ARTICLE

Molecular Dynamics Simulation System for Structural Analysis of Biomolecules by High Performance Computing

Hisashi ISHIDA^{1,2,*}

¹Center for Computational Science and Engineering ²Quantum Beam Science Directorate Japan Atomic Energy Agency, 8-1-7 Umemidai, Kizugawa, Kyoto 619-0215, Japan

We have developed an integrated molecular simulation system for biological macromolecules, called SCUBA (Simulation Codes for hUge Biomolecular Assembly) which is designed to run a biomolecule system composed of more than one million particles efficiently on parallel computers. SCUBA is now being used not only for molecular dynamics (MD) simulations but also to determine the 3D structure of supra-macromolecules. For the latter purpose, a new method to fit a high-resolution X-ray structure at a certain reaction state into low-resolution electron microscopy (EM) data at different reaction states has been developed. In our method, the fitting is carried out using MD simulations in explicit water medium; the target function considers restraints to fit the X-ray structure into an EM density map, and the all-atom interactions for all the molecules including the water medium. This method was implemented into SCUBA and applied to ribosome, one of the supra-biomolecules utilized for translating genetic information to the amino acid sequence. The system was composed of more than 2 million atoms. The method showed that SCUBA can carry out the fitting simulation with a high parallelization efficiency ratio of more than 50% using 512 CPU cores of PRIMERGY BX900 supercomputer at the Japan Atomic Energy Agency.

KEYWORDS: molecular dynamics simulation, determination of 3D biomolecule structure, electron microscopy, ribosome, high performance computing

I. 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 have developed an integrated molecular simulation system for biological macromolecules, called SCUBA (Simulation Codes for hUge Biomolecular Assembly) which is designed to run a biomolecule system composed of more than one million particles efficiently on parallel computers.¹⁾ SCUBA showed an excellent performance for a variety of biomolecular systems such as Holliday junction of recombinant DNA and the 70S ribosome.²⁻⁴⁾

SCUBA is now being used not only for MD simulations but also to determine the 3D structure of supra-macromolecules. For the latter purpose, a new method to fit a high-resolution X-ray structure at a certain reaction state into low-resolution electron microscopy (EM) data at different reaction states has been developed. There have already been several refinement methods developed for this purpose.⁵⁻¹⁰⁾ Molecular dynamics flexible fitting (MDFF) is one of the major programs used for this aim. In this program, the target function used considers restraints to fit the X-ray structure into an EM density map, and the atomic interactions among biomolecules. However, this method does not necessarily consider accurate atomic interactions such as in water medium because the calculation of such interactions is very time consuming. Moreover, to retain the secondary structure such as α-helices, β-sheets and base-paired nucleotide residues, strong dihedral-angle restraints are imposed on α -helices and β -sheets, and strong hydrogen-bond distance restraints are imposed on base-pairs during the fitting simulation. The structures determined by these methods are not necessarily accurate at the atomic level. A few studies use a few layers of water medium for the fitting.¹⁰⁾ Although this amount of water may be sufficient to construct the fit structure, it would not be sufficient if dynamics such as the conformational transition between the first atomic model (usually from an X-ray structure) and the final structure (usually fit in an EM structure) is also investigated.

In our new method, the fitting is carried out using MD simulations in explicit water medium not only to construct structural models which fit in the EM density maps but also to investigate the conformational transitions. This method was implemented into SCUBA and applied to ribosome, one of the supra-biomolecules utilized for translating genetic information to the amino acid sequence. We used EM data which have been observed by cryo-EM and single-particle reconstruction,¹¹⁾ and an atomic X-ray structure of the 70S ribosome complexed with tRNAs and elongation factor G

^{*}Corresponding author, E-mail: ishida.hisashi@jaea.go.jp

^{© 2011} Atomic Energy Society of Japan, All Rights Reserved.

(EFG).¹²⁾ Our aim is to build atomic structures of the 70S ribosome and analyze the conformational change during translation. The system was composed of more than 2 million atoms including water molecules. The new method showed that SCUBA can carry out the fitting simulation efficiently on PRIMERGY BX900 supercomputer at the Japan Atomic Energy Agency (JAEA).

