

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によるタンパク質立体構造予測
小椋 信弥

1 はじめに

自然に存在するタンパク質は、そのエネルギーを最小とするような立体構造を最適手法を用いて求めることで立体構造予測が可能である。これまでタンパク質立体構造予測においては、シミュレーテッドアニーリング (Simulated Annealing : SA) がよく使用されてきた。しかし、タンパク質のエネルギー関数は非常に複雑であるため、従来の逐次 SA では容易にタンパク質立体構造予測を行うことができず、立体構造予測のための新たな手法の開発が必要とされている。

そこで、本研究では、並列 SA と遺伝的アルゴリズム (Genetic Algorithm : GA) のハイブリッドアルゴリズムである、遺伝的交叉を用いた並列 SA (Parallel Simulated Annealing using Genetic Crossover : PSA/GAc) を提案する。タンパク質の持つエネルギー関数の特徴から、局所的な探索が得意かつ計算の高速化を図ることができる並列 SA に、大域的な探索が得意かつ部分解の組み合わせで最適解が得られる問題に有効な GA のオペレータを取り入れた本アルゴリズムは、タンパク質立体構造予測に対して有効であると考えられる。

2 遺伝的交叉を用いた並列 SA の概要

遺伝的交叉を用いた並列 SA (PSA/GAc) では複数の SA を並列の実行する。さらに、Fig. 1 に示したように、一定時間ごとの SA の解の伝達時に、GA のオペレータである遺伝的交叉を用いたものである。遺伝的交叉では、もとの親と生成した子との 4 個体間のうち評価値の高い 2 個体を選択して、選択された 2 個体から次の探索を続ける。ある設計変数の最適値が求まっている場合、遺伝的交叉によってその設計変数の最適値を他の SA 探索に伝達できるため、アニーリングの収束を早めることができる。これらの処理を繰り返し、終了条件を満たすまで探索を続ける。GA のオペレータを用いた SA であるため、SA の探索点を個体 (Individual)、探索点の総数 (SA の並列数) を個体数 (Population Size) と呼ぶ。

3 タンパク質立体構造予測

本研究では、構造予測の対象としたタンパク質は、Met-enkephalin である。Met-enkephalin は、5 個のアミノ酸残基 Tyr (Tyrosine), Gly (Glycine), Gly, Phe (Phenylalanine), Met (Methionine) からなる。Met-enkephalin は、 $E \leq -11kcal/mol$ の領域で最小エネルギー

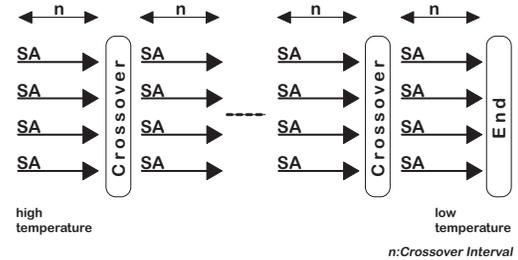


Fig. 1 PSA/GAc

構造をとることが知られている。Met-enkephalin の主鎖における 10 個の二面角と、側鎖における 9 個の二面角をそれぞれ設計変数とした。

本研究との共同研究を行っている、岡崎分子研の岡本先生らの逐次 SA で得られた結果と、PSA/GAc で得られた結果を Table 1 に示す。なお、Success Rate は、全試行数に対して、最適解が得られた試行回数の割合を示す。

Table 1 Comparison of Two Algorithms

Algorithms	Success Rate
PSA/GAc	0.90
Conventional SA	0.50

Table 1 より、最適構造を得る確率は、従来の SA を用いる場合よりも PSA/GAc を用いた場合の方が高いことが明らかとなった。Met-enkephalin の最適構造を Fig. 2 に示す

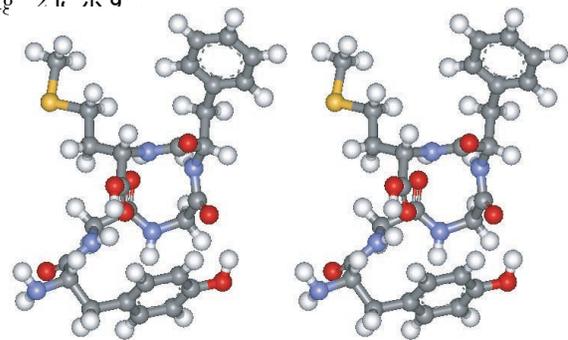


Fig. 2 Structure of Met-enkephalin

4 まとめ

本研究では、PSA/GAc を提案し、Met-enkephalin という小規模なタンパク質の構造予測を行った。その結果、従来構造予測に用いられていた逐次 SA よりも良い性能を示すことが明らかとなった。