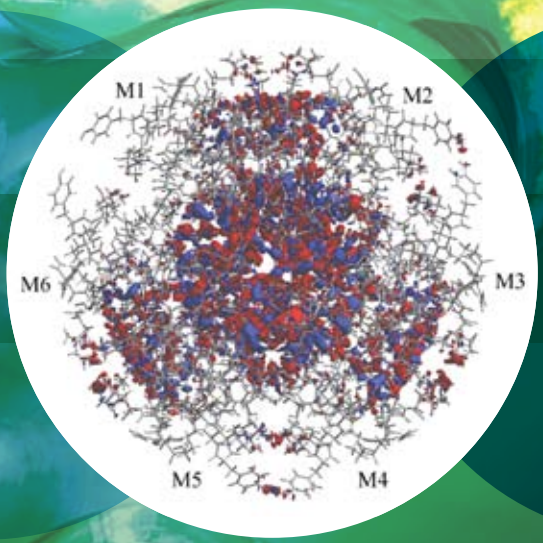
ISLiM
Integrated Simulation of Living Matter

BioSupercomputing Newsletter

Vol.2 2010.3

Next-Generation Integrated Simulation of Living Matter



Example of the results from ProteinDF:
Redistribution of electrons according to the agglomeration of insulin
Showing the difference between hexamers and monomers in all-electron distribution

＊See P.3 and P.10 for ProteinDF.

CONTENTS

SPECIAL INTERVIEW

Biosupercomputing: Carving out the Future of Life Sciences Research

Aiming to Become a Global Trendsetter in the Life Sciences by Making the Best Use of the Next-generation Supercomputer!

Computational Science Research Program
Deputy Program Director
Ryutaro HIMENO

As the base of life sciences research in developing "Grand Challenges" applications for the next-generation supercomputer project, RIKEN has been tackling "Next-Generation Integrated Simulation of Living Matter" and proposed "biosupercomputing" as a new field of study. Currently, RIKEN and researchers in the fourteen participating organizations have been developing software that makes the best use of the petaflop-scale next-generation supercomputer. Many people have great expectations for "biosupercomputing" to lead the world by establishing new methods in the life sciences based on advanced computational science.

Searching for explanations of life phenomena through computational scientific approaches

Until recently, the royal road to understanding the life sciences has been laboratory work. Computers have mainly been used to organize experimental data. Experimental technologies and methods, such as high-throughput experimental equipment, super high-speed sequencers, genetic modification technology and single molecule imaging, have rapidly improved and have been involved in the clarification of many factors. On the other hand, since many detailed factors have become clear, researchers grappling with experimental research have started to feel that the distance to an essential understanding of life phenomena has been increasing. This is why we started to explore the use of computers to understand life phenomena.

The computational performance of supercomputers has continued to improve by approximately ten times every 3.8 years for the last twenty years. As performance improves, the fields of research that yield practical applications have changed. First, supercomputers made it possible to solve the problems of architectural structure and then fluid mechanics. Supercomputers can now process large-scale experimental data, such as the calculation of nano-materials and the analysis of the human genome. Ultimately, with the development of the next-generation supercomputer, which will have computational performance measured in petaflops, the day is coming when we can probe and explain various phenomena inside living matter based on the physical equations of atoms and molecules.

What we found by the development of experimental research is that life phenomena are far too complicated for the human brain to comprehend. To deal with this, we can input the rules of life phenomena observed into computers and organize and test them. This will allow us to be able to identify the life phenomena that can be explained and those that cannot. In addition, it will be possible to predict and control life phenomena based on prediction. As a tool of computational scientific approaches related to life phenomena, the use of supercomputers is now grabbing attention. We can say that this is a change in thinking from "biology describing phenomena" to "biology predicting a new phenomenon." This is also a proposal of new ways of using supercomputers that are radically different from the conventional, and orthodox methods used in engineering and physics.

Leading the world by new thinking

Currently, we are taking on the following three challenges by using the next-generation supercomputer. First, as described above, the challenge of reorganizing obtained knowledge and information, sorting out what is not understood, making predictions and controlling phenomena to understand life phenomena in total. Second, the challenge of explaining various phenomena inside the human body based on the laws of physics for atoms and molecules. Third, the challenge of simulating a human body in a computer, in the same way car crashes and earthquake wave simulations are simulated, so that we can probe the human body according to its physical properties and each organ (such as the heart) and find surgical methods that

MD Core Program for Large-scale Parallel Computers

Person in charge of development: Makoto Taiji (RIKEN)

Aiming for the Gordon Bell Prize for the world's fastest molecular dynamics calculation

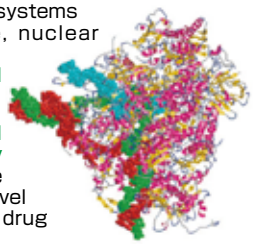
Computational technique for simulating the functions of proteins at high speed over long periods of time

Feature of software

- Molecular dynamics calculation software: super parallelization of 100,000 or more cores (in one-million to ten-million atom systems)
  Realize a simulation for 100 million atoms at effective some petaflops and aim for the fastest effective performance

- Long-term simulation for the motion of biomolecules
  Aim a sub-millisecond simulation for one-million atom systems
  Example of subjects: Cell membrane structure, nuclear membrane transportation, viruses, etc.
  [100 times greater (or more) than the conventional timescale]

- By capturing large, physiologically essential structural changes in proteins at a previously impossible timescale, we aim to understand the structural principles of life systems at the atomic level and achieve the calculation accuracy required for drug discoveries.



