

Development of a Molecular Dynamics Simulation System for Large-Scale Supra-Biomolecules and its application to ribosome

Hisashi Ishida¹

¹ Center for Computational Science and e-system
Japan Atomic Energy Research Agency
6-9-3 Higashi-Ueno, Taito-ku, Tokyo 110-0015, JAPAN
ishida.hisashi@jaea.go.jp

We are developing an integrated molecular dynamics 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 [1]. SCUBA has several special features:

1. Algorithm for non-cutoff electrostatic interactions: SCUBA utilizes the Particle-Particle Particle-Mesh (PPPM) algorithm, which calculates the full Coulomb electrostatic interactions. This algorithm reduces the computational time required to calculate the electrostatic forces from the conventional $O(N^2)$ to $O(N\log N)$.
2. Dynamics load balance: To overcome the load imbalance associated with irregular atomic distribution, a dynamic load-balancing algorithm has been implemented.
3. High performance: SCUBA achieves both a high parallelization efficiency ratio and a high vectorization ratio. The techniques employed to achieve such a high performance are explained in detail below.

Parallelization of SCUBA: We employed a domain decomposition method. In the DD method, the volume of the physical system is divided into rectangular subcells. These subcells are then allocated to processors in such a way that neighboring subcells are located in the same processors.

Vectorization of SCUBA: We concentrated on the vectorization of van der Waals interactions, and the PP and PM interactions because these non-covalent interactions are the most time consuming part of an MD simulation.

Performance of SCUBA: We have reached a vectorization ratio of more than 95% and a parallelization efficiency ratio of 50% even when 512 (360) CPUs were used on the Altix 3700 Bx2 of JAEA and the Earth Simulator of JAMSTEC, respectively.

Application of SCUBA to the MD simulations of the 70S ribosome

Using SCUBA, molecular dynamics (MD) simulations were performed on *Thermus thermophilus* 70S ribosome with and without the nascent polypeptide inside the exit tunnel (a system containing approximately 2 million atoms) for a total time of 20 nanoseconds. The simulations revealed that Arg92 of L22 plays the role of a gate for the nascent polypeptide, and that global motions involve the relative movement of the 50S and 30S subunits for protein synthesis [2].

References

- 1) Analysis of the function of a large-scale supra-biomolecule system by molecular dynamics simulation system, SCUBA. Ishida, H., M. Higuchi, Y. Yonetani, T. Kano, Y. Joti, A. Kitao, and N. Go. Annual Report of the Earth Simulator Center April 2005 - March 2006:237-240.
- 2) Path of nascent polypeptide in exit tunnel revealed by molecular dynamics simulation of ribosome, Ishida, H. and Hayward, S., Biophys. J, **95**,5962-5973 (2008)

Development of a molecular dynamics simulation system for large-scale supra-biomolecules and its application to ribosome

Quantum Bioinformatics Team,
Center for Computational Science and Engineering,
Japan Atomic Energy Agency

