Analysis of the Function of a Large-scale Supra-biomolecule System by Molecular Dynamics Simulation System, SCUBA (Simulation Codes for hUge Biomolecular Assembly)

Project Representative
Atsushi Matsumoto Japan Atomic Energy Agency

Authors
Hisashi Ishida *1, Kei Yura *1, Takuma Kano *1 and Atsushi Matsumoto *1
*1 Japan Atomic Energy Agency

The Earth Simulator has the highest power ever achieved to perform molecular dynamics simulation of large-scale supra-molecular systems. We are developing a molecular dynamics simulation system, called SCUBA, which is designed to run a system composed of more than a million particles efficiently on parallel computers. This fisical year, the arrays used in the program of SCUBA were intensively optimized to reduce the amount of memory use. This optimization enabled SCUBA to perform molecular dynamics simulations of large-scale supra-molecular systems comprised of more than a million atoms on the Earth Simulator. Moreover, the Martyna-Klein-Tuckerman algorithm was extended to utilize the multiple time step method, which increases the time step length significantly. Then, in order to elucidate the dynamics of the 70S ribosome, molecular dynamics simulation of the 70S ribosome including its explicit solvent, a system which is composed of about two million atoms, has been performed using SCUBA. A model of a nascent polypeptide was included in the system to investigate how the nascent polypeptide passes through the exit tunnel within the large subunit of the 70S ribosome. The ratchet-like motion of the 70S ribosome, which is thought to be important for the genetic translation, was successfully observed.

Keywords: large-scale supra-biomolecular MD simulation, multiple time step method, 70S ribosome

1. Introduction

Molecular dynamics (MD) simulation not only provides dynamic descriptions of molecules on the atomic scale, but also provides valuable information for the interpretation of experimental data. The rapid development of computer power and the elucidation of the structures of biological macromolecules by X-ray crystallography and other experiments have increased the need for large-scale MD simulations in the field of biology.

We are developing an integrated molecular simulation system for biological macromolecules, called SCUBA (Simulation Codes for hUge Biomolecular Assembly), which is designed to run a system composed of more than a million particles efficiently on parallel computers.

SCUBA has several special features:

1. Topology of biomolecules
   The structure of SCUBA's program code is optimized for a system in which the topology of a biomolecular structure is considered.
2. A variety of force field parameters
   At present SCUBA can use the AMBER, CHARMM and GROMOS force field. The parameter file for the topology of biomolecules can be obtained from the PDB file by using the input module in SCUBA.
3. A variety of simulation methods
   A variety of simulation methods, such as energy minimization, molecular dynamics, free energy calculations, normal mode analysis, principal component analysis and so on, are included.
4. Algorithm for non-cutoff electrostatic interactions
   SCUBA utilizes the Particle-Particle Particle-Mesh (PPPM) algorithm, which efficiently calculates all the Coulomb electrostatic interactions [4]. This algorithm reduces the computational time required to calculate the electrostatic forces from the conventional $O(N^2)$ to $O(N\log N)$.
5. Input and output compatibility
   SCUBA's input and output file format is currently compatible with those used by AMBER [1].
6. Portability
   Written in Fortran90, SCUBA is designed to be easy to read, modify and extend. Users can easily maintain the existing code, and develop the current algorithms and integrate new ones.