Large-scale molecular dynamics simulation

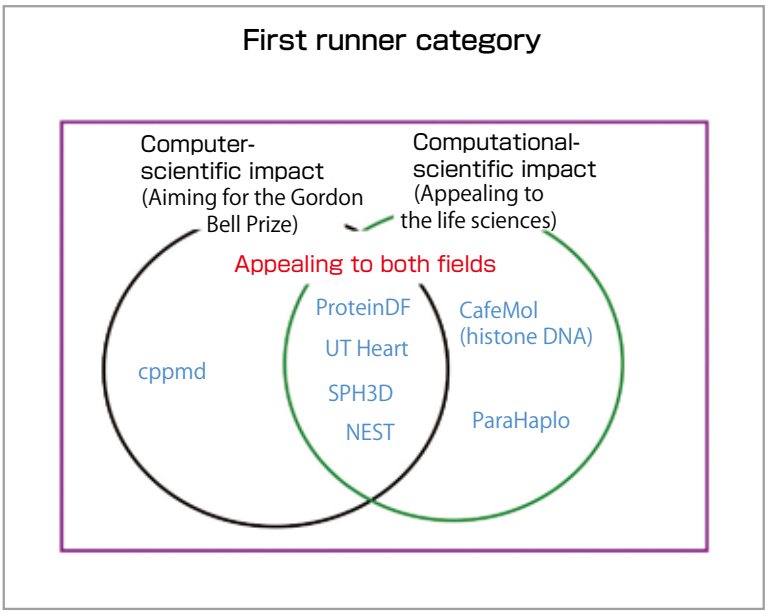
Documentation of the first and second runners for the development of Life Grand Challenges applications (partial)

"The First Runners" for the Development of Life Grand Challenges Applications

Application name	Person in charge of development	Description	Purpose
MD core program for large-scale parallel computers (cppmd)	Makoto Taiji (RIKEN)	Computation technique for simulating the functions of proteins at high speed over long periods of time	Aim to win the Gordon Bell Prize for the world fastest molecular dynamics calculation
All-electron wave function calculation for proteins based on density functional theory (ProteinDF)	Fumitoshi Sato (The University of Tokyo)	Computational science and technology for accurately elucidating the reaction of proteins at the electronic level	Aiming at the world's largest all-electron calculation for proteins
Multi-scale, multi-physics heart simulation(UT Heart)	Toshiaki Hisada (The University of Tokyo)	Globally unprecedented, world's largest simulation of a virtual heart in a computer	Aiming at the simulation of an entire heart from the myocardial cell level
Whole-body voxel simulation (Voxel structure fluid coupling analysis program SPH3D)	Shu Takagi (RIKEN)	Computational science and technology suitable for analyzing and predicting a soft human body for medical applications	Aiming at the simultaneous simulation of creation, transportation and infarction of a thrombus
Neural Simulation Tool(NEST)	Marcus Diesmann (RIKEN)	Elucidates whether it is possible to explain the activities of an actual brain based on a simple neuron model	Aiming at creating the world's largest action simulation of a cerebral cortex focal neurological circuit
Coarse-grained model calculation (CafeMol)	Shoji Takada (Kyoto University)	Elucidates the mechanism of the winding and unwinding of DNA onto the histone in a nucleus	Explains the mechanism of a great length of DNA stored in and read from a nucleus when necessary
Statistical test software for haplotype analysis (ParaHaplo)	Naoyuki Kamatani (RIKEN)	Elucidates the relation of individual genetic information to diseases and drug responses	Explains the relationships of diseases and drug responses to gene and environmental factors for forty-seven diseases

will yield practical applications for the medical field.

The challenge of an overall understanding of life phenomena is one of the most important assignments for the life sciences in the 21st century. Many researchers around the world feel the necessity of computational scientific approaches in the life sciences. Under these circumstances, what we have to do is to make a major course correction that changes the conventional research methods used in the life sciences. By quickly starting to take on the challenge of this assignment and by being a trendsetter, we believe that our research will have a considerable ripple effect on the world. Japan is about to be the first nation in the world to start down this path by having this great tool, the next-generation supercomputer. In addition, the many researchers of the "All Japan team" are cooperating to achieve this significant goal. "Biosupercomputing" has tremendous potential to lead the world in this field. At the very least, it will be able to play the role of trendsetter in leading us in a new direction.



Example of categorizing applications for tuning and saving machine time for Life Grand Challenges

Developing software for actual operation

Aiming for the full-operation of the next-generation supercomputer (in 2012), applications for thirty-some themes have currently been developed in the life sciences. Among them, we focus on seven applications as the "first runners" to achieve a great result as soon as the full-operation begins. To select the first runners, we considered the following criteria: calculation is possible only by utilizing the performance of the next-generation supercomputer, the outcome is scientific and has high value for application to the life sciences and the application can have an impact immediately after the start of operations. This does not mean that we are ranking the applications in terms of the scientific meaning. What is more important is the level of perfection of the application and whether the calculation is making the best use of the next-generation supercomputer.

For example, the "MD core program for large-scale parallel computers (cppmd)" is a computation technique to simulate the functions of proteins at high speed over long periods of time based on the physical laws in the atomic and molecular world. We would like to achieve the world's fastest molecular dynamics calculation to win the Gordon Bell Prize. The Gordon Bell Prize is given to the project that achieves the best results through the use of parallel computers for practical scientific computing.

We have 6 more first runners other than the MD core program. Some of these are expected to achieve great results, including those that have a chance of winning the Gordon Bell Prize and others that will show that Japanese research is leading the world. We hope that these first runners prove the superior computational performance of the next-generation supercomputer as full-operation begins and promote the high performance and future potential provided by "biosupercomputing." After the first runners, twelve "second runners" are waiting. Some of them are expected to have significant impact on the development of medical treatment and pharmaceuticals, while others still are research that is unique to Japan. Finally, the "third runners" are waiting to aim for results that will have global impact in a wider range of areas.

