

# グローバルコンピューティング環境における PSA/GAc の検討

## Examination of Parallel Simulated Annealing using Genetic Crossover on the Global Computing Environment

青井 桂子

Keiko AOI

**Abstract:** This paper examines the search by PSA/GAc on the global computing environment. In the case of prediction of huge protein structure, we should use global computing. This study uses NetSolve as implementation model of global computing environment, and makes sure of effective of PSA/GAc on the global computing environment.

### 1 はじめに

これまでの研究で, PSA/GAc はタンパク質の立体構造予測において, 従来用いられていた SA よりも有効であることが明らかとなっている<sup>1)</sup>. 大規模なタンパク質の立体構造予測には並列化による計算時間の短縮が不可欠であり, 我々は PSA/GAc の並列モデルの研究も行ってきた<sup>2)</sup>. 本研究では, 今後更に大規模なタンパク質の立体構造予測を行うために, ネットワークを介して遠隔地にある計算資源を利用する, グローバルコンピューティング環境下での PSA/GAc の検討を行う.

### 2 グローバルコンピューティング化のための PSA/GAc モデル

#### 2.1 グローバルコンピューティング<sup>3)</sup>

グローバルコンピューティングとは, 広域ネットワーク上に分散した計算資源や情報資源を積極的に活用し, 科学技術計算に代表されるような大規模な計算を高速に処理することである. 計算資源とは高性能な計算サーバや大規模なストレージをもつファイルサーバである. 情報資源とは計算を行うのに必要なデータやプログラムそのものである. これらの資源が広域ネットワークに接続され, グローバルコンピューティング環境を形成する.

#### 2.2 グローバルコンピューティング環境の特徴

グローバルコンピューティング環境は, 従来のコンピューティング環境に比べて, 様々な障害により, ネットワークに大きな遅延が発生したり, 完全に切断されるなどして, 資源の構成が動的に変化する. 計算が複数の資源にまたがって行われるため, ある資源で計算した途中結果を別の資源で計算する場合に利用することが必要となる. 大規模なタンパク質の構造予測を行う場合, 解探索過程に要する時間が長くなるため, 途中で資源に何らかの問題が発生した場合に, それまでの計算を無駄にしないことが重要となる.

#### 2.3 PSA/GAc による断続的な探索

PSA/GAc での探索の途中結果を生かして次の探索に引き継ぐモデルを考える. タンパク質の主鎖の構造には,  $\alpha$ -helix などの水素結合により安定した二次構造が存在し, これらが部分解となっていることが多く見受けられる. この空間構造は設計変数である二面角からなっているため, 前の試行の探索を引き継ぐものとして二面角の値を採用する構造を設けた.

PSA/GAc において 1 試行目の出力ファイルの一つに, 2 試行目の初期パラメータファイルとなるものを作成することで, より効率の良い探索が可能になると思われる. 2 試行目以降の初期パラメータファイルでは, 前の試行で最も良かった解の二面角 (設計変数) の値と SA のパラメータである温度を用いて再び探索を続けるモデルを作成した. 概念図を Fig. 1 に示す.

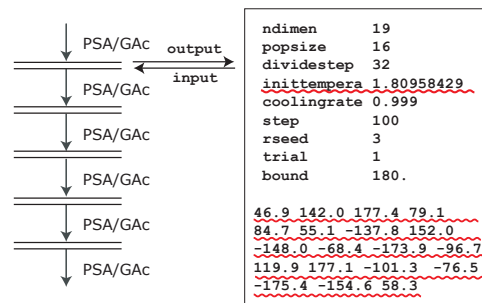


Fig. 1 グローバルコンピューティング化のための PSA/GAc

#### 2.4 数値実験

探索期間の短い PSA/GAc を複数回行ったときと, 探索期間の長い PSA/GAc を行ったときに, 変わらない解精度, 成功率が得られるかどうかの実験を行った.

PSA/GAc において 16 個体を用いて 6000MCsweep の計算を行い, Met-enkephalin の構造予測を行った. Met-enkephalin は -11kcal/mol 以下のエネルギー値のときに最適構造をとることが知られている<sup>4)</sup>. 6000MC-

sweep を連続的に探索を行ったもの、1000MCsweep、250MCsweep、100MCsweep ごとの探索を断続的に合計 6000MCsweep となるように探索を行ったものとを比較するパラメータの値は設計変数 19, 交叉周期 32, 初期温度 2.0, 冷却率 0.999 とした。この時の解精度の推移を Fig. 2 に示す。また、この時の成功率を Fig. 3 に示す。成功率とは、最適構造を得られた回数を試行回数で割ったものである。これらの結果は 10 回試行平均である。

