

## タンパク質および遺伝的アルゴリズムの基礎学習

藤本 万里子

### 1 課題の達成状況

今月は、今後の研究活動で必要となる基礎学習を中心に行った。本章では、それらの詳細について述べる。

#### 1.1 文献調査報告

##### ● タンパク質の基礎

タンパク質は、アミノ酸が多数結合したポリペプチド鎖が折り畳まれて立体構造 (Fig. 1(a)) を作ることで、生物学的な機能を果たす。結合の回転角である二面角 (Fig. 1(b)) が立体構造を決定する設計変数となる。タンパク質分子がなす立体構造はタンパク質分子のエネルギー最小状態に対応しているため、最適化手法を用いてエネルギー最小状態を探索する分子シミュレーション法と呼ばれる手法が注目されている。

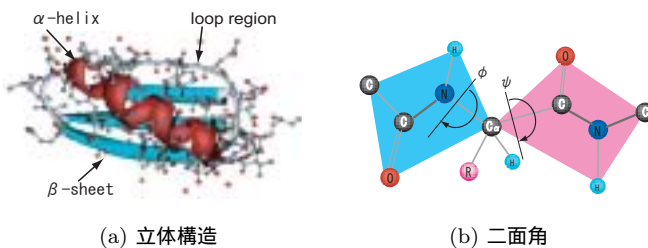


Fig. 1 タンパク質

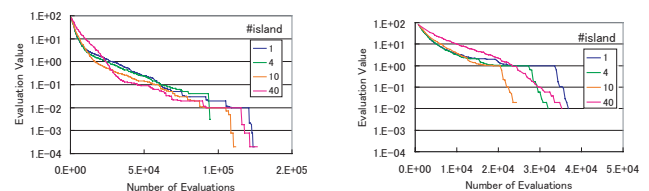
##### ● 遺伝的アルゴリズムの基礎

単一母集団 GA、および並列化モデルの 1 つである分散 GA の基本概念や各種オペレータについて調査した。単一母集団 GA は、膨大な反復計算を必要とするため計算負荷が高く、また探索初期に局所解に収束する場合があるという短所を持つ。この問題点を解決すべく分散 GA が開発された。分散 GA は、母集団を複数の分割母集団に分割し、各分割母集団で遺伝的操作を独立に行う。また、分割母集団間で移住と呼ばれる個体の交換操作を行う。分散 GA は単一母集団 GA と比較して、高品質な解が得られるという利点を持つ。

#### 1.2 GA のパラメータの検討 (データ処理実習)

本研究室で開発された分散 GA のライブラリである ga2k を数学的テスト関数に適用し、数値実験を行った。

本実験の目的は、GA のパラメータが解探索に与える影響を考察することである。分割母集団数を変化させた場合の解探索履歴を Fig. 2 に示す。Fig. 2(a) に平均値、Fig. 2(b) に中央値による結果を示す。なお、個体数 400、交叉率 1.0、選択手法はトーナメント選択とした。実験の結果、分割母集団数によって解探索性能が変化することや、各パラメータの最適値は他のパラメータの値に依存することが分かった。また、数値実験を行う際には、平均値や中央値など、様々な視点からの考察が必要であるということも学んだ。



(a) 平均値

(b) 中央値

Fig. 2 分割母集団数の変化による解探索履歴

なお、その他の詳細な実験結果については、ISDL レポート No.20030704003 を参照されたい。

#### 1.3 GA の作成 (プログラミング実習)

単一母集団 GA のプログラムを C 言語で作成している。なお、プログラミングを行う際は、モジュールプログラミングを作成することを心がけている。

### 2 今後の課題

#### ● 並列分散 GA の作成

作成中の GA のプログラムを分散 GA に拡張する。また、MPI を用いて並列プログラミングを行う。完成後は自作 GA と ga2k を比較し、動作確認を行う。

#### ● シミュレーテッドアニーリングの基礎学習

シミュレーテッドアニーリングの基礎に関する文献調査を行う。

#### ● 試行数と信頼性に関する調査

コイン投げ実験において、試行回数と得られる値の信頼性について検討する。