A Message from the Team Leader

Cell Scale Team

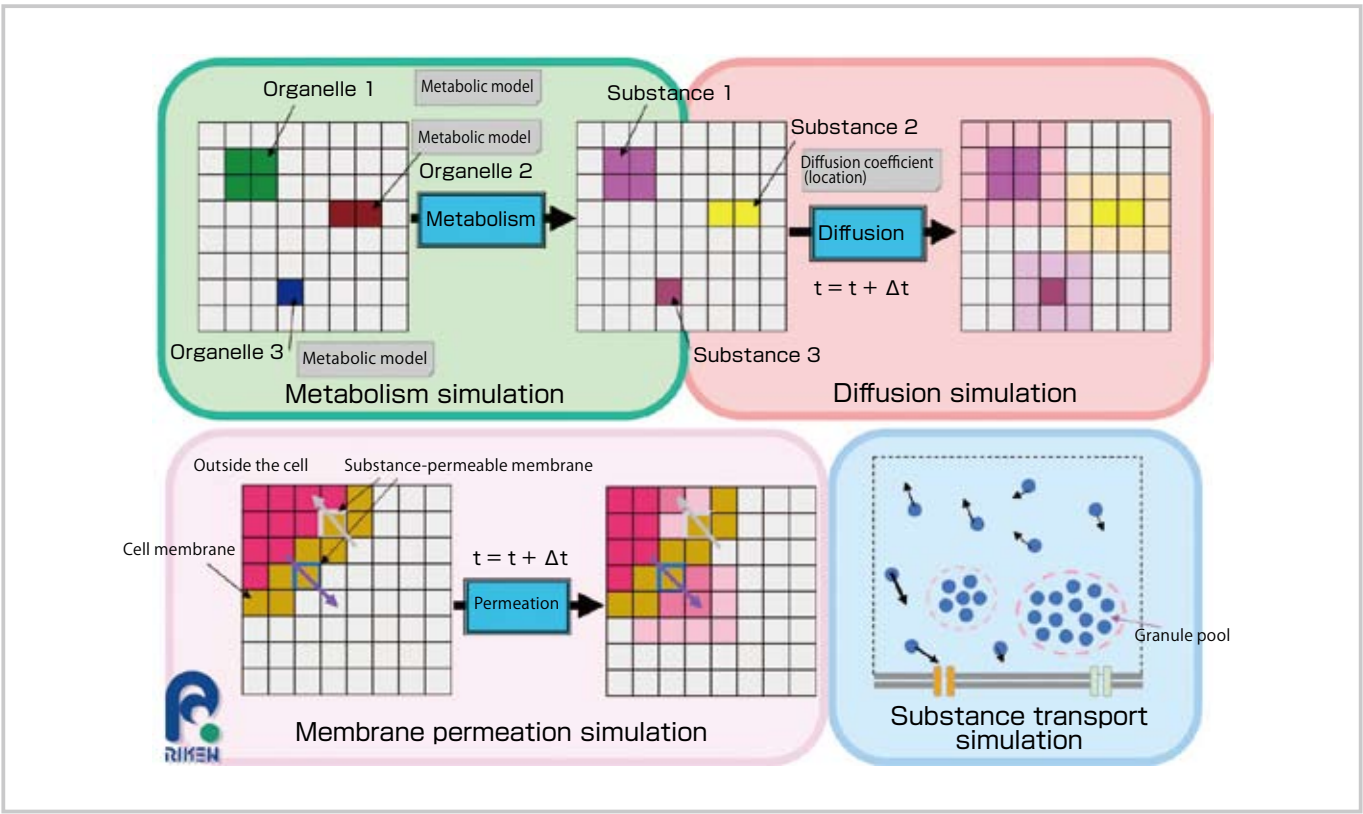
Simulating Cell Phenomena by Recreating Cells as They Exist in Living Organisms

Cell Scale Team
Team Leader
Hideo YOKOTA

There are many approaches that can be applied to theoretical biology, ranging from the micro-level method based on molecular dynamics calculation to macro-level methods such as mechanical simulation. Our cell research takes an approach that falls under the latter category. Our ongoing research in the digitization of a living organism's shape information is now at the stage where whole-body profile data of a mouse has been obtained at a 10-μm resolution. Furthermore, advances in live-cell imaging technology have enabled us to collect and digitize the data of a living cell and study its dynamic process. This means a variety of cell phenomena, including cell shapes, can be now observed as four-dimensional digitized data. Imaging technology, however, is not without its limits. Accordingly, many intracellular phenomena remain unexplainable. Our team is working on the development of virtual cell modeling, which will allow us to understand what happens inside a cell through digital simulation.

What are the key factors in understanding cell phenomena? The answers are intracellular chemical reaction networks and membranes that serve as the structural components of the cell. The computing of chemical reactions (i.e., metabolism) incurs enormously-high costs, although in the future molecular dynamics-based calculation may be applied. To bypass this problem, we intend to create a simulated cell model by using certain chemical reactions that appear to be dominant in the cell. There are several precedent studies in this field, such as the E-CELL system by the Institute for Advanced Biosciences at Keio University and the Cell Illustrator by the Human Genome Center at the University of Tokyo. However, these systems simulate chemical reactions by assuming the cell to be a closed bag filled with homogeneous materials or by making a cell model that is extremely simplified, thus ignoring the cell's location. Living cells have shapes and organelles, which perform different vital functions. The reactions of organelles, such as Golgi bodies and mitochondria, vary depending on the environment. As such, simulation requires an environmental context, which will allow us to recreate cells on a computer as they exist in a living orgasm. In doing so, we also need to be able to describe the diffusion of substances and the cytoskeleton-based active transport process in addition to intracellular chemical reactions. The functions of cell membranes, the other key factor, include a channel that allows specific ions to permeate, a pump that moves substances by expending energy called ATP and a receptor that receives signals from outside the cell. Our aim is to create a comprehensive simulation model that addresses all factors, including the cell's location, in an effort to understand all the phenomena that occur within living cells.

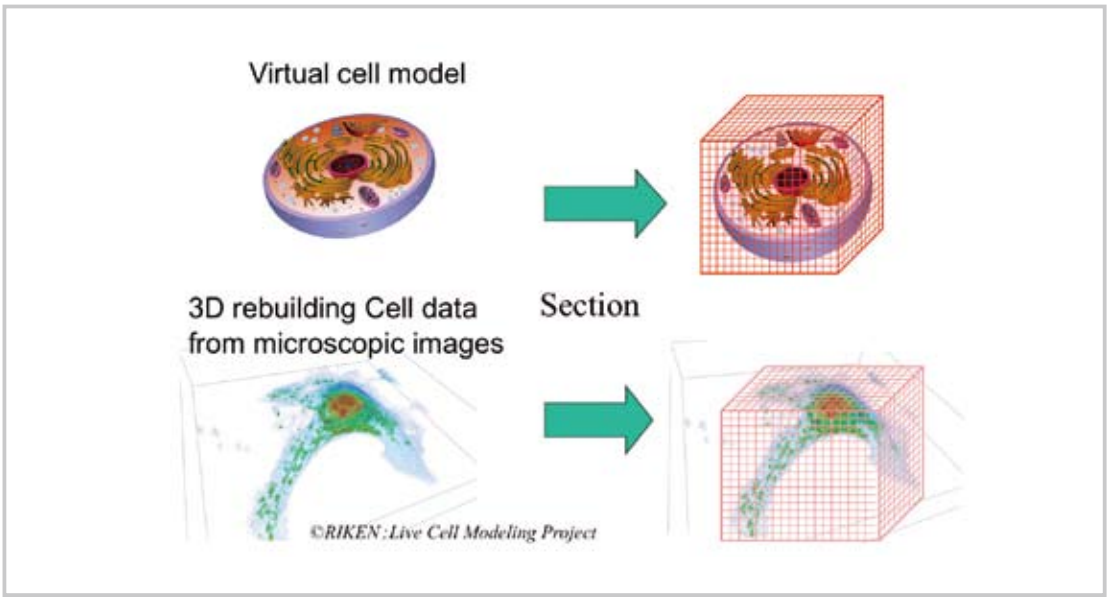
Another important note on the cell's location is the fact that a cell does



