

論文の内容の要旨

応用生命工学専攻

平成 11 年度博士課程進学

氏 名 関 嶋 政 和

指導教官名 清水 謙多郎

論文題目

分子動力学法の並列計算方式と拡張アンサンブル法の開発

1 はじめに

分子動力学法 (Molecular Dynamics Method, MD) とは、多原子系における原子の運動を、原子間の相互作用を計算しながら、個々の原子に対する Newton の運動方程式を積分することにより求める方法である。このような積分は解析的に計算することが不可能であるため、有限差分法を用いて数值的に計算する。MD の基本手順は有限差分法の各タイムステップにおいて力の計算、原子座標の更新を繰り返し、時々刻々の原子の位置・速度を計算するというものである。MD はタンパク質、核酸、糖、脂質のような生体分子の構造、動力学、熱力学などを理解するための重要な手段となっている。MD では原子レベルでフェムト秒単位の動きをシミュレートすることができるので、分子の機能の予測・解析に利用される。また、タンパク質中の残基を置換したときの自由エネルギー変化等、活性や安定性の変化の解析にも利用することができる。NMR や X 線結晶解析から得られた情報から 3 次元構造を決定するのにも用いられている。

2 生体分子シミュレーションの効率的な実行

2.1 大規模な系に対する詳細な解析への試み

このような生体分子のシミュレーションにおいては、大規模で複雑な系を対象とした計算が必要であり、また、詳細な時間単位での精密な計算が望まれる。しかしながらその計算量は

膨大であり、多大な計算時間を必要とする。

1 台のコンピュータの処理速度を向上させるには、物理的・技術的な限界があり、計算速度を向上させるには計算処理を分割して複数のプロセッサ (処理装置) で同時に実行させる並列計算が必要不可欠となる。このため、MD でも多数の並列計算アルゴリズムが開発されてきた。これらのアルゴリズムのほとんどは原子分割法、空間分割法と呼ばれる手法に大別される。原子分割法は原子数が N 、プロセッサ数が P であるとき、原子を配座に関係なく N/P 個ずつ各プロセッサに割り当てて計算を行う方式で、各プロセッサの負荷を分散させることができるが、タイムステップごとにあるプロセッサが計算した結果を他のすべてのプロセッサに対して通信する必要がある、データ通信量が問題になる。空間分割法は系を格子状に分割し、各格子に含まれる原子を同一プロセッサに割り当てて計算を行う方式で、あるプロセッサは、自身が計算した結果をその結果を必要とする周辺の格子を担当するプロセッサにのみ通信することにより通信量の低減を図っている。空間分割法は原子分割法より一般的に高い性能を示すという結果が得られているが、系が不均一な場合、タイムステップごとにプロセッサ間で計算時間にばらつきが生じ、あるプロセッサではすでに計算が終了しているにもかかわらず、残りすべてのプロセッサで計算が終了しなければ次のタイムステップの計算に移ることができないという同期の問題が生じる。

2.2 本研究で開発した並列計算方式とソフトウェア

本研究では、これらの問題に対処するため、空間分割法を基礎としつつ、各プロセッサの計算処理が終了し次第、次の計算処理を開始できる並列計算手法を開発し、さらにこのような並列計算を支援するソフトウェアシステム Parsley を開発した [1] [2] [3]。MD のプログラムを記述する際、並列処理可能なタスクと呼ぶ基本単位 (空間分割された各格子の力の計算や原子座標の計算に対応) とそれらの依存関係を指定すれば、本システムが自動的にこの依存関係に従って、プロセッサを割り当てる。これにより、従来のようなタイムステップごとのプロセッサ間の待ち合わせが不要となり、しかもシステム利用者は、このようなプロセッサの割り当ての詳細に関知する必要がなくなる。また、従来の空間分割法では、シミュレーション実行中、分子が格子の境界を越えて動いた場合、担当するプロセッサを変更するという操作を人手で (プログラムとして記述して) 行わなければならなかったが、本手法では、システムがこの種の作業を自動的に (しかも最適に近い方式で) 行うことができる。

2.3 結果

ウシ膵臓トリプシンインヒビター (Bovine Pancreatic Trypsin Inhibitor, BPTI) 1 分子と多数の水分子からなる系 (総原子数 16735) に対し分子動力学シミュレーションを実行した。従来の空間分割法と Parsley によって並列化した手法を大規模並列計算計算機 HITACHI SR2201 上で実行し性能を比較した結果、本研究で開発した Parsley による方法が、125 プロセッサのとき 1 プロセッサのときの 30.4 倍となり、125 プロセッサのとき従来の空間分割法より 3.49 倍早く処理が終了するという高い性能を示した。とくに、プロセッサ数によらず、プロセッサ

