

プロフィール比較法の更なる発展

世界的規模のタンパク質立体構造予測コンテストで第3位

これまで開発を行ってきたプロフィール比較に基づくタンパク質の立体構造予測システムを更に発展させ、その性能検証のために昨年実施されたタンパク質立体構造予測コンテストに参加し、世界最高水準の評価を得た。今回のわれわれの成果は、実際的なタンパク質立体構造予測技術の発展と向上に大きく寄与したものとえよう。われわれの開発したシステムは、今後ゲノム配列データの生み出す大量遺伝子情報の解析に広く応用されていくものと考えられる。

ゲノム解析におけるプロフィール比較法のメリット

近年のゲノム配列データの急激な蓄積は、計算機によるタンパク質の構造/機能推定に対する切迫した需要を生み出している。予測対象のタンパク質と構造既知のタンパク質との配列類似性の検証は、実際的な立体構造推定の根幹をなすものである。そのため、配列からの情報抽出の効率が非常に重要な鍵となる。

本誌上で既報の¹⁾通り、これまでに我々は、より効果的な配列情報の活用と既知構造情報の有効利用を目指し、プロフィール比較と呼ばれる手法に基づくタンパク質の立体構造予測システム²⁾ FORTE1を開発してきた。FORTE1では、予測対象タンパク質と構造既知タンパク質の両方についてプロフィールが作成され、それら相互の比較を通して、明らかな配列類似性を欠く場合でも、タンパク質の立体構造類似性が検出される。ここでプロフィールとは、進化的に関連があると推測されるタンパク質配列群のマルチプルアラインメントでの各残基位置におけるアミノ酸種毎の出現頻度をスコア化したものであり、一般に配列データの増大につれてその信頼度も上

昇し、予測結果の質、量の向上に寄与するものと考えられる。また、構造ゲノムプロジェクトの進展による構造情報の増大につれ、構造予測におけるFORTE1をはじめとするプロフィール比較法の貢献もますます大きくなるものと考えられる。

比較法の発展とコンテストにおける成果

このような背景から、タンパク質立体構造予測において既知構造情報の利用を更に推進するために、FORTE1の技術を応用した複数のシステムの開発を行った。FORTE1システムは、様々な手法で作成されたプロフィールを無理なく使用できる柔軟性を持つ。この利点をいかし、FORTE1による立体構造類似性の検索結果を利用したFORTE2と、マルチプルアラインメント作成に別手法を使用したFORTE1Tという2手法を新たに構築した。FORTE2では、FORTE1によって検出された、より遠縁もしくは類似構造をもつタンパク質のアミノ酸配列情報をプロフィールに含めることで、より遠縁の構造類似性を検出することが期待される。FORTE1Tでは、より良い精度をもつマルチプルアラインメントがまず作成され、それらを種としたプロフィールを構築することで、より広範な構造類似性検出能力と、より正確なアラインメント作成能力を有することが期待される。FORTE1, FORTE2, FORTE1Tの3手法に加え、タンパク質の2次構造情報を考慮したFORTE-H、ローカルアラインメント用のFORTE-SSという2手法を更に準備し、これら開発した技術の性能の検証を行うためにCASP (Critical Assessment of Techniques for Protein Structure Prediction) と呼ばれるタンパク質立体構造予測コンテスト³⁾に昨年参加した。

CASPは、タンパク質立体構造予測法の発展および、

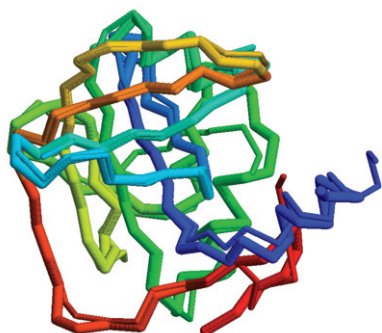


図1 CASP6においてFORTE法を用いて行った予測の一例 (PDB ID:1WGB)