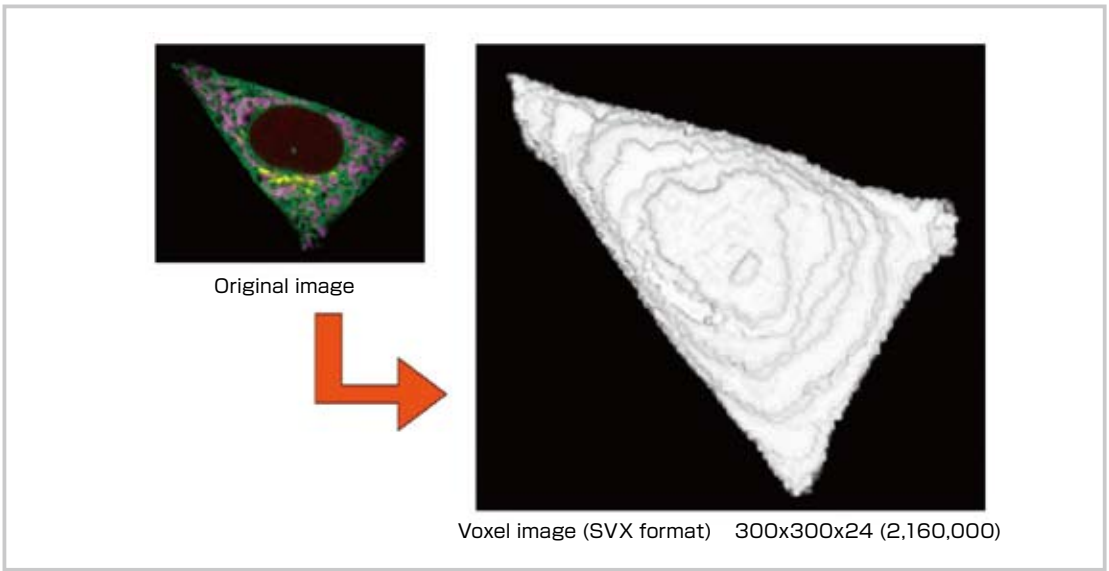
Integrated platform of cell simulation

not exist as an independent unit. Cells in the liver, for example, each have roles to play according to their location, enabling the liver to perform its functions properly. The location within the organ also determines types of cell reactions among the liver cells. Consequently, a simulation that ignores this factor lacks merit when attempting to replicate living cells influenced by their surrounding environment.

In order to conduct a simulation with a cell model that authentically replicates cell behavior in a living organism, we are working on the development of application software called RICS. The software creates a cell model in a lattice space with 100-nm sections and applies experiment data on intracellular phenomena to conduct coupled calculation of chemical reactions, substance diffusion and membrane permeation on a one million-voxel volumetric display (100^3). The system is designed to be a common simulation platform, inclusive of all types of cells. A development project currently targeting hepatocytes is planned to shift its focus onto hepatic lobule simulation next. However, simulation itself is not the goal; rather, it is a tool that will lead us to an understanding of the phenomena and functions of life. It is also important that these simulations contribute to medical science. We aim to carry our development further to the simulation and analysis of cell reactions to drugs and the removal of organs and cancerous tumors. Potential applications of the simulation model also include the design of artificial and regenerated organs, for which the calculation of cell locations is critical.



Section of a cell in voxels



Cell image in 3-D voxels

A Message from the Team Leader

Brain and Neural Systems Team

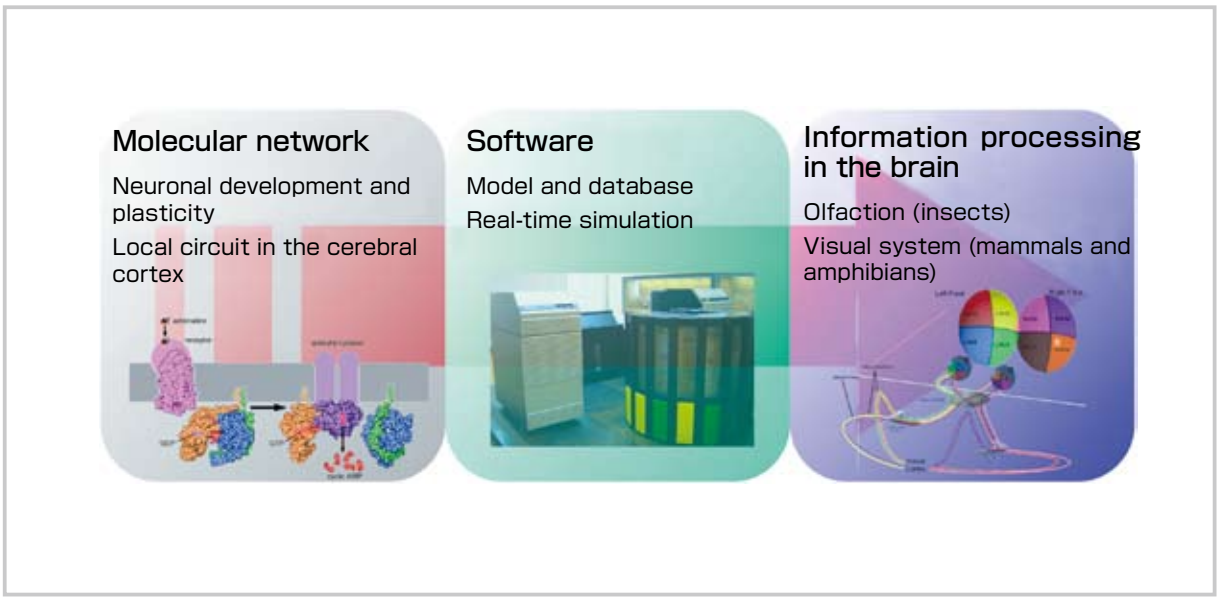
Create a Brain on a Supercomputer to Unravel the Functions of the Brain and Nervous System

