

地球シミュレータによる蛋白質の高速シミュレーション

齋藤 稔[†] 佐谷野 健二^{††}

JSP2002 において、最新のベクトル型並列計算機によって蛋白質のシミュレーションが、いかに高速化されるかについての考察を行った。同時に、現在、世界最高速のスーパーコンピュータ(地球シミュレータ)によるシミュレーションの高速化について予測した。その後 SACSIS2003 において、実際に地球シミュレータにソフトウェア(COSMOS90)をインストールして蛋白質のシミュレーションの実行スピードを計測した。その結果、1 ノードでの実行スピード(0.027 sec/step)は、予測した値(0.029 sec/step)と一致した。また、8 ノードで実行することによって、初めて 0.01 sec/step を超えた実行スピード(0.008 sec/step)を達成することができた。その後の、更なる高速化の結果について報告する。

High performance simulation of proteins by the earth simulator

MINORU SAITO,[†] KENJI SAYANO^{††}

In JSP2002, we estimated the execution speed of protein simulation on the earth simulator which is the fastest computer at the present time. In SACSIS200, we measured the execution speed of COSMOS90 on the simulator and compared with the predicted speeds. The execution speed using a node (0.027 sec/step) is almost the same as the predicted one (0.029 sec/step). We successfully executed a step of simulation within 0.01 sec using 8 nodes. In the presentation, we will show recent results of further speed up.

1. はじめに

生命科学の分野は、ヒトのすべての遺伝子が解読されるにつれて、激しい国際競争の時代に突入した。遺伝情報から作り出される膨大な数の蛋白質の構造と機能を解析するために、放射光からコンピュータサイエンスに至るまで、あらゆる科学技術が動員されている。我々の目的は、現在の世界最高速のスーパーコンピュータ(地球シミュレータ)によるハイパフォーマンスコンピューティングによって、蛋白質のシミュレーションがどの程度高速化されるかを明らかにすることである。そのために、パフォーマンスを出すためだけに作られたアプリケーションソフトでなく、蛋白質の研究で実績を積み重ねてきたソフトウェア(COSMOS90¹⁾)を地球シミュレータにインストールして高速化した。

一般に、蛋白質のシミュレーションは、膨大な計算量になる。その理由は、蛋白質やその周りの一部を単純化して計算量を省くことができないからである。蛋白質は、数千~数万個の原子からなる長い直鎖状分子が水溶液中(あるいは生体膜

中)で折り畳まれている。蛋白質は、様々な原子間相互作用が微妙にバランスして形を保っており、単純化はこのバランスを壊すことになる。したがって、蛋白質をリアルにシミュレーションするためには、蛋白質を取りまく水分子や膜分子を含めた数万~数十万原子に対してハイパフォーマンスコンピューティングを行わざるを得ない。

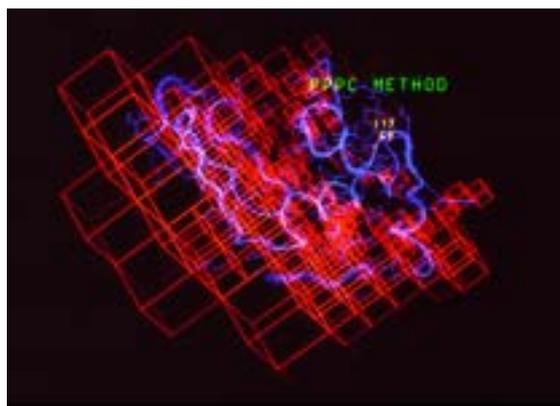


図 1 Particle-Particle and Particle-Cell (PPPC)法によって長距離クーロン力を $N \log N$ で計算する。図は、ヒトリゾチム蛋白質(130 アミノ酸:青色のリボンで表示)の 117 番目の C 炭素原子に加わるクーロン力を計算する際の空間分割を示す。

Fig.1 Long-range Coulomb interactions are calculated by the order $N \log N$ using Particle-Particle and Particle-Cell (PPPC) method. This figure shows space subdivision based on PPPC to calculate Coulomb forces acting to the 117th α -carbon atom of human lysozyme (130 amino acids).

[†] 弘前大学 理工学部

Faculty of Science and Technology, Hirosaki University,
E-mail: msaito@si.hirosaki-u.ac.jp

^{††} 産業技術総合研究所

National Institute of Advanced Industrial Science and Technology (AIST)

表1 スーパーコンピュータのCPUと通信のピークスピード

Table 1 Peak speed of CPU and communication of super-computers.

