

2 個体分散遺伝的アルゴリズムによるタンパク質の構造予測

岩橋 崇史

1 はじめに

タンパク質の機能の解明は生命現象の仕組みの解明につながる。タンパク質の機能は、その立体構造によって決定されるため、タンパク質の立体構造を解明することが重要となる。タンパク質の立体構造の研究は、従来は実験による解析が中心であったが、近年の計算機能力の発達を受け、アミノ酸の配列のみを用いて、最適化手法により立体構造を予測するコンピュータシミュレーションによるタンパク質の立体構造予測の研究が注目を集めている¹⁾。本研究では、多点探索の最適化手法である 2 個体分散遺伝的アルゴリズムによるタンパク質の立体構造予測を行い、小規模なタンパク質立体構造のエネルギー最小化が GA により可能であることを示す。

2 コンピュータシミュレーションによるタンパク質の立体構造予測

タンパク質の最適な立体構造はそのタンパク質がもつエネルギーの最小状態と対応している。このため、最適化手法によってタンパク質がもつエネルギーが最小となる二面角の組み合わせを求めることにより、最適な立体構造を予測することが可能となる。従来のコンピュータシミュレーションによるタンパク質の立体構造予測では、最適化手法にシミュレーテッドアニーリング (Simulated Annealing : SA) が用いられてきた¹⁾。しかし、タンパク質のエネルギー関数は、大域的にいくつかの、局所的には無数の極小値を有している。そのため SA では大規模なタンパク質の立体構造予測を行うことは困難であると報告されている²⁾。

一方で、多点探索の最適化手法の一つである遺伝的アルゴリズム (Genetic Algorithm : GA) がある。GA は生物の進化の過程を模倣したアルゴリズムであり、各探索点のある環境に棲息する生物個体とみなす。そして、個体の集合 (母集団) に対し、遺伝的操作を繰り返し適用する。これにより、個体集団全体が成長し、最適解への到達が期待できる。GA はタンパク質の立体構造予測に有効であると考えられるが、GA によるタンパク質の立体構造予測についても困難であると報告されている³⁾。原因として、GA は早期収束によって局所解へ収束してしまうという問題点が挙げられる。本研究では GA の早期収束に注目し、早期収束の回避を考慮したモデルである分散遺伝的アルゴリズム (Distributed Genetic Algorithms : DGA)⁴⁾ および 2 個体分散遺伝的アルゴリズム (Dual individuals Distributed Genetic Algorithm

: Dual DGA)⁵⁾ を使用する。

3 2 個体分散遺伝的アルゴリズム

3.1 分散遺伝的アルゴリズム

DGA は、個体の母集団を複数のサブ母集団 (島) に分割し、分散処理を行う。具体的には、サブ母集団ごとに GA の遺伝的操作を適用し、ある間隔で他のサブ母集団と個体を交換する。この個体の交換を移住 (migration) といい、移住を行う間隔と移住を行う個体数の割合を、それぞれ移住間隔 (migration interval)、移住率 (migration rate) という。DGA は、島内で局所探索を行うことができ、また島間の個体の移住によって GA の問題点である早期収束を防ぎ、多様性を維持した探索を行うことができる。

3.2 2 個体分散遺伝的アルゴリズム

Dual DGA⁵⁾ は DGA において 1 島あたりの個体数を 2 としたものである。島内の個体数を極限まで少なくすることによって、島数を増やし、より多様性を維持した探索が期待される。しかし、島内の個体数を極端に少なくすることによって、通常の遺伝的操作では島内の多様性が急速に失われると考えられる。そこで Dual DGA では、遺伝的操作に多様性を維持する機構を組み込んでいる⁵⁾。DualDGA は、1 島あたりの個体数を 2 とすることより、設定すべきパラメータが、個体数と移住間隔のみとなり、従来の DGA と比較して、パラメータ設定の困難さを解消している (Fig. 1)。

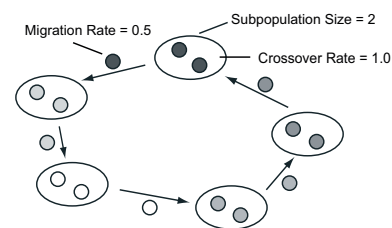


Fig. 1 Fixed parameter of Dual DGA

4 数値実験

4.1 実験内容

本研究では、立体構造予測を行うタンパク質として、小規模なタンパク質である Met-enkephalin を用いた。GA による Met-enkephalin の立体構造予測では、早期

収束を防ぐために個体数をいくつか用意し、実験を行った。用意した個体数を下記に示す。

- 個体数
800, 1600, 3200, 6400

また、DGA, Dual DGA による Met-enkephalin の立体構造予測では個体数、島数、移住間隔をいくつか用意し、実験を行った。用意した個体数と移住間隔を下記に示す。