Hisashi Ishida

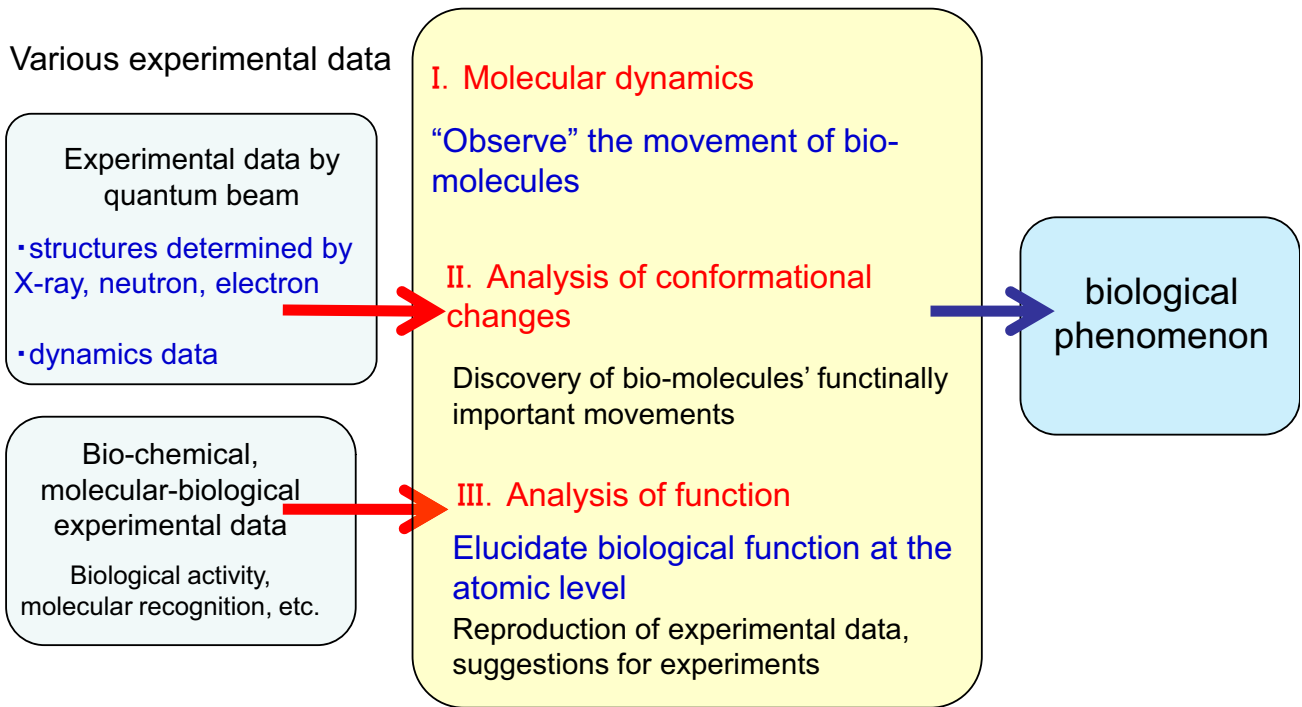
Advanced Computing Technologies toward PetaFLOPS
24 Apr 2009

Content

- Introduction to molecular dynamics (MD) simulations in the field of biology
- Development of a MD simulation program, SCUBA
- MD simulations using SCUBA (Functional analysis of ribosome)
- Summary
- In future

Significance of molecular simulation

Tool to elucidate biological functions at the atomic level, and to understand biological phenomenon

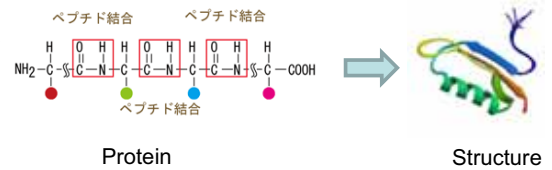


Target: supra biomolecules

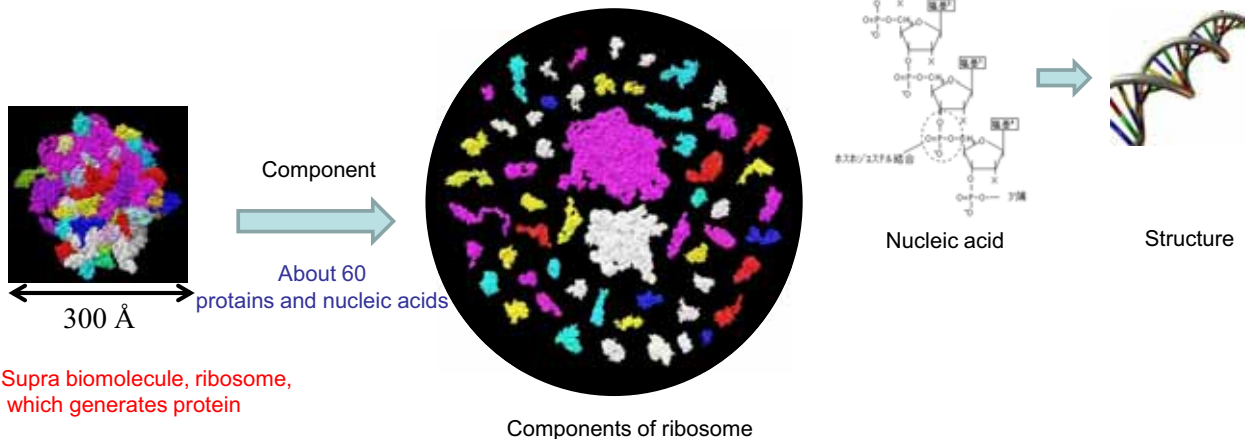
Biomolecules (proteins, nucleic acids) are composed of amino acids and nucleotides, and have unique structures.