Brain and Neural Systems Team
Team Leader
Shin ISHII

Since the time of ancient Greece, the origin of intelligence, i.e., the function of the brain, has been one of the most intriguing mysteries to mankind. Since the advent of computers, the brain has often been compared with computers and the possibility of creating intelligence on a computer has been discussed repeatedly. In 1952, Hodgkin and Huxley conducted electrophysiological experiments on the giant axon of the squid and, based on the results, developed a model of the ion permeability property of the nerve-cell membrane as an electrical circuit (deterministic equation). They repeated experiments and introduced hypotheses such as the inactivation of sodium channels in order to explain the quantitative data they obtained. Later, it was actually demonstrated that the inactivation process was initiated by the chain of amino acid residue contained in the channel molecule. This was the world's first achievement in the field of systems biology and retains its scientific significance even today. They evaluated their model's validity through simulation using the then-state-of-the-art computational technology, a mechanical calculator. This successful study is suggestive in terms of the following two aspects. Firstly, the then-state-of-the-art computational technology played a crucial role in a hypothesis-driven biological study, that is, a study in systems biology. Secondly, the field of quantitative biology based on computational simulation has historically been led by studies on the brain and nervous system.

The conductance-based model of membrane potential constructed by Hodgkin and Huxley was later integrated into software such as "NEURON" and "GENESIS," both of which make it possible to simulate the use of a multi-compartment model. "Kinetikit," a toolbox for biochemical simulation run on "GENESIS," was developed by Bhalla. With this toolbox, pioneering studies were conducted to describe synaptic plasticity, the fundamental function of the nervous system, by simulating biomolecular signal transduction. These represent the past history that the brain and nervous system has been an important target in "simulation-based biological studies." Recent advancements in measurement technology in life sciences have produced an enormous amount of experimental data, enabling modeling studies to extensively expand the range of their research targets. Even at the time of Hodgkin and Huxley, the modeling studies on the brain and nervous system stood as authentic "science" in which regularity is inferred from the tangible numbers obtained by measurement experiments.

Coming back to the present, the scale of simulation studies on the brain and nervous system is becoming increasingly larger in Europe and the United States. In Switzerland, Ecole Polytechnique Fédérale de Lausanne has been working on the "Blue Brain Project" with the full cooperation of IBM since June 2005. In 2007, it connected 8,192 CPUs together using a 4-rack-system IBM Blue Gene computer and succeeded in running simulations of a functional column of the cerebral cortex composed of 10,000 neurons. In the United States, the Pittsburgh Supercomputing Center and Salk Institute are conducting the "MCell" project to construct a detailed 3-D neuron model and examine the dynamics of each signal transduction molecule through Monte Carlo simulations. Unfortunately, studies on the brain and nervous system through a computational scientific approach have not attracted much interest in our country. The missions of our Brain and Neural Systems Team are, therefore, to launch a project unique to our country using

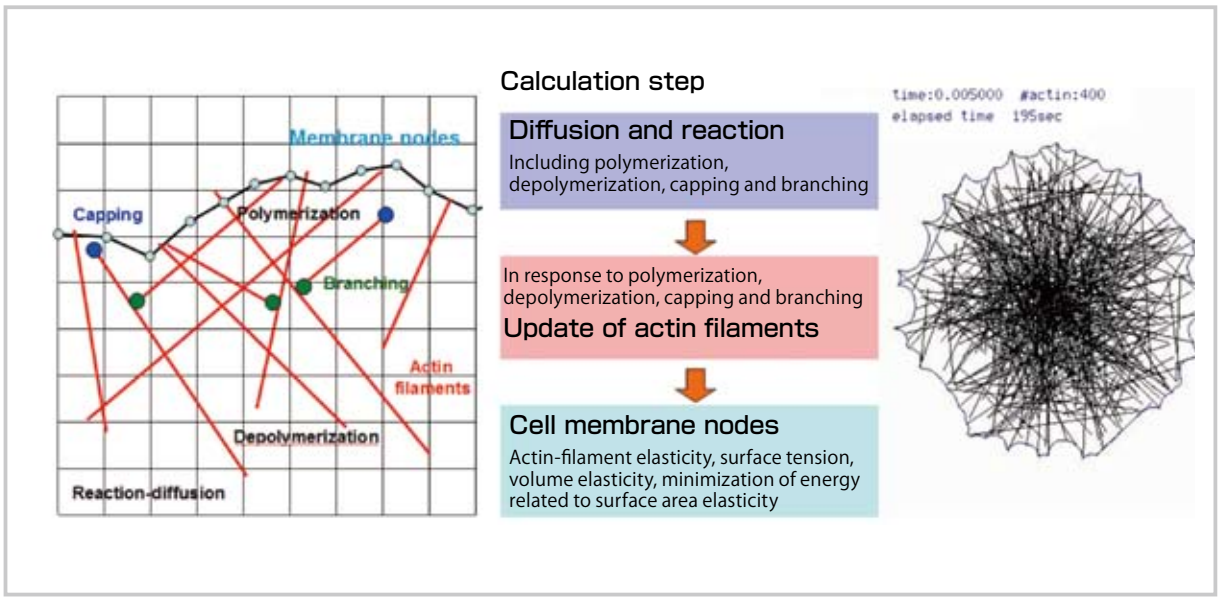


A large-scale simulation of brain functions by the Brain and Neural Systems Team

one of the world's fastest next-generation supercomputers and promote computational scientific studies on the brain and nervous system that are performed by making full use of technology for large-scale computer simulations.

Let me introduce our projects using the next-generation supercomputer. The studies in Europe and the Unites States have been conducted under the dogma that the roles of functional elements in the brain and nervous system can be unraveled by re-creating "the microdynamics of a neural circuit or neuron," while ignoring the computational theory of "the macro-scale function of the brain," that is, how information from the outside world is processed. Our team believes that it is important to reconsider this dogma, which is based on a constructive approach, and introduce a stance of computational theory. In other words, the brain functions as a whole under the interaction with the outside world (environment) and changes its functions by adapting to the environment, which is dynamic. The brain should primarily be considered to be such an information-processing and learning device. In such circumstances, the roles of elements such as neural circuits and neurons are dynamically determined in the processes of interaction with and adaptation to the environment. The behavior of the elements, therefore, should be dependent on information from the outside world as well as genetic information.

