

グリッド計算機によるタンパク質foldingのバイオインフォマティクス

●太田 元規

東京工業大学学術国際情報センター

＜研究の目的と進め方＞

グリッドコンピューティングとは、ネットワークに接続された複数のコンピュータを利用した計算技術を指す。部屋にある情報コンセントに自分のパソコンをつなぐだけで、リモートにおかれたコンピュータの計算パワーやディスクの恩恵にあずかることができるような分散環境を目指している。グリッドコンピューティングの具体的な応用分野として、ナノテクノロジーとともにバイオインフォマティクスが注目されている。グリッドコンピューティングの特質である分散環境がデータベース管理などに有利に働くこと、また、タンパク質のfolding問題などには莫大な計算機資源が必要だと目されているからである。グリッドコンピューティングはバイオインフォマティクスのために構築された技術ではないが、計算機科学と融合することにより相乗的な発展が期待される。本研究はグリッド計算機を利用してfolding問題にアタックすることで計算ノウハウを蓄積し、バイオ／情報両分野の更なる発展に貢献することを目指す。

＜研究開始時の研究計画＞

グリッドコンピューティングはまだ新しい計算技術であるため、どの程度の威力があるのかを試す。そしてfolding計算を実行し、本当にfoldする軌道を与えることができるかを確かめる。成功の場合には詳細な軌道の解析が可能となる。

＜研究期間の成果＞

- 1) Trp-Cage という20アミノ酸からなるタンパク質について50ナノ秒の分子動力学計算を200本実施した。総CPU数が800台のTitechGridにCONDORというスケジューリングツールを通じて計算依頼を出した。200本のジョブは 短いものは2週間、長いものは1ヶ月で計算を終了した。総計算時間が8CPU年におよぶ計算を約1ヶ月で終えることができた。
- 2) 軌道中には実験で決めた構造とのずれの平均が1Å以内となる構造も多く含まれていた(図)。よってシミュレーションは無事成功した
- 3) フォールドする様子を詳細に解析するために軌道アラインメント法という新しい手法を開発した。軌道アラインメント法を利用すると軌道ペアの類似性が評価できる。類似性スコア（アラインメントスコア）をもとにして約90本のフォールディング、アンフォールディング軌道の樹形図を作成した。
- 4) 樹形図を考察することで以下の点を見出した。
 - ・ フォールディング軌道とアンフォールディング軌道を樹形図から区別することはできない。
 - ・ フォールディング構造には天然構造と側鎖の状態（ロータマー）が同一の、本当の天然構造と、ロータマーが異なる擬天然構造がある。
 - ・ 天然構造、擬天然構造のフォールド状態は、主鎖と側鎖の運動が異なる固有の経路をへて実現される。

＜国内外での成果の位置づけ＞

グリッドコンピュータに代表される計算機科学の進展によりタンパク質のフォールディングシミュレーションは現実味を帯びてきたが、得られた軌道の情報処理については適当な手法が存在していなかった。大計算を実行し、フォールディングを成功させたことに加え、軌道アラインメント法を開発し、フォールディングの詳細を解析したことの評価は極めて高い。

＜達成できなかったこと、予想外の困難、その理由＞

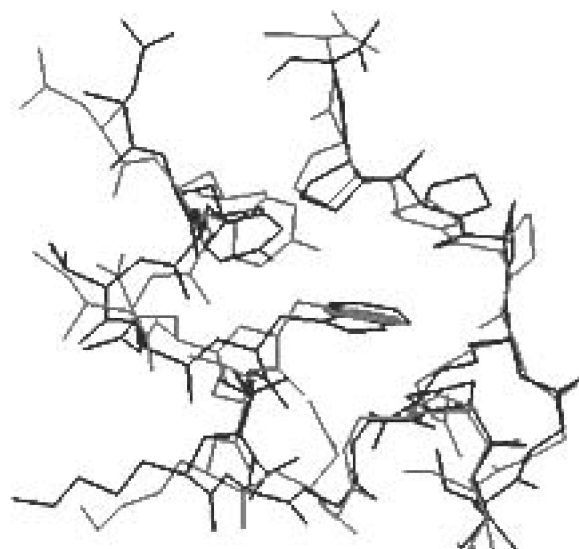
当初はともかくグリッドを使い、天然構造を再現することが念頭にあった。シミュレーションが成功してからしばらくは、軌道の処理法を見出すことができずに苦しい時期を過ごしたが、軌道アラインメント法の考案でその困難も突破することができた。成果は十二分に得られた。

＜今後の課題＞

軌道アラインメント法では処理が難しい大域的な構造空間中の運動の解析法を確立すること。Trp-Cage以外の、より大きなタンパク質で折りたたみを解明することがあげられる。

＜研究期間の全成果公表リスト＞

- 1) 論文
1. 412221357
Ota M, Ikeguchi M., Kidera A. Phylogeny of protein-folding trajectories reveals a unique pathway to native structure Proc. Natl. Acad. Sci. USA , 101, 17658-17663 (2004)



分子動力学計算で得られたfoldした形（黒）とNMRで得られた立体構造（灰）。主鎖のRMSDは1.00Å