Machine	CPU (Gflops)	通信 (GB/s)
SR2201	0.3	0.3
VPP500 (Scalar)	0.2	0.4
(Vector)	1.6	0.4
VPP5000 (Scalar)	1.2	1.6
(Vector)	9.6	1.6
E.Simulator(Scalar)	1.0	
(Vector)	8.0	16.0
	64.0 (1 node)	12.3

蛋白質を高速にシミュレーションするためには、高速なアルゴリズムに基づいたアプリケーションソフトを高速なハードウェアに実装することである。筆者の一人(斎藤)は、シミュレーションのうちで最も時間のかかる原子間クーロン力の高速計算アルゴリズム(PPPC法と呼ぶ)を考案し、このアルゴリズムを採用したシミュレーションのプログラムCOSMOS90(Fortranで1万8千行)を開発した¹⁾。PPPC法は、Barnes&Hut tree codeで空間を祖視化することによって、 N 個の原子のクーロン力を $O(N^2)$ にかわって $O(N \log N)$ で処理する(図1参照)。COSMOS90は、蛋白質の標準シミュレーションプログラムAMBERに比べて約4倍高速であり、ベクトル化とVPP-Fortranによる並列化がなされている。

前回、COSMOS90を最新のベクトル型並列計算機VPP5000で実行した結果、わずか10PEsで標準的ソフト(AMBER)のSR2201上での最高スピード(0.2 sec/step: 256プロセッサで頭打ち)²⁾を上回った(0.018 sec/step)³⁾。最近、VPP5000よりも更に高速なベクトル型並列計算機NEC SX6が発表された。また、これに基づく地球シミュレーターが稼働を始めた。地球シミュレーターでCOSMOS90を実行することにより、更なる高速化が期待できる。

地球シミュレーターは、2003年7月現在、世界で最高速のスーパーコンピュータである(640ノードで35.86 Tflops)。気候変動や固体地球変動などのシミュレーションを高精度に行うために開発された^{4, 5)}。地球シミュレーターは、640個のノードからなる。1ノードは、8個のベクトルプロセッサからなる。ベクトルプロセッサのピーク演算性能は、8Gflopsである(スカラー性能は

loop for MD step

calculating bonded forces
(bond, angle, torsion)

calculating nonbonded forces based on PPPC
subdividing a system into hierarchical cubic cells
making an interaction table for P-P and P-C
calculating Coulomb and vdW forces from P and C

updating positions according to eq. of motion

communicating new positions

next step

図2 COSMOS90の構造。COSMOS90は、PPPCに基づいて長距離クーロン力を高速に計算してシミュレーションを行う。

Fig.2 Structure of COSMOS90. COSMOS90 simulates protein dynamics by efficiently calculating long-range Coulomb interactions using PPPC method.

1Gflops)。各ベクトルプロセッサは、互いに32GB/secのバンド幅でメモリを共有しており、メモリコピーの最大スループットは16GB/secである。一方、640個のノードは、単段クロスバススイッチで結合しており、理論最大スループットは12.3GB/secである⁶⁾(表1参照)。

地球シミュレーターセンターは、次のような分野に対してセンターを利用する共同プロジェクトを公募している。1.地球科学(大気・海洋)、2.地球科学(固体地球)、3.計算機科学、4.先進的の科学分野。ここで、先進的の科学分野とは、地球シミュレーターを使って画期的な成果が期待される分野を指している。筆者らの申請課題は、先進的の科学分野の中のバイオ分野からバイオシミュレーション委員会によって選抜され、地球シミュレーターセンターの課題選定委員会によって正式に採択された(2003年1月から継続中)。

2. プログラムの構造と高速化の方針

図2にCOSMOS90の処理の流れを示した。通常の分子動力学シミュレーションのソフトウェアと同様に、時間刻みのループの中で、原子間相互作用に起因する力の計算と運動方程式に基づいた座標の更新を行う。一般に、蛋白質の分子動力学シミュレーションでは、タイプの異なる多数の原子間相互作用を計算しなければならない。なかでも、クーロン相互作用は、長距離力であるため、系のすべての原子に及び、従って、計算時間の大部分を費やす。COSMOS90は、PPPC法によって、これを高速に処理している。空間を入れ子になった大小のセルに分割し(図1)それをもとに原子とセルとの相互作用テーブルを作る。このテーブルを基にして力の計算を行っている(図2)。