However, the human brain consists of at least 10 billion of neurons. Most of the parameters such as the connectivity between neurons and the expression or activation of various functional molecules in each neuron are unknown. As it is virtually impossible to construct a model of such a complicated system and to simulate the use of the model in a setting in a real external environment, the targets have to be limited. Therefore, the research projects our team focuses on are the visual system of mammals (humans in particular) and the olfactory system of invertebrates (insects in particular). Our goals by the time the next-generation supercomputer becomes available are to re-create the following: the dynamics of a local circuit of the cerebral cortex containing 10^5 neurons and 10^9 synapses, its local function to learn, and the spatiotemporal dynamics of the molecular network in a single neuron, which is involved in learning and development. For these purposes, we are currently working on the development of software such as "NEST" (a large-scale parallel neuron simulator) and "NeuroMorphoKit" (a membrane-skeletal-system simulator) and are conducting the research on the basic technologies necessary for software development. These will enable us to simulate the interaction of the whole visual system with the environment, by associating with the retina model and the eye movement model, both of which are also our research projects. Such simulations require combining many element models, so we are developing "PLATO," which is a common underlying platform for the construction of a large-scale model. Because the olfactory system in insects is composed of a relatively small number of neurons (10^5 neurons), we have built an environment including the database and software necessary to simulate the system of multi-compartment neuron models since the start of our research. In the aforementioned projects, RIKEN functions as the core facility. The studies are conducted at Kyoto University and the University of Tokyo in cooperation with the Okinawa Institute of Science and Technology and the Nara Institute of Science and Technology. Thus, the projects are being implemented under an all-Japan research system.



A framework of the membrane-skeletal-system simulation by NeuroMorphoKit

A Message from the Team Leader

High-performance Computing Team

High-performance Computing Environment to Maximize the Potential of the Next-generation Supercomputer

High-performance Computing Team
Team Leader

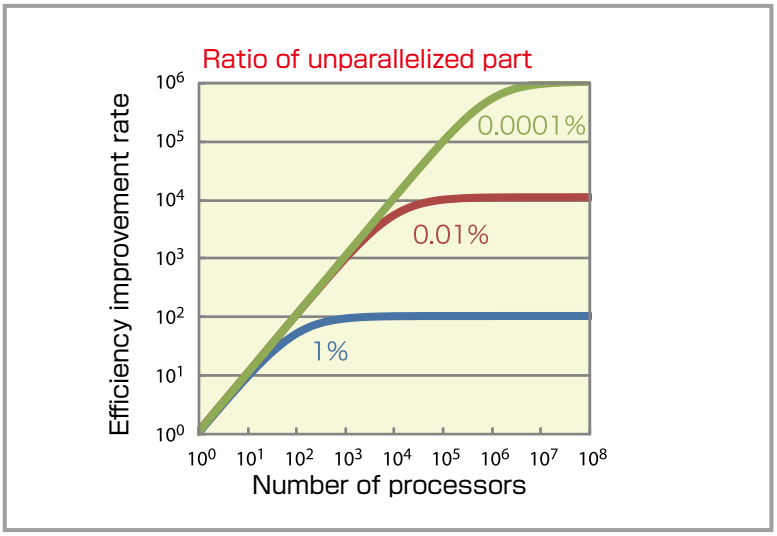
Makoto TAIJI

The key role of the High-performance Computing Team is to boost the performance of applications and to establish a platform to support the industrial use of applications from the perspective of high-performance computing and mathematical engineering. In other words, our job is to optimize the applications developed by other teams for the next-generation supercomputer to help the research and development in the Next-Generation Integrated Simulation of Living Matter.

The next-generation supercomputer is an unprecedentedly massive parallel computer that will consist of more than 80,000 processors, 640,000 cores and 5 million arithmetic units. To maximize its potential, it is indispensable to develop software that operates on the supercomputer as efficiently as possible. To realize it, the deep-parallelization far beyond the level of parallelization used in current parallel computers, which usually consist of 100 or 1,000 processors, is now required. Among the models that represent the rate of efficiency improvement expected from parallel computing, "Amdahl's law" is known as the simplest one. In accordance with this law, the rate of efficiency improvement is estimated by the proportion of the part of a program that can be executed in parallel to the part that cannot be executed in parallel. Obviously, if the ratio of the parallelizable part is small, computing speed cannot be improved even if many processors are used. When 1,000 processors are used, computing performance will degrade if the non-parallelizable part exceeds 0.1%. When it comes to 10,000 and 100,000 processors, the ratio shrinks to 0.01 and 0.001%, respectively. In this manner, the ratio reaches an advanced digit, 0.0001%, with the next-generation supercomputer. In other words, almost perfect parallel computing of 99.9999% is required. To achieve this, we must make every effort, including the introduction of state-of-the-art knowledge in computer science and further modify the algorithms based on the knowledge of applications. That is where the significance of our team's existence lies.

One of the key issues is how to boost the efficiency of communication, where data is transmitted between computing nodes. Such factors as network composition, which are essential factors for our goal, depend on the details of hardware. Therefore, we are aiming at developing optimum software by going as far back as detailed information on the hardware. As specific challenges, we are addressing support for massive parallel computing and the development of core software, as well as common infrastructure libraries and visualization software, etc., with a view to realizing higher-performance applications (See Report on Research of BioSupercomputing Newsletter Vol. 1 for the common infrastructure libraries and visualization software).

We are currently evaluating each team's representative application as their first endeavor. Each application's bottlenecks for massive-parallel computing are understood gradually now. We will report on how their performance can be improved along with our evaluations and collaborating with developers

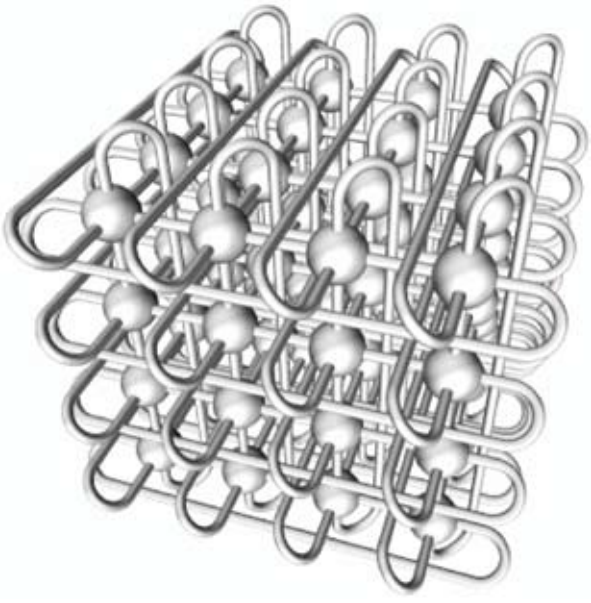


Amdahl's law: As the degree of parallelization rises, the impact on computing efficiency by a small number of unparallelized processors increases.

