

# 部分構造最適化の組み合わせによるタンパク質立体構造予測の提案

宇野 尚子  
Naoko UNO

## 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。タンパク質はアミノ酸が複数連なって構成される物質で、自然界ではある決まった構造に折りたたまれた状態で存在している。この構造をタンパク質の立体構造と呼び、タンパク質の機能と密接に関わっているとされている。そのため、立体構造を解明することによって病理の解明や新薬の開発につながる事が期待されている。

タンパク質はエネルギーの低い安定した構造に折りたたまれるので、エネルギー最小化問題と捉えることができる。本研究では「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)<sup>1)</sup>」を用いて立体構造予測を行っている。本報告では、タンパク質の立体構造予測の難しさと、PSA/GAc が抱えている問題点を挙げ、今後の研究方針について述べる。

## 2 タンパク質立体構造予測

### 2.1 エネルギー関数

目的関数は、タンパク質の系をモデル化したエネルギー関数を用いる。タンパク質のエネルギー関数は非常に複雑で、Fig. 1 のように大域的にいくつかの、局所的に無数の極小値を持つと考えられている。

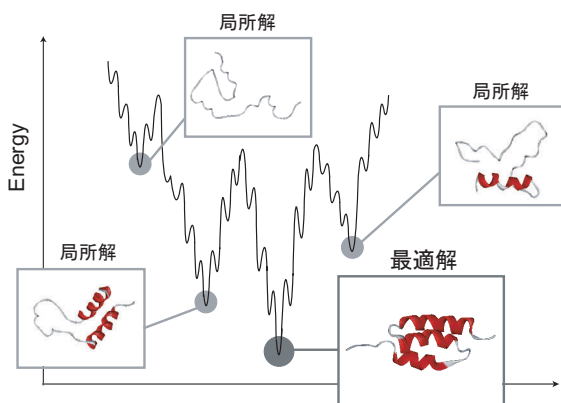


Fig. 1 エネルギー関数の模式図

本研究では、TINKER<sup>2)</sup> という分子動力学計算プログラムパッケージを元に名古屋大学の岡本先生が手を加えたものをエネルギー関数として使用する。設計変数は、原子間の回転角である二面角を用いる。

### 2.2 予測結果の評価基準

タンパク質の立体構造予測を行う際に、評価する基準は2つある。1つはエネルギー値、もう1つは構造の形である。構造が既知のタンパク質ならば、シミュレーション結果がどの程度その構造に似ているかが重要となる。

2つの構造の差異を定量化するために用いられる量がRMSD(Root Mean Square Deviation)である。RMSDは2つの分子構造を重ね合わせて、対応する各原子のずれの二乗を平均したものの平方根で定義される。式(1)にRMSDの求め方を示す。

$$RMSD(A, B) = \sqrt{\frac{1}{N} \sum_{i=1}^N (a_i - b_i)^2} \quad (1)$$

RMSDの単位はÅで、値が小さいほど2つの構造がよく似ていることになる。

## 3 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 (PSA/GAc)

### 3.1 概要

PSA/GAcはFig. 2のように並列に実行しているSAの解の伝達時に遺伝的アルゴリズムのオペレータである遺伝的交叉を用いたものである。

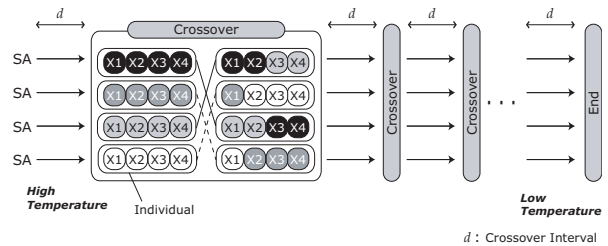


Fig. 2 PSA/GAcの模式図

このモデルでは、解の伝達時に並列に実行しているSAから親としてランダムに2個体を選択し、設計変数交叉を行う。そして、親個体と生成された子個体を合わせた4個体の中から良好な2個体を選択し、次の探索点とする。

### 3.2 PSA/GAcの性能

現段階でPSA/GAcで解くことができるのは非常に小規模なタンパク質に限られている。たとえば、10残基からなるChignolinや13残基からなるC-peptideなど