ベクトル化は、図2の時間刻みのループ内のすべての処理にわたって行っている。また、並列化は、空間分割を除いたすべての処理に対して、主にサイクリックなデータ分割に基づいて行っている。運動方程式によって更新された原子座標は、全通信によってすべてのプロセッサに送られる。このような方針に基づいて、まず、COSMOS90をVPP-Fortranによって並列化した。その後、地球シミュレータに移植するために、MPIによって並列化しなおした。ノード内もMPIで並列化するフラットプログラミングですべての計測を行った。また、VPP5000と地球シミュレータとの結果を比較するために、VPP-FortranからMPIに変更する際に、計算アルゴリズムを変えないようにした。

3. 計測の方針

計算時間の計測は、図3の水中におけるDNAと蛋白質(-repressor)との複合体(16034原子)について行った。最近、筆者の一人(斎藤)は、この系のシミュレーションをCOSMOS90を用いて行い、DNAと蛋白質(-repressor)との結合親和力の変化を高い精度で求めることに成功した(実験値 -1.8 kcal/molに対して計算値 -1.5 ± 0.4 kcal/mol)⁷⁾。この時と全く同じ計算をして、計算時間を計測することにした。パフォーマンスを出すために特殊な計算条件に設定して測定することを避けた。したがって、今回の計測結果のスピードは、実際の応用計算の実施の際に保証される性能である。

4. 計測結果

COSMOS90を地球シミュレータにインストールしてスピードを計測した結果を、表2に示す。表には、AMBERをSR2201で実行した結果、COSMOS90をVPP5000で実行した結果、更に、これらの結果と地球シミュレータのハードウェアの性能とを基にして、COSMOS90を地球シミュレータで実行する場合の予測結果(括弧内の数値)³⁾を示している。まず、COSMOS90によって地球シミュレータのベクトルプロセッサ自身の性能がどの程度引き出されているかを明らかにするために、ノード内の単一のプロセッサで実行した。実行スピードは、ベクトルで 0.189 sec/stepであり、スカラーで 1.593 sec/stepであった。すなわち、ベクトル化による加速は8.5倍であった。したがって、COSMOS90は、地球シミュレータのベクトルプロセッサのピーク性能(8 Gflops: スカラー性能1 Gflopsの8倍)を十分に引き出していると言える。

次に、COSMOS90が地球シミュレータの1ノ

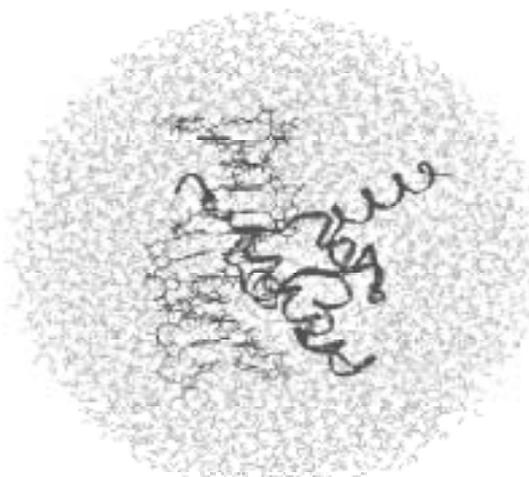


図3 水中のDNAと蛋白質の複合体(16034原子)
Fig.1 Protein-DNA complex in water (16034 atoms)

表2 図3のシミュレーションを1ステップ実行するのに要する時間(秒)(*)はJSPP2002での予測
Table 1 Execution time (sec) for 1step MD simulation of a protein in water (Fig.3). (*)Estimated in JSPP2002.

Machine/ No. of PE	Scalar	Vector
SR2201/ 1	21.88	-
128	0.30	-
256	0.20	-
VPP500/ 1	9.254	0.886
4	2.382	0.233
8	1.279	0.125
VPP5000/ 1	1.801	0.146
4	0.457	0.040
8	0.233	0.023
10	0.189	0.018
E.Simulator/ 1	1.593	0.189
1 node	-	0.027
	-	(0.029)
8 nodes	-	0.008
	-	(0.005)

ードの性能をどの程度引き出しているかを調べるために、8個のプロセッサからなる単一のノードで実行した。実行スピード(0.027 sec/step)は、単一ノードのピーク性能(64 Gflops)から予測した結果(0.029 sec/step)とよく一致した。したがって、COSMOS90は、地球シミュレータの1ノードの性能を十分に引き出しているといえる。次に、COSMOS90を地球シミュレータの8ノード(64プロセッサ)で実行して、スピードを計測した。実行スピードは 0.008 sec/stepであり、 0.01 sec/stepを超えた。

