

PSA/GAc によるタンパク質の立体構造予測と表示

鍵谷 武宏

1 はじめに

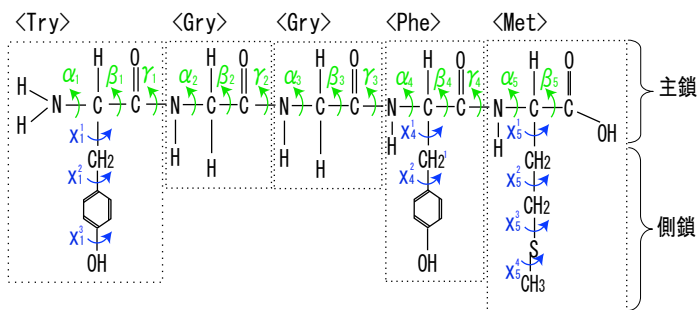
タンパク質は生命現象に直接関わる重要な物質であり、その機能は構造に依存すると言われている。そのため、タンパク質の立体構造を予測することにより、病理の解明、新薬の開発、特定の機能を持った人工タンパク質の設計など様々な分野への応用が期待されている。また、自然界のタンパク質はエネルギーが最も低い安定した状態で存在することが知られている。したがって、系の正しいエネルギー関数さえ与えられれば、タンパク質の立体構造予測はエネルギー最小化問題として取り扱うことができる。そこで、本研究室では、タンパク質のエネルギー最小化問題を解く手法として PSA/GAc を用いている。今回の演習では、C++を用いて PSA/GAc を実装した。また得られたタンパク質の立体構造を表示するシステムを構築した。

2 タンパク質立体構造予測

2.1 タンパク質の立体構造を決定するパラメータ

タンパク質の立体構造は二面角で表現することができる。そのため、最適化計算によるタンパク質の立体構造予測では二面角が設計変数として用いられる。今回の演習では、対象問題として Met-enkephalin を用いた。Met-enkephalin は Try-Gly-Gly-Phe-Met の五個のアミノ酸から成り、それぞれのアミノ酸は Fig. 1 のように、Try が 6 個、Gly が 3 個、Phe が 5 個、Met が 6 個の設計変数を持つ。

<分子構造>



<設計変数>

$X_1, X_1', X_1'', \alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2, \alpha_3, \beta_3, \gamma_3, X_4, X_4', X_4'', \alpha_4, \beta_4, \gamma_4, X_5, X_5', X_5'', \alpha_5, \beta_5, \gamma_5$

Fig. 1 Met-enkephalin のアミノ酸配列

(出典：参考文献 1 より引用)

3 PSA/GAc

3.1 概要

PSA/GAc とは並列に実行している SA の解の実行時に、遺伝的アルゴリズムのオペレータである遺伝的交叉を用いたものである。このモデルでは、Fig. 2 のように解の伝達時に並列に実行している SA から親としてランダムに 2 個体を選択し、設計変数交叉を行う。設計変数間交叉はアミノ酸を一つの固まりとし、各アミノ酸ごとに行う。そして親個体と生成された子個体を合わせた 4 個体の中から良好な 2 個体を選択し、次の探索点とする。

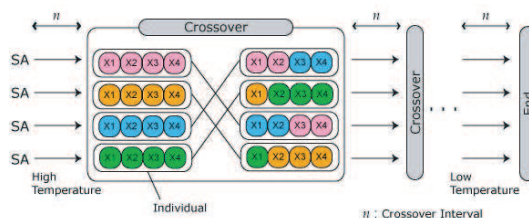


Fig. 2 PSA/GAc の模式図 (出典：参考文献 1 より引用)

3.2 アルゴリズム

PSA/GAc では一度交叉するまでに、クーリングまでの過程を繰り返す。この繰り返しの過程を Monte Carlo Sweep (MCSweep) と呼ぶ。1MCSweep では、Fig. 3 のように設計変数毎に「生成」「受理判定」「遷移」という操作を行い、全ての設計変数においてこの操作を行う。

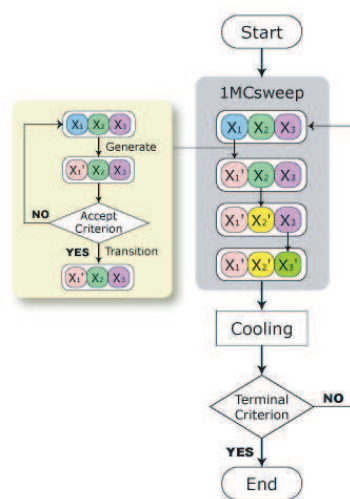


Fig. 3 アルゴリズム (出典：参考文献 1 より引用)