Target: supra biomolecules

Proteins and nucleic acids often function as supra biomolecules

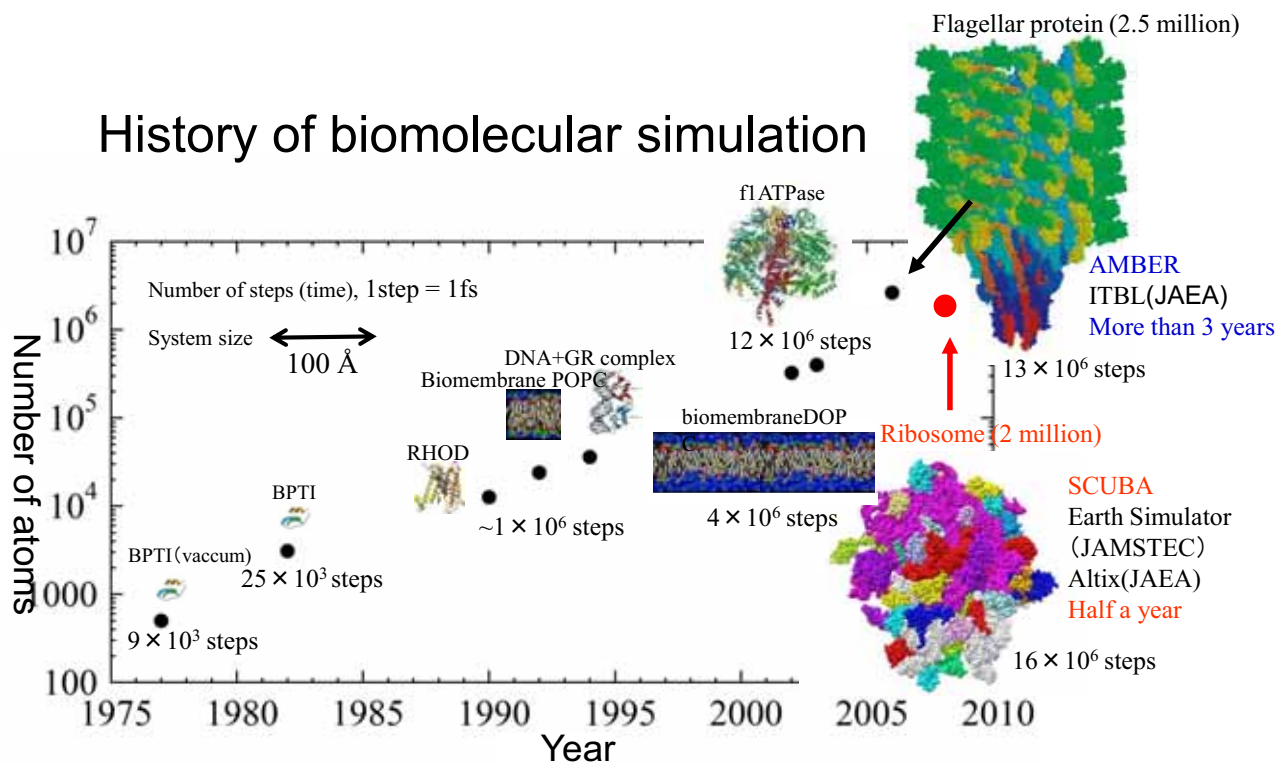


(Previously, the main target of MD simulations was single molecules due to computational limitations)



Development of molecular dynamics simulation program

History of biomolecular simulation



Conventional program (non-parallel): limited to 10^4 atoms, 10^6 steps

A parallel program is necessary to deal with more than 10^4 atoms

Development of a molecular dynamics simulation program

Design of SCUBA

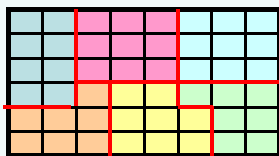
Development of SCUBA to utilize parallel computers efficiently

