは、<http://biunit.naist.jp/ghecom> において無料公開している。

### (2) 同じ分子が結合するポケットの比較解析

ポケットの特徴量から、それに結合する分子種を予測するためには、同一の分子が結合するポケットに何らかの類似性があることが前提となる。本研究では、多くの蛋白質での複合体構造データが入手可能な三種類の分子 (PLP, FAD, ADP) に注目し、それらの分子を取り囲む蛋白質の原子配置に有意な類似性があるかどうか検討した。その結果、相異なる蛋白質間では顕著に高い類似性が見られた。しかしながら、同一な分子が結合する非相異なる蛋白質間では、分子が直接結合している少数の原子の配置のみ、有意な類似性が見られた。この結果は、ポケットを構成する蛋白質の原子配置の類似性から、結合分子予測を行う場合、最近接原子からなる構造テンプレートの使用が有効であることを示唆する。以下に PLP の場合の最近接原子のパターンを示す。

### リガンド各原子に対する最近接原子種 (PLP)



非相異なる蛋白質の同一リガンド結合部位における類似性  
... リガンドと物理化学的に相互作用する蛋白質原子の空間配置<sup>31</sup>

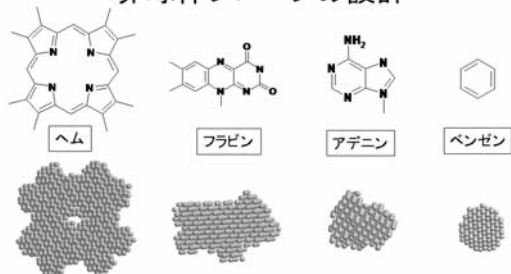
この成果は国内学会④において発表した。

### (3) 非球体プローブを用いた結合部位予測とポケットの特徴量を用いた結合分子種の予測

本研究では、プログラム GHECOM を拡張することによって、ポケットの形状を記述し、それを用いて結合する分子種が判別できるかどうかを試みた。GHECOM では、球状のプローブ

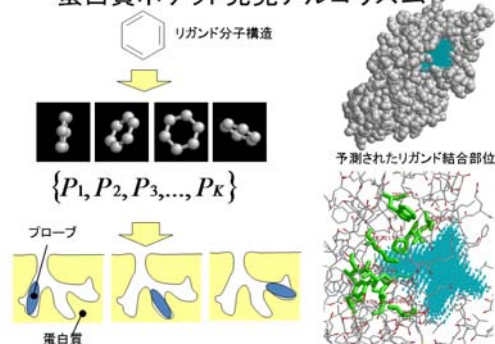
ブ球を用いたが、より実際の化合物に近づけるよう、本研究では非球状のプローブ球を用いた。具体的には、ヘム、フラビン、アデニン、ベンゼンの形状を模した4種のプローブを用いた。

### 非球体プローブの設計



非球状プローブを用いる場合、プローブの位置以外に回転自由度があるため、あらかじめ、回転させたプローブを多数用意し、これらを繰り返し適用し論理和をとることでポケットを抽出するようアルゴリズムを拡張した。

### 回転可能な非球体プローブを用いた蛋白質ポケット発見アルゴリズム



これらのプローブを用いて、まず結合サイト予測を行ったところ、予測性能を向上させることができた。また、これら複数のプローブを用いて抽出したポケットのグリッド数を特徴量として、結合するリガンドの種類が予測する線形判別分析を行った。その結果、ヘム、FAD結合蛋白質については、一定の判別性能が得られた。この成果は国内学会②において発表した。

### (4) 病因性の nsSNP と蛋白質のポケット部との関連の解析

アミノ酸置換を引き起こす非同義一塩基多型 (nsSNP) は、表現型に影響のない中立的な変異が多くある一方、遺伝性疾患に関与する変異も数多く報告されている。本研究では、Uniprot データベースに記載された中立性と病気に関連がある nsSNP について、アミノ酸変異のパターン、進化情報、ポケット性を含む立体構造の特徴を調べ、病気関連 nsSNP を予測するシステムの開発を行った。線形判別分析を用いて、様々な特徴量を試した結果、

アミノ酸出現頻度および溶媒露出度の二つが最も高い予測性を示すことがわかった。

**GHECOM** で計算されたポケット度も試したところ、有意な予測性能を示したものの、前述の二つの特徴には及ばなかった。このことは、蛋白質の機能に重要なポケット部が多く存在する一方で、そうではないポケット部も数多く存在することを示唆する。本研究結果は、国内学会③において発表された。

### 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 4 件)

- ① 川端 猛、モルフォロジーを利用した蛋白質表面の様々な深さのポケットを同時に計算するアルゴリズムの開発、日本生物物理学会第 46 回年会、2008.12.4, 福岡国際会議場
- ② 宮久保博幸、川端 猛、回転可能な非球体プローブを用いた蛋白質ポケット発見アルゴリズムの開発、第 8 回日本蛋白質科学会年会、2008.6.10, タワーホール船堀
- ③ 吉井悠喜、川端 猛、変異サイトの溶媒露出度、ポケット度、アミノ酸頻度等に基づいた nsSNPs の表現型に与える影響の予測、日本生物物理学会第 45 回年会、2007.12.21, パシフィコ横浜
- ④ 渡邊潤也、川端 猛、非相同な蛋白質の同一リガンド結合部位における原子配置の類似性解析、第 7 回日本蛋白質科学会年会、2007.5.24, 東北大学

[その他]

GHECOM プログラムの WEB ページのアドレス  
<http://biunit.naist.jp/ghecom>

### 6. 研究組織

(1) 研究代表者

川端 猛 (KAWABATA TAKESHI)

奈良先端科学技術大学院大学・情報科学研究科・准教授

研究者番号：60343274

(2) 研究分担者

(3) 連携研究者