4 タンパク質の立体構造可視化システム

4.1 システムの概要

最適化計算により生成される立体構造の数は膨大である。そのため生成された構造を効率的に表示するシステムが必要となる。そこで今回の演習では、条件に合うタンパク質を表示できるシステムと、エネルギー履歴の表示システムを構築した。

4.2 タンパク質可視化システム

4.2.1 Jmol

タンパク質可視化システムでは、Jmol というオープンソースソフトウェアを用いている。Jmol は JavaApplet であり、分子モデルを立体表示可能である。また Jmol は JavaScript のライブラリを提供しており、そのライブラリを用いてタンパク質の可視化システムを構築した。

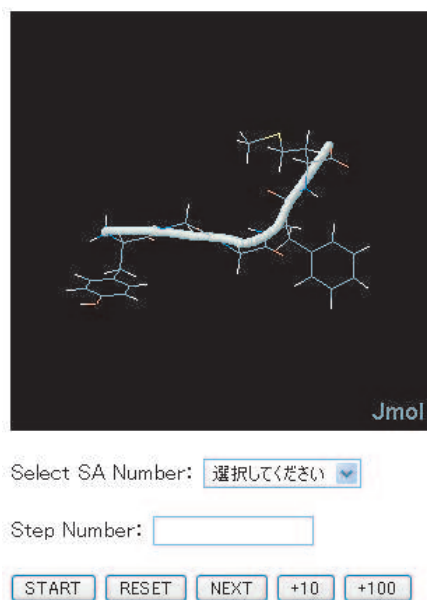


Fig. 4 タンパク質の可視化 (出典：自作)

4.2.2 機能

Fig. 4 に示すタンパク質表示システムでは、並列に実行された SA の番号 (Select SA Number) を選択し、ステップ数 (Step Number) を入力後、「START」ボタンをクリックすることで、その条件に合うタンパク質を表示することができる。また「START」ボタンの他に、「NEXT」ボタンで、次のステップ、「+10」ボタンで 10 ステップ後、「+100」ボタンで 100 ステップ後のタンパク質を表示できる。また「Reset」ボタンでステップ数をクリアすることができる。

4.3 エネルギー履歴表示システム

4.3.1 JFreeChart

エネルギー履歴表示システムでは、JFreeChart というオープンソースソフトウェアを用いている。JFreeChart はグラフ生成用のライブラリであり、グラフの基となるデータセットを引き渡すだけで、棒グラフや円グラフ等、様々な図を作成することができる。

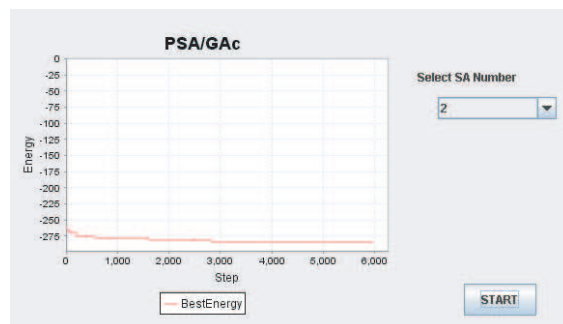


Fig. 5 エネルギーの推移 (出典：自作)

4.3.2 機能

Fig. 5 に示すエネルギー履歴表示システムでは、並列に実行された SA の番号 (Select SA Number) を選択し、「START」ボタンをクリックすることで、その条件に合うタンパク質のエネルギー履歴を表示することができる。グラフは縦軸が Energy、横軸が Step 数となり、エネルギーがステップ数によってどのように推移するかを見ることができる。

5 まとめ

今回の演習では、C++ で PSA/GAc を実装し、タンパク質の可視化とエネルギー履歴を表示するシステムを構築した。PSA/GAc では、タンパク質の立体構造予測に効果的な、MCsweep というアルゴリズムを用いた。このシステムにより、PSA/GAc により生成された構造を効率的に表示することが可能となった。今回の演習では実現できなかったが、エネルギー履歴のグラフとタンパク質の可視化の連携ができれば、さらに効率のよいシステムになると考えられる。

参考文献

- 1) 自作 PSA/GAc の性能検証, 宇野尚子
<http://mikilab.doshisha.ac.jp/dia/research/report/2003/0719/006/report20030719006.html>
- 2) サーブレットでグラフを描く (JFreeChart 活用)
<http://www.atmarkit.co.jp/fjava/javatips/092jspervlet03%6.html>