II. Methods

1. Molecular Dynamics Simulation System, SCUBA

MD simulation for biomolecules is performed by solving the equation of motion using an energy function:

$$U_{str} = \sum_{bonds} K_r \left(r - r_{eq}\right)^2 + \sum_{angles} K_{\theta} \left(\theta - \theta_{eq}\right)^2 + \sum_{dihedrals} \frac{K_{\varphi}}{2} \left(1 + \cos\left[n\varphi - \gamma\right]\right) + \sum_{i < j}^{atoms} 4\varepsilon_{ij} \left[\left(\frac{R_{ij}^{\min}}{r_{ij}}\right)^{12} - \left(\frac{R_{ij}^{\min}}{r_{ij}}\right)^6 \right] + \sum_{i < j}^{atoms} \frac{q_i q_j}{Dr_{ij}}$$
(1)

 U_{str} consists of both bonded interaction terms (including bond stretching, bond angle bending and torsion angle twisting potentials) and non-bonded interaction terms (including van der Waals and electrostatic potentials).

r and θ are the values of the bond lengths and angles and r_{eq} , and θ_{eq} are their equilibrium values, respectively. K_r and K_{θ} are bond and angle force constants respectively. The dihedral term includes parameters for the dihedral force constant K_{φ} , the dihedral angle φ , the dihedral multiplicity *n*, and the phase shift γ . Non-bonded interactions between pairs of atoms *i* and *j* at a relative distance r_{ij} are represented by the last two terms. The van der Waals energy is calculated with Lennard-Jones 6-12 (LJ) potential parameters of R_{ij}^{\min} and ε_{ij} : R_{ij}^{\min} is the distance between atoms *i* and *j* at which the LJ potential is equal to zero, and ε_{ij} is the well depth of the LJ potential. Coulomb electrostatic energy is calculated by *D* (a dielectric constant) and q_i (the partial atomic charge on atom *i*).

The programming of Eq. (1) demands meticulous work due to the complicated unique topology of bimolecules. To realize such high performance computing, State-of-the-art techniques were used for the development of SCUBA.

SCUBA has several special features:

(1). A variety of force field parameters

At present SCUBA can use the AMBER, CHARMM and GROMOS force fields.

(2). 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.

(3). 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.

(4). 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 evaluates the interactions between the atoms in the subcell and between the atoms in the neighboring subcells.

(5). Vectorization

In order to improve the performance of SCUBA on vector machines such as the Earth Simulator, the algorithm to calculate the interactions among the atoms is intensively vectorized.

(6). Dynamic load balance

To overcome the load imbalance associated with irregular atomic distribution, a dynamic load-balancing algorithm to keep the subcells equally reassigned to the processors during simulation is implemented.

(7). High performance

By intensive parallelization and vectorization, and by using the dynamic load balance mentioned in (4), (5) and (6), SCUBA has achieved both a high parallelization efficiency ratio and a high vectorization ratio on the Earth Simulator. An MD simulation for a system of RuvAB-Holliday junction complex, which consisted of about 5.5×10^5 atoms, achieved a parallelization efficiency ratio of 75.8%, and a vectorization ratio of 96.2% even when 45 nodes (360 CPU cores) were used.¹⁾ At present, using 512 CPU cores of PRIMERGY BX900 supercomputer at JAEA, SCUBA has achieved a parallelization efficiency ratio of more than 50%. (This will be discussed in details at Section IV.)

2. Algorithm for EM-Fitting Simulation

Single particle analysis using an electron microscope has been used to observe the structure of supra-biomolecules. However, the resolution of the EM image has not achieved the atomic level. Several studies have been carried out to determine the 3D structure of supra-biomolecules at atomic resolution by fitting the constituent molecules, determined by X-ray crystallography, into an EM density map. In those attempts, each constituent molecule is often fit into the EM density map manually, and the constituent molecules may have atomic collisions at their interfaces. To avoid the collision, we have developed a flexible fitting method to alleviate the steric strain of atoms while retaining the condition that the constituent molecules fit the EM density map.⁵⁾ However, because this previous method was used under the condition of a vacuum, a strong restraint was required to maintain the secondary structures, which may cause the structure to be unnatural.

