

論文の内容の要旨

論文題目 タンパク質立体構造解析の高速化に関する研究

氏名 鈴木 正昭

タンパク質は 20 種類のアミノ酸がペプチド結合によって鎖状に連結した高分子であり、生理環境下に置かれるとある特定の三次元構造に折り畳まれる。タンパク質分子の生化学的機能はその三次元構造と密接に関係しているためタンパク質立体構造解析はポストゲノムの重要な研究課題であるが、近年の遺伝子解析により存在が明らかにされたタンパク質のほとんどはそのアミノ酸配列が知られているのみで、立体構造は依然未知である。現在最も広く用いられている構造解析手法は X 線回折結晶解析および NMR 解析であるが、それぞれタンパク質の結晶化および大量精製、精度等に問題を残しており万能ではない。

現在も続くコンピュータの著しい発展は、タンパク質の折り畳みシミュレーションによる立体構造予測を可能にし始めている。立体構造予測シミュレーション手法には、既知タンパク質の構造情報を利用する知識的手法と、アミノ酸配列の情報のみをもちいる非知識的手法とがあるが、類縁タンパク質をデータベースにもたない新規タンパク質の構造予測には非知識的手法が必要である。非知識的手法では、「タンパク質の天然立体構造はアミノ酸配列より一意に決定され、その構造は系の自由エネルギー最小状態に対応する」という Anfinsen らの主張に基づいて、主にモンテカルロ法、分子動力学法、遺伝的アルゴリズムを用いた最適化計算により天然立体構造探索が行われる。

タンパク質系は多自由度系であり膨大な準安定状態(自由エネルギー極小状態)が存在する。そのため、構造探索プロセスにおけるエネルギー極小状態へのトラップをいかに回避するかがシミュレーション成功の鍵である。効率的な構造サンプリング手法として、拡張アンサンブル法が近年注目を集めている。マルチカノニカル法やレプリカ交換法に代表される拡張アンサンブル法は、非ボルツマン因子に基づいたシミュレーションによってポテンシャルエネルギー空間上の一次元酔歩を実現し、エネルギー障壁を乗り越える手法である。特にレプリカ交換法は、現在のスーパーコンピュータの主流である分散メモリ型および SMP (Symmetric MultiProcessor) クラスタ型並列計算機への適用性が高い。

レプリカ交換法では、異なる温度を持つ互いに相互作用しないオリジナルの系のコピー(レプリカ)を複数用意する。それぞれのレプリカについてカノニカルアンサンブルシミュレーションを独立かつ同時に進めながら途中で温度を交換することで、ある一つのレプリカに注目すれば温度が酔歩(エネルギーが酔歩)するため準安定状態を抜け出すことができる。レプリカ交換法の問題点としては、系の自由度が大きくなるにつれて多くのレプリカを必要とする点が挙げられる。

以上のような背景のもと、本研究ではアミノ酸配列情報のみを用いた大規模タンパク質立体構造予測を目的として、レプリカ交換分子動力学法によるタンパク質立体構造予測コードを開発し、SMP クラスタ型ベクトル超並列計算機である地球シミュレータ向け最適化、および大規模系における従来のレプリカ交換法の問題点を克服する改良レプリカ交換法の提案を提案する。本論文は五章で構成される。

第一章では研究の背景および目的を述べる。

第二章では、拡張アンサンブル法としてレプリカ交換法を用いた、並列分子動力学計算によるタンパク質立体構造予測について概要を述べる。

第三章では、地球シミュレータ上で高いパフォーマンスを発揮するための最適化について述べる。立方体領域に一様分布した荷電粒子系による性能評価の結果、レプリカ数が 512、レプリカあたりの粒子数が約 11 万のレプリカ交換分子動力学計算において、2,048 プロセッサ使用時に約 92% の並列化効率を達成した。その際、3.9 TFLOPS の実行性能（理論ピーク性能の約 24%）が得られた。4,098 プロセッサ使用時の並列化効率は約 85% と見積もられ、高い並列性能を持つことを確認した。また、レプリカ数が 256、レプリカあたりの原子数が 52,143 である水中のタンパク質の折り畳みシミュレーションにおいて、2,048 プロセッサ使用時に約 88% の並列化効率を達成した。4,096 プロセッサ使用時の並列化効率は約 79% と見積もられ、実問題においても良好な並列性能を持つことを確認した。

第四章では、大規模系においても効率的なレプリカ交換計算を実現させるための、タンパク質のフラグメント分割に基づく改良レプリカ交換法を提案する。レプリカ交換法の問題点の一つに、系が大規模化（系の自由度が増加）するにつれてシミュレーションに必要なレプリカ数も増加する点がある。オリジナルのレプリカ交換法に対して、提案手法は部分構造毎のリファインメントを順次繰り返していき全体構造の最適化を図るものである。シミュレーションに必要なレプリカ数はフラグメントに含まれる自由度数にのみ依存することになり、必要レプリカ数を減らすことができる。提案手法の有効性検討のためのいくつかの数値実験の結果を示した。ポリアラニン(ALA)₁₆を用いて提案した改良レプリカ交換法とオリジナルのレプリカ交換法の比較を行った。(ALA)₁₆は比較的単純な系であるが、250K以下の低温の平衡分布を正しく得ることはそれほど簡単ではない。オリジナルREM、FREMおよび温度一定MDによる計算を実施し、そこから得られた 200Kにおける平衡分布を比較した。ここでは、提案手法を用いた場合にオリジナルのレプリカ交換法と比較して 1/4 あるいは 1/8 のレプリカ数で正しい平衡分布を得ることに成功した。

第五章では、本研究によって得られた結論を述べる。