

タンパク質立体構造予測時における全立体構造を表示するシステムの開発

瀬戸川 滋彰

Shigeaki SETOGAWA

1 はじめに

近年、タンパク質の立体構造予測が注目されている。タンパク質は生命現象に直接関わる重要な物質であり、その機能は構造に依存していると言われている。そのため、タンパク質の立体構造予測はタンパク質の立体構造に起因する病気の病理の解明、新薬の開発、特定の機能を持った人工タンパク質の設計など様々な分野への応用が期待されている。自然界に存在するタンパク質はエネルギーが最も低い安定した状態に折りたたまれているため、系の正しいエネルギー関数さえ与えられれば、タンパク質の立体構造予測はエネルギー最小化問題として取り扱うことができる。本研究室では「遺伝的交叉を用いた並列シミュレーテッドアニーリング (Parallel Simulated Annealing with Genetic Crossover:PSA/GAc)¹⁾」を用いて立体構造予測を行なっている。

しかし、現在の手法にはエネルギー関数の不完全さに起因する問題や計算に起因する問題が存在する。そこで、従来の計算機を用いた予測手法に加えて、人の目による評価、修正を行なうことでこれらを解決するシステムを実装する。本研究では、その前段階として予測中に生成された立体構造を Web 上で効率的に表示させるシステムを実装する。

2 従来の手法の問題点

従来の手法の問題点は、エネルギー関数の不完全さに起因する問題と、計算に起因する問題の 2 つに分類される。以下、それぞれについて詳しく述べる。

2.1 エネルギー関数の不完全さに起因する問題

系のエネルギー関数は様々なものが提案されている。しかし、自然界のものと完全に一致するエネルギー関数は存在しない。そのため、予測中の立体構造が、実際にはエネルギーが増加するであろう構造変化をしてしまうことがある。また、各エネルギー関数によってどのような構造になればエネルギー値がより低いと判断されるかが異なるため、エネルギー関数によって出やすい部分構造や出にくい部分構造が存在する。さらにどのエネルギー関数を用いても出すことが難しい部分構造が存在する。

2.2 計算に起因する問題

タンパク質の立体構造予測は計算量が多く非常に時間が掛かる。また、人が見れば少し構造を変化させればエネルギー値の低い構造になることが明らかな場合でも、計算で立体構造を予測させようとするとその構造に遷移するまでに時間がかかる、あるいは遷移しない場合がある。

3 最終的な目標システム

3.1 システム概要

PSA/GAc を用いての立体構造予測をする際に、並列毎、ステップ毎に生成された全ての立体構造をファイルに出力し、それを見やすい形で利用者に提示する。利用者が提示された立体構造を見て予測がどのように進んでいるかを判断し、場合によっては提示された立体構造に手を加えてそれを元に再び立体構造予測を開始させることのできるシステムを開発する。本システムのプロットを Fig.1 に示す。