Our new method utilizes an MD simulation method. First, from a given set of atomic coordinates, the simulated EM density at a *k*-th voxel point, \mathbf{r}_{M_k} can be calculated as:

$$\rho_{\rm sim}(\mathbf{r}_{M_k}) = \sum_i \rho_i G(\mathbf{r}_{M_k} - \mathbf{r}_i), \qquad (2)$$

where ρ_i is the atomic mass in the atomic mass unit and \mathbf{r}_i is the atomic coordinate of atom *i*. *G* is the Gaussian function which distributes ρ_i to the voxel points around the atom. $\rho_{sim}(\mathbf{r}_{M_k})$ is calculated as the sum of the contributions from the neighboring atoms. The resolution of the EM density map we used here is defined by half the width of a Gaussian

in Fourier space.

In order to fit the atomic model into the EM density map, we set a target function which is used to constrain the whole structure into an EM density map, in addition to the standard all-atom energy function which imposes physical constraints on the structure.

$$U_{total} = k_{EM} \cdot U_{EM} + U_{str}, \qquad (3)$$

where, U_{str} is the standard all-atom energy function of Eq. (1), and k_{EM} is a constant weight value. U_{EM} is the function to fit the modeled structure into the EM density map. Here we employ the correlation coefficient (CC) of the simulated and computed densities times -1 as U_{EM} :

$$U_{EM}\left(\{\mathbf{r}_{M}\}_{\exp},\{\mathbf{r}_{M}\}_{\sin}\right) = -\frac{\sum_{k=1}^{n} \rho_{\sin}(\mathbf{r}_{M_{k}}) \cdot \rho_{\exp}(\mathbf{r}_{M_{k}})}{\sqrt{\sum_{k=1}^{n} (\rho_{\exp}(\mathbf{r}_{M_{k}})^{2}) \sum_{k=1}^{n} (\rho_{\sin}(\mathbf{r}_{M_{k}})^{2})}} = -CC, \quad (4)$$

where $\{\mathbf{r}_{M}\}_{exp}$ and $\{\mathbf{r}_{M}\}_{sim}$ represent the sets of voxel points, $\mathbf{r}_{M_{1}}, \mathbf{r}_{M_{2}}, \dots, \mathbf{r}_{M_{n}}$, for the experimental and simulated EM densities respectively. $\rho_{exp}(\mathbf{r}_{M_{k}})$ and $\rho_{sim}(\mathbf{r}_{M_{k}})$ are the experimental and simulated EM densities at a voxel point $\mathbf{r}_{M_{k}}$, respectively. *n* is the total number of voxel points in the system. The fitting is then performed by minimizing the target function. The force derived from the derivative of CC is described as:

$$\mathbf{F}_{EM}(\mathbf{r}_{M_k}) = -\frac{\partial U_{EM}}{\partial r} = \frac{\sum_{k=1}^{n} \rho_{\exp}(\mathbf{r}_{M_k}) \cdot \nabla \rho_{\sin}(\mathbf{r}_{M_k})}{\sqrt{\sum_{k=1}^{n} \left(\rho_{\exp}(\mathbf{r}_{M_k})^2\right) \sum_{k=1}^{n} \left(\rho_{\sin}(\mathbf{r}_{M_k})^2\right)}}$$
(5)
$$+\frac{\sum_{k=1}^{n} \rho_{\exp}(\mathbf{r}_{M_k}) \rho_{\sin}(\mathbf{r}_{M_k})}{\sum_{k=1}^{n} \left(\rho_{\sin}(\mathbf{r}_{M_k})^2\right)} \cdot \frac{\sum_{k=1}^{n} \rho_{\sin}(\mathbf{r}_{M_k}) \cdot \left(-\nabla \rho_{\sin}(\mathbf{r}_{M_k})\right)}{\sqrt{\sum_{k=1}^{n} \left(\rho_{\exp}(\mathbf{r}_{M_k})^2\right) \sum_{k=1}^{n} \left(\rho_{\sin}(\mathbf{r}_{M_k})^2\right)}}$$