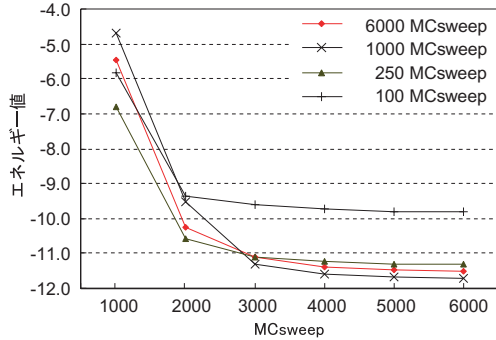


Fig. 2 解精度

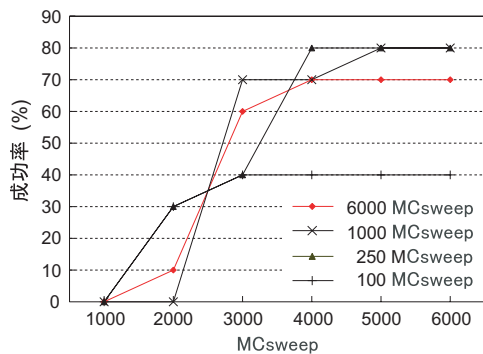


Fig. 3 成功率

これらの結果から、短い探索を断続的に行っても比較的良い割合で解を求めることができることがわかった。よって、グローバルコンピューティング環境で PSA/GAc を実装しても解精度及び成功率が極端に落ちることがないことが示された。

### 3 NetSolveにおけるグローバルコンピューティング環境の実装

NetSolve はクライアントサーバーモデルに基づいて設計されているミドルウェア層のシステムである。

#### 3.1 NetSolve システムの構成

NetSolve システムでは Client, Server, Agent の 3 つの要素の概念がある。NetSolve で実行できる問題は、Server 側に予めライブラリと PDF (Problems Description Files) で記述しているもののみである。

Client は指定した Agent に、計算の要求をする。(call fnetsl("function", a, b, ierr)) 要求を受けた Agent はその Agent に接続されている Server の中から必要なマシ

ンを選択し、そのマシンに Client から送られた入力データ (関数名と引数の数値データ) を送る。Server は、必要なデータと用いる関数が送られると、実行する関数が Server 側で定義されているものであればプログラムが実行され、数値データが返される。Fig. 4 に NetSolve の概念図を示す。

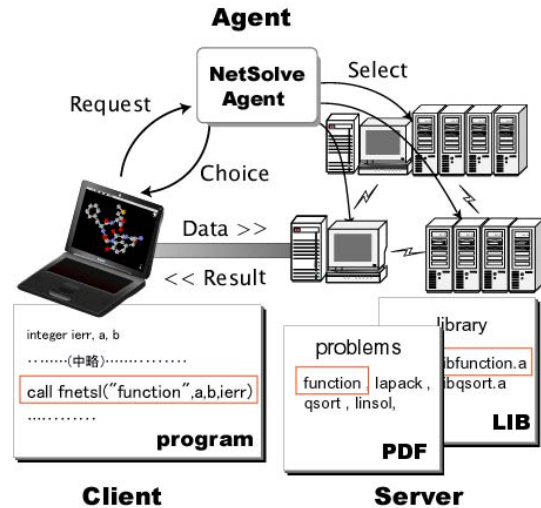


Fig. 4 NetSolve

#### 3.2 NetSolve でのプログラムの実行

NetSolve を使って実際に Server 側にライブラリと PDF を作り、プログラムを実行させた。

### 4 考察と今後の課題

PSA/GAc を断続的に行った場合と連続的に行った場合の解精度や成功率を比較したところ、ほぼ同等の性能が示された。これによって、計算量が膨大なものに対しても、グローバルコンピューティング環境で様々な資源を利用して断続的に探索することで最適構造が得られることが示された。

今後の課題としては、NetSolve の MPI の動作の調査を行い、MPI が動くことが確認されれば NetSolve による PSA/GAc の非同期モデルの構築を行い、これを NetSolve で実行する。

#### 参考文献

- 1) 小掠真貴, 廣安知之, 三木光範, 角美智子, 岡本祐幸. 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の検討. 情報処理学会研究報告, Vol. 2001, No. 27, pp. 57-60, 2001.
- 2) 廣安知之, 三木光範, 角美智子, 小掠真貴. 遺伝的交叉を用いた並列 SA (分散メモリ型並列計算機への実装モデルの検討). 情報処理学会第 62 回全国大会 講演論文集, 2001.
- 3) 谷村勇輔. グローバルコンピューティング環境における遺伝的アルゴリズム. 修士論文, pp. 300-303, 2001.
- 4) Yuko Okamoto, Takeshi Kikuchi, and Hikaru Kawai. Prediction of Low-Energy Structures of Met-Enkephalin by Monte Carlo Simulated Annealing. CHEMISTRY LETTERS, pp. 1275-1278, 1992.