

構造プロテオミクスに基づく蛋白質間相互作用の解析と予測

●中村 春木¹⁾ ◆木下 賢吾²⁾ ◆塩生 くらら¹⁾

1) 大阪大学蛋白質研究所 2) 東京大学医科学研究所

<研究の目的と進め方>

蛋白質間の相同性という概念を蛋白質間の相互作用における相同性という概念に発展させ、さらに構造プロテオミクスによる立体構造情報を加えた解析を行って、各遺伝子産物の生化学的および生物学的機能の同定を行うことを目的とする。具体的には、複数の異なる生物種間で保存されたホモログ/オソログ蛋白質間の相互作用を同定し、Homologous Interaction Database (HINTdb) を構築・公開し、新しい相互作用の推定と機能未知の蛋白質の機能推定を行い、進化的情報を利用した Evolutionary Trace(ET) 法と蛋白質分子表面上の相互作用部位の特徴を基に、新たな蛋白質間ドッキング・シミュレーション法を開発する。さらに flexible ループの構造予測を行い、立体構造情報を加えた蛋白質間相互作用機序の高精度解析と複合体構造予測を行う。

<研究開始時の研究計画>

(i) 蛋白質間の相同相互作用を同定し HINTdb を構築・公開する。

種々の蛋白質間相互作用データを収集して相同性があるホモログを同定する。そして、ドメイン情報を加えて新たなデータベース (HINT db) を開発する。蛋白質の相同相互作用のデータベースとして立体構造上 fold に類似性が有意に認められるものについてはファミリーのメンバー候補と考えることにし、相同相互作用データベースに登録される件数を増やす。fold の類似性の抽出には、我々が最近開発した NER (Number of Equivalent Residues) を利用する。この HINTdb に登録された相互作用データに対して、その信頼性スコアを開発しスコアに応じた信頼度付の蛋白質機能アノテーションを行う。

(ii) ET法と蛋白質分子表面上の相互作用部位の特徴を基に、蛋白質間のドッキング・シミュレーション手法を開発し、立体構造情報を加えた蛋白質間相互作用メカニズムの解析と複合体構造予測を行う。

蛋白質間相互作用が発現される場合は蛋白質分子表面であることから、蛋白質分子表面の形状とそこでの物理化学的特性に注目して、従来より蛋白質分子認識機構の解析を行ってきた。具体的には、蛋白質機能部位における分子表面の形状と物性のデータベース (eF-site) の構築と、グラフ理論を基にしたクリーク探索法による分子表面の高速識別手法を開発・応用した。さらに、DNA 二重鎖の結合部位を、蛋白質分子表面の曲率と静電ポテンシャルの分布とから計算機に学習を行わせ、86% 以上の高精度で結合部位の推定を実現している。これらに基づき、ET 法を相互作用する蛋白質ペアの双方に対して同時に実施し、ペアの双方に共通な進化的特長を抽出して蛋白質双方の相互作用部位を同定する。複合体形成におけるスコア関数を用いて、2つの分子表面で記述された蛋白質ペアのドッキング・シミュレーションを行う。その際、flexible loop の構造アンサンブル予測を、GB/SA 法と Force-biased マルチカノニカル分子動力学 (McMD) 法を組み合わせて実施する。

(iii) 以上をまとめ、複合体立体構造が未知の複数の蛋白質に対し

ても蛋白質間相互作用を原子レベルで推定する一連のシステムを構築して応用し、データベース化を図る。

<研究期間の成果>

本年度の研究計画で述べた蛋白質間の相同相互作用データベース (HINTdb) を開発し、インターネットでアクセスできる Web page を公開した。また、それをもとにした相互作用の評価サイト (HitPredict) を公開した。ヒトの HPT な相互作用データも新たに解析・公開されたので、それらのデータについても解析を追加した。一般に、質量分析による解析は精度が高く 70% ほどの信頼性があるが、Y2H による解析は 30% 程度の信頼性しか得られないという傾向がある。

