

タンパク質立体構造データベースに基づく分子間相互作用の解析

●川端 猛

奈良先端科学技術大学院大学 情報科学研究科

〈研究の目的と進め方〉

[平成14年度] 研究課題「立体構造を中心とした統合タンパク質データベースの開発」

本研究は立体構造のデータを比較・分類し、さらに、それにゲノム由来の配列データをリンクしたデータベースの開発を目的とする。完成したデータベースはWeb上で広く研究者に公開し、タンパク質の進化、物性、分子機能、さらには、より高次の生命現象を理解するための基礎となることを目指す。既に海外では、SCOP, CATH, FSSP等類似のデータベースがあるが、本研究では、既存のデータベースの長所を取り入れ、さらに次の2つの点で特徴を出していきたい。一つは研究代表者が開発した立体構造比較プログラムMATRAS (Kawabata and Nishikawa, 2000)の使用である。このプログラムは他の類似の比較プログラムと異なり、マルコフ連鎖の進化モデルを用いているため、進化的に関係のある立体構造とそうでない構造を分離する能力が高いとされている。このプログラムを用いることで、他の自動分類より生物学的意味の高い分類が達成できる可能性がある。もう一つは他の生物学の情報との積極的なリンクである。ゲノム配列、リガンド結合、タンパク質間相互作用など関係するデータベースを取り込んでいくことで様々な視点からタンパク質を考えることができるはずである。

[平成15年度] 研究課題「立体構造比較によるタンパク質機能解析用データベースの開発」

「構造ゲノム科学」プロジェクトに象徴されるように、近年、立体構造データが急速に蓄積しており、それらを生かしたタンパク質分子機能の理解するための方法論の整備が望まれている。本研究の目的は、膨大なタンパク質の立体構造の情報を、その分子機能の理解や推定に役立てるような手法やデータベースを開発することにあり、最終的にはWEB等の形で広く一般に公開することを目指す。

[平成16年度] 研究課題「タンパク質立体構造データベースに基づく分子間相互作用の解析」

本研究では、豊富なタンパク質立体構造データベースを生かして、その分子機能、特にタンパク質間およびタンパク質-低分子間の結合(相互作用)を理解し、相互作用が未知のタンパク質について立体構造からその機能を予測することを最終目的とする。「構造ゲノム科学」プロジェクトの推進により、機能未知の立体構造データの増加が予想され、構造からの機能推定の重要性は高い。また、幅広いタンパク質の機能の中で「他の分子との結合」は最も基本的で立体構造との関連が明確な機能であるといえる。従来、申請者は立体構造からの機能推定の方法としてタンパク質の立体構造比較の手法の開発に従事し、立体構造比較プログラムMATRASを開発し、WEB上で一般公開を行った。大域的な立体構造の類似は進化的な類縁関係を意味する場合が多いため、MATRASのような構造比較プログラムは構造からの機能推定の強力なツールになるが、基質特異性などより詳細な機能の推定

には不十分である。本研究では、まず1.既知の複合体の立体構造データの特徴解析を行う。このとき、申請者が従来研究を進めてきた進化的・博物学的な側面を重視しながらも、物性的・幾何学的な側面にも注目することで、より精密な分子機能の理解を目指す。そうして得られた特徴量を用いて、2.機能未知の単量体構造データからの分子間相互作用予測を行う。具体的には、それぞれの結合分子と関係の高い特徴量を用いて、結合サイトの予測、結合する分子種の予測、会合の強さの予測などをニューラルネットワーク等の統計的学習法を用いて行う予定である。

〈研究開始時の研究計画〉

[平成14年度]

- 1) 統合データベースの基礎となる立体構造比較の計算サービスを行うWEBサーバの整備
- 2) Matrasをフリーウェアとして配布する作業
- 3) 立体構造の比較分類結果をまとめたWEBデータベースの作成。各構造ごとに類似立体構造のリストとマルチプルアライメントを掲示。
- 4) 立体構造以外の情報として、リガンドの結合サイト、進化的な保存サイトの情報を含め、さらにタンパク質間相互作用にまで、拡大していくことを考えている。

[平成15年度]