7. Control files
   The control files for running SCUBA are described in a user-friendly manner.
8. Time-integral algorithms for long time steps
   SHAKE, and RATTLE which allow the time step taken to be larger by fixing the bond lengths and angles in the system are available.
9. Time-integral algorithms with high accuracy
   The Martyna-Klein-Tuckerman (MKT) algorithm which produces the correct ensemble thermodynamically is available. This fisical year, the MKT algorithm was extended to utilize the multiple time step (MTS) method, which increases the time step length significantly [5].
10. Parallelization
    SCUBA employs the domain decomposition (DD) method, which divides the volume of the physical system into rectangular subcells with a length longer than the potential cutoff radius. The processor assigned to a subcell needs to evaluate the interactions between the atoms in the subcell and between the atoms in 26 neighboring subcells. SCUBA employs the method for minimizing communication between processors proposed by D. Brown [6], which enables the number of processors between which data must be transferred to be reduced to only 7 of the neighboring subcells.
11. Vectorization
    In order to improve the performance of SCUBA on the Earth Simulator, the algorithm to calculate the interactions among the atoms is intensively vectorized.
12. Dynamic load balance
    To overcome the load imbalance associated with irregular atomic distribution, a dynamic load-balancing algorithm is implemented. Moreover, the number of processors used to calculate the PM part of PPPM can be optimized to minimize the computation by allowing the number to change.
13. High performance
    By intensive parallelization and vectorization, and by using the dynamic load balance mentioned in 10, 11, and 12, SCUBA has achieved both a high parallelization efficiency ratio and a high vectorization ratio. SCUBA has achieved a parallelization efficiency ratio of 75.8 %, and a vectorization ratio of 96.2 % even 45 nodes (360 processors) were used to perform an MD simulation for a system of RuvAB-Holliday junction complex which consisted of 546,725 atoms.
14. Optimization of memory use
    This fisical year, the arrays used in the program of SCUBA were intensively optimized to reduce the amount of memory use. This optimization enabled SCUBA to perform molecular dynamics simulations of large-scale supra-molecular systems comprised of more than a million atoms on the Earth Simulator.

2. Multiple time step method implemented in SCUBA

This fisical year, the Martyna-Klein-Tuckerman (MKT) algorithm was extended to utilize the multi-time step (MTS) method, which increases the time step length significantly [5].

In MD simulations, the forces associated with different degrees of freedom evolve on different time scales. In the biological system, the amount of computation of the short-range forces is light while that of the long-range forces is heavy. The problem with integrating the equations of motion is that the short-range interactions, whose calculations are not heavy, determine the time step limit.

To overcome this limitation and to increase the time step, several algorithms have been developed. One method, used by SHAKE and RATTLE, is to fix the bond lengths and angles in the system. Another approach is to use a multiple time step (MTS) method devised by the decomposition of the Liouville operator. The time evolution of the system can be written as a transformation from t = 0 to t = τ,

$$\Gamma(\tau) = e^{iL\tau}\Gamma(0)$$

where Γ is a multidimensional vector of independent coordinates and momenta of the system. L is the Liouville operator, which consists of three components, iL_1, iL_2, and iL_{bath}. iL_1 updates coordinates $\mathbf{q}$, and the size of the system. iL_2 updates momenta $\mathbf{p}$. iL_{bath} updates the parameters of barostat, thermostat, and adjusts the momenta corresponding to the changes of the bath variables. A detailed description of these equations is included in reference [5].

The basic idea of the MTS method is to integrate the fast motions at a small time step and the slow motions at a long time step. To realize the MTS in the framework of the MKT algorithm, the operator iL_2 is divided as:

$$iL_2 = iL_{\text{fast}} + iL_{\text{slow}}$$
$$iL_{\text{fast}} = \sum_{i=1}^{N} \mathbf{F}_i^{\text{fast}} \cdot \nabla_{\mathbf{p}_i}$$
$$iL_{\text{slow}} = \sum_{i=1}^{N} \mathbf{F}_i^{\text{slow}} \cdot \nabla_{\mathbf{p}_i}$$

Among the several decomposition schemes of the operator L, we utilize XO-RESPA [5], in which the time-evolution operator is factorized as:

$$\exp(iL\Delta t) = \exp\left(iL_{\text{bath}}\frac{\Delta t}{2}\right)\exp\left(iL_{\text{slow}}\frac{\Delta t}{2}\right) \times\left[\exp\left(iL_{\text{fast}}\frac{\delta t}{2}\right)\exp(iL_1\delta t)\exp\left(iL_{\text{fast}}\frac{\delta t}{2}\right)\right]^n \times\exp\left(iL_{\text{slow}}\frac{\Delta t}{2}\right)\exp\left(iL_{\text{bath}}\frac{\Delta t}{2}\right) + O(\Delta t^3)$$

where a small time step, δt = Δt/n (n is an integral which is more than 1), is used to update momenta of atoms due to the short-range forces.

In the MTS method, forces are classified into fast and slow

forces according to the frequency of the vibration. In general, forces associated with bonded interactions including bond stretches, torsional, short-range nonbonded interactions (van der Waals, electrostatics) within the cutoff length are classified as fast forces. In SCUBA, the electrostatic interactions outside the cutoff length are efficiently calculated as reciprocal-space interactions by the PPPM method. Although the reciprocal-space interactions are usually classified as long-range forces, an additional elaborate separation of slow forces from the reciprocal-space forces is implemented in SCUBA to further improve the accuracy of the MTS method. A detailed explanation of the separation of the reciprocal-space forces of PPPM is described in reference [7].

