

## 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 24 年 6 月 1 日現在

機関番号：32607

研究種目：若手研究（B）

研究期間：2010～2011

課題番号：22700314

研究課題名（和文） タンパク質相互作用パターン解析と予測法の開発

研究課題名（英文） Analysis of Protein-protein interactions and development of protein-protein docking method

研究代表者

寺師 玄記 (TERASHI GENKI)

北里大学・薬学部・講師

研究者番号：40383658

研究成果の概要（和文）：タンパク質間相互作用パターン解析と予測法の開発のため、タンパク質間相互作用部位データベースの作成、および三次元座標で表現された相互作用部位検索に使用する検索方法を新たに開発した。タンパク質の三次元構造を問い合わせにして、局所的な部分構造検索(3-D substructure search)の新たなアルゴリズムを開発し、既存の手法で最も探索速度の速い手法と比較して9～23倍高速な検索が可能となった。

研究成果の概要（英文）：In order to analyze the protein-protein interactions (PPIs) and new docking method, we have developed the PPIs database which consists the three dimensional coordinates of interacting sites and a new fast structure based search algorithm.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	500,000	150,000	650,000
2011年度	300,000	90,000	390,000
年度			
年度			
年度			
総計	800,000	240,000	1,040,000

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：生体生命情報学、バイオインフォマティクス、タンパク質間相互作用

## 1. 研究開始当初の背景

タンパク質—タンパク質間相互作用 (Protein-Protein Interactions: PPIs)の研究は、生物学的機能解析や医薬品開発にとって重要な課題である。タンパク分子同士がどのように相互作用しているのか、すなわち PPIs を予測する研究は数多くなされており、現在タンパク質立体構造解析の分野で新たに注目を浴びる分野である。ヨーロッパバイオインフォマティクス研究所は、PPIs の研究を早くから重要視し、2001年から世界の PPIs 予測手法の客観的比較を行う研究を Critical

## Assessment of Prediction of Interactions (CAPRI)

<http://www.ebi.ac.uk/msd-srv/capri/capri.html>を開催することで行っている。CAPRIに参加する研究者グループの手法は様々であり、

- (1) 原子の座標を格子点上に表現し、高速フーリエ変換を使用した物理学的エネルギー計算によるドッキング解析などの Physical based な手法、
- (2) タンパク質相互作用部位のアミノ酸傾向や PPIs 部位のデータベース化による解析と

いった知識知見に基づく Knowledge based である手法、

(3) アミノ酸配列検索情報からアミノ酸残基の保存性とドッキング構造の分子間相補性をスコア化した Physical based + Knowledge based の組み合わせを用いた手法。などがある。

研究代表者は、上記 Knowledge based エネルギー計算手法を開発する過程で、タンパク質間相互作用部位の中に特徴的座標パターンが存在する可能性があることに着目した。例として(図1) Protein Data Bank(PDB)に登録されたタンパク質 A と B の複合体(A-B)と、A と C の複合体(A-C)を比較すると、B と C の間にタンパク質構造全体での構造的類似性が観察されないが PPIs 部位に限りその相互作用パターンは同じである、というものである。このことから本研究代表者は、現在解析されているタンパク質複合体、及びタンパク質ドメイン間相互作用パターンを、座標パターンに変換後、データベース化し高速に PPIs 予測することを着想した。

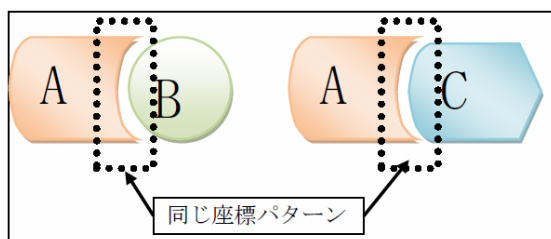


図 1

## 2. 研究の目的

本研究の課題は、「タンパク質間相互作用パターン解析と予測法の開発」である。現在解析されているタンパク質分子構造単体から、タンパク質複合体の相互作用状態を予測するシステムの開発を目指す。

複合体状態を予測するのに指標となるデータは、複合体状態が既知の三次元構造である。タンパク質間相互作用部位の座標情報を高速に検索できるようにデータベース化

このシステムにより、複数のタンパク分子からなるタンパク質複合体の相互作用状態が解析されていないものについて、コンピュータープログラムを使用し、既知のタンパク質複合体情報を検索し、高速かつ簡便にタンパク質—タンパク質相互作用の状態を推測することが可能となる。

## 3. 研究の方法

(1) タンパク質—タンパク質間相互作用データベース構築

本研究では、異なるタンパク分子間のタンパク質間相互作用情報だけでなく、タンパク質ドメイン間の相互作用情報も含めてタン

パク質—タンパク質間相互作用データベースの構築を行った。2012年現在の最新のドメインデータベース(SCOP1.75)からドメイン単位で、異なるタンパク質鎖間の相互作用と、同じタンパク質鎖内での異なるドメイン間相互作用部位の座標を抽出した。

