# Large Scale Simulations of Proteins on the Earth Simulator: Acceleration Performance by Vectorization and Parallelization

Project Representative
Minoru Saito Hirosaki University

Author
Minoru Saito Hirosaki University

The purpose of subgroup 1 is to computationally demonstrate and visualize the structural changes of hemoglobin using COSMOS90 which can efficiently simulate proteins in water with all degrees of freedom and long-range Coulomb interactions. The purpose of our second stage (in 2004) was the preparation of hemoglobin under the same conditions as biological studies and the acceleration of COSMOS90 for the hemoglobin system by vectorization and parallelization. We successfully obtained the performance speed 0.023 sec/step for hemoglobin in water (120036 atoms), where all subprocesses of COSMOS90 including Barnes-Hut tree code were vectorized and parallelized. This performance speed is faster than other typical software widely used in the world, CHARMM, AMBER7, and NAMD2.4. We are performing short time simulations for hemoglobin to check the reliability of our simulations.

**Keywords**: Molecular dynamics, Allosteric effect, Benchmark, Hemoglobin, Barnes-Hut tree.

## 1. Introduction

Various modern technologies (from computers to the accelerators) are utilized in the life science to study proteins. The accelerators become a powerful tool to explore 3-dimensional structures of proteins. High-speed computers become a necessary tool to simulate proteins and reveal their dynamical features. Proteins are large molecules consisting of thousands of atoms and have complicated structures. Furthermore, they largely fluctuate and easily change the whole structure even at the room temperature. Our purpose is to demonstrate large conformational changes of hemoglobin by performing realistic simulations. To perform the realistic simulations of proteins, we must include all atoms of proteins in water and their all interactions from chemical bonds to long-range Coulomb interactions.

The purpose of our project at the next stage is to computationally demonstrate large structural changes of proteins on the Earth Simulator using COSMOS90 tuned up in this study. As a target protein, we chose a hemoglobin molecule (Fig. 1). A hemoglobin molecule can efficiently transfer oxygen molecules from the lungs to the muscles. The binding of an oxygen molecule enhances additional oxygen bindings on other sites. Various experimental studies revealed that this cooperative binding is associated with large structural change. However, the experimental studies could not reveal the dynamical features of the structural changes, although they observed the structural difference between the initial and final states. The purpose of our group is to computationally demonstrate and visualize such structural changes of proteins using the Earth Simulator and software COSMOS90. We have installed the software COSMOS90 on the Earth Simulator to demonstrate large conformational changes of hemoglobin by performing long-time simulations.

COSMOS90 was developed by the author (M. Saito) in 1990 and made it possible to simulate a protein in water with all degrees of freedom and with long-range Coulomb interactions using the Particle-Particle and Particle-Cell (PPPC) method[1]. The PPPC method was proposed also by the

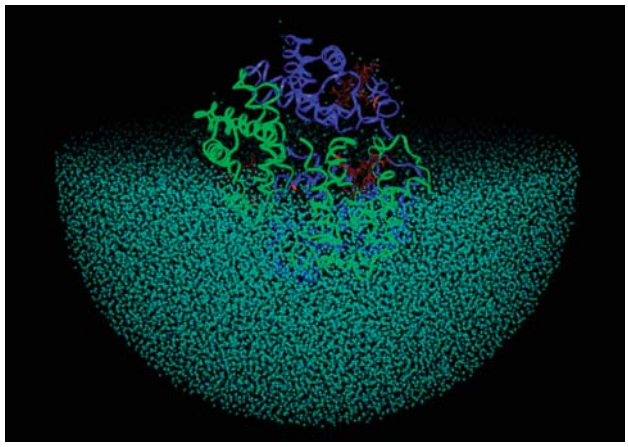

Fig. 1 Hemoglobin in water (120036 atoms). Radius of water sphere is 66 angstrom.[2]

author to efficiently calculate long-range Coulomb interactions between atomic charges in the order NlogN instead of $N^2$ by dividing a system into hierarchical cubic cells based on the Barnes & Hut tree code. In 2004, the author has tuned up COSMOS90 on the Earth Simulator by vectorizing and parallelizing its all subprocesses including the Barnes-Hut tree construction.

## 2. Parallelizaion of COSMOS90

A computation flow in COSMOS90 was shown in Fig. 2. COSMOS90 has a large loop for MD time step. In the MD time step loop, bonded (bond, angle, and torsion) and nonbonded (Lennard-Jornes and Coulomb) forces acting to atoms are computed. The nonbonded forces are efficiently computed in COSMOS90 by preparing an interaction table based on the PPPC method. The vectorization and parallelization tuning was performed for all subprocesses including the Barnes-Hut tree construction in the time step loop. For the parallelization based on MPI, processors inside a node were treated with the same manner as those between nodes (flat MPI programming).