It should be noted that the force in Eq. (5) is not calculated at the atomic position but at the voxel points. The force at each particle is given as:

$$\mathbf{F}_{EM}(\mathbf{r}_i) = \sum_{k=1}^{n} \left\{ \mathbf{F}_{EM}(\mathbf{r}_{M_k}) \times \frac{\rho_i \cdot G(\mathbf{r}_{M_k} - \mathbf{r}_i)}{\rho_{sim}(\mathbf{r}_{M_k})} \right\}.$$
 (6)

where $\rho_i \cdot G(\mathbf{r}_{M_k} - \mathbf{r}_i) / \rho_{sim}(\mathbf{r}_{M_k})$ is the contribution ratio of an *i*-th atom to $\rho_{sim}(\mathbf{r}_{M_k})$ (see Eq. (2)). If there is no contribution of the *i*-th atom to $\rho_{sim}(\mathbf{r}_{M_k})$ then $\rho_i \cdot G(\mathbf{r}_{M_k} - \mathbf{r}_i) / \rho_{sim}(\mathbf{r}_{M_k})$ is zero, while if only the *i*-th atom contributes to $\rho_{sim}(\mathbf{r}_{M_k})$ then $\rho_i \cdot G(\mathbf{r}_{M_k} - \mathbf{r}_i) / \rho_{sim}(\mathbf{r}_{M_k})$ is one. In this study, we call \mathbf{F}_{EM} an EM density force. The total force which should be imposed on the atomic positions can be described as:

$$\mathbf{F}_{total}(\mathbf{r}_{i}) = k_{EM} \cdot \mathbf{F}_{EM}(\mathbf{r}_{i}) + \mathbf{F}_{str}(\mathbf{r}_{i}), \qquad (7)$$

where $\mathbf{F}_{str}(\mathbf{r}_i) = -\nabla_{\mathbf{r}_i} U_{str}$ is the atomic interaction force derived from Eq. (1). Thus, the atomic positions at the next step can be calculated using the force of Eq. (7) during either the minimization or MD simulation. The updated atomic positions are evaluated by the target function in Eq. (3) and the stepwise procedure from Eq. (3) to Eq. (7) is repeated until the target function has been sufficiently minimized. Thus, fitting of the simulated structure can be carried out so that the simulated coordinates fit into the experimental EM density map without causing steric strain. Practically, instead of using k_{EM} , it is convenient to use w_{EM} :

$$w_{EM} = \left\langle k_{EM} \cdot \left| \mathbf{F}_{EM}(\mathbf{r}_i) \right| \right\rangle / \left\langle \left| \mathbf{F}_{str}(\mathbf{r}_i) \right| \right\rangle$$
(8)

where $\langle \rangle$ shows the average among atoms. If $w_{EM} = 1$. 0 then k_{EM} is set to be $\langle k_{EM} | \mathbf{F}_{EM} | \rangle$, the averaged amplitude of the EM-fitting force used in EM-fitting simulations, is the same as that of \mathbf{F}_{str} .

III. Materials

1. X-Ray and EM Structure of Ribosome

We used the target EM data of *Escherichia coli* 70S ribosome, which were observed under reaction conditions in tRNA translocation. The data were selected because these data suggest a global and concerted motion of the 70 S ribosome.^{9,11} The data were obtained from the electron microscopy database at the European Bioinformatics Institute; EMD-1365 and EMD-1363¹¹ corresponding to the intermediate and pre-translocational states respectively (**Table 1**).

To prepare the initial atomic structure for the fitting, we used the atomic structure of *Thermus thermophilus* 70S ribosome (containing the large subunit 50S, the small subunit 30S, EFG, two tRNA molecules and an mRNA) which was determined by X-ray crystallography (PDB code: 2WRJ, resolution: 3.6 Å)¹²⁾ (**Fig. 1**). Hydrogen atoms were added to the structure in order to construct an all-atom structure, which is composed of $\approx 2.67 \times 10^5$ atoms.