さらに、これらデータベースを利用して、蛋白質間相互作用ネットワークにおけるハブ蛋白質の構造的特徴を解析した。この際、5つ以上の異なる蛋白質と相互作用することが知られているものをハブ蛋白質、相互作用する相手の蛋白質として知られているものが1つのみの場合を仮にノンハブ蛋白質として、その構造的な差を解析した。ハブ蛋白質ではフレキシブルなループ構造がノンハブ蛋白質に比較して多いかと直感的に思われたが、統計的には逆の結果が得られた。一方、興味深いことに、ハブ蛋白質のアミノ酸配列から予測される Naturally disordered region および、その立体構造解析によって観測されない広い領域として考えられる Naturally disordered region が、ノンハブ蛋白質の場合に比較して有意に多いことがわかった。また、この現象は、大きなハブ蛋白質において観測される現象であり、小さなハブ蛋白質では Naturally disordered region を持つという特徴は顕著でなく、むしろ、多くの荷電残基をもつという特徴が有意に多く観測された。これらの知見は、ハブ蛋白質がどのように複数の異なる蛋白質と、それなりの特異的な相互作用をしながら分子認識をしているのか、という問題を解く鍵を与えていると思われる。

一方、ET 法を利用した蛋白質間のドッキング・シミュレーションを行うため、まず ET 法専用のサーバを構築し、インターネット上に公開した。2つの蛋白質にそれぞれ別々に ET 法を適用して、進化的に有意な機能部位の可能性があると思われる部位の表面をまず抽出し、ET 法のスコアに応じた重み付けを行う。次に、2つの蛋白質の表面同士が最も法線方向ベクトルが反対を向くように、ET 法スコアによる重み付けを掛け合わせたドッキングスコアを計算し、このドッキングスコアが最大になるように、モンテカルロ法と GA 法を利用して、高速かつ exhaustive な探索を行う。ドッキングスコアには、分子同士の衝突を避けるために、原子間の van der Waals ポテンシャルを模したエネルギーを加えておく。このようなアルゴリズムによるドッキング・シミュレーション・プログラムをほぼ完成させ、CAPRI に提出された問題とシグナル伝達系の蛋白質間相互作用の系に応用した。高速性を得るためには、ET 法のスコアが低い所については表面同士のドッキングスコアを計算せず、探索を減らす工夫をした。このようにして蛋白質複合体モデルを算出した結果、多くの場合に良好なドッ

キング結果が得られた。

さらに、flexible loop の構造アンサンブル予測を精度良く行うため、蛋白質分子のループ領域以外の所はフィックスしておき、ループ領域についてのみ自由度を与えて、そのフラグメントの構造多形を探索した。この際、ループ領域のペプチド・フラグメントの N 末と C 末の残基には、それぞれ position restraint を与え、構造に束縛を加えた。具体的な構造探索計算では、explicit な水分子を用いる計算よりも高速化を図るため、GB/SA 法を用いた implicit な溶媒モデルを用い、Force-biased McMD 法によって 300K から 700K までの広い温度領域でエネルギーが平坦となるマルチカノニカル・アンサンブルを作成した。このアンサンブルから 300K によるカノニカル・アンサンブルを再構築し、サンプリングしたループ領域のペプチド構造間の類似性を表す尺度から、主成分解析によって自由エネルギー地形を描いた。この手法をまず NMR と X 線結晶解析の双方で立体構造が得られている RNase A の 64-71 の 8 残基のループと 13-24 の 12 残基のループの双方について実施した。その結果、300K でのアンサンブルから得られた自由エネルギー地形は、それぞれ 3 つおよび 2 つのクラスターに分離されたが、それぞれのクラスターの中で最も主要なものは、X 線結晶解析および NMR 構造解析によるループ構造と類似したものであった。さらに、DHFR (ジヒドロ葉酸還元酵素) のリガンド・フリーの立体構造では、リガンド結合部位付近に存在する M20 ループ構造がフレキシブルで電子密度が観測されないため、M20 ループの構造が決められていない。ランダムな M20 ループ構造から出発し、上記のシミュレーション手法で 300K における M20 ループ構造の自由エネルギー地形を描いたところ、ほぼ 4 つのクラスターに分離された。そのうち 1 つのクラスターは、リガンド (葉酸) 結合時に観測される M20 ループ構造と重なったが、主要なクラスターはそれらのホロ体のループ構造とは異なっており、アポ体の主な構造であることが示唆された。さらに、各クラスターの中心構造を基に、リガンドとのフレキシブル・ドッキングを実行したところ、ホロ体の M20 ループ構造と重なるクラスターとのドッキングが良いスコアを与えることを見いだした。すなわち、この方法によって、リガンド結合状態のループ構造を推定し、リガンドとのドッキング計算により induced fitting が起こるシステムにも応用できることを示した。