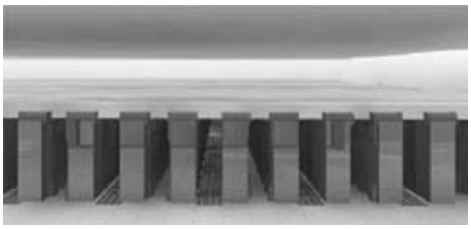
from now on. In the field of life science, I feel that there still remains a distance between the computational goal of the maximum utilization of the next-generation supercomputer and the scientific goal of solutions to actual problems. We are planning to discuss these issues with the project members and consider not only how to simply advance parallel computing, but also how we can achieve the maximum scientific results by most effectively leveraging computational resources. In addition, we consider that it is also an important function of our team to make information on hardware and operational constraints easy to understand and pass it on to developers. After the next-generation supercomputer actually starts operating, we should consider on the targets appropriate for the machine, while feeding back what problems and solutions can bring about higher performance to the user side. As we cannot know many aspects of the supercomputer until we actually run it as its parallel computing extends to such a massive scale, we believe that it is also one of our functions to convey expertise such as "usability."

While supporting each application, our team is currently developing the "MD (molecular dynamics) core program for massively parallel computing" at the same time. Raising the efficiency of each team's applications is of course an important challenge, but we have to build up our expertise in massively parallel computing in the first place. We also have another goal of verifying the next-generation supercomputer's parallel computing efficiency. Therefore, we are addressing this challenge to establish the art of computing for simulating the function of proteins at a high speed for many hours.

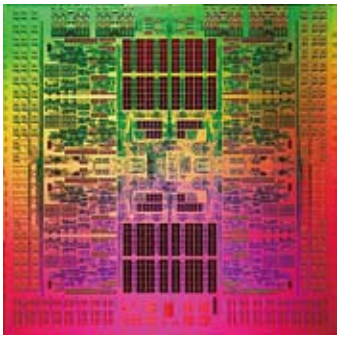
It can probably be said that other research and development teams are in their busiest phase now when they are advancing application software development, but our team feels that we will enter our busiest phase from now when the next-generation supercomputer actually starts operating. At the phase, probably we will face a number of challenges that we must tackle or haven't even thought about.



Direct-coupled network: three-dimensional torus network (conceptual diagram)



Next-generation supercomputer to be installed in computer room
(conceptual image, source: Fujitsu Ltd.)



Next-generation supercomputer CPU
(SPARC64 VIIIfx, source: Fujitsu Ltd.)

Protein Reaction Simulation Based on All-electron Calculation

Institute of Industrial Science, the University of Tokyo (Molecular Scale WG)
(From top) Fumitoshi SATO, Toshiyuki HIRANO, Noriko UEMURA, Naoki TSUNEKAWA, Junichi MATSUDA

Our group belongs to the Molecular Scale Team, which is trying to obtain a better understanding of the molecular world – the foundation of the activities of life – and is in charge of the most fundamental measures. For details of the team, please see the article written by Akinori Kidera, the team leader, in the previous issue.

The goal of our research is, in a word, to develop all-electron simulation software for proteins, which will bring out the best performance in the next-generation supercomputer. As target software, we selected ProteinDF. This software has achieved all-electron wave function calculation for various proteins using the density functional theory that is a standard electronic state calculation method. Toshiyuki Hirano is in charge of the migration of ProteinDF to ISLiM and the tuning of massive parallelization. He plays a key role in expanding the ability and functions of ProteinDF and will develop amazing simulations that make the best use of the next-generation supercomputer. Junichi Matsuda is a professional that has tuned various programs on various processors and joined our team on November 1, 2009. His assignment was about to be excluded from the budget for the next fiscal year two weeks after first participating, but he was able to join us as planed to offer his best in tuning the programs for the Venus chip. For the electronic state calculation, we need a technology that succeeds self-consistent calculation under complicated conditions, as well as brings out hardware performance. The required task for this is sort of like controlling the space shuttle from launch to landing and the larger the molecule is, the more difficult it becomes. The person in charge of this task is a specialist, Noriko Uemura. As a touchstone of her newly developed method, she is grappling with the calculation of the cytochrome c_3 molecule, which has four hemes (iron-porphyrin) in a single body and is the motif for the ProteinDF logo. (Figure 1). Finally, Naoki Tsunekawa is the researcher that is taking on the challenge of the free-energy calculation using all-electron calculation. As a strategy, he is trying to connect different energy functionals. To prove the validity of this strategy, he would not mind tackling several million quantum chemical calculations (and he actually has).

Having these members, we have been developing software that helps us understand protein molecules from an electronic state by bringing out the best in the next-generation supercomputer. We will put our energy into not only basic research, but also into research for more practical applications, such as a drag-resistance simulation for cytochrome P450 (CYP) (Figure 2). We also have great expectations for applications to energy and environmental issues, such as the CO_2 fixation of RubisCo, artificial photosynthesis and biomimetic industrial catalysts. Please join us in our research.

ProteinDF_ISLiM has already succeeded in obtaining an MPI/OpenMP hybrid parallel calculation structure and achieved the calculation in 2,500 parallel cores using a Cray XT-5. As a result of this achievement, ProteinDF_ISLiM was chosen as one of the first runner applications. We are currently in the development phase where ProteinDF_ISLiM is being optimized for the open specifications of the next-generation supercomputer. Helped by the knowledge from RIKEN and Fujitsu Ltd., we are going to catch up with the hardware development process. Please look forward to our future activities.