It is also beneficial for the parallel computations to use a MTS method which avoids the computation of the slow forces in PPPM at every small time step because the calculation of the reciprocal-space forces includes massive the fast Fourier transform (FFT) calculations, which have a lower parallelization efficiency ratio, and massive data communication among the parallel processors, which also reduces the parallelization efficiency ratio. This MTS method requires no more CPU time than SHAKE or RATTLE, and can lead to even larger improvements in speed.

As a natural enhancement of the separation of the forces into fast and slow forces, the factorization of the fast or slow forces into more forces can easily be done. In SCUBA's input file, users can assign the forces into three categories, fast, middle and slow forces.

3. Molecular dynamics simulation of the 70S ribosome

3.1 Introduction

Ribosome is one of the supra-biomolecules used in the process of translating genetic information to the amino acid sequence. It is composed of a small (30S) and a large (50S) subunit. The 30S subunit decodes genetic information, and the 50S subunit is responsible for polypeptide bond formation, and the elongation of the nascent polypeptide. The nascent polypeptide leaves ribosome through an exit tunnel.

A number of the structures of the 70S ribosome have been determined by X-ray crystallography and electron microscopy (EM). Before we started the molecular dynamics simulation of the 70S ribosome, we analyzed the conformational changes in the 70S ribosome by fitting an X-ray structure of the 70S ribosome into EM images which were observed under various reaction conditions from the initial to the final stage of the translation of messenger RNA. The fit structures were compared, and it was found that there were several exit tunnels with a diameter of more than 6 Å. The structure of a tunnel was changeable and its entrance was closed at the initial stage of protein synthesis [8].

However, the view of the conformational change of the tunnel by fitting an X-ray structure into EM images is static, not dynamic. Therefore, the dynamic properties of the tunnel are not yet well understood. In order to understand the dynamic properties of the tunnel, we started to perform an MD simulation of the 70S ribosome with a nascent polypeptide inside the tunnel.

3.2 Modeling of the nascent polypeptides

We used the atomic structure of *Thermus thermophilus* 70S ribosome, which was determined by X-ray crystallography (PDB code: 1YL3 and 1YL4, resolution: 5.5 Å) [9]. However, the atomic structure of the nascent polypeptide has not yet been determined by X-ray crystallography. Therefore, we modeled the structure of the nascent polypeptide. First, a number of straight polyalanine, which have 36 amino acids of alanine, were located along the tunnel. Figure 1 shows one of these polyalanines. Then, ten conformations, which have less atomic collisions between the modeled polypeptide and the wall of the tunnel, were chosen as appropriate candidates. In order to optimize the conformation of the polyalanine, a simulated annealing (SA) was performed for each polypeptide in vacuum. The distance-dependent dielectric constant of $4.0r$ with the value of r in Angstrom was used, and nonbonded interactions were evaluated with a cut-off radius of 14 Å. A time-step of 0.5 fs was used throughout the simulated annealing. After the SA simulations, the structure which had the minimum energy (summation of the internal energy of the polyalanine and interaction energy between the polyalanine and the ribosome) was selected as the model of the optimal structure.

