

# 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 using Genetic Crossover

小掠 真貴

Maki OGURA

**Abstract:** In this paper, we propose Parallel Simulated Annealing using Genetic Crossover for solving protein structures. Protein have an energy function there are many local solutions, so present algorithm cannot solve protein structures. In case of using the proposed algorithm for the continuous test problems, it is found that this algorithm can search the solution effectively. So, we expect that the proposed algorithm be effectively for solving protein structures.

### 1 はじめに

昨年度の研究では、数値計算を用いてタンパク質の構造を解析するため、新たなシミュレーテッドアニーリング (Simulated Annealing : SA) の手法を提案した。

タンパク質は生物のさまざまな生命現象を直接担っているという点で重要な物質であり、その構造を解明することは生命現象の仕組みを解明することへとつながっている<sup>1)</sup>。実験的手法では構造を解析することが可能であるが、解析できるタンパク質の量や規模に限界があるため数値計算による解析が必要となる。タンパク質の構造同定は最適化問題の一つとして考えられ、自然なタンパク質の立体構造はエネルギーの最小状態に対応している。数値計算を用いた構造解析では SA を中心に研究が進められているが、そのエネルギー関数は大域的に幾らかの、局所的には無数の極小値を持っていることから、現在の手法では探索が困難である。

この問題点を解決するにあたり、局所的な探索が得意な SA に、大域的な探索が得意で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である遺伝的アルゴリズム (Genetic Algorithm : GA) のオペレータを取り入れることで、タンパク質の構造同定に対して有効な手法を提案できると思われる。昨年度の研究では、SA に GA のオペレータを取り入れた手法として遺伝的交叉を用いた並列 SA を提案し、連続関数最小化問題に適用してその有効性を検証した。

### 2 並列シミュレーテッドアニーリング

SA は最適化問題を解く有効な手段であるが、最適解を得るためには非常に多くの計算量を必要とするという欠点が存在する。そのため並列計算機が利用可能となるに伴い、処理負荷の重い SA も並列処理の研究対象となった<sup>2)</sup>。しかし、SA はマルコフ連鎖を次々とたどる処理であるため、本来強い逐次性があり並列化は容易ではない。そのため並列 SA では、計算の効率化と解品質

の向上のどちらかを犠牲にすることが多い。

SA ではマルコフ連鎖をたどり無限に冷却を行うことで、理論上は探索が最適解に到達するため、大半のモデルはマルコフ連鎖が分断されないように並列化されている。しかし計算機で SA を実装するときには探索時間が限られており、限られた時間の中では理論を満たす探索は不可能に近い。マルコフ連鎖を分断しないことに注意しすぎるため、かえって探索の速度向上を妨げているアルゴリズムもあるといえる。速度を向上させ、また得られる解の品質を向上させるため、本研究ではマルコフ連鎖を分断した並列 SA を提案することとした。

### 3 昨年度の研究の成果

本研究では、遺伝的交叉を用いた並列 SA を提案し、連続値最小化問題を対象として提案手法の性能を評価した。

#### 3.1 遺伝的交叉を用いた並列 SA の概要

提案手法は Fig. 1 に示すように、一定間隔のアニーリングを繰り返し、遺伝的交叉を行って最適解を求めようとするアルゴリズムである。GA のオペレータである遺伝的交叉を用いた SA であるため、SA の探索点の総数 (並列数) を個体数、アニーリングステップ (計算繰り返し回数) を世代数と呼ぶこととする。

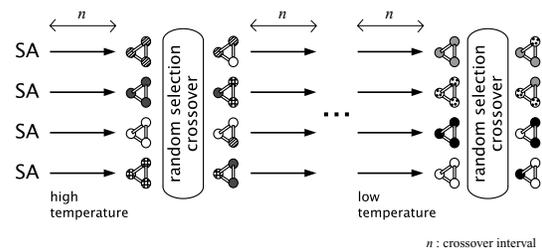


Fig. 1 Simulated annealing with genetic crossover

このモデルでは、解の伝達時に並列に実行している SA からランダムに親として 2 個体を選択し、設計変数

間交叉を行う。もとの親と子との4個体のうち評価値の高い2個体を選択して、この2個体から次の探索を続けるという方法である。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探索に伝達することができるため、アニーリングの収束を早めることができる。