- 1) [進化的側面] 全立体構造を比較分類し、相同なタンパク質をすべてマルチプルアライメントしたデータベースを作成する。その際に研究代表者らが開発したプログラムMATRASを利用する。
- 2) [物性的側面] 立体構造に由来する様々な物理化学的な特徴、たとえば、分子表面の凹凸解析、静電相互作用、基準振動解析などを計算し、分子機能との関連を調べる。
- 3) [複合体解析] タンパク質-タンパク質や、タンパク質-他分子の複合体の立体構造データの統計解析を行い、前述の進化的、物性的側面との関連を調べる。
- 4) [相互作用予測] 上記の知見を用いて、単量体の立体構造から、リガンド結合やタンパク質間相互作用等の分子機能を予測するプログラムを開発する。

[平成16年度]

まず、1.既知の複合体の立体構造データの特徴解析を行う。タンパク質どうし、タンパク質と他の分子(DNA、ヌクレオチド、金属等)の複合体の立体構造データを収集し、それらの結合サイト付近を様々な特徴量を計算し、それらと分子結合との統計的な相関があるかどうかを調べる。特徴量としては、以下の2つのアプローチを用いる。(1)進化的・博物学的側面：相同なタンパク質を全てマルチプル3Dアライメントを行いサイトごとの配列・構造の保存の度合いを計算する。機能的に重要なサイトはファミリー内で保存されている可能性が高いはずだ。3Dアライメントは申請者が開発した立体構造比較プログラムMATRASを用いる。(2)物性的・幾何学的側面：分子表面の凹凸の度合い、静電相互作用、基準振動解析

などを計算する。特に、化学反応を触媒する酵素においては、溶媒からの遮蔽、基質の自由度の固定のために凹部が結合サイトになる可能性が高いはずだ。また、電荷を持つ分子及び極性の大きい分子は、静電相互作用が結合に大きく影響を及ぼすと思われる。これらの特徴とそれぞれの分子の結合との統計的な相関を調べる全ての知見を用いて、2.機能未知の単量体構造データからの分子間相互作用予測を行う。具体的には、それぞれの結合分子と関係の高い特徴量を用いて、結合サイトの予測、結合する分子種の予測、会合の強さの予測などをニューラルネットワーク等の統計的学習法を用いて行う予定である。

〈研究期間の成果〉

[平成14年度]

- 1)WEBサーバは2002年夏から一般公開を始めた。アドレスは<http://biunit.aist-nara.ac.jp/Matras/>である。それ以来、少しずつ修正・機能拡張を続けている。現在、主な機能として、ペアワイズ3Dアライメント、マルチプル3Dアライメント、3Dライブラリ検索がある。ペアワイズ3Dアライメントは2つの構造の比較であり、ユーザーがローカルマシンからアップロードした構造を使用できる。マルチプル3Dアライメントは5つまでの構造を同時に比較する機能である。3Dライブラリ検索はユーザーが指定したクエリの立体構造をPDBあるいはSCOPの代表データセットと比較し、類似構造を返すサービスである。PDBの代表リストは毎週更新される。既に国内外から1日1件程度の頻度で使われている。
- 2)Matrasのソースプログラム配布については現在準備中である。英語マニュアルの作成に手間取っている。個人的に依頼があった場合にはβ版を配布している。
- 3)スーパーファミリごとのマルチプルアライメントのデータは既に作成しており、現在その精度・有用性について検討中である。
- 4)リガンド等の付加的な情報の付加については、リガンド結合部位の認識プログラム、リガンド結合表面積計算のプログラムを開発を終了しており、現在結果を検討している段階である。

[平成15年度]