<国内外での成果の位置づけ>

分子間相互作用を、分子表面に着目して解析し予測する手法は、国際的にも、我々が先手を打って進めており、低分子リガンドとの相互作用を解析するためのデータベース：eF-site については、*Curr. Opin. Struct. Biol.* 誌等でも紹介され、蛋白質機能部位のデータベースとして広く利用されている。蛋白質間相互作用への応用では、A. Olson による SurfDock というプログラムもあるが、必ずしも完成度は高くなく、多くの各論による蛋白質間相互作用に対して、様々なアプローチで対応する必要がある。

ペプチド・フラグメントを拡張アンサンブル法によって構造予測を行う研究は、我々のグループも含めて、特に β -ヘアピンに対して多くなされているが、ループ構造を McMD 法でモデル化する手法は我々のグループ以外にはほとんどなされていない。今回の計算手法は、計算時間の短縮のための GB/SA 法と、自動的にマルチカノニカル・アンサンブルを構築できる FBMcMD とを組み合わせたものであり、応用領域は広い。

<達成できなかったこと、予想外の困難、その理由>

既に CAPRI に参加しているが、必ずしも複合体構造の「予測」という意味では成功していない。その原因は、ET 法によって機

能部位と思われる部分を抽出する際に、複数の機能部位を持つ蛋白質の場合に、対象の蛋白質ペア以外の機能部位も一緒に抽出されてしまい、時に、他の部位の方が主要と見なされてしまうケースがあったためである。これについては、分子生物学の実験結果を念頭に置いて判断する者のバイアスもかかるため、ET 法による機能部位の抽出は、低い可能性でもそれを尊重する姿勢によって今後は改善されよう。

<今後の課題>

蛋白質間相互作用のデータベース HINTdb の情報をもとに、立体構造既知の単体の蛋白質構造から出発し、今回開発・応用したループ構造のモデリングを加えてホモロジー・モデリングの精度もあげることにより、複合体の立体構造が知られていない蛋白質ペアについて、その複合体の立体構造モデルを構築する。これらのある程度網羅的に行うことにより構造インタラクトームの研究を行い、分子認識機構を網羅的・俯瞰的に理解する。

<研究期間の全成果公表リスト>

- 0602011920 Watanabe, Y. S., Fukunishi, Y., Nakamura, H.: Generation of a flexible loop structural ensemble and its application to induced-fit structural changes following ligand binding. *BIOPHYSICS* 2, 1-12 (2006)
- 0602011932 Tsuchiya, Y., Kinoshita, K., and Nakamura, H.: PreDs: a server for predicting dsDNA-binding site on protein molecular surfaces. *Bioinformatics* 21(8), 1721-1723 (2005)
- 0602011909 Patil, A., Nakamura, H.: Filtering high-throughput protein-protein interaction data using a combination of genomic features. *BMC Bioinformatics* 6, 100 (2005)
- 0602011654 Shionyu-Mitsuyama, C., Ito, Y., Konno, A., Miwa, Y., Ogawa, T., Muramoto, K., Shirai, T.: In vitro evolutionary thermostabilization of congerin II: a limited reproduction of natural protein evolution by artificial selection pressure. *J Mol Biol.* 347(2) 385-397 (2005)
- 0602011537 Patil, A., and Nakamura, H.: HINT: a database of annotated protein-protein interactions and their homologs. *BIOPHYSICS* 1, 21-24 (2005)
- 0602011937: eF-site, <http://ef-site.hgc.jp/eF-site/>
- 0602011941: HINTdb, <http://helix.protein.osaka-u.ac.jp/hint/>
- 0602011946: HitPredict, <http://helix.protein.osaka-u.ac.jp/http/>
- 0602011949: PreDs, <http://pre-s.protein.osaka-u.ac.jp/~preds/>
- 0602011952: Evolutionary Trace Server, <http://pdbjets.protein.osaka-u.ac.jp/>