

配列情報からの蛋白質立体構造予測

近年のゲノム配列データの急激な蓄積は、蛋白質の計算機による自動的な構造/機能の推定、あるいは予測の必要性を増大させるだけでなく、実は、計算機を用いた配列情報からの立体構造予測に恩恵をもたらしている。

予測したい蛋白質と構造既知の蛋白質との配列類似性の検証は、立体構造推定の根幹をなす作業であり、配列からの情報抽出の効率がその成否の重要な鍵を握っている。強力な類似配列検索手法である PSI-BLAST¹⁾は、局所配列比較法を予測配列についてのプロフィールと逐次的に組み合わせることで配列情報を有効に活用し、ペアワイズの配列比較法をはるかに凌ぐ構造予測を可能にした。ここでプロフィールとは、相同であると推測される類似配列群での各残基位置における20種類のアミノ酸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、量も共に向上するものと考えられる。

ここ数年、より効果的な配列情報の活用のために、予測したい蛋白質と構造既知蛋白質の両方についてのプロフィールを作成し、比較を行うプロフィール比較と呼ばれる手法が盛んになってきている。今回新たにプロファ

イル比較による立体構造予測システム FORTE1を開発するに際しては、従来とは異なるプロフィール間類似性尺度の導入²⁾により、立体構造類似性の認識感度上昇による予測可能範囲の拡大とアラインメント精度の改良を目指した。FORTE1では、予測配列、構造既知蛋白質の配列の両方について、PSI-BLASTを利用して作成される配列プロフィールが用いられている。プロフィール作成には、構造情報を考慮するなど様々な方法が提案されているが、FORTE1ではゲノム配列データに後押しされた大量の配列情報を非常に有効に利用することで、他手法以上の結果も得られている。現在6千3百余の既知構造ライブラリの配列プロフィールは、大規模PCクラスター³⁾を利用して迅速に準備されており、データベースの日次更新も夢ではなくなってきている。またFORTE1は、作成法に係わりなく他手法で作成されたプロフィールも無理なく使用できる柔軟性も併せ持っている。これらは非常に強力な利点であるといえよう。

FORTE1は、計算機による自動的立体構造予測コンテスト⁴⁾への参加ならびにウェブサービスの試験的運用⁵⁾も行っている。



図1 FORTE1のwebインターフェイス

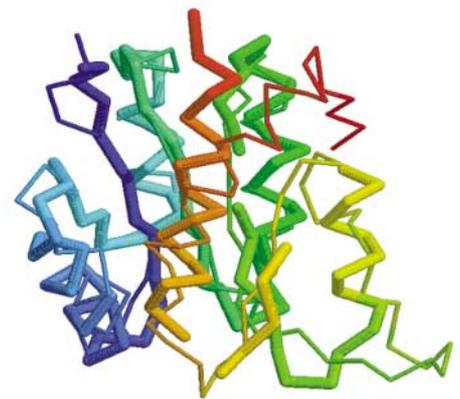


図2 FORTE1による予測の一例
太線は予測構造、細線は実際の構造を示す。



とみい けんたろう
富井健太郎
k-tomii@aist.go.jp
生命情報科学研究センター

関連情報

- 1) S.F. Altschul et al., Nucleic Acids Research, Vol.25, No.17 (1997).
- 2) 特許出願：2002-377704 「蛋白質立体構造予測システム」(富井健太郎)。
- 3) 秋山泰：AIST Today Vol.1, No.10, 15 (2001).
- 4) <http://bioinfo.pl/LiveBench/>
- 5) <http://www.cbrc.jp/forte1/>