

課題名 (タイトル) :

生体分子粗視化モデルの開発と応用

利用者氏名 : ○木寺 詔紀  
 検崎 博生

所属 : 社会知創成事業 次世代計算科学研究開発プログラム  
 次世代生命体統合シミュレーション研究推進グループ 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

理研を中心に行っている文部科学省プロジェクト「次世代生命体統合シミュレーションソフトウェアの研究開発」において、分子スケールチームに属し、「生体分子系の粗視化シミュレーション技術の開発」(研究代表者、高田彰二(京大理))を行ってきた。そこで、タンパク質、DNA、生体膜の複合系の粗視化分子モデルの分子動力学(MD)シミュレーションを行うソフトウェアCafeMolを開発中である。

CafeMol は、これまでに、タンパク質、DNA、生体膜の複合系でシミュレーションできるようになっている。前年度までには、分子動力学計算の時間発展の openMP によるスレッド並列化とレプリカ交換法の MPI 並列化を実装し、数千コア規模での並列性能を確かめることができた。

今年度は、単体性能の向上とレプリカ交換法の並列化のより有効な実装とテストを行い、次世代スパコン実機での実質的な大規模計算への準備が進める。

2. 具体的な利用内容、計算方法

テストに使う系としては、ヌクレオソームが20個フィラメント状につながった核酸/タンパク質複合体を用いる。ヌクレオソームはDNAがヒストンタンパク質におよそ2回転程巻きついたもので、細胞内でのDNAの最小構造単位である。そして、ヌクレオソームは集まってフィラメントを形成するが、その構造やダイナミクスは生物的に大変興味が持たれている。

測定は単体性能の向上についてはFX1を用い、レプリカ交換法の並列化の改変については超並列PCクラスタを用いて行った。コンパイラは富士通コンパイラを用いた。

3. 結果

単体性能の向上については、力場計算、エネルギー計算、ネイバリングリスト計算の3つについて行った。FX1での測定を行い表1に示すように、1割程度計算時間を短縮することに成功した。測定区間については、力場計算のforceとネイバリングリスト計算のneighbor、座標の更新のupdate、エネルギー計算のenergy、レプリカ交換のreplica、計算結果の出力outputに分かれる。また、forceとneighbor、energy、replicaについてはMPI通信時間により時間がかかるので、演算部分(ope)とMPI通信部分(mpi)に分けている。forceの演算部分については、さらに、force(local)、force(nlocal)、force(pnl)に分けている。このforce演算が最も計算時間が掛かる部分であり、特にforce(pnl)は静電相互作用の計算を含むために、もっとも計算量が多い部分である

	修正前 (秒)	修正後 (秒)
force	883.538	793.071
force (ope)	883.538	793.071
force (mpi)	0	0
force (local)	1.834	1.827
force (nlocal)	0.525	0.527
force (pnl)	881.172	790.716

neighbor	167.328	148.488
neighbor (ope)	167.328	148.488
neighbor (mpi)	0	0
update	0.214	0.209
multigo	0	0
energy	99.491	89.093
energy (ope)	99.491	89.093
energy (mpi)	0	0
random	1.19	1.204
replica	0.009	0.007
replica (mpi)	0	0
output	0.024	0.025
total	1151.856	1032.156
ope	1151.857	1032.158
mpi	0	0
main_loop	1151.857	1032.158

表1：単体性能の測定結果

レプリカ交換による並列化については、通常の温度ではなく、ハミルトンレプリカ法を用いて、イオン濃度を交換するために起こるために起こる問題を回避するために変更を行った。

