

複数のタンパク質構造の重ね合わせ計算

タンパク質は、線状に連鎖された多数のアミノ酸からなる分子であり、様々な形状(立体構造)を有している。それらの立体構造を重ね合わせ比較することにより、進化的な近さや機能的な類似性を知ることができる。ところが、タンパク質は、進化の過程でアミノ酸が挿入、脱落されるため対応するアミノ酸は分からない。その上、立体構造は向きや位置もばらばらな状態で得られるため、単純な方法では重ね合わせることができない。そこで、進化的に遠いタンパク質の立体構造を比較するため、これまでにいくつかの方法が研究されてきたが、それらは発見的なアプローチで開発されてきており、得られる結果の物理的意味合いが不明確といわざるを得ないものであった。

本研究では、確率モデルによる立体構造の数理表現に焦点を当て、隠れマルコフモデル(HMM)によってタンパク質の形状を計算機上に記述した。ここで、モデルのパラメータと剛体変換(回転と平行移動)をいかに求めるかが課題となる。まずモデルパラメータが分かっていた場合から考えると、モデル上と立体構造上の対応点分かっている、それらすべてがユークリッド測度という統一的な計量(距離を測るための目盛り)で誤差を測ることができる場合、剛体変換のパラメータを容易

に求めることができる。しかし、ここでは、対応点も分からず、かつ、局所ごとに計量が異なる場合を考えなくてはならない。本研究では、モデルと立体構造との間に不均一な計量とユークリッド計量からなる階層的な構造を導入し、EM法と呼ばれる統計的手法に則って対応点と剛体変換を交互に求めることで確率的に最適な値を推定する方法を開発した。さらに、この方法を拡張することによって、複数の立体構造からモデルパラメータとそれぞれの剛体変換を同時に推定する算法を開発した。

図は、提案する方法を実際のタンパク質の立体構造に適用した結果である。これを見ると、それぞれのタンパク質の向きと位置を正しく推定し、それらの値から立体構造をほぼ完全に重ね合わせができていないことが確認できる。

立体構造の形状は確率モデルによって表現されているので、タンパク質の分類、事前知識の導入、欠損情報の補完など様々な応用や拡張が容易である。また、よりタンパク質の立体構造に特化したモデルを構築することにより、さらなる精度向上も期待できる。本研究はそれらの礎となる方法論を新たに創製したことになる。

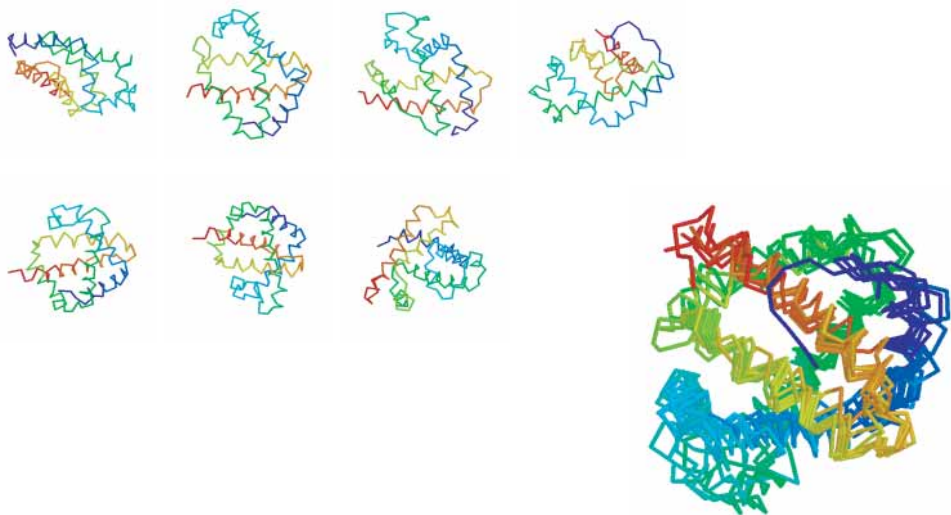


図 タンパク質立体構造の重ね合わせ

左図に示す7つの立体構造を最尤推定と呼ばれる確率的な推定手法を使って向きと位置を計算した結果、右下図のようにほぼ完全な重ね合わせが得られた。



かとう つよし
加藤 毅

kato-tsuyoshi@aist.go.jp
生命情報科学研究センター

関連情報

- T. Kato, K. Tsuda, K. Tomii, and K. Asai: CBRC Tech. Report, AIST02-J00001-2, 1-12 (2003).