は、エネルギー最小化で十分良い結果が得られている。しかし、これより大きなタンパク質 (60 残基からなる Protein-A など) においては、いまだ良い結果が得られていない。ただ、全体としては天然構造に似ていなくても、部分的な構造は非常に良く似ている場合がある。これは逆に言うと、他の部分はうまく予測できるが、ある部分構造の予測はうまくいかないということである。そこで、実験でよく得られる天然構造に近い部分構造を固定し、それ以外の部分を設計変数として PSA/GAc で探索を行った。すると、それまで天然構造に近い解が得られなかった部分構造も比較的うまく予測できるということがわかった。これらの実験結果より得られた知見をもとに、新たな立体構造予測の方法について提案を行う。

## 4 部分構造最適化の組み合わせによる立体構造予測の提案

### 4.1 提案手法の概要

現在は完璧なエネルギー関数が存在しないという問題も含め、探索空間が広いために膨大な計算時間がかかるという問題もあり、大規模なタンパク質の立体構造予測は難しい。そこで、大きなタンパク質も小さな部分に分け、それぞれを別々に最適化するという考えを考える。これは前述のように、小規模なタンパク質であれば PSA/GAc で十分良い結果が得られていることから、大規模なタンパク質を扱うよりも良い構造が得られる可能性が高い。また、それぞれの探索結果から、ある部分で決まった構造が多く得られたならば、その部分を固定して、他の部分を最適化するということが可能である。

### 4.2 部分構造の作成

今回は 60 残基からなる Protein-A を大きく 15 残基ずつ 4 つの部分構造に分けた。また、分けたつなぎ目の部分を中心として、16 残基ずつの 3 つの部分構造に分け、合計 7 つの部分構造を用意した。その様子を Fig. 3 に示す。また、対象問題である Protein-A の Native 構造を Fig. 4 に示す。

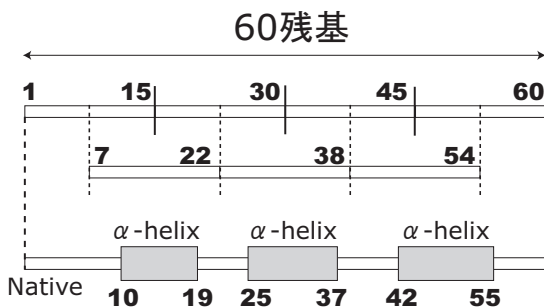


Fig. 3 部分構造への分割

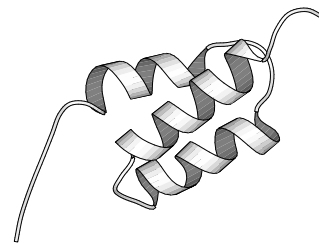


Fig. 4 Protein-A の Native 構造

### 4.3 実験結果

各部分を最適化した結果の一部を示す。Fig. 5 が 16 ~ 30 残基, Fig. 6 が 39 ~ 54 残基の部分構造である。左に Native 構造, 右に実験で得た最小エネルギー構造を示す。

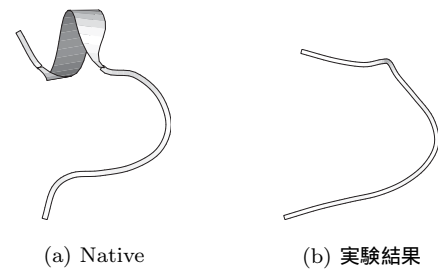


Fig. 5 16 ~ 30 残基



Fig. 6 39 ~ 54 残基

このように、ターンの部分も  $\alpha$  ヘリックスもうまく予測できていることがわかる。

## 5 まとめ

本報告では、部分構造最適化の組み合わせによる立体構造予測の提案を行った。実験により、部分構造の最適化がうまくいくことがわかったので、今後はこれらの分析と組み合わせを行い、大規模なタンパク質の構造予測を完成させたい。

## 参考文献

- 1) 廣安知之, 三木光範, 小掠真貴, 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: 数値モデルと応用, Vol.43, No.SIG10(TOM7). 2002.
- 2) TINKER Home Page . <http://dasher.wustl.edu/tinker/>