

課題名 (タイトル) :

タンパク質・核酸など生体高分子の分子シミュレーション

利用者氏名 : ○木寺 詔紀、寺田 透、森次 圭

所属 : 次世代計算科学研究開発プログラム・分子スケール研究開発チーム

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

本研究チームでは、次世代生命体統合シミュレーション研究開発プログラムの分子スケール研究の一環として、生体分子 (タンパク質等) のシミュレーション法とそのソフトウェアの開発研究 (コード名: $\mu 2lib$)、を行ってきた。特に、全原子シミュレーション法と疎視化モデルとの連成手法を新規開発することを目指した。プロジェクトでの研究目的は以下の 2 点である:

- ・次世代スーパーコンピュータ「京」の全計算機資源を用いて高効率で計算することができる
- ・それによって従来の分子シミュレーションの方法ではできなかったレベルの計算をすることができる

生命活動をタンパク質や核酸などの生体分子のレベルからシミュレーションによって解こうという分野における問題は、その巨大な系の大きさと生命現象の時間スケールの大きさである。その大きさのために、全原子シミュレーション法には巨大な計算機資源を用いても多くの場合、生命現象の解明が可能な系の大きさと計算時間の長さを実現するシミュレーションは不可能である。そこで不可避免的に疎視化モデルの利用が求められるが、そこには精度の制約が生まれる。従って、その両者の利点を併せ持つ連成計算 (全原子シミュレーション法の精度と疎視化モデルの効率) が必要となる。また、数十万コアという並列計算を実現するためには、不可避免的に弱連成のアルゴリズムであることが要請される。これらを可能とするため、新規アルゴリズムである MultiScale Enhanced Sampling (MSES) 法を開発した。

MSES 法は、全原子モデルと低自由度の疎視化モデルによる連成シミュレーションである。疎視化モデルのポテンシャルが規定する運動空間に全原子モデルをドライブし、全原子モデルと疎視化モデルとを接続するバネ強度を 0 に外挿することで、全原子モデルの空間での分布関数を得ることができる。バネ強度の 0 への外挿は、バネ強度を変数としたハミルトニアンレブ

リカ交換法によって行う。従って、MSES 法はバネ強度の異なる多数のコピーを用いた弱連成のシミュレーションであり、高度の並列計算が可能である。レプリカ交換が疎視化モデルの自由度により決まることから、通常温度レプリカ交換法と異なり全原子モデルの自由度の制限なく巨大系のサンプリングが実行可能となる画期的な方法である。昨年度までの研究において、 $\mu 2lib$ への MSES 法の実装および高度化は完了している。まずテストケースとして小タンパク質のフォールディング過程に適用することで方法論の妥当性を示した。また、従来の拡張アンサンブル法では難しかった、天然変性タンパク質である *sortase* の (溶媒水を取り入れた) 生理学的環境下での構造アンサンブル計算を実現し、*sortase* 変性領域の折れたたみ過程やシグナルペプチド結合におけるカルシウムイオンの働きといった生物化学的な知見を得ることができた。

2. 具体的な利用内容、計算方法

(後述のとおり、プロジェクトの最終年度にあたり「京」でのチューニングと高度化に集中的に取り組んだため、RICC で作業を行う時間が取れませんでした。そのため、京での成果を中心に記述します。)

$\mu 2lib$ は生体分子の分子動力学シミュレーションソフトウェアであり、C++ のクラスライブラリを構成している。単体でのシミュレーションの他に、本プロジェクトで開発したマルチコピーシミュレーションの新規アルゴリズム (ストリング法、MSES 法) にも対応している。「京」のシステムに向けた、*mpi* と *openMP* によるハイブリッド並列による実行が可能である。

開発プログラムのチューニングやベンチマークを目的として、サイズの異なるさまざまな系 (モデルタンパク質としてよく使われている比較的小さな蛋白質 *adenylate kinase* や分子スケールチームの共通ターゲットである巨大蛋白質 *AcrB* など) を用いたテスト計算を実行した。