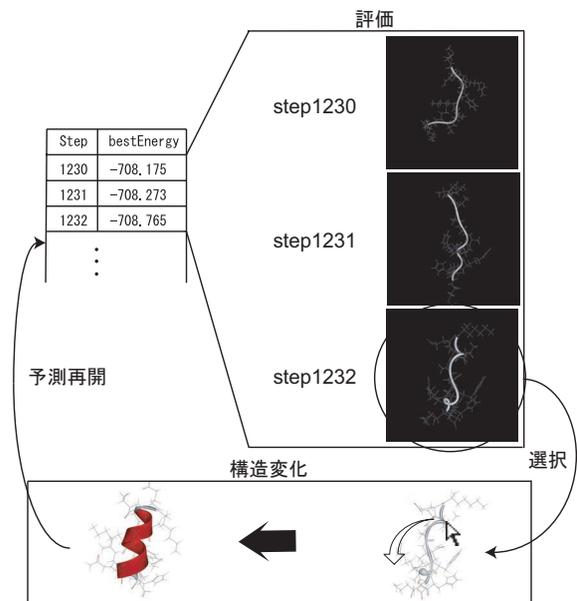


Fig. 1 システムの概念図

3.2 システムの機能

本システムに必要な機能を大まかに分けると、以下の 2 つに分かれる。

- 予測中に生成された全立体構造を表示させる機能
- 立体構造を任意に変化させる機能

このうち予測中の立体構造を表示する部分の実装を瀬戸川が、立体構造を任意に変化させる部分の実装を天白がそれぞれ担当する。

3.3 提案システムによるメリット

本システムをによるメリットは以下の通りである。

- 予測中に生成された立体構造を表示することで、探索の様子を人が見て判断することが可能となる。
- 人による修正が入ることで、従来の方法では出すことが難しかった構造を出すことが可能となる。
- 予測中の立体構造を修正することで、少ない計算量でより良い立体構造が得られる可能性がある。

4 生成された全立体構造を表示するシステム

4.1 システム概要

立体構造予測の最中に生成された全ての立体構造を、その探索の様子が容易に判断できる形で利用者に提示する。表示は Web 上でを行い、Ajax を用いてストレス無く表示できるようにする。

4.2 実装済みの機能

現在、本システムには HTML ファイル全体の再読み込みをせずに立体構造ファイルの表示、及び切り替えを行なう機能を実装済みである。現在の実行結果を、Fig.2 に示す。

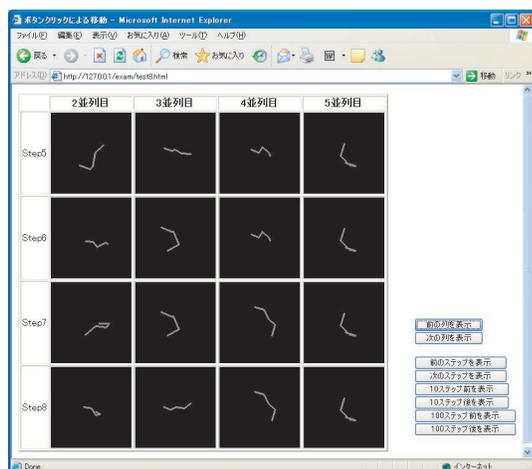


Fig. 2 実行結果

4.3 現行システムの問題点

現行の本システムではいくつか問題点があり、システム概要で述べた機能を満たすことが出来ない。現行システムの問題点を以下に挙げる。

- ストレス対策を施していない
非同期通信、ファイルの先読みなど、待ち時間を軽減するための対策を施していないため、大量の立体構造を見ようとする際には動作が遅くなる。
- 探索の様子の判断が難しい
生成される立体構造の数が非常に多いため、立体構造の遷移だけを見ても探索の様子の把握が難しい。

4.4 これから実装する機能

4.3 で挙げた問題点を解決するためにこれから本システムに導入する機能を以下に挙げる。

- Ajax を用いた立体構造の表示、及び切り替え機能
立体構造の切り替えを行なう際のストレスを軽減するために、Ajax を用いて非同期通信を行い表示や切り替えに伴う画面遷移を無くし、待ち時間を減少させる。
- 予測結果のグラフとの連動機能
立体構造予測結果のエネルギー値と RMSD 値という、生成された立体構造の評価に用いる値をグラフにしたものと本システムを連動させ、グラフ上の点をクリックするとその点に対応する立体構造を見ることができるようになる。立体構造の遷移とエネルギー値の遷移を比較することで、探索の様子の把握を容易にすることが可能となる。

5 今後の方針

本研究では、タンパク質の立体構造予測を行なう際に、従来の計算機を用いた予測手法に今回提案した人による評価・検討を行なう新しい手法を加えることで、従来とは違う観点からの立体構造予測が可能となる。今後は、4.4 で挙げた機能を実装し、さらに立体構造を任意に変化させる機能を持ったシステムと組み合わせて実際に立体構造予測を行なう。

参考文献

- 1) 廣安知之, 三木光範, 小椋真貴, 岡本祐幸.
遺伝的交叉を用いた並列シミュレーテッドアニーリングの検討. 情報処理学会論文誌: 数理モデルと応用, vol.43, No.SIG10(TOM7). 2002.