High performance

1. parallelization

Massive computation using space-decomposition

Distribution of processor load using dynamic load balance



2. high-accuracy calculation

Non-cutoff coulomb interaction calculation

3. Extensibility (maintainability, additional function)

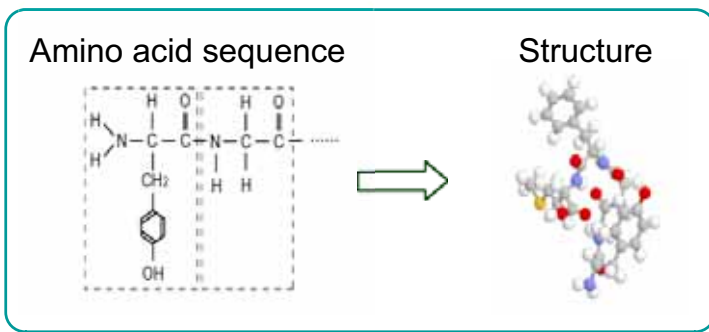
Highly-modularized program according to function

Multi function

1. Time integral algorithms for various ensembles
2. Modeling of structure based on experimental data
3. Energy minimization
4. Free energy perturbation
5. Principal component analysis
6. Force field for AMBER, CHARMM, etc.

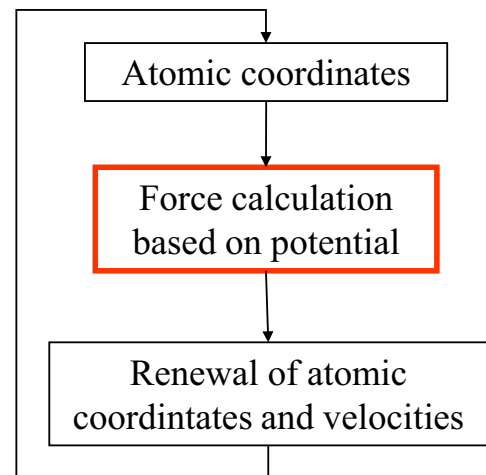
SCUBA enables high-speed simulation for supra biomolecules consisting of more than 1 million atoms

Molecular dynamics simulation



$$E_{total} = \sum_{\text{bonds}} K_r (r - r_{eq})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)]$$

$$+ \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

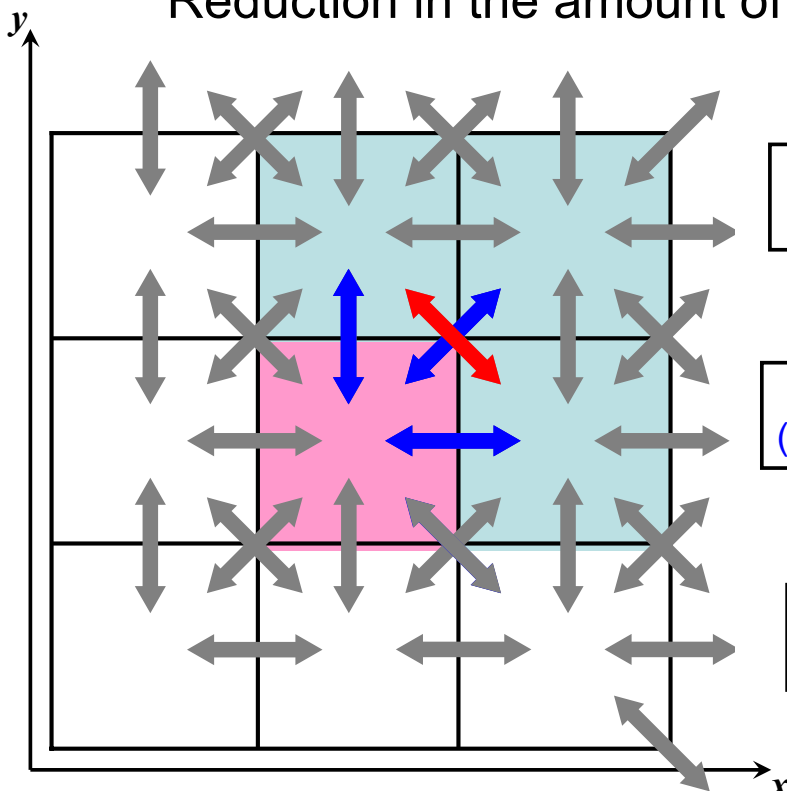


Most of the computation is **calculation of force**

→ reduction of computational time by parallelization

Problem: data communication

Reduction in the amount of communication



[Our design]