2. Energy Minimization

In order to alleviate steric strain in the all-atom structure, energy minimization of Eq. (3) with the weight $w_{EM} = 0.0$ was carried out for the 70S ribosome in a vacuum. Harmonic constraints with a force constraint of 10 kcal/mol Å² were applied to the heavy atoms. The distance-dependent dielectric constant of 4.0*r* with the value of *r* in Angstrom was used for electrostatic interactions. The non-bonded interac-

 Table 1
 The content of the EM data and results of fitting

EM data ID	1365 ¹¹⁾	1363 ¹¹⁾
Functional state	Intermediate	Pre-translo
		cational
Resolution [Å]	11.75	10.9
Voxel size [Å]	2.82	2.82
Initial CC by method in II-3 [5]	0.822	-
Initial CC by method in this study	0.826	0.850*
Final CC	0.884	0.877

* is the value of CC for the structure fit into EMD-1365



Fig. 1 The X-ray structure of the 70S ribosome complexed with two tRNAs and EFG.¹²⁾ 50S, 30S, E-tRNA, P-tRNA, EFG and L10-L12 stalk are shown in white, yellow, purple, green, red and cyan, respectively. They are all depicted as wire models.

tions were evaluated with a cut-off radius of 14 Å. The modeled 3D structures were minimized for 1,000 steps of the steepest descent method followed by 10,000 steps of the conjugate gradient method.

3. Alignment of EM Density Maps to the Energy-Minimized Structure Using Rigid-Body Fitting Method

The initial position and orientation of the whole atomic model should be as similar to the target position and orientation in the EM density map as possible. First, we used a rigid-body fitting method by regarding the whole atomic model as a rigid-body. The high-values voxel points in the EM density map were selected so that the volume occupied by the voxel points was equivalent to the molecular volume. The model was fit into the map by matching moments of inertia between the atomic model and selected voxel points. Finally, the structure was slightly translated and rotated until the position and orientation of the molecule fit into the target EM density map as closely as possible.⁹

4. Rigid-Body Fitting Using MD Simulation in a Vacuum

It should be noted that the target function used in the previous section III. 3 is to fit the data points into the molecular volume of ribosome.⁹⁾ In contrast, the target function for fitting in this study is the correlation coefficient (CC), and all the data which are positive were used for the EM-fitting. Therefore the fit structure obtained by the previous method in Section III. 3 can be different from the structure obtained by the method using MD simulation.

In the rigid-body fitting, a simulated-annealing MD algorithm was employed because the optimization of molecular structures can be achieved without causing them to get trapped at local minima. To perform large-scale simulations rapidly, a dynamic load balance algorithm described in Section II. 1, which optimizes the load balance among the processors, was employed. If the processors are equally assigned to the same sub-space of the system divided by the domain-decomposition method, the irregular distribution of the atoms in a vacuum lowers the efficiency of the domain-decomposition method as time elapses. The dynamic load balance is especially effective for our MD simulations performed in a vacuum, where the distribution of atoms in the system is heavily heterogeneous. By using the dynamics load balance we could increase the computational speed by about 2-fold.

The rigid-body fitting simulation was applied for EMD-1365. The system was heated to be 100,000 K using $w_{EM} = 1.0$. The high temperature of 100,000 K comes from the fact that the number of variables of freedom in the system is small in the first stage. The temperature was gradually decreased from 100,000 K to 0 K. When the CC converged at a maximum, the structure of the 70S ribosome was selected as the initial structure for the following flexible fitting. The total time of the simulated-annealing to optimize the 3D structure of the 70S ribosome was about 100 ps.

The CC for EMD-1365 increased from an initial value of 0.822 up to 0.826. The atoms deviated from the initial positions by less than 1.09 Å on average. Considering that the resolution of EMD-1365 is rather low, 11.75 Å, it was concluded that the rigid-body fitting using the previous method in Section III. 3^{9} and the rigid-body fitting using MD simulation in this study give substantially the same result of the fitting.