### 3.2 遺伝的交叉を用いた並列SAのアルゴリズム

遺伝的交叉を用いた並列SAのアルゴリズムにおけるクーリングまでの過程は、逐次SAと同様である。逐次SAの処理を並列に実行し、ある一定間隔で遺伝的交叉を行って、終了条件に達すれば探索を終了する。アルゴリズムをFig. 2に示す。

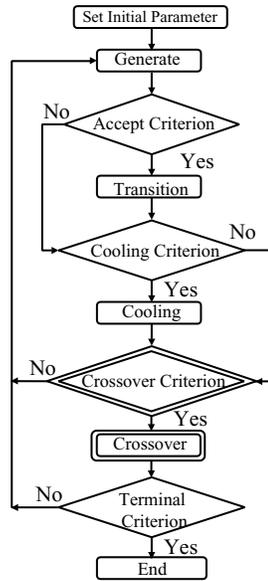


Fig. 2 Algorithm of SA with genetic crossover

実装した並列SAでは、各過程においてアルゴリズムは以下のように決定した。生成処理では、近傍の範囲に依存する一様分布を用いた。近傍の範囲の決定には、受理率によって適応的に大きさが調節されるというCoranaのSA<sup>3)</sup>のアルゴリズムを用いた。受理判定にはMetropolis基準を用いて、次の状態を受理するかどうかを決定した。クーリングには指数型アニーリング $T_{k+1} = 0.93 \cdot T_k$ を用い、次の状態を受理した回数が20回に達するごとにクーリングを行った。また、世代間隔が32になるごとに、並列に実行しているSA間で遺伝的交叉を行い、終了条件を満たしていれば探索を終了することとした。

### 3.3 遺伝的交叉を用いた並列SAの利点

遺伝的交叉を用いた並列SAを通常の逐次SAと比較すると、アニーリングの高速化、解品質の向上という利点がある。この要因として以下の3つのことが挙げられる。まずSAの探索を並列化したことである。並列化することで探索点が増えて収束が早くなり、また探索範囲

が広がって高品質な解が発見できる。次に近傍の範囲を調節するCoranaのアルゴリズムを用いたことである。受理確率を常に50%にすることで無駄な探索が減り、効率よく探索が行える。最後は、GAのオペレータである遺伝的交叉を用いたことである。設計変数ごとに最適値がある問題において、1つの個体がある最適値を見つけていると、遺伝的交叉によってその最適値が他のSA探索に伝達され、最適解の発見が早くなる。また、遺伝的交叉を行うときには評価値の高いものが選ばれることも、解の収束を早め、解品質を向上させることにつながっている。

### 3.4 数値実験

遺伝的交叉を用いた並列SAの有効性を検証するため、遺伝的交叉以外のGAオペレータを解の伝達方法として用いた3つの並列SAと解探索能力を比較した。2設計変数の連続値最小化問題を対象としたとき、提案手法である遺伝的交叉を用いた並列SAの解探索能力が最も優れていることが示された。

次に設計変数を増加させた問題を対象とし、分散GA、逐次SAとの探索能力の比較を行った。この実験結果からは、逐次SAはアニーリングの時間を十分に長くしたり回数を増加させたりするだけでは、遺伝的交叉を用いた並列SAほどの結果が得られないことが示された。また、GAでの探索が困難な設計変数間に依存関係のある問題に関しては、遺伝的交叉を用いた並列SAの探索が有効であることが示された。

## 4 結論と今後の課題

本研究では、タンパク質の構造同定に有効だと思われる手法を提案した。数学的テスト関数を用いた数値実験において提案手法の有効性が示された。今後はタンパク質の立体構造のエネルギー関数に対して遺伝的交叉を用いた並列SAを適用し、現在構造同定に用いられている手法よりも良いふるまいをし、良好な解が得られることを確認する。

### 参考文献

- 1) 池内俊彦. 生命を学ぶ タンパク質の科学. オーム社, 1999.
- 2) Bruce E. Rosen, 中野良平. シミュレーテッドアニーリング -基礎と最新技術-. 人工知能学会誌, Vol. 9, No. 3, 1994.
- 3) M. Marhesi A. Corana, C. Martini, and S. Ridella. Minimizing multimodal functions of continuous variables with the 'simulated annealing' algorithm. *ACM Trans. on Mathematical Software*, Vol. 13, No. 3, pp. 262-280, 1978.