- 1)SCOPの代表ドメイン構造約4000個について、総当りでMATRASで比較計算を行い、クラスタリングして巨大なデンドログラムを作成した。数千のリーフを含むデンドログラムの表示は困難であるため、JAVAアプレットを用いて、ブラウザ上でインタラクティブに木構造の階層を行き来できるツールを新規に開発し、MATRASのWEBサーバ上で公開を始めた。
- 2)タンパク質表面の凹状のポケット部に低分子は結合しやすいことを考え、タンパク質のポケット部を認識するプログラムの開発をほぼ終了した。プローブ球をタンパク質表面に配置することで、単純に幾何学的にポケット形状部を発見する方法である。この結果に関しては、来年度、積極的に研究発表を行うつもりである。
- 3)既知の複合体の立体構造データを解析することで、タンパク質—タンパク質間相互作用のインターフェースに現れやすいアミノ酸の傾向を調べた。疎水性アミノ酸、特に芳香族のアミノ酸が多く、電荷を持つアミノ酸の中ではアルギニンが例外的に多く現れる。この傾向を利用して、シンプルな相互作用サイトの予測プログラムを開発した。この結果に関しては、生物物理学会等で発表した。

4アミノ酸配列からの立体構造予測も依然重要なテーマであり、MATRASサーバにBLASTによるPDB配列への検索サービスの機能を追加するとともに、単純に残記名を入れ替えるだけの簡易モデリングの機能を付加した。また2次構造予測を利用してPSI-BLASTの感度を上げる方法を開発した。この結果に関しても、生物物理学会等で発表した。

[平成16年度]

- (1)立体構造比較サーバMATRASの維持と機能拡張
一昨年から維持している立体構造比較サーバMATRASの維持機能拡張作業を引き続き行った。機能拡張としては、PDBの鎖ごとの機能情報をまとめた"PDB Information"というページを作成した。これは、各PDBごとに、ヘテロ分子の結合サイトと、対応するSWISSPROTに含まれる活性部位などの情報が表示され、構造ベースの分子間相互作用の解析に役立つ。
- (2)タンパク質表面のポケット部分の発見
低分子がタンパク質に結合するサイトは一般に凹状のポケット形状をしているとされる。よって、タンパク質表面上に幾何学的なポケット形状を探すことは、低分子の結合サイトの予測の第一ステップとして重要である。我々は、2つの大きさのプローブ球を表面に配置することで、ポケット形状を認識するアルゴリズムを開発した。現在、実際のリガンドとの対応のチェックを行っている段階である。
- (3)タンパク質間相互作用サイトのアミノ酸出現傾向の調査

分子間相互作用予測の第一歩として、複合体立体構造のデータベースから、タンパク質間相互作用のアミノ酸の出現傾向を調べている。疎水性残基、特に芳香族、荷電ではArgが多いなどの傾向が観察された。この出現傾向指数と他の特徴を組み合わせた相互作用サイトの予測法の開発も進めている。

(4)2次構造予測による遠縁ホモログの認識能の改良

構造がわかっていない蛋白質の立体構造予測は、分子間相互作用の解析の前段階として重要なステップである。ホモロジーモデリングによる予測対象を拡張するため、本研究では、クエリのアミノ酸配列の予測2次構造を用いて、PSI-BLASTの出力のから相同である可能性が高いものを絞り込む手法を開発した。この成果は、出版され、WEBサーバPsiSecとして一般公開をしている。ユーザーがクエリ配列を投げると、サーバはプロフィール作成と同時に2次構造予測を行い、最新のPDBの配列に対してPSI-BLASTを実行したあと、その2次構造の一致度を計算し、E-valueと2次構造のジョイントスコアを出力する。

〈国内外での成果の位置づけ〉

[平成14年度]

立体構造の比較サービスを行っているサイトは欧米ではいくつか存在するが、国内では実稼働しているサーバはほとんどない。よってMatrasサーバを維持する重要性は高いと信じている。

[平成15年度]

我々が維持している立体構造比較サーバMATRASについて、Nucleic Acids Research誌に発表した結果(文献1)、MATRASの国際的に認知度が広がったと考えている。立体構造からの分子機能の解析の研究は、まだ初期的な段階に留まっているが、この方向性は「構造ゲノム科学」の研究の流れの中で極めて重要であり、国内外で同じ問題意識で取り組む研究者も多い、最もホットな構造インフォマティクス話題のひとつである。積極的に研究を

展開し、国際的な研究レベルに比肩できる成果を出して生きたい。

[平成16年度]

MATRASサーバは一日に十数件の実行依頼があり、この種のサーバとして確実に定着したと思われる。ポケット形状を発見するプログラムは、PHECOMと命名された。まだ投稿準備中であるが、いくつかの国内学会で発表したところ、相応の反響があった。