5. Preparation for Flexible Fitting in Water Medium

To neutrize the system optimized by the rigid-body fitting simulation, magnesium ions were used. Moreover, excess ions, including magnesium, potassium, and chloride, were added in the box at concentrations of 0.1 M KCl and 7 mM MgCl₂. A buffer of water around the solute of at least 15 Å in all directions was used. The total system contained $\approx 2.11 \times 10^6$ atoms (including $\approx 6.11 \times 10^5$ water molecules)

The system was first heated from 0 K to 350 K within 500 ps during which the molecules and bound ions were fixed with decreasing restraints and the water molecules and excess ions were allowed to move. After the restraints were removed, the system was equilibrated for 1 ns with no restraint at a constant pressure of one bar and a temperature of 350 K. Then the box-size was fixed, and the system was equilibrated for a further 1ns at a constant temperature of 350 K. During the equilibration the translational and rotational movements were removed. To integrate the equation of motion, the multi-time-step (MTS) algorithm was used with a time step of 1 and 2 fs for short and medium forces with the smooth cut-off of 8 Å and 4 fs for other long-range forces. The time step for the EM-fit algorithm was set at 8 fs. For the EM-fit, all the density values more than zero were used.

6. Optimal Weight for EM-Fitting Force

 w_{EM} should be large enough to fit the atomic model into an EM density map, while w_{EM} should be small enough not to



Fig. 2 (a) The time evolution of the correlation coefficient in Eq. (4). The lines are for $w_{EM} = 1.0 \times 10^{-3}$ (inverted triangle), 5.0×10^{-4} (square), 4.0×10^{-4} (diamond), 3.0×10^{-4} (green circle), 2.0×10^{-4} (red circle), 1.0×10^{-4} (black circle) and 0.0 (triangle), from top respectively. (b) The number of residues forming α -helices in proteins in the 70S ribosome. The orders of the lines are the same as (a).

overfit it too much. If all the atoms are fit into a medium or low-resolution map by using too large w_{EM} , atoms will spread uniformly to make a blurred map.

To assess the optimal weight for EM-force, the CC and the number of residues forming α -helix in proteins in the 70S ribosome was evaluated. Figure 2(a) shows the time evolution of the CC. With $w_{EM} = 0$, the CC decrease slowly because ribosome slid slowly from the EM density map. With $w_{EM} \ge 1.0 \times 10^{-4} (0.01\%)$, the CC increases. Figure 2(b) shows the time evolution of the number of α -helices. α -helices start to corrupt if $w_{EM} \ge 3.0 \times 10^{-4}$, indicating that overfit of the atomic structure into the EM map has occurred. If $w_{EM} \leq 2.0 \times 10^{-4}$ is used, α -helices are maintained. This tendency in the change in the number of residues in α -helices in Fig. 2 was observed with regards to the β -sheets (not shown in figure). It is generally considered that the maintenance of local sub-structures such as secondary structures is favorable during EM-fitting simulations.⁷⁾ Consequently, it is concluded that w_{EM} should not exceed 2.0×10⁻⁴ (0.02%) to maintain the secondary structures such as the α -helices and β-sheets. (In the actual fitting simulation mentioned in Section IV, a long-time simulation of about 100 ns with a smaller w_{EM} of less than 1.0×10^{-4} was performed to increase the CC slowly ont only to obtain the fit structure but also to investigate the conformational changes at the atomic level.)

It should be noted that the EM force with $w_{EM} = 2.0 \times 10^{-4}$, which seems to be much smaller than the atomic interaction force, does not necessarily mean that the total EM force at a certain area is also small. If the area does not match the EM density area, then the direction of the EM force of the atoms in this area will be similar because the atoms in this area should move in the same direction to match the EM density area. In this case, the total EM force could be quite large to move the area into the EM density map. In contrast, the directions of the atomic interaction forces in this area may be varied, and the total atomic interaction force could be small-

er than the total EM density force.