太線は予測構造を示し、細線は実際の構造を示す。

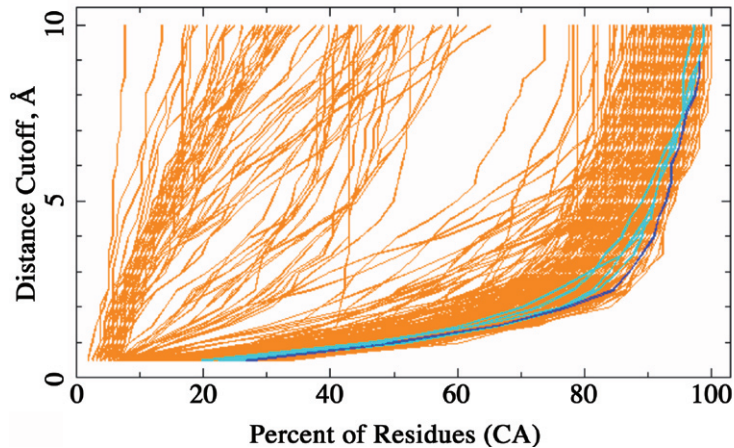


図2 図1の結果の他種法との比較

縦軸は予測構造と実際の構造との類似度の閾値を示し、値が小さい程類似度は高い。横軸は縦軸の閾値の範囲内で、全体の何%が重ね合わせ可能であるかを示す。実線が右下に傾くほど予測構造と実際の構造が一致していることを示す。濃青線が、図1の結果を示す。薄青線は、我々の提出した図1以外の結果を、橙色線は他種法の結果を示す。

予測法に対する客観的評価の場の提供を趣旨として、1994年から2年に一度開催されている。今回はその第6回目であり、2004年の6～8月の約3ヶ月間で、実験的に立体構造が解明される寸前の76タンパク質についてアミノ酸配列からの立体構造予測実験が、世界中の200を超える研究グループによって行われた。我々のグループは、FORTE1の技術を応用して開発した上述の5手法と、構造モデリングおよび検証ツールを統合したFORTE-SUITEシステムを用いて予測を行い、相同性を有する構造認識(Fold Recognition/Homology)と呼ばれる部門において第3位という高い評価を受け、12月4日から8日にイタリアで開催された結果発表会において招待講演を行った^{3)~5)}。フォールド認識部門では、構造既知タンパク質との類似性を配列相同性解析や構造認識技術などを用いていかに感度良く発見し正確にモデル構築するかが競われる。上位2チームは多数の他技術の予測情報を二次的に使うメタサーバーと呼ばれる方式であるため、単独技術としては世界最高の成績と言っても過言ではない。



図3 イタリアで開催された評価会議での招待講演の様子(富井)

プロジェクトの今後

構造ゲノムプロジェクトの本格的稼働によりタンパク質立体構造情報が急増している現状を考慮すると、実験と相補的な形でモデルを提供する我々の手法の利用価値は、例えば創薬の新規ターゲット探索やその検証、あるいはドラッグデザインにおける比較的高精度な予測構造の提供などといった様々な場面を通じますます増大し、ゲノム配列データの生み出す大量遺伝子情報の解析に広く応用されていくものと考えられる。現在FORTE1のウェブサービス^{6,7)}は、非商用の学術的利用に限り一般公開されている。

◆ 関連情報

- 1) 富井健太郎：AIST Today Vol.3, No.5, 19 (2003) .
- 2) 特許出願番号：2002-377704 「蛋白質立体構造予測システム」(富井 健太郎) .
- 3) <http://predictioncenter.llnl.gov/casp6/Casp6.html>
- 4) AIST Today, Vol. 5, No. 1, 35 (2004) .
- 5) プレス発表： http://www.aist.go.jp/aist_j/press_release/pr2004/pr20041202_2/pr20041202_2.html
- 6) K. Tomii, Y. Akiyama：Bioinformatics, Vol. 20, No. 4, 594 - 595 (2004) .
- 7) <http://www.cbrc.jp/forte/>

● 問い合わせ

独立行政法人 産業技術総合研究所 臨海副都心センター
生命情報科学研究センター
分子設計チーム 富井 健太郎
分子設計チーム 研究チーム長 広川 貴次
生体膜情報チーム 本野 千恵

E-mail : k-tomii@aist.go.jp

〒135-0064

東京都江東区青海 2-42