間の同期・通信にかかる時間が大幅に短縮された。

2.4 本ソフトウェアの汎用性について

本研究で開発したソフトウェアシステム Parsley は、C 言語で記述され、科学技術計算の分野で標準的に利用されている MPI (Message Passing Interface) ライブラリを通信の基盤として利用しているため、上記の SR2201 だけでなく、多くの並列コンピュータ、およびワークステーションをネットワークで結合した分散システム上でそのまま利用することができる。本研究では、さらに Parsley とその上で動作する分子動力学シミュレーションを並列計算機 IBM SP/2 及び COMPAQ DS20 (DUAL ALPHA Processor \times 8) クラスタ上に移植した。IBM SP/2 では 32 プロセッサのとき 1 プロセッサのときの 4.95 倍の性能が得られ、COMPAQ DS20 では 16 プロセッサのとき 1 プロセッサのときの 9.8 倍の性能が得られることを示した。また、従来の空間分割法との比較では、IBM SP/2, COMPAQ DS20 でそれぞれ 5.77 倍、7.92 倍の性能が得られており、Parsley による方法が有効であることを示した。

3 スケジューリング方式に関する研究

Parsley には、処理しなければならない複数のタスクのうち、どのタスクを優先的に実行すべきかスケジューリングする機構が備わっている。この機構を用いることで、例えばクラスタ型計算機のように、計算機資源が複数利用者に共有されている環境で、各利用者が利用できる計算機資源が他の利用者の影響を受け動的に変化したりする場合に、動的に負荷を分散させることができ、計算処理速度の異なる計算機においては処理速度に応じて負荷を分散させることができる。本研究ではまた、MD のように同じ処理を繰り返し実行する性質を持つプログラムに対しタスクの実行時間を計測し、その履歴を用いることにより自動的にスケジューリング方式を改善する方式を開発した。これにより、プロセッサのアイドルタイムを 35%~55% 減少させるのに成功した。

4 拡張アンサンブル法に関する研究

MD は上述のように時間軸に沿って生体分子のシミュレーションを行い、例えばタンパク質の折りたたみ機構を解明のような時間と共に分子の構造が変化していく系の解析に用いることができる。これに加えて、MD を用いて構造空間のサンプリングを行うことで、分子認識機構の解明や、タンパク質の安定性の変化など、多くの自由エネルギー曲面を求める問題にも適用することができる。これらを解析するためには、タンパク質のエネルギー変化を追ってシミュレーションを行い、広大な構造空間を探索する必要がある。これまで、MD でカノニカルアンサンブルを生成することによって、このような自由エネルギー曲面が研究されてきた。しかしながら、この方法では、特に室温時において、構造空間に多数のエネルギー障壁が存在するため、初期構造に依存したローカルミニマムに留まる傾向が強いという問題があった。

このため、構造空間をランダムウォークで探索できる非カノニカルアンサンブルを生成するシミュレーション方法が開発されてきている。レプリカ交換法 (Replica Exchange Method,

REM) は、求めたいエネルギー空間を覆うように複数の温度で MD を実行してカノニカルアンサンブルを生成し（これら MD の実行結果一つ一つをレプリカという）、これらレプリカ同士の温度をルールに従って交換し合い、個々のレプリカが様々な温度を経験することで構造空間の幅広いサンプリングを実現する。この結果に対して WHAM (Weighted Histogram Analysis Method) を用いることで任意の温度における各状態の出現確率を求めることができるというものである。

本研究では、Parsley REM MD' というレプリカ交換法を実装し、WHAM を用いることでこれらの各レプリカによるサンプリングから各状態の出現確率が正しく求められることを確認した。

5 おわりに

本研究で開発したソフトウェア Parsley は、MD だけでなく、エネルギー最小化計算、モンテカルロシミュレーション、分子軌道計算など、計算化学の諸問題に適用可能な高い汎用性を有する点が特徴である。また、分子動力学シミュレーションと FMM (Fast Multipole Method) による静電ポテンシャルの計算などのように、異なる種類の処理を同時に実行する場面は、実際の生体分子の構造予測・機能解析において頻繁に生じるが、このような計算において、開発したシステムは大きな威力を発揮するものと考えられる。本研究では、ある特定の並列計算機での処理の高速化を達成しただけでなく、様々なシステムや異なる処理速度を持つ計算機が混在する環境でもスケジューリング機構などの実装により効果があることを示した。

本研究ではさらに拡張アンサンブル法の一つであるレプリカ交換法を実装し、非カノニカルアンサンブルシミュレーションを実現した。現在レプリカ交換マルチカノニカルアンサンブル MD を Parsley 上に実装中である。

参考文献

- [1] M. Sekijima, S. Takasaki, S. Nakamura, M. Ikeguchi, and K. Shimizu. Parsley: a scalable framework for dependence-driven task scheduling in distributed-memory multiprocessor systems. In *Proceedings of the Eleventh IASTED International Conference on Parallel and Distributed Computing and Systems*, pp. 800–805, 1999.
- [2] 関嶋政和, 高崎慎也, 中村周吾, 池口満徳, 清水謙多郎. サブタスク間の依存関係に基づくスケジューリング機構を備えた並列プログラミング環境の開発. 情報処理学会論文誌 プログラミング, Vol. 41, No. SIG 2 (PRO 6), pp. 65–77, 2000.
- [3] M. Sekijima, S. Takasaki, S. Nakamura, M. Ikeguchi, and K. Shimizu. A parallel programming environment with dependence-driven task scheduling in distributed-memory multiprocessor systems. In *Proceedings of the ISCA 13th International Conference on Parallel and Distributed Computing Systems*, pp. 348–354, 2000.