

Grid 環境におけるタンパク質立体構造予測  
青井 桂子

1 序論

自然に存在するタンパク質の立体構造は系の自由エネルギーの最小状態に対応している。このため、アミノ酸配列情報からコンピュータシミュレーションにより最小化問題としてタンパク質分子の立体構造が予測可能であると考えられる。タンパク質のエネルギー関数は局所的に無数の、大域的にも複数の極小値を持つ。このため、我々は SA と GA のハイブリッドアルゴリズムである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 (PSA/GAc) による立体構造予測を行っている。

一方で、ネットワーク上につながれた広い地域に配置された計算資源を結びつけて、広域的に分散 / 並列処理を行う Grid と呼ばれる新しい計算モデルが研究されるようになった。

本研究では、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ための Grid モデルウェアである NetSolve を用いて Grid 環境を構築し、Grid 環境に適した PSA/GAc のモデルを実装する。

2 NetSolve の PSA/GAc 実装モデル

NetSolve は、Tennessee 大学の Jack Dongarra らによって開発された Grid RPC System である。NetSolve システムはネットワーク上にあるハードウェアとソフトウェアの両方の計算資源にリモートアクセスを可能にする。GA と SA のハイブリッドである PSA/GAc は、複数の逐次 SA を並列に実行し、一定間隔で遺伝的交叉を行う。遺伝的交叉の処理では、もとの親と生成した子との 4 個体のうち評価値の高い 2 個体を選択して、選択された 2 個体から次の探索を行う。

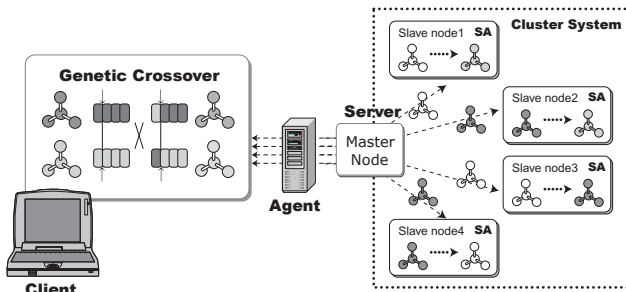


Fig. 1 NetSolve のための PSA/GAc の実装モデル

タンパク質のエネルギー計算は、その 1 つ 1 つの計算に時間がかかるのではなく、エネルギー計算回数が膨大になり、計算時間がかかることがわかっている。このた

め、Fig. 1 に示すように、Server 側に逐次の SA を実行させ、交叉周期になると Client 側に個体を返すモデルを考案し、構築した。

3 数値実験

本研究では、ヒト副甲状腺ホルモン (Human parathyroid hormone : HPTH) のフラグメントを立体構造予測の対象とした。提案モデルである NetSolve のための PSA/GAc モデルと従来の PSA/GAc モデルの単位時間あたりの計算量を比較する。

1 台の PC を用い、従来の PSA/GAc を用いてタンパク質の立体構造予測を行ったものと、1 台の PC を NetSolve の Client として 2 台のクラスタシステム (Cambria と Gregor) を Server として起動した環境を構築し、提案モデルでタンパク質の立体構造予測を行ったものを Fig. 2 に示す。

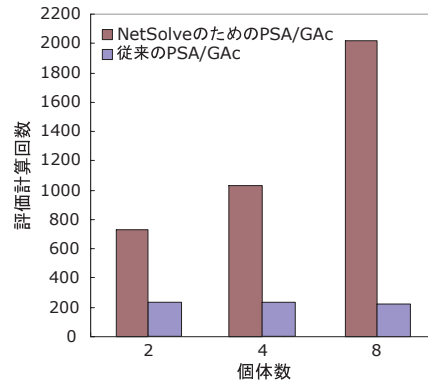


Fig. 2 実験結果

Fig. 2 のグラフでは縦軸に評価計算回数、横軸に個体数を表す。2, 4, 8 個体と個体数を変化させて、1 時間あたりの評価計算回数の比較を行ったところ、従来の PSA/GAc を 1 台の PC で行ったものは、計算能力が一定であるため、評価計算回数は個体数に関わらず一定であった。また、提案モデルで探索を行ったものは、逐次 SA の部分を NetSolve の Server に送るため、従来の PSA/GAc で探索したものに比べ、評価計算回数は格段に増加した。

4 結論

本研究では Grid 環境における NetSolve を用いた PSA/GAc の実装モデルを構築し、従来の PSA/GAc モデルとの比較を行った。Fig. 2 より NetSolve を用いて、外部 Server に逐次 SA を実行させることで、遥かに多くの評価計算回数を行うことができた。