3. 結果

水溶液環境中における分子動力学シミュレーションの計算コストの大部分は、**particle mesh Ewald** 法の実空間相互作用計算ルーチンが占めている。ここでは、相互作用ペアについて距離の関数であるエネルギーと力を計算する。京で高い計算性能を得るには、コンパイラで生成されるコードがソフトウェアパイプライン化される必要がある。このため、コードがソフトウェアパイプライン化されるよう、ペアリスト作成法の変更、if 文の排除、障害となる関数の置き換え、配列アクセス方法の変更などを行った。昨年度までのコードでは、このルーチンの実効性能はピーク性能の 4.2%であったが、この最適化によって 25.8%まで向上した。また、プロセス間の負荷の平準化を支援するため、主な計算ルーチンの計算時間を、プロセス毎に表示する機能を付け加えた。ユーザーが、このデータをもとに、**particle mesh Ewald** 法の実空間および逆空間相互作用計算ルーチンに割り当てるプロセス数を調整できるようにプログラムを改良した。

マルチコピーシミュレーションのアプリケーションのうち、レプリカ交換分子動力学シミュレーションのプログラムでは、エネルギーと温度の情報をマスタープロセスに集め、マスタープロセスが交換判定を行っていた。この方法は、交換すべき情報が少ない（温度のみ）という点で、優れているが、集団通信が必要という点で、3D トーラスを特徴とする京の高速なネットワーク性能を十分生かしていないという問題があった。そこで、交換すべき情報が多い（座標データ）代わりに、隣接通信のみで交換判定が実現可能な方法を採用したアプリケーションを新たに作成した。2つのプログラム（集団通信版、隣接通信版）の計算性能を比較した結果、コピー数が 32 以上になると隣接通信版の方が優位になることが明らかとなった。また、MSES 法についても隣接通信版を作成し、同様の結果が得られた。

京利用の準備段階において、京のマルチコア CPU に対応するための相互作用計算ループ部分に対する OpenMP 化を完了した。京の試験利用では、OpenMP/隣接通信版のマルチコピーシミュレーションプログラム（レプリカ交換法、ストリング法、MSES 法）のチューニングとベンチマーク計測を行なった。まず、京の高速なネットワークを以てしても数百を超えるコピー数ではスケラビリティが落ちることが明らかにな

ったため、大規模コピー計算に向けた 1 プロセスマルチコピー版を作成した。通常のマルチコピーシミュレーションでは 1 プロセスが 1 コピーを担当する。コピー数を固定した **strong scaling** を計測する場合、使えるコア数を 2 倍にすると 1 プロセスあたりのコア数も 2 倍になるが、これは単体 (1 コピー) での **strong scaling** を計測しているのと同様である。分子動力学シミュレーションでは系のサイズによりスケールするコア数に上限があり、それ以上のコア数ではスケラビリティが急激に落ちる。1 プロセスマルチコピー版では、1 プロセスが 1 コピーを担当して使うコア数を変動するかわりに、コア数を固定しコピー数を変える。使えるコア数を 2 倍にすると、1 プロセスあたりのコア数は変わらないがコピー数が半分になる点で、単体での **strong scaling** 計測とは異なり大規模コピー計算でも高いスケラビリティを維持することができる。

1 プロセスマルチコピー版ストリング法について京でベンチマーク計測を行った。計算対象としては、**adenylate kinase** (3,343 原子) を用い、相互作用計算には 2 つの手法を適用した。一つは GB/SA と呼ばれる溶媒和モデルであり、溶媒自由度を陽に含まない。No-cutoff のため、各プロセスが担当する相互作用数は不変である。もう一つは **smooth particle mesh Ewald** 法 (SPME) で、水分子 (TIP3P モデル) ・イオンを陽に含んだ計 62,475 原子系で計算を行う。Cutoff 半径を行き来する水分子の存在により各プロセスが担当する相互作用数が変化しロードバランスにばらつきが生じる。コピー数を固定した (1,536 コピー) **strong scaling** のベンチマークを実行した結果、GB と Ewald 両方について、最大ノード数である 24,576 ノードまでに対して高いスケラビリティを達成した (図 1)。