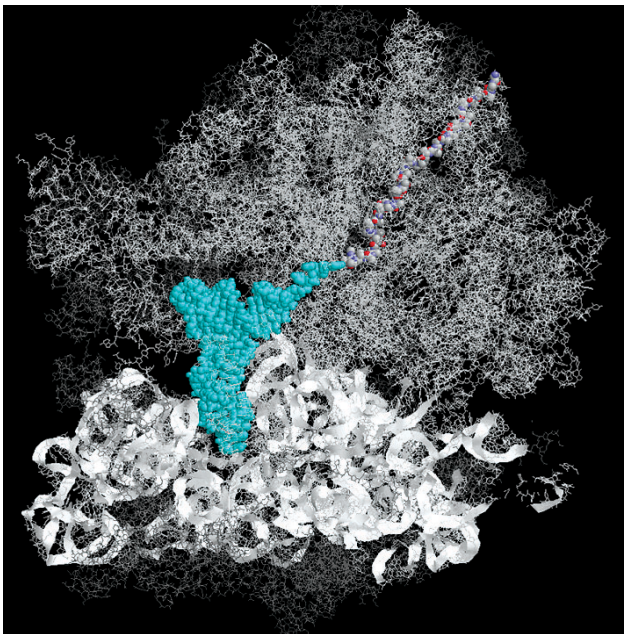

Fig. 1 Model of the nascent polypeptide inside the exit tunnel. Blue is the tRNA at P-site. From the PTC, the nascent polypeptide (CPK model) passes through the exit tunnel in the large subunit of the 70S ribosome. Large subunit, 50S, is drawn in wire-frame (upper side). Small subunit, 30S, is drawn in ribbon model (down side).

3.3 Molecular dynamics simulation of the 70S ribosome

Using SCUBA, we have carried out a molecular dynamic simulation of the 70S ribosome with the optimal polypeptide located inside the tunnel. The 70S ribosome is placed in a rectangle box with the size of 286 Å, 286 Å and 263 Å as shown in Fig. 2. In this box all the atoms of the 70S ribosome are separated more than 15 Å from the edge of the box. To neutralize the charges of the 70S ribosome, counter-ions were placed at the positions with large negative electrostatic potential. The 70S ribosome was then surrounded by 539,992 water molecules. Finally, the systems of the 70S ribosome were composed of 1,878,425 atoms (The 70S ribosome = 253,897, polyalanine = 363, counter-ion = 4,399 and water = 539,992 × 3). The dielectric constant used was 1.0 and the van der Waals interactions were evaluated with a cut-off radius of 8 Å. The PPPM was used for the electrostatic interactions. In the PPPM the reciprocal interactions were evaluated with grid spacing around 1 Å. The size of the charge grid was chosen to be $2^8 = 256$ for each dimension so that the fast Fourier transform could be applied to increase the speed of the calculation of the reciprocal-space term. The MD simulation of the system was carried out for several nanoseconds at a constant pressure of one bar and a temperature of 350 K. The temperature of 350 K was chosen to match the optimal conditions for the growth of *Thermus thermophilus*. MKT's constant temperature and pressure algorithm was used to control the temperature and pressure of the system. The MTS algorithm with a time step of 1 fs for short forces and 2 fs for long forces was used throughout the simulation to integrate the equations of motion.

3.4 Results

The average atomic fluctuations of the 70S ribosome in the thermal equilibrium were calculated as shown in Fig. 3. It was found that the regions near the boundary between the 50S subunit and the 30S subunit have unique features. In the 50S subunit, the L7/L12 stalk and L1 stalk were found to be especially flexible. The L7/L12 stalk consists of the L7/L12 dimer, protein L11 and helices H42-H44, and it is considered that the L7/L12 stalk is involved in the interaction with EF-G in the EF-G·GTP bound state. The L1 stalk consists of protein L1 and helices H76-H78, and the L1 stalk is considered to facilitate the release of the E-site tRNA and to be actively involved in the translational movement of tRNA from the P-site to the E-site. Therefore, it is considered that the flexibility of the L7/L12 stalk and L1 stalk is important for the function of ribosome. In the 30S subunit, there were no significant features. The atomic fluctuations in the 30S subunit were smaller over all as compared with those of the 50S subunit, although some outward regions have large fluctuations such as h6 at the region of spur, helix 33 and 37 in the region of head (h33–h40). It is considered that the 30S subunit moves rather like a rigid-body.

The atomic fluctuations at the boundary between the 50S subunit and the 30S subunit were also evaluated. It was found that the intermolecular B1a and B1b bridges had especially large fluctuations. B1a, which has been called the "A-site finger", consists of a long helical RNA arm (helix 38 of 23S rRNA which comes from the central protuberance) and S13 which is situated at the head of the 30S subunit. B1b is formed of only proteins:S13 and L5 which are situated near the P-site tRNA. Therefore, it is considered that the flexibili-

