

課題名 (タイトル) :

長時間分子動力学シミュレーションのデータ解析プログラムの開発と応用

利用者氏名 : 小山洋平

所属 : 生命システム研究センター 計算分子設計研究グループ

1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

報告者の所属研究室では分子動力学 (MD) 専用計算機である MDGRAPE-4 の開発が行われており、10 万原子系で一日当たり数マイクロ秒の計算が可能になる予定である。このような時空間スケールでは小型タンパク質のフォールディングや化合物の結合などの重要な過程が見え始めると期待される。フォールディング・アンフォールディングや結合・解離などの複雑な構造変化を解析するための手法として原子間相互作用のエネルギー (ポテンシャルエネルギー) に対して主成分分析 (PCA) を実行する手法 (PEPCA) を開発してきた。長時間の MD シミュレーションに対して PEPCA を適用するためには次元数とデータ数の双方が大きいデータに対して PCA を実行する必要がある。このような大規模データに対する PCA を実行するために、近年、応用数学などの分野で randomized PCA と呼ばれる手法が提案されている。前年度までに randomized PCA の実装を行うことで、小型タンパク質に対して PEPCA が適用できるようになった。本年度は残りの計算量が多い箇所である周期境界条件下でのファンデルワールス/静電相互作用エネルギーの高速化を行った。

2. 具体的な利用内容、計算方法

座標データの読み込みを高速化するため、バイナリデータである NetCDF フォーマット (Amber 形式) を読み込めるようにした。周期境界の単位セルとして、立方体や直方体だけではなく truncated octahedron や rhombic dodecahedron を包含した triclinic cell も取り扱えるようにした。また、周期境界条件中で、あるカットオフ距離以内にある原子間の距離を求めるために、単位セルの座標に変換した後最近接の原子ペアを計算する手法と、セルリストアルゴリズムを用いた手法の二通りの実装を行った。両者で原子間距

離の計算を行う部分を OpenMP により並列化を行った。また、セルリストでは空間分割を行う部分の並列化も行った。

3. 結果

小型タンパク質であるリボヌクレアーゼ A (124 アミノ酸) の周期境界条件でのデータを用いてテストを行った。原子間相互作用の数は 200 万ほどで、randomized PCA に用いる power iteration の回数が 2 回という条件 (データとランダム行列との積 $\times 1$ 、power iteration $\times 2$ 、低次元空間への射影 $\times 1$ の計 4 回、各データにつきポテンシャルエネルギーの計算が必要になる) で、100 個のデータに対して 8 コア並列を用いた結果 22.6 秒で終了した。これは 1 日あたり 38 万データの解析に相当する。ポテンシャルエネルギーおよび行列演算に要する計算時間の比は 2 : 1 程度であった。

4. まとめ

本年度は周期境界条件でのカットオフ距離を考慮した非共有結合エネルギーの高速化を行った。現時点で、リボヌクレアーゼ A に対して原子間相互作用の PEPCA を 100 万構造用いて実行するには 3 日以内で終了するという見積もりとなった。

5. 今後の計画・展望

現在開発している C++ のプログラムは Amber フォーマットのみに対応しており、MDGRAPE-4 が採用している Gromacs フォーマットに未対応である。本年度中に対応予定であったが、他の実装などに時間がかかってしまったため、今後対応を行う。また、現在の実装では SIMD の効率的な利用などがあまり考慮されていないため、SIMD 演算が有効に機能するような改善を目指す。