The vectorization and parallelization of the subprocesses except for the Barnes-Hut tree construction[3] were described in the previous report in 2003. In this report, the vectoriztion and parallelization of the Barnes-Hut tree construction were described. The Barnes-Hut tree construction for the PPPC method in COSMOS90 was highly vectorized by Makino's algorithm[4]. We parallelized the Barnes-Hut tree construction subprocess in keeping with the vector acceleration by Makino's algorithm, as follows.

Processors of the Earth Simulator divide large cells in the lower level of the Barnes-Hut tree to construct a part of the Barnes-Hut tree (local tree), independently (Fig. 3). The

Loop for MD step

(1) Calculating bonded forces (bond,angle,torsion)

(2) Communicationg bonded forces

(3) Calculating nonbonded forces based on PPPC

- (3.1) Subdividing a system into hierarchical cubic cells
- (3.2) Communicating child cells and next cells
- (3.3) Making an interaction table for P-P and P-C
- (3.4) Communicating dipoles
- (3.5) Calculating charges and dipoles of cells
- (3.6) Communicating charges and dipoles of cells
- (3.7) Calculating Coulomb and vdW forces from P and C

(4) Updating positions according to eq. of motion

(5) Communicationg new positions

Next step

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

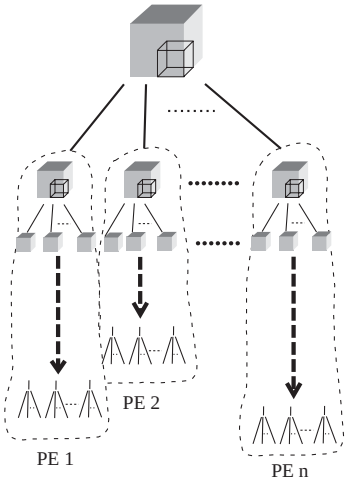


Fig. 3 Parallelization of Barnes-Hut space subdivision.[2]

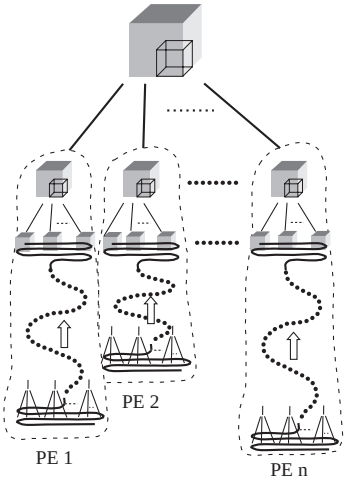


Fig. 4 Parallel calculation of charges and dipoles for cells. [2]

processors have the information to link the cells in the local tree as tables of the next and child cells. Each processor communicates each other the tables of the next and child cells by using MPI_allgather. If the number of processors is ≥ 8 and < 64, the 8 processors independently divide 8 child cells of the root cell. If the number of processors is ≥ 64, the 64 processors independently divide 64 grand child cells of the root cell. The processors that do not divide the cells simply receive the table of the next and child cells of the local trees constructed by the other processors. All processors utilize the tables of the next and child cells to search cells interacting with the particles from the root cell to particles in the whole tree and make the interaction tables between particle and cells. Then, the interaction table that is the largest array in COSMOS90 is distributed to the processors. The distribution of the interaction table clears the memory bottle neck occurred to simulate large scale MD simulations. The processors calculate charges and dipoles of the cells in the local Barnes-Hut tree from the particle toward the root cell and communicate the charges and dipoles with each other by using MPI_allgather (Fig. 4).

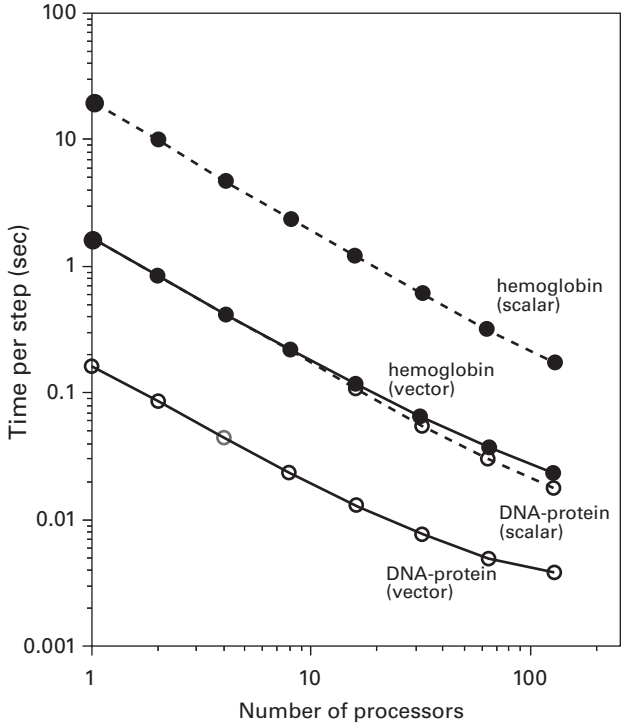


Fig. 5 Calculation speed of COSMOS90 on the Earth Simulator.[2]