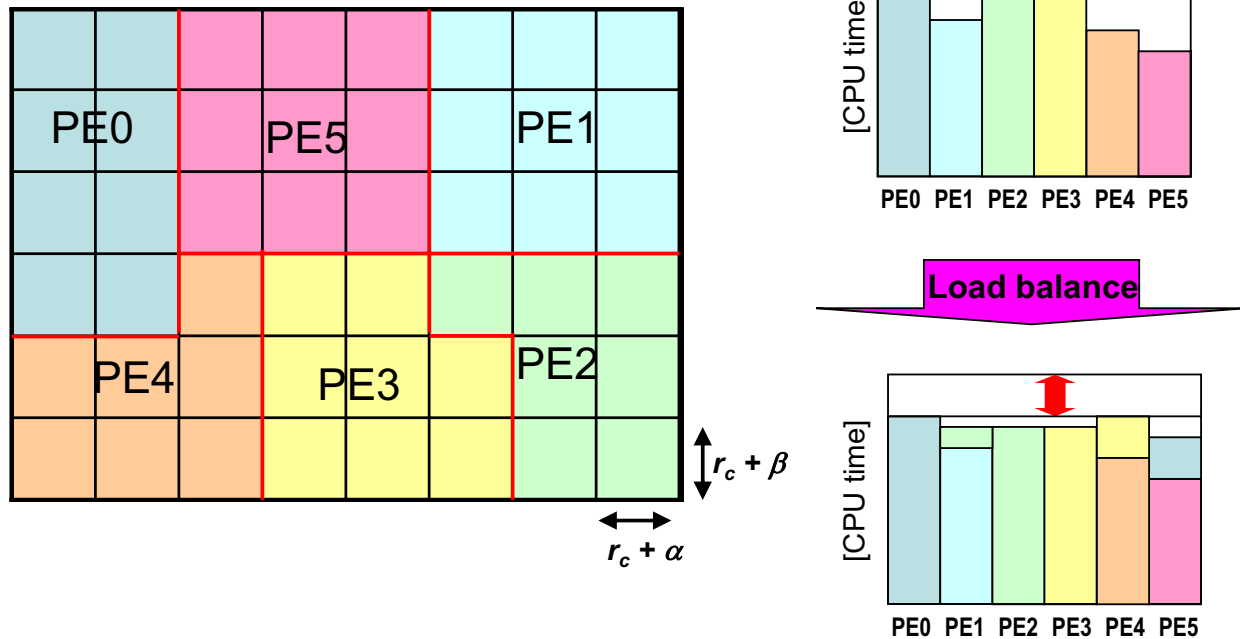
Receive data from **7** subcells
(**3** subcells in 2-dimension)

Calculation of
(**direct** and **indirect**) interaction

Send data to **7** subcells
(**3** subcells in 2-dimension)

Reduction of communication by 7/13

Dynamic load balance



- A biological system is heterogenous.
- It is important to optimize the distribution of the load

Calculation of long-distance electrostatic interaction

Total energy of electrostatic interaction

$$E = \frac{1}{2} \sum_{i,j=1}^N \sum_n \frac{q_i q_j}{|\mathbf{r}_{ij} + n\mathbf{L}|}$$

$$E = E^{(r)} + E^{(k)} + E^{(s)} + E^{(d)}$$

Real-space energy

$$E^{(r)} = \frac{1}{2} \sum_{i,j=1}^N \sum_m q_i q_j \frac{\text{erfc}(\alpha |\mathbf{r}_{ij} + m\mathbf{L}|)}{|\mathbf{r}_{ij} + m\mathbf{L}|}$$

Reciprocal-space energy

$$E^{(k)} = \frac{1}{2L^3} \sum_{k \neq 0} \frac{4\pi}{k^2} e^{-k^2/4\alpha^2} |\tilde{\rho}(k)|^2$$

Self-energy

$$E^{(s)} = -\frac{\alpha}{\sqrt{\pi}} \sum_i q_i^2$$

Fourier transformation of electron density

$$\tilde{\rho}(\mathbf{k}) = \int_{V_b} d\mathbf{r} \rho(\mathbf{r}) e^{-i\mathbf{k} \cdot \mathbf{r}} = \sum_{j=1}^N q_j e^{-i\mathbf{k} \cdot \mathbf{r}_j}$$

Fast calculation of reciprocal-space energy by Particle-Particle Particle-Mesh (PPPM)