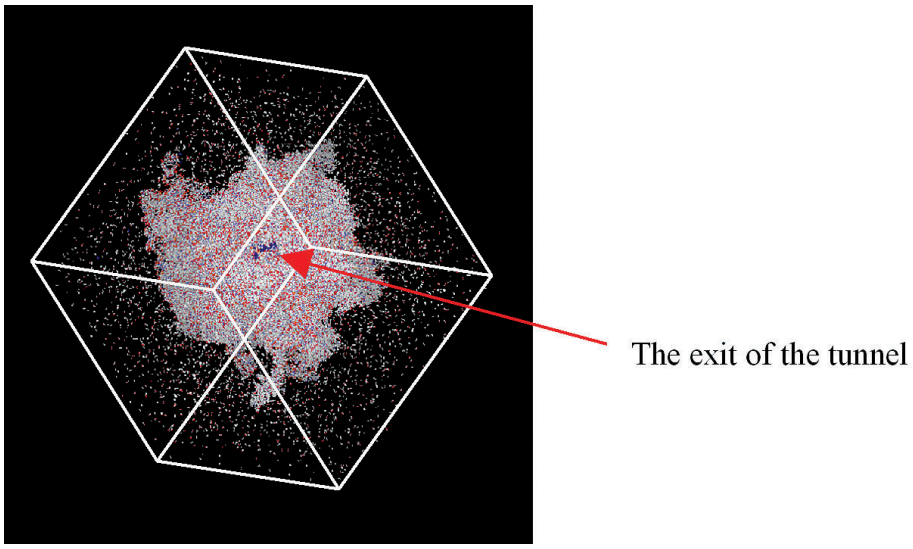

Fig. 2 70S ribosome, a supra-biomolecular complex, consisting of 4 RNA molecules and about fifty protein molecules. The size of the 70S ribosome system and the number of atoms in the system are 286 Å × 286 Å × 263 Å and 1,878,425 atoms, respectively. The exit of the tunnel can be seen in the center in this figure. Only 1% of water molecules are shown to show the 70S ribosome clearly.

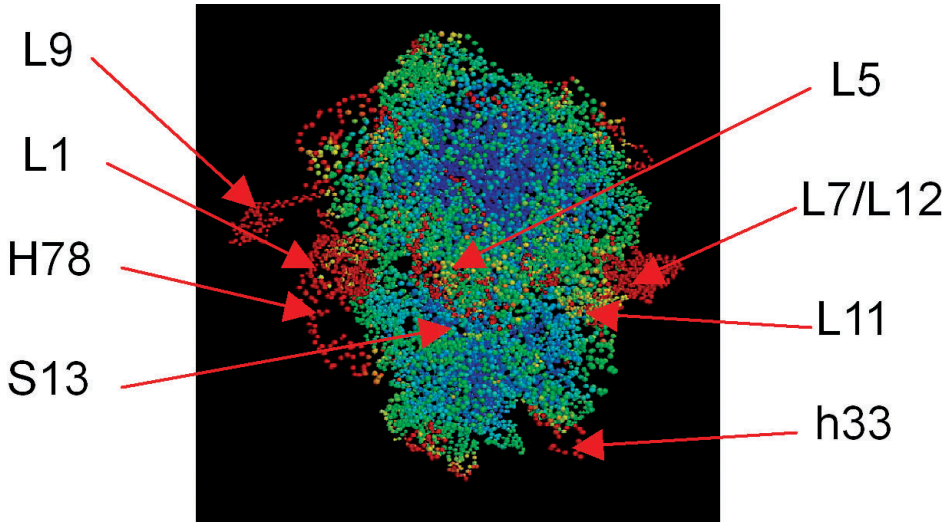

Fig. 3 The amplitudes of the thermal fluctuations of atoms in the 70S ribosome are color coded. The top half is 50S and the bottom half is 30S. The high amplitude of the thermal fluctuations is shown in red, and the low amplitude is shown in blue. Only the C_α atoms in the ribosomal proteins and phosphor atoms in the ribosomal RNA are shown.

ty of B1a and B1b is involved in a large-scale inter-subunit movement of the 50S subunit and the 30S subunit, which has been called "ratchet-like motion" [10].

Moreover, principal component analysis (PCA) was performed to measure the directions of the collective motions of the 70S ribosome. Even though our simulation time was rather short (about 1 nanosecond), it was revealed that a low frequency mode showed a so-called ratchet-like motion.

The structure of the polypeptide in the exit tunnel had a partially folded conformation. The part of the polypeptite which was in the middle of the tunnel was extended and the parts of the polypeptite that were near the entrance and exit of the tunnel were partially folded. This observation is consistent with the result obtained by an experiment which showed that the formation of polypeptide in the tunnel significantly depended on its location [11].