イオン濃度を交換する場合、レプリカ間で静電相互作用の到達範囲が変化するため、レプリカ間で力場計算に掛かる時間が大きく変化することが問題となる。そこで、力場計算に時間が掛からないレプリカについては、レプリカ交換までのステップ数を増やすことによって、レプリカ間の計算時間を均等化することにした。具体的には、各レプリカでネイバリングリストの大きさを調べ、ネイバリングリストの大きさに反比例するようにステップ数を変化させた。128 レプリカの場合について、変更前の結果を図 1 に、変更後の結果を図 2 に示す。青は、各ノードでの演算時間を示しているが、node 数が大きいところで、演算量が少なくなっているのが図 1 で分かる。変更後の結果である図 2 では、演算量の偏りが減っているのが分かる。ただし、変更後では全ノード共通の通信時間が増えており、計算時間が増えているのが

問題であり、今後これをいかに減らしていくかが問題点となる。

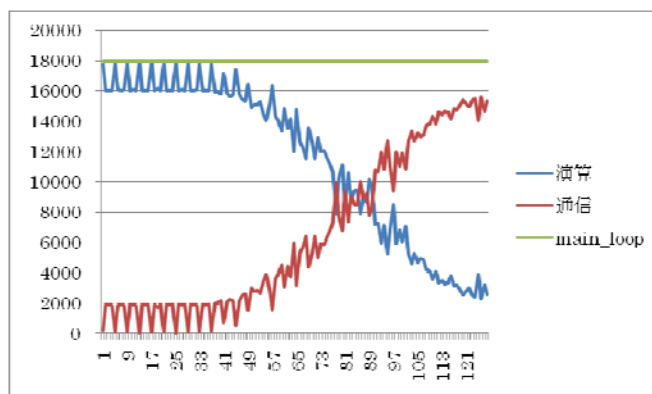


図 1：レプリカ毎の演算時間と通信時間

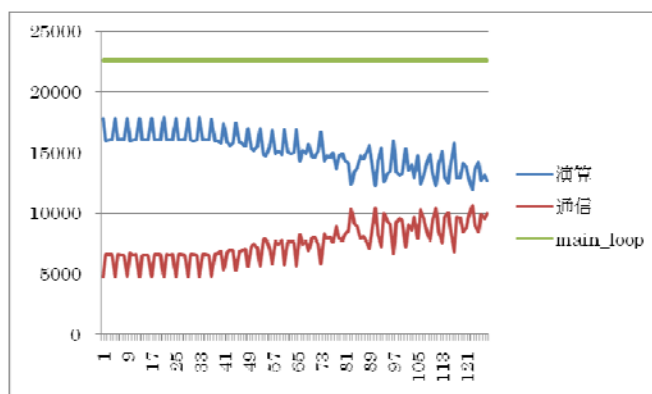


図 2：ネイバリングリストを用いてレプリカ間のインバランスを修正した結果

#### 4. まとめ

単体性能の向上とレプリカ交換法の並列計算の変更を行い、次世代スパコンで性能を発揮するための準備が整いつつある。

#### 5. 今後の計画・展望

- RICC の継続利用を希望の場合は、これまで利用した状況（どの程度研究が進んだか、研究においてどこまで計算出来て、何が出来ていないか）や、継続して利用する際に行う具体的な内容

今年度では、レプリカ交換の並列化の変更を行ったが、ノード間の演算量のインバランスは改善されたが、全ノード共通の通信時間が増えることによって、計算時間も増えてしまった。来年度は、この通信時間の増加について改善を行いたいと

## 平成 22 年度 RICC 利用報告書

考えている。

また、今年度は、8000 コアレベルの計算がほとんどできなかった。理由として、年度前半は、単体性能のチューニングを行い、後半レプリカ並列の改変を行ったが、8000 コアレベルのテスト計算をできるようになるまで時間が掛かり、テスト計算を行う機会が十分でなかったからである。

よって、来年度は、年度最初から積極的に 8000 コアレベルのテストを行い、大規模並列化計算の効率的な並列化を行っていきたいと考えている。

7. 一般利用で演算時間を使い切れなかった理由

8. 利用研究成果が無かった場合の理由

ソフトウェア CafeMol の大規模並列化テストが目的であったので、現状では成果発表の段階までいっていないため。