(2) 三次元座標を問い合わせにしたタンパク質—タンパク質間相互作用データベース検索方法の開発

入力したタンパク質座標データと座標配置が類似しているタンパク質—タンパク質間相互作用部位データを検索するアルゴリズムの開発を行った。データベース検索には、アミノ酸配列非依存の座標群同士を高速に重ね合わせるシステムが必要である。本研究で、RMSD値が閾値よりも低い類似した局所座標を検索するアルゴリズムが開発された。タンパク質表面座標を問い合わせ情報とした、データベースの座標群の重ね合わせによるタンパク質—タンパク質間相互作用部位データベース検索システムが開発された。

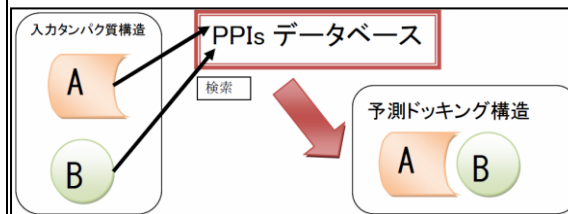


図 2

## 4. 研究成果

(1) タンパク質—タンパク質間相互作用データベース

ドメインデータベース(SCOP1.75)からドメイン間相互作用部位の座標を抽出した。これにより総計1700万座標以上で構成される相互作用座標パターンが抽出された。各相互作用座標パターンには、SCOPから得られるファミリー情報、PDB ID、次に記述される高速検索のための座標インデックスが付加されている。

タンパク質—タンパク質間相互作用データベース作成のためのプログラムが開発されたため、任意のデータベースを構築することが可能となっている。これにより、任意の相互作用部位の定義によってデータベースの構築が可能である。本研究では、アミノ酸残基のC $\alpha$ 原子間の距離が閾値以下である残基ペアを相互作用部位と定義している。

(2) 三次元座標を問い合わせにしたタンパク質—タンパク質間相互作用データベース検索方法の開発

入力したタンパク質構造の表面に存在する座標から、タンパク質—タンパク質間相互作用データベースの検索を行うアルゴリズムを開発した。(1)で作成されたデータベースに対して、類似した構造を高速に検索するのに重要なアルゴリズムである。本研究で開発された部分構造検索アルゴリズムの主な特徴は、任意の数の三次元座標を6つに分割し、そこから三次元座標群の特徴を、15のベクトルで表現するというものである。類似した部分構造を巨大なデータベースから検索するとき一般的に用いられる RMSD の計算は、計算コストがかかることが分かっている。そのため、RMSD の計算の代わりにベクトルの比較を行い計算の高速化を図っている。このベクトルの比較から計算される数値(Lower bound of a RMSD)は、最適な重ね合わせの結果得られる RMSD 値よりも必ず小さいという特徴を持っていることが本研究から明らかにされた。従って、二つの部分構造それぞれのベクトルの比較によって得られる数値が、指定した RMSD 値の閾値よりも大きい場合、その構造間の RMSD 値は閾値よりも必ず大きいことが高速に判断することが可能となった。

本研究で開発された部分構造検索アルゴリズムの計算速度比較の結果を図3に示す。図3は RMSD 値<1.0 の部分構造を検索した結果を図示した。横軸は問い合わせに使用する座標数(10, 20, 40, 60)を表している。縦軸は検索速度(1秒間に検索される相互作用部位の局所構造×100万)を表している。図1中の A3 は現実に利用可能な従来の類似部分構造検索法のなかで最も高速な手法である(Shibuya T. Searching protein 3-D structures in linear time. Journal of computational biology : a journal of computational molecular cell biology 2010;17(3):203-219.)。図3中の LB3D は、本研究によって開発された新たな類似部分構造検索法である。問合せ座標が少ないほど、RMSD 値が低いものが存在しやすいことから、問合せ座標数が少ないほど検索速度が低くなる傾向がある。本研究で開発された部分構造検索法は、図3に示されているように従来法(A3)よりも明らかに高速な検索が可能であることが示された。

