

研究開発課題別事後評価結果

1. 研究開発課題名

タンパク質の構造・機能予測法の開発とヒトゲノム配列への適用

2. 代表研究者名

太田 元規 (名古屋大学大学院情報科学研究科 教授)

3. 代表研究者による成果開発概要

3-1. 研究目的

DNAの担う遺伝情報はセントラルドグマに従いタンパク質配列として発現する。タンパク質は自発的な折り畳み反応によって固有の立体構造を形成するが、この立体構造がタンパク質の担う、特異的な分子機能の基盤となる。この一連の生命情報の流れ、つまりタンパク質の配列・構造・機能相関を解明することは、構造バイオインフォマティクスの主要課題である。現在、ドメイン単位での立体構造予測としてフォールド認識法が実用化されてきたが、長大な天然変性領域を含み、マルチドメイン構造をとる高等生物由来のタンパク質には対応しきれていない。また立体構造を通じたタンパク質分子機能の研究はまだ成熟しておらず、基質結合、構造変化、活性部位などの研究は個別に行われている。そのため、ゲノム配列決定を受けて直ちにその生物由来のタンパク質の構造と機能をインフォマティクスの力で列挙する、ということは未だ遠い未来像である。本研究開発では立体構造予測、分子機能予測に関する研究開発を基礎から応用まで幅広く推進し、その成果をアノテーションシステムとして統合する。それをヒトゲノム由来の配列に適用し、結果をデータベースとして構築、公開する。

3-2. 研究開発概要

上記目的を遂行するために以下の4グループからなる体制を組み、研究開発を実施した。

1) タンパク質の構造予測システム開発

高等生物由来のタンパク質の立体構造予測に関する基礎研究および応用研究を実施する。2004年に実施されたタンパク質立体構造予測コンテスト(CASP6)で生命情報工学研究センターのチームは成績優秀と認められた。その際、主要予測ツールとして利用したフォールド認識法:FORTEを中心に、ドメイン単位の構造予測法の精度向上を図る。タンパク質の天然変性領域予測法を研究開発し、ドメイン同定法に応用する。CASP7に参加し、予測法を評価する。配列比較でBLASTよりも高精度な確率的アラインメント法をプロフィール比較に適用し性能を評価する。

2) タンパク質の分子機能予測システム開発

タンパク質の分子機能予測に関する基礎研究および応用研究を実施する。タンパク質の立体構造データベース(PDB)を様々な視点から分類した二次データベースを作成し、タンパク質の相互作用と構造変化、基質結合に関する基礎研究を実施し、その成果をふまえて機能予測法を開発する。可能であればタンパク質複合体構造の予測コンテスト(CAPRI)に参加し方法を評価する。酵素反応データベースを拡充し、機能アノテーションに活用する。PDBから二次データベースを自動作成するシステムを開発する。

3) 実験検証

既に開発済みの、高速に発現系構築が可能な特殊なベクター：PRESAT-vector を利用して、タンパク質のフォールドの予測、ドメイン境界の予測、相互作用の予測について、試験管内結合実験と NMR 法で検証し、結果を各予測法の改良にフィードバックする。天然変性領域の予測を検証するための実験法を開発する。

4) アノテーションシステムおよびデータベース開発

本研究課題で開発した方法、これまでに開発してきた方法、また、他グループが開発した方法などを総動員して、高等生物由来のゲノムにコードされているタンパク質配列について、構造・機能アノテーションを付与するシステムを開発する。それをヒトゲノム由来の配列に適用し、結果を集積したデータベースを構築、公開する。

3-3. 研究成果概要

1) タンパク質の構造予測システム開発

FORTE と構造評価関数:Stability を組み合わせたドメイン構造予測システムを作成した。タンパク質全長にわたる天然変性領域予測法:POODLE-W を開発し、POODLE シリーズを充実させた。これらの方法で CASP7 に参加し安定的な評価を得た。確率的アラインメント法をプロフィール比較に適用し、従来法を凌ぐ方法であることを立証した。