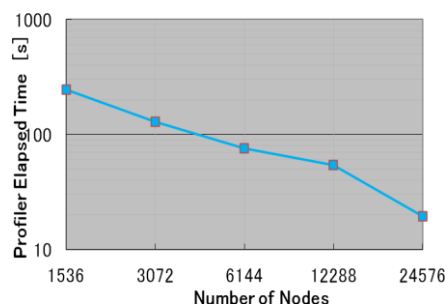


図 1: 京での 1 プロセスマルチコピー版ストリング法ベンチマーク。

また、実行効率 10% の目標を達成するため、分子スケ

ールチームの共通ターゲットである巨大蛋白質複合体 AcrB (溶媒水を含めて約 40 万原子) を用いて温度レプリカ法のベンチマークを実行した結果、高並列時でも高い実行効率を得ることができ、最大ノード数である 24,576 ノード並列時で約 12.3%の実行効率を実現した (図 2)。

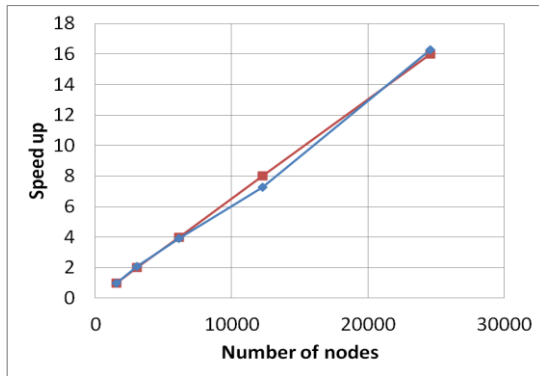


図 2 : 京での AcrB 温度レプリカ法のベンチマーク。赤線が理想値。

4. まとめと今後の計画・展望

プロジェクトで研究開発を行ったプログラムを幅広く使ってもらえるよう、マニュアルなどの整備を引き続き行う。プログラム講習会などの機会には分かりやすい使用方法の説明に努めていきたい。

5. 利用がなかった場合の理由

次世代生命体統合シミュレーション研究開発プログラム・分子スケール研究開発チームでのプロジェクトの最終年度にあたり、京でのチューニングと高度化に集中的に取り組んだため、RICC で作業を行う時間がなかった。また、分子チームが行う応用研究は数カ月にも及ぶ長時間の分子動力学計算を必要とするため、数日のみの特別利用を用いて行うことができなかった。

平成 24 年度 RICC 利用研究成果リスト

【論文、学会報告・雑誌などの論文発表】

1. Tohru Terada and Akinori Kidera, "Comparative molecular dynamics simulations study of crystal environment effect on protein structure", Journal of Physical Chemistry B, 116, 6810 (2012).
2. Kei Moritsugu, Tohru Terada and Akinori Kidera, "Disorder-to-order transition of an intrinsically disordered protein revealed by multiscale enhanced sampling", Journal of the American Chemical Society, 134, 7094 (2012).
3. Yasuhiro Matsunaga, Hiroshi Fujisaki, Tohru Terada, Tadaomi Furuta, Kei. Moritsugu and Akinori Kidera, "Minimum free energy path of ligand-induced transition in adenylate kinase", PLoS Computational Biology, 8, e1002555 (2012).

【国際会議などの予稿集、proceeding】

該当なし。

【国際会議、学会などでの口頭発表】

1. 木寺詔紀

”Multiscale simulation studies on protein structural changes upon ligand binding”

日英ワークショップ 英国大使館 東京 2012 年 12 月

2. 木寺詔紀

「MM/CG プログラム $\mu 2$ -lib によるタンパク質の大規模サンプリング」

ISLiM ソフトウェア研究開発報告会 東大武田ホール 東京 2013 年 1 月

【その他】

該当なし