

遺伝的交叉を用いた並列 SA によるタンパク質立体構造のエネルギー最小化  
小椋 信弥

1 はじめに

本研究では、タンパク質のエネルギー最小化のための新しい手法として、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で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のエネルギー最小化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シミュレーテッドアニーリング (Parallel Simulated Annealing using Genetic Crossover: PSA/GAc) は、並列に実行している各 SA の解の伝達時に、GA のオペレータである交叉を用いた最適化手法である。本研究では、GA のオペレータを用いた SA であるため、探索点を個体と呼び、SA の探索点の総数 (並列数) を個体数と呼ぶこととする。

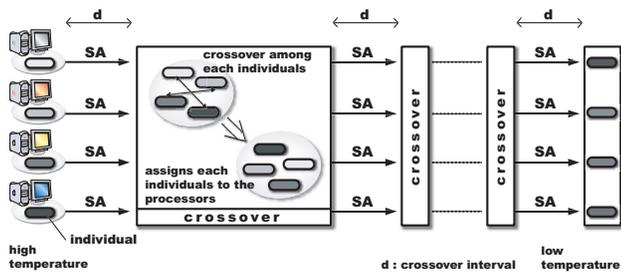


Fig. 1 PSA/GAc のアルゴリズム

PSA/GAc での探索手順を以下に示す。

- step1 初期解を生成し、複数ある探索点が並列に SA の処理である生成処理, 受理判定, クーリングを一定ステップ行う。
- step2 アニーリングが一定期間  $d$  (交叉間隔) に達すると、並列に実行している SA の解からランダムに 2 つずつ解を選びペアを生成する。このときすべての個体がペアを組むため、個体数の半数のペアが生成される。

- step3 ペアを組む 2 つの個体を親として遺伝的交叉を行い、2 個体の子を生成する。この交叉法については、後で詳細を説明する。
- step4 もとの親と生成した子との 4 個体のうち評価値の高い 2 個体を選択する。
- step5 一定期間  $d$  のアニーリングを行う。
- step6 すべてのペアで step3 ~ step5 の処理を行う。
- step7 終了条件を満たすまで step2 ~ step6 を繰り返す。

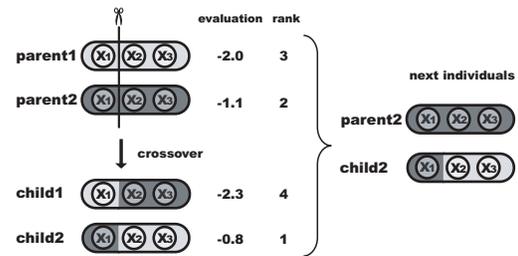


Fig. 2 PSA/GAc における交叉と選択

Step3 における遺伝的交叉を簡単な例を用いて説明する。Fig. 2 には、3 設計変数 ( $x_1, x_2, x_3$ ) の場合が示されている。parent1 と parent2 が交叉の対象として選ばれた個体である。これらの個体に対して交叉が行われるが、これは通常のビット表現における一点交叉や 2 点交叉などとは異なり、交叉点は常に設計変数の境界に存在し、設計変数の値そのものは変化しない。同図では、設計変数  $x_1$  と  $x_2$  の境界が交叉点としてランダムに選択され、交叉点以降の設計変数を個体間で入れ替える。この設計変数間交叉によって新しい 2 個体が生成される。parent1,2, child1,2 の 4 個体それぞれの評価後、評価値の高かった parent2 および child2 が次のステップの探索点に選ばれている。

3 PSA/GAc によるタンパク質立体構造のエネルギー最小化

本研究でエネルギー最小化の対象としたタンパク質は、Met-enkephalin, C-peptide, およびヒト副甲状腺ホルモンのフラグメント (PTH(1-34)) である。それぞれのタンパク質の概要を Table 1 に、立体構造を Fig. 3 ~ Fig. 5 に示す。本研究では PSA/GAc の性能を、共同研究者である岡崎国立共同研究機関分子科学研究所の岡本先生らの逐次 SA を用いた実験結果と比較することによって検証する。

Table 1 対象としたタンパク質の概要

タンパク質	アミノ残基数	二面角数	原子数
Met-enkephalin	5	19	75
C-peptide	13	64	218
PTH(1-34)	34	178	582
Protein G	56	275	859