IV. Performance of SCUBA with EM-Fitting Algorithm

While SCUBA is parallelized by using the domain decomposition method, the EM-fitting method was parallelized by using the slab decomposition method. The length of the slab was set to be the same size as the voxel size, 2.82 Å, of the EM density maps. Each CPU core is assigned to deal with the atoms in a certain range of slabs and calculate the simulated EM density values in Eq. (2) and the EM density force in Eq. (6).

Figure 3 shows the performance of SCUBA. The standard performance of SCUBA (without using the EM-fitting algorithm) has a parallelization efficiency of more than 50% using 512 CPU cores. Although the performance of SCUBA using the EM-fitting algorithm deteriorates slightly due to the decreased parallelization efficiency of the EM-fitting algorithm, the efficiency is virtually the same as the standard performance.

V. Result

1. Optimal Structure Fit into EMD-1365 and EMD-1363

The EM-fitting simulation was carried out until the positions of the center of mass of the constituent molecules in the 70S ribosome were far enough from those at the initial position but still close enough for these to overlap their distribution. For this fitting, the fitting simulation was carried out for about 400 ps on average. Then in order to relax the conformation which might be deformed due to the artificial EM force, the conventional MD simulation was carried out without the EM-force ($w_{EM} = 0.0$) but by imposing harmonic potentials on these centers of mass for 2 ns. This procedure was repeated 50 times, and w_{EM} was increased gradually from 1.0×10^{-5} up to less than 1.0×10^{-4} . After the atomic



Fig. 3 Parellelization efficiency of SCUBA on PRIMERGY BX900 supercomputer; standard performance (black line with closed circle) and performance with the EM-fitting algorithm (red line with closed circle) As a reference, the parallelization efficiency and vectorization ratio on the Earth Simulator are shown with red and blue dotted lines, respectively.¹⁾

structure was fit into EMD-1363, the fitting simulation to fit it into EMD-1365 was carried out. The final CCs for EMD-1365 and EMD-1363 were 0.884 and 0.877, respectively (Table 1).

The total time for this calculation was 120 ns (20 ns for the fitting simulations and 100 ns for standard MD simulations). It was confirmed that the secondary structures in the 70S ribosome were maintained throughout the simulations. The total CPU core time used was about 1 million hours on PRIMERGY BX900 and FX1 supercomputers at JAEA.

Figure 4 shows the optimal structure fit into EMD-1365 and EMD-1363. A clear difference can be seen around EF-G, the L11 region and the L10-L12 stalk. In the X-ray structure the L10-L12 stalk is bent towards the L11 region, while it stretches outwards in the fit structures as the EM density map implies. It is considered that the L10-L12 stalk is bent due to the crystal packing in the crystal structure. During the fitting simulation, S13 in 30S, and L5 in 50S approached each other from the right and left and leave to the left and right, respectively. Both S13 and L5 interact with tRNAs, indicating that they are important for the translocation of tRNA molecules. This movement implies the ratchet-like movement between 50S and 30S occurred.^{9,11}

2. Conformational Change during EM-Fitting Simulation

To confirm whether ratchet-like movement had occurred or not, conformational changes during the EM-fitting were analysed using DynDom3D.¹³⁾ **Figure 5** clearly shows the ratchet-like motion was successfully captured by Dyn-Dom3D.

VI. Conclusion

We showed that it is possible to construct 3D structures at the atomic level which can be fitted into an EM density map. It should be noted that this method, performed in water me-



Fig. 4 Constructed structures of the 70S ribosome (ribbon models in blue and red) fit into EMD-1365 (blue cloud) and EMD-1363 (red cloud). The X-ray structure is shown in yellow. (a) The view is from the side. The L10-L12 stalk and 30S are encircled with small and large circles, respectively. (b) The view is from the bottom. For clarity, only 30S is shown. The arrows indicate the fitting movement during the simulation.



