

L1113-09	タンパク質における立体構造比較法の新規開発					
	氏名	鈴木 洋二	主査	黒田	副査	大野 ・ 長澤 ・ 稲田 ・ 津川

【はじめに】タンパク質の立体構造はその機能と密接なつながりを持っているため、立体構造を比較することはタンパク質を理解するうえで重要な課題である。近年、Protein Data Bank には22,000 以上ものタンパク質の立体構造が登録されており、その数は増加し続けている。これらの膨大な数のデータを活用するための計算ツールは必須のものである。

【目的】既存の立体構造比較方法よりも効率的な計算手法の新規開発を目的とする。

【アルゴリズム】本研究では、配列比較において最適なアライメント（類似した残基同士を対応づけること）を求めるための動的プログラミング法（DP 法）を構造比較に応用する。DP 法は異なる種類のアミノ酸間の類似性を比較することによって、アミノ酸配列をアライメントする手法である。そして DP 法はアミノ酸間の類似性の代わりにアミノ酸間の距離（ベクトル）を用いることで、構造比較を行うことができる。この手法は DP 法を拡張したダブル動的プログラミング法（DDP 法）で行われている。DDP 法は立体構造を比較するとき、計算精度は優れているが、その一方で計算量が多い。本研究は DDP 法とは全く異なる手法でベクトル、DP 法、スコア行列処理法（フィルタリング）を組み合わせ、より効率的な手法を開発する。

【結果】本研究は大きく分けて4つの段階からなる。

1. 2次元モデルでのアルゴリズムの設計

タンパク質の3次元立体構造は非常に複雑であり、その情報は膨大である。そのため本研究では、まず平面に簡単な2次元モデルを作成した。そしてこれらの2次元モデルにおけるスコア行列、DP法の計算結果などを解析することにより、アルゴリズムを決定した。

2. 3次元モデルによる計算精度の検証

1で開発されたアルゴリズムを用いて、3次元モデルによる計算精度の検証を行った。この3次元モデルは、計算結果の成否が判断しやすいように、実在するドメイン構造を切り出して単純化したものである。検証では分子ビューアーによるアライメント箇所を目視・RMSDの計算・SSAP(信頼性の高い計算ツール)との比較により、計算結果の成否の判断を行った。ここで得られた結果は構造比較における本研究の有効性を示した。

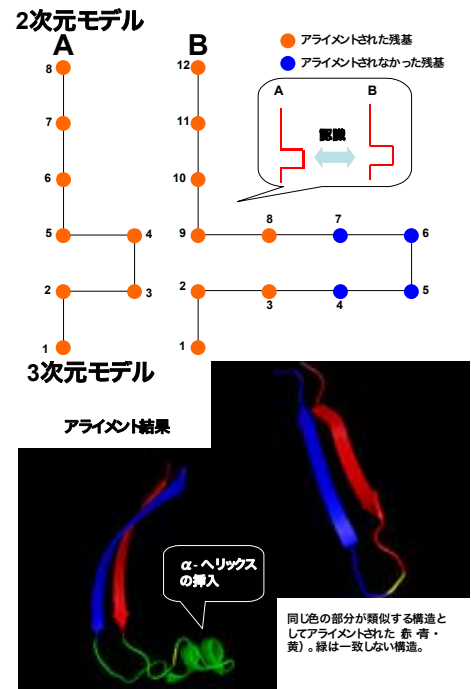
3. 実在ドメインによる計算精度の検証

実在するドメインの立体構造情報を用いて構造アライメントを行い、その計算精度を検証した。

4. 大規模計算による検証

CATH データベースを基準として、161 個のドメインからなるデータセットを作り、Topology レベルでの構造分類を行った(ROC 曲線)。そして Sensitivity:82.301%、Specificity:87.358% という結果を得た(resi_all)。resi_all における計算量を 1/9,1/25 に減らした場合の計算精度も計算した(resi_3, resi_5)。

以上の結果から、本研究により開発されたアルゴリズムはタンパク質の立体構造を比較する上で、有効な手段であることが示された。また、アルゴリズムを改善することにより、さらなる精度の向上、計算量の減少が期待される。



3-AにおけるROC曲線