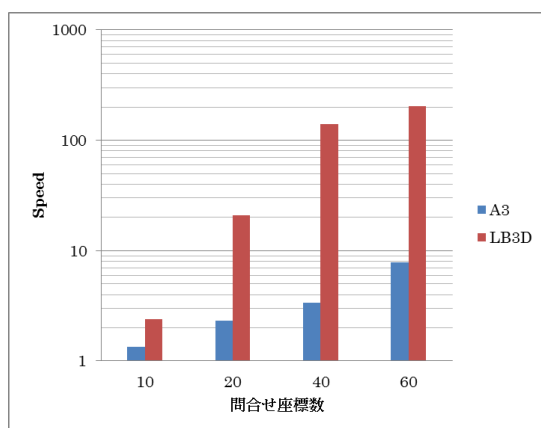


図3

図4は、100万部分構造の検索にかかる時間を縦軸にし、問合せ座標数を横軸で表したものである。グラフ中の Naive は、すべての部分構造に対して RMSD 値を計算する方法を表しており、N をデータベース中の部分構造数、m を問い合わせ座標数とすると  $O(Nm)$  時間の計算コストがかかる。図4にあるように、問合せ座標数が増大するほど計算時間が線形に増加している。一方、本研究で新たに開発された部分構造検索アルゴリズムは、 $O(m + N/m^5)$  時間の計算であることが解析された。Naive と比較して、問合せ座標数が増大しても計算時間が大きく影響を受けることがなく非常に高速な検索アルゴリズムであることが示された。

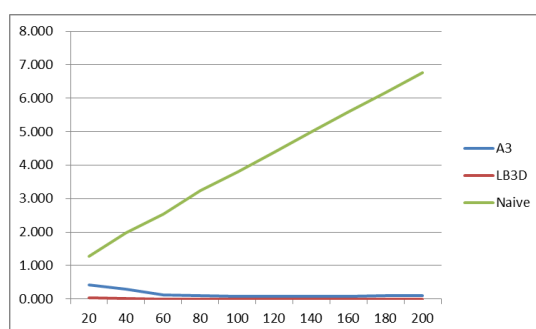


図4

詳細な部分構造検索アルゴリズムの計算手法、解析結果は、本研究により発表された論文「Terashi G, Shibuya T, Takeda-Shitaka M. LB3D: A Protein Three-Dimensional Substructure Search Program Based on the Lower Bound of a Root Mean Square Deviation Value. Journal of computational biology : a journal of computational molecular cell biology 2012;19(5):493-503.」に記載されている。

### (3) タンパク質複合体構造予測法の開発

本研究で開発されたタンパク質—タンパク質相互作用データベースと部分構造高速検索アルゴリズムを用いて、タンパク質複合体予測法の開発を行った。複合体状態が未知の2分子のタンパク質座標を入力し、その表面座標と類似した相互作用部位の座標パターンを検索する手法である。

図5はその適用例の一つである。2分子の入力座標(赤・青)から、タンパク質間相互作用データベースに対して検索を行った。その結果、タンパク質分子全体の構造は異なるものの相互作用部位の部分構造が類似しているタンパク質間相互作用の座標パターンが検出された(緑・シアン)。図5で示した例において複合体状態のタンパク質複合体

構造と比較を行ったところ、相互作用部位の主鎖のRMSD値が6.0Åであった。

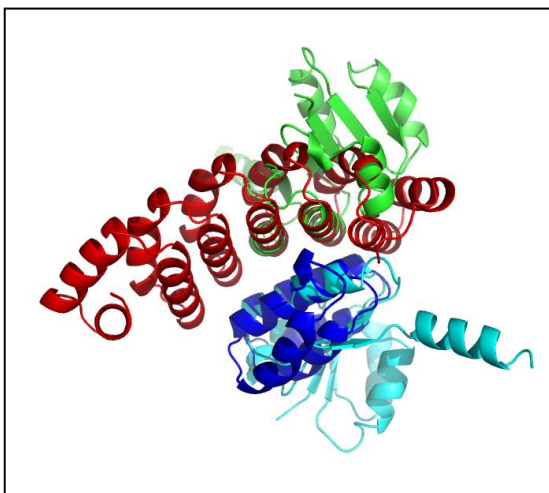


図5

本研究で開発されたタンパク質複合体構造予測法は、既知のタンパク質複合体の相互作用パターンに基づき、予測を行うことが可能である。しかし、既知のタンパク質複合体の相互作用パターンと類似した座標を有していても、必ずしも同様に相互作用しているとは限らない例が数多く存在している。これら偽陽性に相当する予測結果を判別する手法を開発することが今後の課題である。

#### 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)

Terashi G, Shibuya T, Takeda-Shitaka M. LB3D: A Protein Three-Dimensional Substructure Search Program Based on the Lower Bound of a Root Mean Square Deviation Value. *Journal of computational biology : a journal of computational molecular cell biology* 2012;19(5):493-503. 査読有  
DOI:10.1089/cmb.2011.0230  
<http://online.liebertpub.com/doi/abs/10.1089/cmb.2011.0230>

[学会発表] (計 1 件)

① 寺師 玄記、LB3D: Fast protein 3D substructure search program、2011 年日本バイオインフォマティクス学会年会、2011 年 12 月 8 日、神戸

[その他]

#### 6. 研究組織

##### (1) 研究代表者

寺師 玄記 (TERASHI GENKI)

北里大学・薬学部・講師

研究者番号：40383658