1. Assign the charges to the mesh points.

$$\rho_M(\mathbf{r}_p) = \frac{1}{h^3} \sum_i q_i W(\mathbf{r}_p - \mathbf{r}_i)$$

2. Calculate FFT of distributed charges

$$\hat{\rho}_M(\mathbf{k}) = h^3 \sum_{\mathbf{r}_p \in M} \rho_M(\mathbf{r}_p) e^{i\mathbf{k} \cdot \mathbf{r}_p}$$

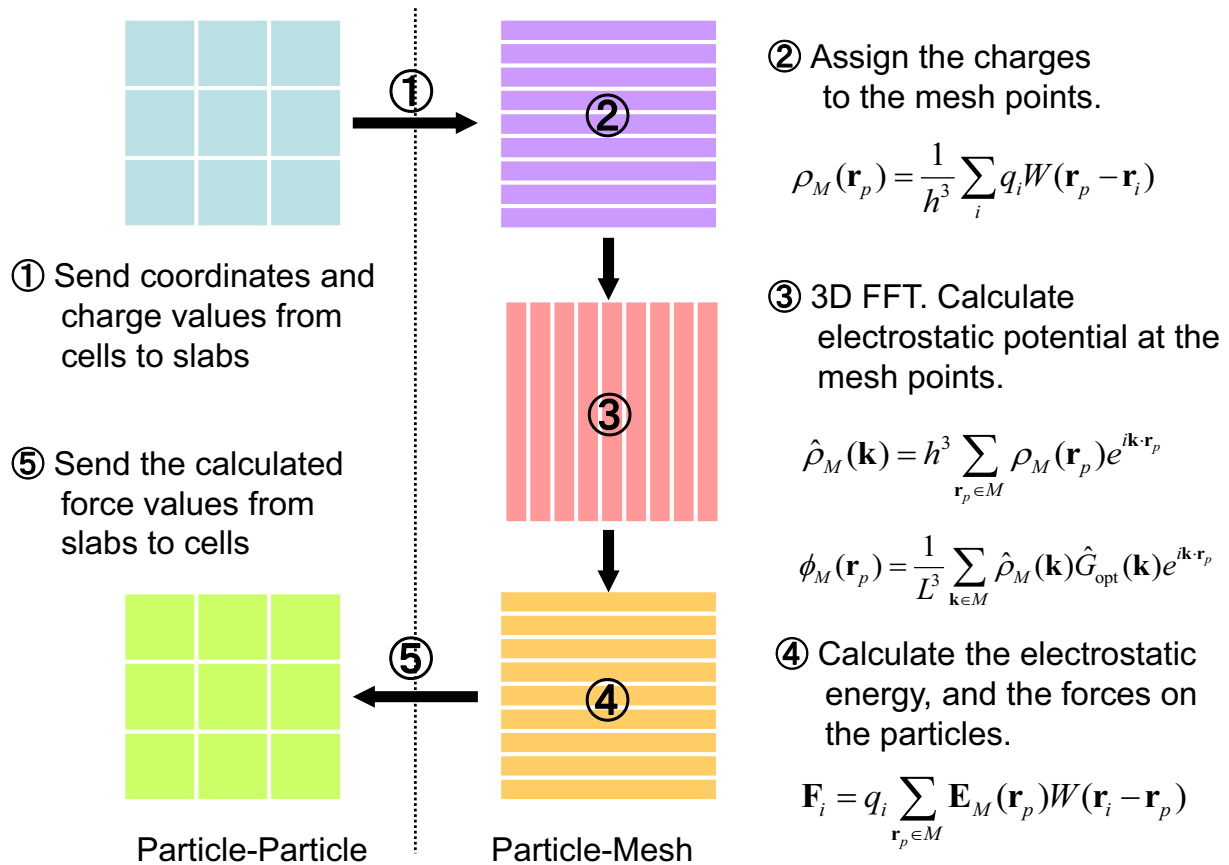
3. Calculate electrostatic potential at the mesh points.

$$\phi_M(\mathbf{r}_p) = \frac{1}{L^3} \sum_{\mathbf{k} \in M} \hat{\rho}_M(\mathbf{k}) \hat{G}_{\text{opt}}(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{r}_p}$$

