

NetSolve を用いた PSA/GAc による小規模なタンパク質の立体構造予測における各処理時間の測定
青井 桂子

1 今月の課題

NetSolve の API として, "isd1_netsl_farm" を用いて PSA/GAc を構築した. NetSolve を用いた PSA/GAc による小規模なタンパク質の立体構造予測における各処理時間 (通信時間, Server での計算時間, システムの待機時間) の測定を行った.

2 NetSolve Farming の問題点

NetSolve Farming は非常に便利な機能であるが, 実験より, 次のような問題点が確認されている.

1. 大量の RPC 要求を繰り返し Farming で実行した場合, CLOSE WAIT 状態のポートが大量に発生し, 処理が中断する.
2. Farming のオーバヘッドが通常の RPC よりも長い.

1. は, Farming がノンブロッキング通信の RPC を利用していることが理由となっている. ブロッキング通信とノンブロッキング通信の RPC の違いは, 処理要求の送信と結果の受信の間に, 受信すべきデータが相手から送られているかどうかの受信状態のチェックを行うか, 行わないかである. CLOSE WAIT のポートが発生するのは, この受信状態をチェックする部分である. これらの問題点を解決するような API(isd1_netsl_farm) を下坂が作成したので, 本実験では "isd1_netsl_farm" を用いて新たに各処理時間を測定する.

3 数値実験

実装したシステムでは, Server での計算時間のほかに通信時間, NetSolve システムの NetSolve の Farming 機能を用いており, Farming のさいにはシステムの待機時間が生じる.

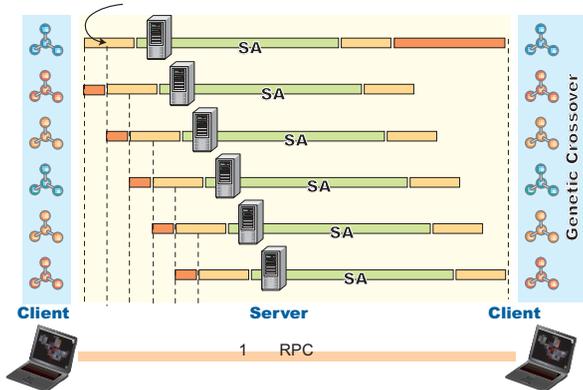


Fig. 1 1 回の RPC の処理時間の内訳

本実験で対象とした小規模なタンパク質は 5 つのアミノ酸からなり, 設計変数となる二面角を 19 個持つ Met-enkephalin と, 10 個のアミノ酸からなり, 二面角を 30 個持つ (Ala)₁₀ である.

本実験では, 各対象問題に対して, Server での SA のステップ数を 32MCsweep, 64MCsweep, 128MCsweep, 256MCsweep にしたパターンのもので実験を行う. また, PSA/GAc における個体数 (並列数) を 2, 4, 8, 16, 32, 64, 128 個体として実験を行った. 各 MCsweep 数, 個体数で比較を行う. 実験では Cambria Cluster を用い, NetSolve Server として 224 ノードを割り当てた.

3.1 Server での計算時間の測定

Fig. 2 と Fig. 3 に Met-enkephalin と (Ala)₁₀ における各 MCsweep の時の Server での計算時間の比較を示す. Fig. 2, Fig. 3 より MCsweep 数の増加に応じて計算時間も長くなることが示された.

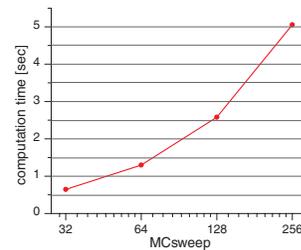


Fig. 2 Met-enkephalin 計算時間

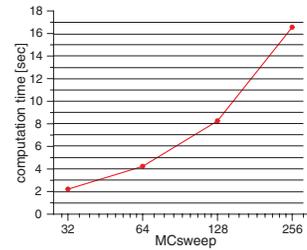


Fig. 3 (Ala)₁₀ 計算時間

3.2 システムの累積待機時間の測定

"isd1_netsl_farm" を用いた場合の各個体数に要する累積待機時間 (待機時間) を Fig. 4 と Fig. 5 に示す. 1 回の RPC に要する RPC は約 0.5sec であり, 個体数が増えるほど累積待機時間は増加する. 対象問題が大規模で, Server での計算時間が待機時間に比べて極端に長い時間を要するものでなければ, 非同期にすることなどを考える必要がある.

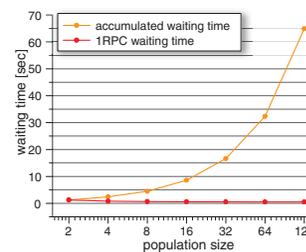


Fig. 4 Met-enkephalin 待機時間

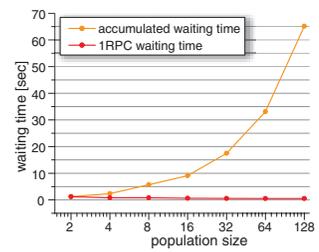


Fig. 5 (Ala)₁₀ 待機時間

4 今後の課題

NetSolve を用いた PSA/GAc の非同期モデルの作成