

## タンパク質の立体構造プロフィールと極値分布を用いた構造認識法の開発

●太田 元規<sup>1)</sup> ◆富井 健太郎<sup>2)</sup>

1) 東京工業大学学術国際情報センター 2) 産総研生命情報科学研究センター

### 〈研究の目的と進め方〉

タンパク質の立体構造予測法の1つである構造認識法について、配列プロフィール、立体構造プロフィール、構造—配列適合性関数を利用することによって独自の方法を作り上げる。この際、自作のライブラリセットを利用した徹底的なベンチマークテストを行う。作成した方法について予測サービスを行うサイトを立ち上げる。立ち上げたサーバを利用し、自動立体構造予測のコンテンツに参加し評価を仰ぐ。

### 〈研究開始時の研究計画〉

配列プロフィールを用いた相同性検索であるPSI-BLASTでは、プロフィールとE-valueを利用することが精度向上に貢献している。配列プロフィールはタンパク質の構造情報を排除したものであるため、構造情報から作成された立体構造プロフィールを付加することで更に精度が向上する可能性があり、実際そのような報告もある。IMPALAというreverse-PSI-BLASTのプログラムを利用することで配列プロフィールをファイル出力できるので、立体構造プロフィールと混合し、それを再度IMPALAの入力とすることで、E-valueを利用する効果も保持される。

### 〈研究期間の成果〉

- 1) 立体構造プロフィールと配列プロフィールをブレンドした結果、3:7程度の混合率の時に効果的であることを発見した。この比率の場合、スーパーファミリーレベルの認識能が、配列プロフィールによるサーチに比べ若干向上した。ファミリー、フォールドレベルの場合、認識能にはあまり変化が見られなかった。構造プロフィールはスーパーファミリー程度のあらい類似性に対しては効果的な場合もあることがわかった。また、フォールドほど類似性が希薄になってしまうと現時点では認識することがほぼ絶望的ということも判明した。
- 2) プロフィールサーチの結果、E-valueが10以下の場合、適合性関数を利用してZ-scoreを求めて結果を再評価するシステムを作成した。
- 3) 混合プロフィールのサーチと適合性関数による再評価を実行するサーバを作成し、公開した。
- 4) サーバを利用して自動立体構造予測のコンテスト(CAFASP3)に参加した。
- 5) CAFASP3の結果はあまり良くなかったが、今後の課題について知見をうることができた。特に、配列プロフィールを作成する場合に、iteration回数を多くすることが安定的な解を得るためには重要であることが判明した。

### 〈国内外での成果の位置づけ〉

国際会議であるPacific Symposium on Biocomputingにおいてポスター発表を行い、本研究の成果について議論した。生命情報科学研究センターにて立ち上げた解析サーバには毎週5件ほどの解析依頼が来ている。

### 〈達成できなかったこと、予想外の困難、その理由〉

当初期待していたよりも、立体構造プロフィールと配列プロフィールの混合プロフィールが機能しなかった。サーチ法としてはE-valueの計算が容易なIMPALAを使っていたが、IMPALAではプロフィールとプロフィールの比較ができない。(複数の構造認識法の結果を総合的に判断して利用する、いわゆるメタサーバの研究を除けば)世の中はプロフィールとプロフィールの比較に研究の焦点が移動した結果、現在はほとんど全ての手法でプロフィール同士の比較が採用されている。この流れに初期過程で合流できなかった原因の1つはIMPALA (E-value)への固執が強すぎたせいであろう。同様に、“出来の良い方法を作る”という意味では、構造—配列適合性関数を利用することへの固執が強すぎたと思う。

### 〈今後の課題〉

現在、研究協力者が作成した構造認識法の出力モデルを評価する時に、構造—配列適合性関数のスコアを利用することを考えている。

### 〈研究期間の全成果公表リスト〉

#### 1) 論文

1. Tomii K, Hirokawa T, Motono C; Protein structure prediction using a variety of profile libraries and 3D verification. *Proteins*. 61, Suppl. 7, 114-121 (2005).
2. Ikeda K, Tomii K, Yokomizo T, Mitomo D, Maruyama K, Suzuki S, Higo J; Visualization of conformational distribution of short to medium size segments in globular proteins and identification of local structural motifs. *Protein Sci*. 14, 1253-1265 (2005).
3. Suenaga A, Takada N, Hatakeyama M, Ichikawa M, Yu X, Tomii K, Okimoto N, Futatsugi N, Narumi T, Shirouzu M, Yokoyama S, Konagaya A, Taiji M; Novel mechanism of interaction of p85 subunit of phosphatidylinositol 3-kinase and ErbB3 receptor-derived phosphotyrosyl peptides. *J Biol. Chem.* 280, 1321-1326 (2005).
4. Shiozawa K, Maita N, Tomii K, Seto A, Goda N, Akiyama Y, Shimizu T, Shirakawa M, Hiroaki H; Crystallographic characterization of the N-terminal domain of PEX1. *Acta Crystallogr D Biol. Crystallogr.* 60, 2098-2099 (2004).
5. Shiozawa K, Maita N, Tomii K, Seto A, Goda N, Akiyama Y, Shimizu T, Shirakawa M, Hiroaki H; Structure of the N-terminal domain of PEX1 AAA-ATPase. Characterization of a putative adaptor-binding domain. *J Biol. Chem.* 279, 50060-50068 (2004).
6. Tomii K, Akiyama Y; FORTE: a profile-profile comparison tool for protein fold recognition. *Bioinformatics*. 20, 594-595 (2004).
7. Tomii K, Ota M, Noguchi T, Akiyama Y; Construction of a fold recognition server, PILOT, based on PSI-BLAST, IMPALA and LIBRA\_rotamer, Pacific Symposium on

Biocomputing, 133, (2003).

2) ソフトウェア

FORTE: [http://www.cbrc.jp/htbin/forte-cgi/forte\\_form.pl](http://www.cbrc.jp/htbin/forte-cgi/forte_form.pl)

PILOT: [http://mbs.cbrc.jp/pilot/pilot-cgi/pilot\\_form.pl](http://mbs.cbrc.jp/pilot/pilot-cgi/pilot_form.pl)