

タンパク質に関する基礎学習  
米田真純

1 前月からの課題

1. タンパク質の基礎勉強
2. SA の基礎となるアルゴリズム・理論の学習
3. C で SA のプログラムを作成
4. Fortran の基礎勉強として, Fortran で SA のプログラムを作成

2 課題の達成状況および研究成果

2.1 タンパク質

2.1.1 タンパク質とは

タンパク質はわずか 20 種類のアミノ酸がペプチド結合によって鎖状(ポリペプチド鎖)に連結している生体高分子である。20 種類のアミノ酸から構成されるといっても膨大な数の並び方があり, 極めて多数のタンパク質が存在する。タンパク質は, 固有のアミノ酸配列および立体構造を持ち, その結果として独自の特異的な働きをする。このようにタンパク質は全体としての多様性と個々としての特異性を持っている。タンパク質は複雑な生命現象に関わっている物質であり, タンパク質の機能はその立体構造によって決定される。このため, タンパク質の立体構造を解析・予測することは生命現象を知る上で重要なことである。

2.1.2 タンパク質の階層構造

タンパク質は Fig. 1 に示す階層構造を持っている。タンパク質の姿である立体構造が三次構造, アミノ酸配列が一次構造にあたる。二次構造は三次構造中にみられる特徴的な部分構造であり,  $\alpha$ ヘリックス・ $\beta$ シートが代表的なものとしてある。タンパク質の中にはポリペプチド鎖が複数より集まった構造のものも見られる。この集合している個々のポリペプチド鎖はサブユニット構造と呼ばれ, タンパク質の四次構造にあたる。

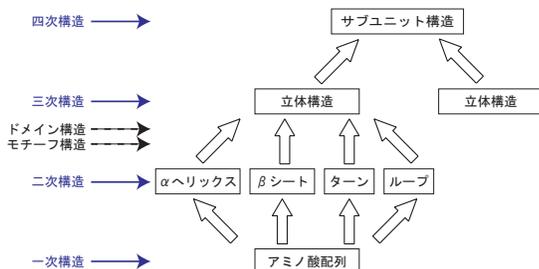


Fig. 1 タンパク質の階層構造

2.2 SA

SA は最適化問題 (組合せ最適化問題) を解く汎用近似解法の一つである。汎用近似解法は, 個々の最適化問題に広く適応可能な一段上位の解法に位置づけられるので, メタ戦略とも呼ばれる。ほとんどの組合せ最適化問題は, 問題が大きくなると解空間が指数的に増大するため, 真の最適解を求めるのは実質的に不可能である。

SA は解が改良される方向のみだけでなく, 改悪の方向にも探索が進むという仕組みにより, 容易に局所解に陥らず, 理論的には真の最適解が得られるということが証明されている。

2.3 プログラミング

Fortran に関して, プログラムを作成し, 基礎的な文法・サブルーチンの使い方などを学習した。

C と Fortran で SA のプログラムを実装した。対象問題を離散問題である TSP の eil51 とし, 数値実験を行った。近傍として 2-change, 受理判定として Metropolis 基準を用いた。Table 1 に各パラメータ設定, Table 2 に実行結果を示す。

Table 1 各パラメータ設定

パラメータ	値
最高温度	110
最低温度	1
クーリング周期	1020
総アニーリング数	168000
冷却率	0.97

Table 2 実行結果 (eil51, Optimum=426)

	最小値	5 回試行の平均
C の実行結果	440	468
Fortran の実行結果	449	472

3 今後の課題

1. 並列 SA のプログラムを実装する。
2. タンパク質の構造解析・予測に関する学習。
3. 昨年の卒業論文・修士論文を読み, 遺伝的交叉を用いた並列 SA を理解する。