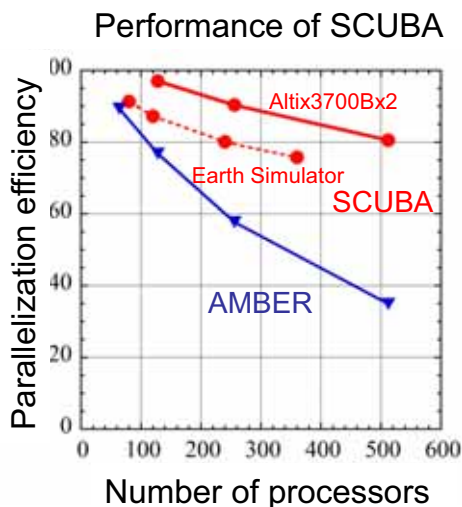
4. Calculate the forces on the particles.

$$\mathbf{F}_i = -q_i \sum_{\mathbf{r}_p \in M} \phi_M(\mathbf{r}_p) \nabla_{\mathbf{r}_i} W(\mathbf{r}_i - \mathbf{r}_p)$$

Parallelization of PPPM



Performance of SCUBA



Altix3700Bx2 (JAEA)

Parallelization efficiency: more than 80% (512 processors)



Earth Simulator (JAMSTEC)

Parallelization efficiency: more than 75% (360 processors)
Vectorization ratio: more than 95% (360 processors)

50% of parallelization efficiency using 100 processors:
100 processors can compute 50 times faster than 1 processor

The performance of SCUBA is better than that of the world standard MD program, AMBER

Analysis of the functions of ribosome using SCUBA

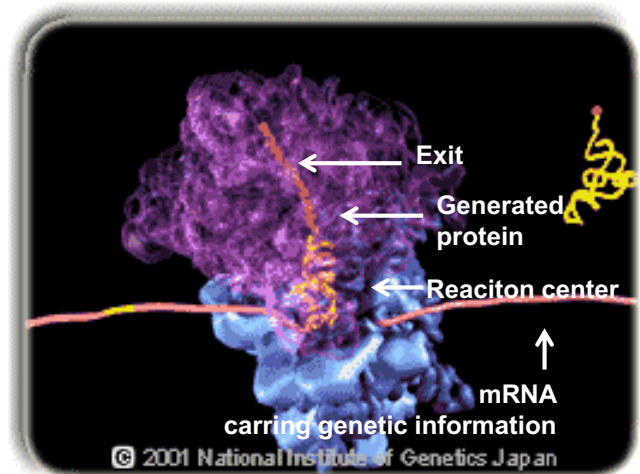
Ribosome functions in the expression of the genetic code from nucleic acid into protein, in a process called translation

Aim:

Elucidation of the dynamics of ribosomal functions (regulation of the movement of protein, and protein synthesis)



Understanding the molecular mechanism of ribosome helps us to develop new antibiotics which inhibit the function of ribosome



Ribosome generating protein (image)
(From National Institute of Genetics)

Regulation of the movement of protein through the ribosomal tunnel

Modeling of a protein structure in the tunnel



Discovery of two paths through which a protein can pass
Which path can protein pass through?

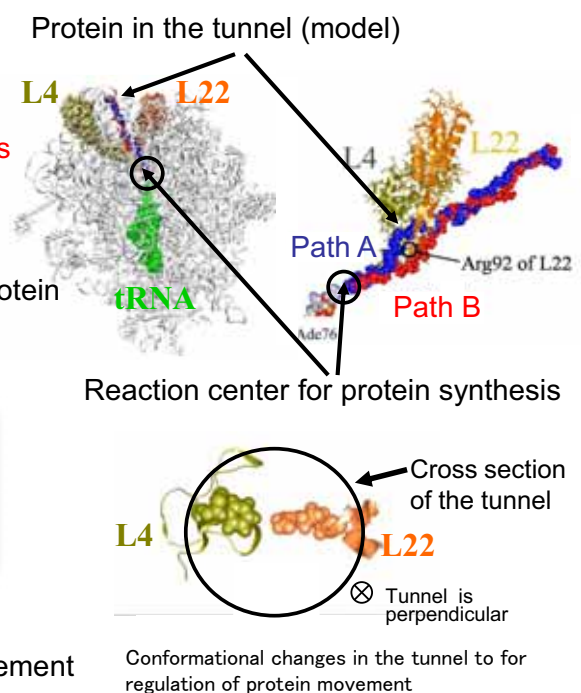
Method: insert a protein into each path, and perform molecular simulations to induce the movement of the protein