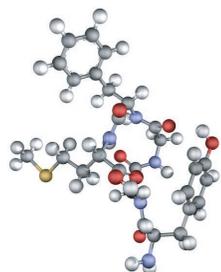


Fig. 3 Met-enkephalin

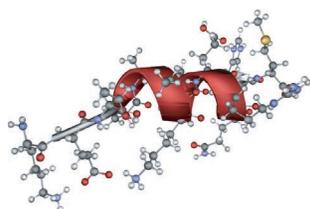


Fig. 4 C-peptide

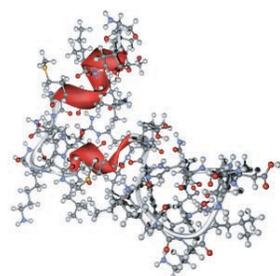


Fig. 5 PTH(1-34)

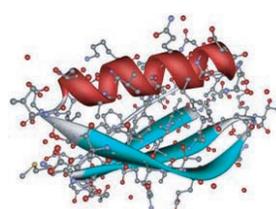


Fig. 6 proteinG

それぞれのタンパク質のエネルギー最小化に PSA/GAc を適用した結果と岡本らの実験結果の比較を Table 2 に示す．Met-enkephalin については、エネルギー値が  $-11.0 \text{ kcal/mol}$  以下で最適な立体構造をとることが知られているため、これについては全試行中において  $-11.0 \text{ kcal/mol}$  以下のエネルギー値を得た割合（成功率）を示した．これは値が 1.0 に近い方がより高い解探索能力を持っているといえる．一方、C-peptide および PTH(1-34) については全試行中で得た最も低いエネルギー値を示した．本問題はタンパク質のエネルギー最小化問題であるため、得られたエネルギー値が低いほどより高い解探索能力を持っているといえる．

Table 2 エネルギー値による逐次 SA と PSA/GAc の比較 (C-peptide, PTH(1-34))

	Met-enkephalin	C-peptide	PTH(1-34)
逐次 SA	0.5	-42.0	-210.0
PSA/GAc	0.9	-53.9	-246.0

Table 2 より、いずれのタンパク質においても、PSA/GAc は岡本らの逐次 SA よりも高い解探索能力を示していることが分かる．

## 4 タンパク質の部分構造に着目したローカルサーチアルゴリズム

前述のように、タンパク質の立体構造のエネルギー最小化においては、PSA/GAc が有効である．しかし、PSA/GAc は最適化対象であるタンパク質の性質を考慮していないため、Fig. 6 に示した ProteinG のような大規模かつ複雑なタンパク質の立体構造のエネルギー最小化を行う場合には、その計算に膨大な時間がかかると予想される．そこで本研究では、タンパク質の部分構造である  $\alpha$  ヘリックスと  $\beta$  シートが持つ特徴を利用し、PSA/GAc によるタンパク質の立体構造のエネルギー最小化を効率化するアルゴリズムを提案する．

### 4.1 $\alpha$ ヘリックス構造

$\alpha$  ヘリックスは、1 本のポリペプチド主鎖が規則正しくらせん状に巻いた構造になっている． $\alpha$  ヘリックスは、主鎖における 2 つの二面角 ( $\phi, \psi$ ) の値が ( $-70 \pm 30^\circ, -37 \pm 30^\circ$ ) であるアミノ残基が 4 個以上連なるときに形成される．提案するアルゴリズムは、タンパク質のエネルギー関数の解探索中に、あるアミノ残基において  $\alpha$  ヘリックス構造の一部が見つかったとき、その前後のアミノ残基に  $\alpha$  ヘリックスを拡張するというものである．この操作を行うことにより、探索の初期で  $\alpha$  ヘリックスが形成され、結果として探索の効率化につながると考えられる．このローカルサーチアルゴリズムを PSA/GAc に導入する．

### 4.2 $\beta$ シート構造

$\beta$  シートは、並行に配置された 2 本のタンパク質主鎖が水素結合で固定された構造である． $\beta$  シートも多くの大規模なタンパク質に含まれる部分構造であるため、これについても探索の効率化を行う必要がある． $\beta$  シートについては、対象とするタンパク質のアミノ酸配列と、タンパク質の部分構造データベースとの間で何らかの形でマッチングを行うことで探索の効率化が図れると考えられるが、現在これについては今後の課題である．

## 5 まとめ

本研究では、PSA/GAc を提案し、タンパク質のエネルギー最小化問題に適用した．その結果、Met-enkephalin, C-peptide, および PTH(1-34) のエネルギー最小化において PSA/GAc が高い解探索能力を示すことが明らかとなった．しかし、Protein G などのより大規模なタンパク質のエネルギー最小化においては探索の効率化が必須であるため、本研究ではタンパク質の部分構造に着目したローカルサーチアルゴリズムを考案した．今後は、考案したアルゴリズムを大規模なタンパク質のエネルギー最小化に適用し、その性能の検証を行うことが課題となる．