4. Conclusion

We have started to perform an MD simulation of the 70S ribosome using SCUBA, which achieves a high parallelization efficiency ratio and a high vectorization ratio. In the near future, we are going to perform an MD simulation of the nascent polypeptide as it passes through the exit tunnel to elucidate the dynamics of the polypeptide and the function of the 70S ribosome.

References

[1] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.R. Ross, T.E. Cheatham, III, S. DeBolt, D. Ferguson, G. Seibel and P. Kollman. AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules. *Comp. Phys. Commun.*, **91**, 1–41 (1995).

[2] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations, *J. Comp. Chem.*, **4**, 187–217 (1983).

[3] W. F. van Gunsteren and H. J. C. Berendsen. GROMOS: GROningen MOlecular Simulation software. Technical report, Laboratory of Physical Chemistry, University of Groningen, Nijenborgh, The Netherlands, (1988).

[4] R.W. Hockney and J.W. Eastwood, Computer Simulation Using Particles. McGraw-Hill, NY, (1981).

[5] G. J. Martyna, M. E. Tuckerman, D. J. Tobias and M. L. Klein, Explicit reversible integrators for extended systems dynamics, *Mol. Phys.*, **87**, 1117–1157 (1996).

[6] D. Brown, J. H.R. Clarke, M. Okuda and T. Yamazaki. A domain decomposition parallelization strategy for molecular dynamics simulations on distributed memory machines. *Comp. Phys. Commun.*, **74**, 67–80 (1993).

[7] R. Zhou, E. Harder, H. Xu and B. J. Berne, Efficient multiple time step method for use with Ewald and particle mesh Ewald for large biomolecular systems, *J. Chem. Phys.*, **87**, 1117–1157 (1996).

[8] H. Ishida, A. Matsumoto, Y. Tsutsumi and K. Yura, Conformational analysis of the structure of ribosome fit into electron microscopy density maps with normal mode analyses and molecular dynamics simulations, 16th International Microscopy Congress, Proceedings, 242, (2006).

[9] L. Jenner, P. Romby, B. Rees, C. Schulze-Briese, M.

Springer, C. Ehresmann, B. Ehresmann, D. Moras, G. Yusupova. and M. Yusupov, Translational Operator of mRNA on the Ribosome: How Repressor Proteins Exclude Ribosome Binding. *Science*, **308**, 120–123 (2005).

[10] J. Frank and R. K. Agrawal, A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature*, **406**, 318–322 (2000).

[11] J. L. Lu and C. Deutsch, Folding zones inside the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.*, **12**, 1123–1129 (2005).

分子動力学シミュレーションシステムSCUBA(Simulation Codes of hUge Biomolecule Assembly)を用いた大規模生体超分子系の機能解析

プロジェクト責任者
松本　淳　日本原子力研究開発機構

著者
石田　恒*1，由良　敬*1，叶野　琢磨*1，松本　淳*1
*1　日本原子力研究開発機構

地球シミュレータは従来にはない大規模生体超分子系の分子動力学シミュレーションを可能とする計算能力をもつ。我々は大規模な生体超分子系を扱う分子動力学シミュレーションシステムSCUBAを開発している。SCUBAは長距離相互作用を高速かつ高精度に計算するPPPM計算法、系のエネルギー、温度、圧力を一定に保つ様々な時間積分アルゴリズムなど、最新のアルゴリズムを採用した計算性能に優れたシミュレーションシステムである。SCUBAは、地球シミュレータ360プロセッサ使用時でベクトル化率95%以上、並列化効率50%以上の優れた性能を達成している。本年度はSCUBAが必要とするメモリ使用量を最適化するためのチューニングをすることにより、百万原子以上からなる生体超分子の分子動力学シミュレーションを地球シミュレータ上で実行可能とした。更に、マルチ時間ステップ法を装備することにより長時間シミュレーションを実行可能とした。そして、約200万原子からなるリボソーム(遺伝情報を翻訳する生体超分子)と新生ポリペプチドの複合体の系について分子動力学シミュレーションを実行した。結果、リボソームの機能に重要と考えられているねじれ運動を再現することに成功した。

キーワード:大規模生体超分子動力学シミュレーション, マルチ時間ステップ法, リボソーム