- 個体数
800, 1600, 3200, 6400
- 移住間隔
1, 2, 3, 4, 5, 6, 7, 8, 9, 10

DGA については、1 島あたりの個体数を 4, 8, 16 の 3 つのパターンを用意した。その他のパラメータにおいては、Table 1 に示す。試行回数は 30 回である。

Table 1 パラメータ

model	DGA	Dual DGA
Sub Population Size	16,8,4	2
Number of Design Variables	19	
Chromosome Length	171 (= 19 × 9 Design Variable)	
Selection	Tournament	-
Tournament Size	2	-
Crossover Rate	1.0	
Crossover	1pt. crossover	
Mutation	0.006 (= 1 / 171)	
Migration Rate	0.25	0.5
Number of Elites	1	-
Terminal Criterion	1,900,000 evaluations	

4.2 実験結果

実験結果より、岡本らの結果^{2, 6)}と比較する。Fig. 2 に、GA, DGA, Dual DGA ならびに、SA, PSA/GAc²⁾の最適解発見率を示す。なお、GA, DGA, Dual DGA の最適解発見率は、今回の実験での最良値を示す。Fig. 2 より、GA は SA よりも最適解発見率が低いことが分かる。しかし、DGA, Dual DGA は SA よりも最適解発見率が高く、PSA/GAc と最適解発見率がほぼ等しいことが分かる。以上より、本研究では適用しているタンパク質は小規模であるが、タンパク質の立体構造予測において、DGA, Dual DGA は SA より解探索能力が高いといえる。また、PSA/GAc とほぼ等しい最適解発見率を示したため、DGA, Dual DGA はタンパク質の立体構造予測に有効な手法であると考えられる。

5 結論

本研究では、Dual DGA によるタンパク質の立体構造予測を行い、その性能の検証を行った。その結果、GA によるタンパク質の立体構造予測は困難であることが確

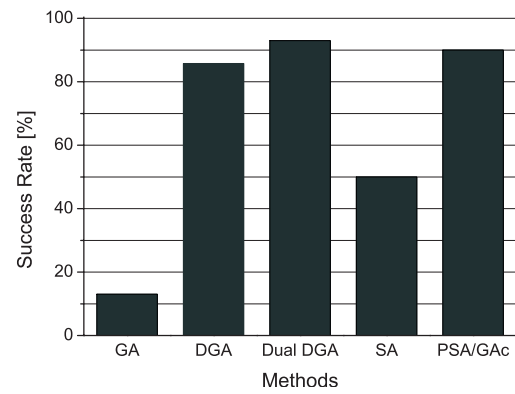


Fig. 2 Success Rate for tertiary structure prediction of Met-enkephalin

認された。しかし、DGA と Dual DGA においては、従来用いられていた SA よりも解探索能力が高いことが明らかとなった。Dual DGA と DGA はほぼ等しい解探索能力ではあったが、Dual DGA は DGA よりもパラメータ設定が簡易であり、並列計算環境に適している。以上より、Dual DGA はタンパク質の立体構造予測に有効な手法である可能性が示されたといえる。

6 今後の課題

小規模なタンパク質立体構造のエネルギー最小化に有効な GA モデルとして Dual DGA を示すことに成功した。しかし、Met-enkephalin よりも大規模なタンパク質である c-peptide には Dual DGA は有効ではないことをその後の実験により確認された。よって、今後は大規模なタンパク質に対しても有効な GA モデルおよび、その並列モデルを提案することを目指す。

参考文献

- 1) 岡本祐幸. モンテカルロシミュレーションで探るタンパク質の折り畳み機構. 物性研究. Vol.70, No.6, pp.719-742, 1998.
- 2) 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.
- 3) Laurence D.Merkle, Gray B.Lamont, Jr. George H.Gates, and Ruth Pachter. Hybrid genetic algorithms for minimization of a polypeptide specific energy model. IEEE Conf on Evolutionary Computation, 1996.
- 4) Reiko Tanese. Distributed genetic algorithms. Proc. 3rd International Conference on Genetic Algorithms, pp.434-439, 1989.
- 5) 廣安知之, 三木光範, 佐野正樹, 谷村勇輔, 濱崎雅弘. 2 個体分散遺伝的アルゴリズム. 計測自動制御学会論文集, Vol.38, No.11, pp.990-995, 2002
- 6) Ulrich H.E. Hansmann and Yuko Okamoto. Numerical Comparisons of Three Recently Proposed Algorithms in the Protein Folding Problem. JOURNAL OF COMPUTATIONAL CHEMISTRY, Vol.18, No.7, pp.920-933, 1997.