### 3. Performance speed of COSMOS90

Hemoglobin was immersed in a water sphere of 66Å radius and consisted of 120034 atoms (number of protein atoms is 9066 and number of water molecules is 36990.). Ordinary super computers such as VPP5000 at Research Center for Computational Science (RCCS) in Okazaki do not have enough performance speed to simulate the hemoglobin system. Because of the limitation of the computer speed, previous simulations for hemoglobin were performed in the vacuum environment but not in the water environment.

The performance speed of COSMOS90 on the Earth Simulator was measured for hemoglobin in water (Fig. 1) for the various numbers of processors (from 1 to 128 processors). The performance speed measured was shown in Fig. 5. The performance of the present study guarantees us to carry out biologically productive simulations on the Earth Simulator with the same performance speed as in this study. The wall-clock time consumed to compute a step of MD was measured by MPI_Wtime of the MPI libraries for the Earth Simulator. The performance speed was measured for two load modules obtained by the vector and scalar compiles of a source code to clarify the acceleration ratio by the vectorization.

Results of the present study were shown in Fig. 5, where the scalar and vector performance speeds were plotted by dotted and solid lines, respectively. The performance speed of COSMOS90 on the Earth Simulator was continuously accelerated upon 128 processors. The maximum performance speed for hemoglobin was 0.023 sec/step for 128 vector processors. This performance speed is about five times faster than NAMD2.4 which consumes the same wall-clock time (0.023 sec) to perform a step of MD simulation for 23558 atoms on an alpha cluster (Lemieux at Pittsburgh Supercomputing Center)[5]. COSMOS90 on the Earth Simulator can performs biologically important simulations in a week under the realistic environmental condition.

The performance speed of a single processor was 1.59 sec/step for the vector module and 19.57 sec/step for the scalar module. Thus, the vectorization accelerated the performance speed for a single processor by 12.3 times faster than the scalar performance. This result means that COSMOS90 adequately brought out the vector ability of the single processor.

### 4. Purpose at the next stage

At the next stage, we will perform biologically productive simulations for hemoglobin in the same condition as the present study by submitting hundred L jobs on the Earth Simulator.

### Reference

(1) M. Saito: Molecular dynamics simulations of proteins in water without the truncation of long-range Coulomb interactions, *Molecular Simulation*, vol.8, pp.321-333 (1992).

(2) M. Saito: Large Scale Simulations of Proteins on the Earth Simulator: Acceleration Performance by Vectorization and Parallelization, *IPSJ Transactions on Advanced Computing Systems*, vol.46, No.SIG7 (ACS10), pp.9-17 (2005).

(3) Barnes, J. and Hut, P.: A hierarchical O (NlogN) force-calculation algorithm, *Nature*, Vol.324, pp.446-449 (1986).

(4) Makino, J.: Vectorization of a tree code, J. Comput. Phys., Vol.87, pp.148-160, (1990).

(5) http://www.scripps.edu/brooks/Benchmarks/
The performance speeds were measured on Lemieux (Hewlett-Packard alpha-server SC ES45/667MHz at Pittsburgh Supercomputing Center) for a protein, dihydrofolate reductase (DHFR), in a cubic box (62.23 Å). This system contains 23,558 atoms, 159 amino acids, and 7023 water molecules. Long-range Coulomb interactions were calculated by PME method.

# 地球シミュレータによる蛋白質の大規模シミュレーション：ベクトル化と並列化による加速性能

プロジェクト責任者
斎藤　稔　弘前大学
著者
斎藤　稔　弘前大学

我々のサブグループの目的は、生命を維持するうえで重要な蛋白質であるヘモグロビンの立体構造変化（アロステリック効果）を、分子動力学シミュレーションプログラムCOSMOS90を用いて追跡し可視化することである。COSMOS90は、水中の蛋白質を長距離クーロン力をカットオフせずに高速にシミュレートするプログラムである。2004年度の目的は、COSMOS90のすべてのコードを地球シミュレータ上でベクトル化し並列化することであった。次に、ヘモグロビンをシミュレートするためのセットアップを行うことであった。COSMOS90のコードについて、Barnes-Hut tree codeを含んだすべてをベクトル化し並列化することに成功した。セットアップを行った水中のヘモグロビンについて計測したところ、0.023 sec/stepであり、米国のNAMD2.4よりも約5倍高速であった。

キーワード：分子動力学シミュレーション，アロステリック効果，高速化，ヘモグロビン，Barnes-Hut tree