2) タンパク質の分子機能予測システム開発

酵素の構造変化を調べ、加水分解酵素の構造変化は小さいが、転移酵素は大きいことを示した。同一ファミリーに属すが会合状態の異なるタンパク質の相互作用面のアミノ酸残基置換を調べ、会合状態を変化させる置換法則を見出した。時間や場に応じて相互作用の相手を変える社会的ハブタンパク質を調べ、その特性が構造の柔軟性にあることを見出した。CAPRI に参加し、評価を得た。酵素反応データベース:EzCatDB を拡充した。PDB の更新時に、構造変化、基質結合に関する二次データベースを自動生成するシステムを開発した。

3) 実験検証

古細菌タンパク質の立体構造を NMR で 2 件決定した。ヒト・マウス由来の 49 配列についてドメイン境界の決定を行い、2 件について NMR 構造を決定した。49 組のタンパク質間相互作用を検証した。改良型 PRESAT-vector と NMR を利用した天然変性領域の実験検証法を開発し、40 配列について発現を確認し NMR 測定を実施している。

4) アノテーションシステムおよびデータベース開発

POODLE と FORTE を組み合わせ、タンパク質配列から構造ドメインと天然変性領域を予測する構造予測パイプラインを作成した。これにより、高等生物のタンパク質について高精度の立体構造予測が可能となった。アポ体とホロ体で鑄型のあるタンパク質は、同時モデリングを実施することで、基質結合と構造変化を予測した。パイプラインにヒト 22 番染色体にコードされている 600 本ほどのタンパク質配列を入力し、総計 1, 025 個の立体構造モデルを作成した。アポ体のみモデルがある場合は、基質結合とホロ体を予測した。酵素反応、複合体モデルについてもアノテーションを付与し、データベース:SAHG を構築した。SAHG はマウスクリック主体のインターフェースを有し、立体構造上で構造変化、基質結合と反応がどう実現しているのかを視覚的に理解できる。

4. 事後評価結果

4-1. 当初計画の達成度

○当初の研究開発計画から見た進捗状況や達成度等

タンパク質の構造予測システムを開発し、従来よりも高い確率で予測できることを確認

した。また、これまで個別に実施された構造予測、分子機能に関する研究結果を集約し、構造予測パイプラインを核とした構造・機能アノテーションシステムとその結果を格納するデータベースが構築された。

○当初計画では想定されていなかった新たな展開

研究開発を通して、新たに発見された事をツールの機能に追加し、改良した。

4-2. 知的財産権、外部発表（論文等）等研究開発成果の状況

○知的財産権

開発した予測ツールを利用しやすく結びつけたアノテーションシステムやそれを用いた解析結果を格納したデータベースを構築した。

○外部発表（論文等）

論文発表だけでなく、構造予測やドッキング予測の国際コンテストに参加し、優秀な成績を挙げた。

4-3. 研究開発成果の公開による波及効果

○データベース、プログラムや論文、学会発表など、これまでの成果公開によるバイオインフォマティクス領域およびライフサイエンス分野全体への影響や効果について

データベースの公開がまだなされていないため、その影響や効果は未知数である。今後の早急なデータベースの公開が望まれる。

○国内外の類似研究開発の動向と比較して本課題の意義

構造・機能予測法の性能を上げ、従来よりも高い精度での予測を可能にした意義は大きい。予測精度が高いほど、その貢献度が高い。

4-4. 成果の実用化の可能性及び成果から予想される波及効果

○今後、成果の生命科学、研究基盤、産業への貢献度

開発されたデータベースやツールは、基礎から応用まで幅広い部分で活用され、構造予測研究における研究基盤となっていく可能性が高い。ゲノムワイドで構造・機能予測の解析が高性能で実現できれば創薬のターゲットになる可能性があるタンパク質等、有用なタンパク質の探索を加速できる可能性が高い。