Results:

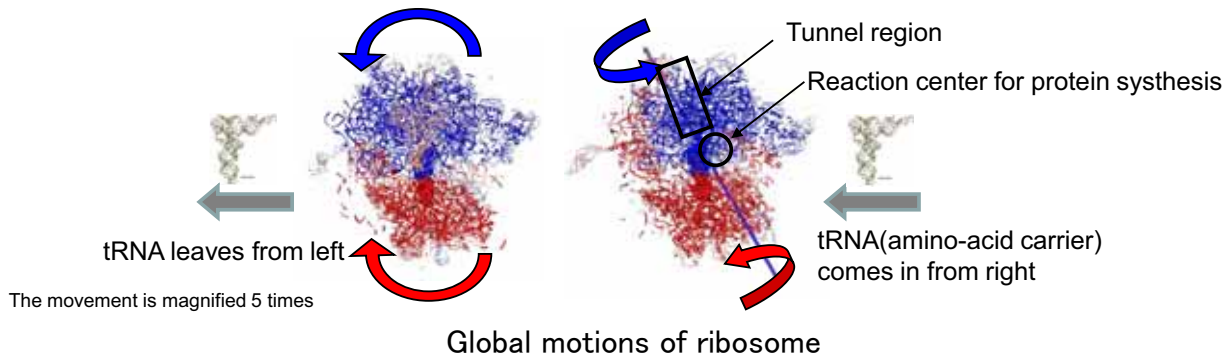
Location of protein	Path A	Path B
Condition of tunnel	narrow	wide
Movement of protein	no	yes

1. Protein passes through path A widening the tunnel
2. Identified the residues for regulation of protein movement

⇒ important information on the development of antibiotics which affects the tunnel and inhibit the ribosomal function



Functional dynamics of ribosome



Important functional movement for protein synthesis was observed.

Rotational axis passes through the tunnel region and the atomic fluctuation is small (1 Å).



Movement for protein synthesis does not contribute to protein movement.

Two theories about the force to move protein

1. Conformational change of tunnel (×)
2. Diffusion of protein (○)

Ribosome is a molecular machine **composed of soft and hard parts.**

1. Control of protein synthesis: **global motion** between two subunits (blue and red)
2. Control of protein movement: **local conformational change** in tunnel

Biophys. J., **95**,5962-5973 (2008)

Summary

- Development of molecular dynamics simulation program, SCUBA

- It is possible to perform large-scale (> 1 million atoms) simulations of supra biomolecules

- Functional analysis of ribosome

Ribosome is a molecular machine **composed of soft and hard parts.**

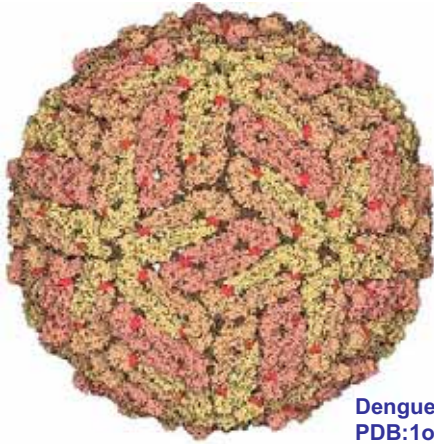
1. Control of protein synthesis: **global motion** between two subunits
2. Control of protein movement: **local conformational change** in tunnel

In the future

Possible targets for the next-generation supercomputer

Mechanism of virus infection

← 500Å →

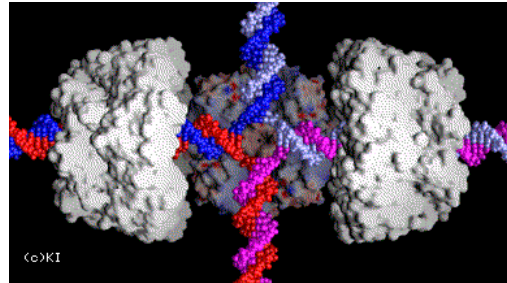


Dengue virus
PDB:1ok8

membrane

Mechanism of DNA recombination

← 300Å →



↑
150Å
↓

RuvAB-DNA complex

<http://www.sdsc.edu/journals/mbb/ruva.html>

supra large-scale MD simulation (more than 10 million atoms) for more than 1mini second

Direct observation of the dynamics of supra macro-molecules