以下のホームページ

(<http://www.scripps.edu/brooks/Benchmarks/>)には、蛋白質の分子動力学シミュレーションソフトのベンチマークが掲載されている(表3)。これによれば、最高速は、NAMD2.4をピッツバーグ

表 3 他の蛋白質 MD ソフトウェアの計算スピード (sec/step) . [http://www.scripps.edu/brooks/Benchmarks/Pittsburgh Supercomputing Center の Lemieux \(Hewlett-Packard AlphaServer SC ES45/667MHz\)](http://www.scripps.edu/brooks/Benchmarks/Pittsburgh%20Supercomputing%20Center%20の%20Lemieux%20(Hewlett-Packard%20AlphaServer%20SC%20ES45/667MHz)%20による%20測定) による測定 Table 3 Performance speed of other MD software for proteins (sec/step)

processors	CHARMM c29b1	AMBER7	NAMD2.4
1	1.332	1.020	1.385
2	0.685	0.460	0.750
4	0.356	0.250	0.390
8	0.217	0.150	0.198
16	0.142	0.100	0.105
32	0.108	0.070	0.061
64	0.098	0.060	0.040
128	0.104	-	0.023

dihydrofolate reductase (DHFR) in a cubic periodic box (62.23 Å dimension). This protein consists of 159 residues. The number of water molecules is 7023. Total number of atoms is 23,558. Long-range Coulomb interactions are calculated by PME method.

スーパーコンピュータセンターの Lemieux (128PEs) で実行した結果である。水中の DHFR 蛋白質(23558 原子)で 0.023 sec/step である。我々が今回計測した系のサイズ(16034 原子)に換算すると、0.016 sec/step であり、我々の結果よりも遅い。我々の計測結果(0.008 sec/step) は、蛋白質のシミュレーションの実行スピードとしては、十分に高速であるといえる。しかし、地球シミュレータのハードウェアの性能から期待した予測スピード(0.005 sec/step) には、まだ達していない。その理由は、現在の MPI 版 COSMOS90 において MPI プロセス間の通信に必要以上の時間がかかっているためである。COSMOS90 の VPP-Fortran の通信制御文(UNIFY, MOVEWAIT) をそのまま MPI の集団通信関数に置き換えたために、MPI プロセス間の通信が十分に効率的に行われていない。今後、通信を効率化して、更なる高速化を行う。

5. 今後の展開

今回、実行スピードの計測に用いた COSMOS90 は、十三年間、実際の蛋白質の研究に用いられて数々の実績を積み重ねてきたソフトウェアである。このようなソフトウェアで、一ステップの実行時間が 0.01 秒を切る(0.008 sec/step) 高速化が、達成できたことによって、バイオ研究分野への波及効果は大きい。しかし、我々は現在のバージョンを使って、アプリケーションを実施することをせずに、更なる高速化を行う計画である。

参考文献

- 1) Saito, M.: Molecular dynamics simulations of proteins in water without the truncation of long-range Coulomb interactions, *Molecular Simulation*, Vol.8, pp.321-333 (1992).
- 2) 斎藤 稔, 三十尾潔高, 志澤由久, 丸川一志, 秋山 泰, 野口 保, 鬼塚健太郎, “分子動力学法プログラム AMBER と Barnes-Hut tree code の並列化による高速化,” 情報処理学会論文誌, Vol.40, No.5, pp. 2142-2151, (1999).
- 3) 斎藤稔, 佐谷野健二, “最新のベクトル型並列計算機による蛋白質のシミュレーションの高速化”, JSP2002, pp.179-180, (2002).
- 4) 谷啓二, 横川三津夫, “地球シミュレータ計画”, 情報処理, Vol.41, No.3, pp.249-254, (2000).
- 5) 横川三津夫, 谷啓二, “地球シミュレータ計画”, 情報処理, Vol.41, No.4, pp.369-374, (2000).
- 6) 上原均, 田村正典, 板倉憲一, 横川三津夫, “地球シミュレータの MPI 性能評価”, 情報処理学会論文誌, Vol.44, No.SIG1(HPS6), pp. 24-34, (2003).
- 7) Saito, M., Sarai, A., Free energy calculations for the relative binding affinity between DNA and -repressor, *Proteins*, Vol.52, pp.129-136 (2003).
- 8) 斎藤稔, 佐谷野健二, “地球シミュレータによる蛋白質の高速シミュレーション”, SACSIS2003, pp.169-170, (2003).