4-5. 総合評価

研究開発計画通りの成果が見られ、バイオインフォマティクス研究の進展に貢献した。タンパク質の構造予測や相互作用等、焦点を絞った研究開発により、従来よりも高精度の予測システムを構築したことは評価される。今後、実験検証や利用者からのフィードバックが行われ、開発されたツールやデータベースがよりよいものとなっていくことを期待する。

5. 主な論文発表等

Koike, R., Kinoshita, K. & Kidera, A. Probabilistic alignment detects remote homology in a pair of protein sequences without homologous sequence information. *Proteins*, 66, 655-663, 2007

Higurashi, M., Ishida, T. & Kinoshita, K. Identification of transient hub proteins and the possible structural basis for their multiple interactions, *Protein Sci*, 17, 72-78, 2008

Kanamori, E., Murakami, Y., Tsuchiya, Y., Standley, D.M., Nakamura, H. & Kinoshita, K. Docking of protein molecular surfaces with evolutionary trace analysis, *Proteins*, 69, 832-838, 2007

Shimizu, K., Muraoka, Y., Hirose, S., Tomii, K. & Noguchi, T. Predicting mostly disordered proteins by using structure-unknown protein data. *BMC Bioinformatics*, 8, 78, 2007

Koike, R., Amemiya, T., Ota, M. & Kidera, A. Protein structural change upon ligand binding correlates with enzymatic reaction mechanism. *J. Mol. Biol.*, 379, 397-401, 2008

参考

論文発表 国内 25 件、海外 28 件

口頭発表 国内 50 件、海外 3 件

ポスター発表 国内 42 件、海外 8 件

公開されている成果プログラムおよびデータベース等

・ EzCatDB

<http://mbs.cbrc.jp/EzCatDB/>

酵素反応の触媒機構を階層的に分類する新しい枠組み (RLCP) を導入し、立体構造が解かれている酵素を分類したデータベース。現在、706 のエントリーが登録され、欧州バイオインフォマティクス研究所 (EBI)、ケンブリッジ大学の酵素関連データベース (Catalytic Site Atlas, MACiE)、大阪大学蛋白質研究所の PDBj と相互リンクを行っている。

・ POODLE (Prediction Of Order and Disorder by machine Learning)

<http://mbs.cbrc.jp/poodle/>

配列的特徴の異なる 3 種類 (配列全体、30 残基程度、数残基程度) の長さのディスオーダーについて、各々に最適な機械学習法を用いて予測を行うプログラム。世界的に権威のある CASP コンテストで高い評価を得ている。

・ classPPI

<http://pre-s.protein.osaka-u.ac.jp/~classppi>

PDB に登録されているホモタンパク質の相互作用部位のデータベース。SCOP での分類を利用し、冗長性を除いた 737 個のエントリーが登録されている。相補性の解析結果を閲覧することができる。

・ piSite

<http://pisite.hgc.jp>

タンパク質立体構造データベース PDB を利用して、タンパク質複合体の相互作用部位を網羅的に集めたデータベース。相互作用部位を集める際に、異なる複数のタンパク質と

相互作用出来るタンパク質の相互作用部位を同時に考慮している点が特徴。また副産物として、多くの異なるタンパク質と一時的な相互作用することが出来る Sociable protein のリストも提供している。110,325 タンパク質鎖が登録されている。

- ・ PreBI

<http://pre-s.protein.osaka-u.ac.jp/~prebi>

タンパク質複合体の立体構造を X 線結晶構造解析で決定した際には、タンパク質間相互作用面として、結晶学的な相互作用面と生物学的な相互作用面が存在する。これまでその決定は、個別に総合的な判断により決定されてきたが、PreBI では相互作用面の物理化学的な性質の相補性を利用して、生物学的な相互作用面を手助けする Web サーバ。