Fig. 5 Stereogram of the ratchet-like movement between 50S and 30S analysed by DynDom3D.¹³⁾ Dynamic domains are mainly composed of two parts, domains I (pale green) and II (red), corresponding to 50S and 30S respectively. The axis for domains I and II is depicted as a pale green arrow in green with a red tip.

dium, does not require restraints to maintain the secondary structure like other methods do. Therefore, the structure constructed by this method is very natural, and it would be reasonable to assume the structures changing on the way to the EM density map are the actual intermediate structures during the translocation.

The next target of the structural analysis of ribosome is to calculate the free energy along the translocation of tRNA through ribosome using sampled structures along the translocational path. The analysis of the free energy landscape will provide further insight into the mechanism of translocation.

Acknowledgment

I would like to thank Yvonne Ishida for reading this manuscript carefully. H. I. was partly supported by a Grant-in-Aid for Scientific Research (C) from the Japan Society of the Promotion of Science (JSPS) KAKENHI (22570165).

References

- H. Ishida, T. Kano, K. Yura, A. Matsumoto, Analysis of Function of a Large-scale Supra-biomolecule System by Molecular Dynamics Simulation System, SCUBA (Simulation Codes for hUge Biomolecular Assembly), Annual Report of the Earth Simulator Center, 257-263, (2007).
- H. Ishida, S. Hayward, "Path of nascent polypeptide in exit tunnel revealed by molecular dynamics simulation of ribosome," *Biophys. J.*, 95, 5962-5973 (2008).
- H. Ishida, "Branch migration of Holliday junction in RuvA tetramer complex studied by umbrella sampling simulation using a path-search algorithm," *J. Comput. Chem.*, **31**, 2317-2329 (2010).
- H. Ishida, H. Kono, "The regulation of the nascent polypeptide in the ribosomal exit tunnel," *RESEARCH ADVANCES IN NUCLEIC ACIDS RESEARCH*, (ed.) R. M. Mohan, Global Research Network, 2009, Kerala, 19-28 (2009).
- H. Ishida, A. Matsumoto, Y. Tsutsumi, K. Yura, "Conformational analysis of the structure of ribosome fit into electron microscopy density maps with normal mode analyses and molecular dynamics simulations," *Proceedings of the 16th International Microscopy Congress*, p.242 (2006).
- 6) K. Noda, M. Nakamura, R. Nishida, Y. Yoneda, Y. Yamaguchi, Y. Tamura, H. Nakamura, T. Yasunaga, "Atomic model construction of protein complexes from electron micrographs and visualization of their 3D structure using VR system," J. Plasma

Phys., 72, 1037-1040 (2006).

- L. G. Trabuco, E. Villa, K. Mitra, J. Frank, K. Schulten, "Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics," *Structure*, 16, 673-683 (2008).
- T. R. Caulfield, S. C. Harvey, "Conformational fitting of atomic models to cryogenic-electron microscopy maps using Maxwell's demon molecular dynamics," *Biophys. J.*, 368A (2007).
- A. Matsumoto, H. Ishida, "Global conformational changes of ribosome observed by normal mode fitting for 3D cryo-EM structures," *Structure*, 17, 1605-1613 (2009).
- 10) K. Murakami, T. Yasunaga, T. Q. P. Noguchi, Y. Gomibuchi, K. X. Ngo, T. Q. P. Uyeda, T. Wakabayashi, "Structural basis for actin assembly, activation of ATP hydrolysis, and delayed phosphate release," *Cell*, **143**, 275-287 (2010).
- M. Valle, A. Zavialov, J. Sengupta, U. Rawat, M. Ehrenberg, J. Frank, "Locking and unlocking of ribosomal motions." *Cell*, 114, 123-134 (2003).
- 12) Y. G. Gao, M. Selmer, C. M. Dunham, A. Weixlbaumer, A. C. Kelley, V. Ramakrishnan, "The structure of the ribosome with elongation factor G trapped in the posttranslocational state," *Science*, **326**, 694-69. (2009).
- G. A. Poornam, A. Matsumoto, H. Ishida, S. Hayward, "A method for the analysis of domain movements in large biomolecular complexes," *Proteins*, 76, 201-212 (2009).