〈達成できなかったこと、予想外の困難、その理由〉

[平成14年度]

本年度はサーバの立ち上げやテーマの開拓に時間を取られて、論文発表につながる学術研究を達成することはできなかったのが反省点である。Matrasサーバについて"Nucleic Acids Research"誌のWeb Software Issueに投稿、現在reviseしている段階である。差読者は、原稿ではなく、Matrasの出力・見栄えに多くの注文を寄せていただいたため、原稿ではなくプログラムをreviseしている。製品を納品する前に検品を受けているプログラマのような気持ちである。プログラム・WEBサーバを使いやすく維持していく作業は、本当に"ガーデニング"のようなものだ。しばらく世話を怠っていると、虫がついたり、霜がおりたりしてどんどん腐っていく。外部の人間から客観的な意見をいただき、常により良いものに維持していくのが理想ではあるが大変なことだ。

[平成15年度]

立体構造比較については、ある水準まで達成できたと考えている。今年度は、特に、分子間の相互作用に注目して研究を進めた。低分子の結合に関しては、単純にポケット形状に注目することで、結合サイトのある程度予測できることがわかったが、タンパク質どうしの結合では、単純な形状から結合サイトを予測することは難しいことがわかってきた。タンパク質間相互作用は、その複合体形成の安定性、特異性などの性質が極めて多様であり、相互作用面の立体構造上の特徴も様々であることが原因であると考える。

[平成16年度]

分子間相互作用において、結合する相手が低分子の場合と、高分子の場合、おのずと考え方を大きく変えざるを得ないようだ。低分子の結合サイトは凹型のポケット形状で特徴づけが可能だが、タンパク質間相互作用では、凹凸、平面が混在した相互作用面を持ち、そう一筋縄ではいかない。

また、複合体立体構造データを基盤とした分子間相互作用の解析は、原子レベルの詳細な相互作用情報がわかるという長所がある反面、細胞内の膨大な分子間相互作用のごく一部しかデータでしかないこと、結合の強さ等の情報が欠如しているなどの限界がある。他の実験手法のデータを補完的に取り込む必要があり、現在プロテオミクス的な手法によるタンパク質間相互作用の実験データとの対応をとっている。

〈今後の課題〉

[平成14年度]

Matrasは完全に進生物学的な理論に基づく構造比較であるが、より物理化学的な視点の構造比較も重要であると考え、少しずつ準備を進めている。表面の凹凸の認識、簡単な弾性モデルの基準振動解析などのプログラムはほぼ完成しており、これらとMatrasの分類、あるいはリガンド、タンパク質などの相互作用面との解析などと組み合わせて、立体構造のインフォマティクスの新たな

局面を切り開いていきたい。

[平成15年度]

引き続き、タンパク質—低分子とタンパク質—タンパク質の結合についての研究を続けていきたいと考えている。低分子については、結合サイトの予測以上に結合する分子の特異性の発見につながる方向に研究を向けていきたい。タンパク質相互作用の扱いは大変難しいが、結合の安定性、特異性などの性質を考え、慎重に問題を設定することで、研究成果をあげていきたいと考えている。

[平成16年度]

ポケット形状部の発見アルゴリズムの論文化が意外に手間取っている。早く、脱稿して、このプログラムを生かしたポケット構造の形状・物性を調査する研究に取り掛かりたい。

〈研究期間の全成果公表リスト〉

- 0404092001
Kawabata, T., MATRAS : a program for protein 3D structure comparison, Nucleic Acids Research, 31, 3367-3369 (2003).
- Uehara K, Kawabata T, Go N. "Filtering remote homologues using predicted structural information". Protein Eng Des Sel., 17, 565-570,(2004).
- 川端 猛「立体構造比較サーバMATRASの使い方」 生物物理 Vol.45,No.1,41-44,2005.
- MATRAS：タンパク質立体構造の比較・検索を行うサーバ
<http://biunit.naist.jp/matras>
- PsiSec：2次構造予測の一致度の情報を使って、PSI-BLASTの感度を上昇させるサーバ
<http://biunit.naist.jp/psisec/index.html>