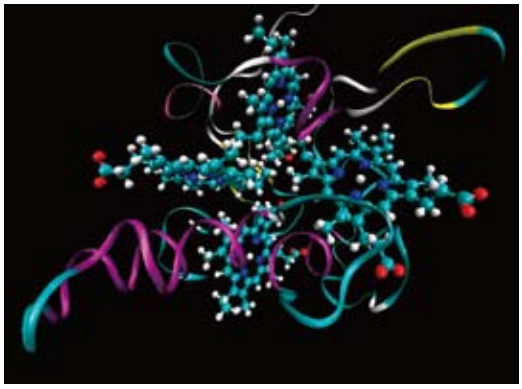


Figure 1: Structure of cytochrome c_3 for *Desulfovibrio vulgaris Miyazaki F*

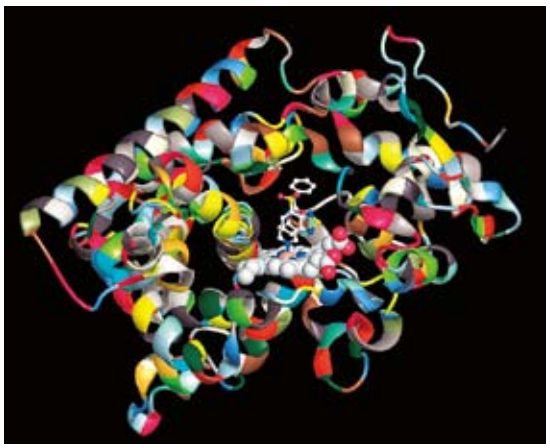


Figure 2: Structure of CYP51

Full Eulerian Fluid-structure Coupled Method

Associate Professor at School of Engineering, the University of Tokyo
(Organ and Body Scale Team)

Kazuyasu SUGIYAMA

Sixty percent of the body consists of water. Body tissues such as organs, blood vessels and blood cells are composed of soft materials. Various motions and transport phenomena, which take place autonomously and regularly throughout the body, are involved in the sustenance of life. Of these, the ones with a characteristic length of a micron-order scale or above can be considered to be the motion of a continuum and conform to the basic equations of fluid dynamics and structural dynamics. The stress characteristics of fluids and solids are mathematically expressed very differently, however. An analysis in which both of these are dealt with together and their dynamic processes are connected to each other is called a "coupled" analysis. Fluid-structure coupled simulation makes it possible to predict the effect of a treatment and help decide the treatment strategy in clinical practice. As a result, there are high expectations for its applications. It is also expected to contribute to the field of life sciences, such as in the understanding of the very essence of life and the demonstration of pathological mechanisms.

Mainly targeting industrial products, the numerical techniques of coupled analyses have been greatly developed. Based on this, many numerical studies on biomechanics are being conducted. However, it is important to further develop numerical techniques suitable for the characteristics of body tissues, which are flexible and complicated in shape, when attempting to rationalize and generalize coupled analyses.

Existing structural analyses are based on the Lagrangian method in which meshes are produced according to the shape of the target and their information is updated in accordance with the deformation of the target (i.e., equations are written based on coordinates fixed to a given parcel of the target). As in the case of industrial products, if a given blueprint can provide precise information on coordinates, the generation of meshes can be automated in many cases and accurate calculations can be performed. However, a blueprint does not exist for the human body and it therefore requires the acquisition of geometric data on blood vessels and organs from medical images obtained by diagnostic equipment such as CT and MRI, before meshes are produced. As the shape of a target becomes more complicated or the size of the system used in computation becomes larger, the automation of mesh production will become more difficult. To enable coupled analysis based on each patient's medical images to be commonly used in clinical practice, the analysis technique should be accessible even without a specialist or expertise on mesh generation.

The Organ and Body Scale Team is developing a coupled analysis technique based on the Eulerian method, in which the process of mesh production is not required (i.e., equations are written based on coordinates fixed in a space)[1]. Our approach to this development is the formulation of basic equations. For example, in the Lagrangian method, the magnitude of deformation of solids can be quantified by the relative change in coordinates of the adjacent material points after the passage of time from the starting point to the present (Figure 1 (a)). In the Eulerian method, on the other hand, the trajectory of a given material points is not tracked and therefore, some inventiveness is needed for the quantification. We define a tensor quantity that describes the deformation on each mesh. By solving its transport equation, the magnitude of deformation of solids is determined (Figure 1 (b)).

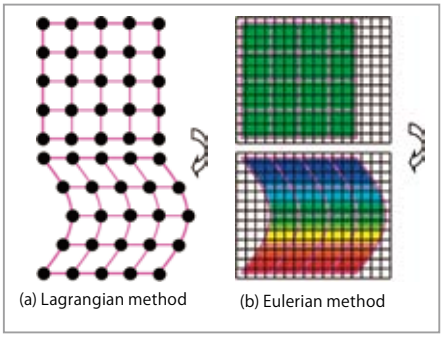


Figure 1: The methods to describe the deformation of solids

Figure 2 shows an example of Eulerian analysis, which deals with a target containing many particles[2]. If the Lagrangian method is used for the analysis of a system in which the boundary between fluid and solid changes with time, it requires a great effort to generate and reconstruct meshes. On the other hand, the use of the Eulerian method, which needs neither mesh generation nor reconstruction, makes it easily possible to perform a coupled analysis (Figure 2 (a), (b), (c)) even on a target with a complicated boundary, if the distribution of the solid volume fraction (the volume ratio of solid per mesh) at the starting point (Figure 2 (a)) is obtained.

Large-scale computation is essential when a realistic system is analyzed. A characteristic of the Eulerian analysis is to make it easily possible to perform a coupled analysis using general computational algorithms for incompressible fluid flow. For parallelization, the expertise that has been cultivated in the field of computational fluid dynamics can be utilized, which becomes a huge advantage in the realization of massively parallel computation. The introduction of this analysis method into the next-generation supercomputer and the consequent performance of massively parallel computation can have an incalculable impact on the field of biomechanics. For example, the expansion of the scale and models of computation in Figure 2 allows us to analyze a series of phenomena starting from the adsorption of platelets through the development of blood clots until their detachment, under a condition in which many red blood cells are present. It is expected that the effect of dynamics in thrombosis will be better understood.

References

[1] Shu Takagi (2009) "Simulation of the human body using the next-generation supercomputer (in Japanese)," Sugaku Seminar, 48, 58-64.

[2] Sugiyama, K., Ii, S., Takeuchi, S., Takagi, S. and Matsumoto, Y. (2010) "Full Eulerian simulations of biconcave neo-Hookean particles in a Poiseuille flow," Comput. Mech. (accepted).

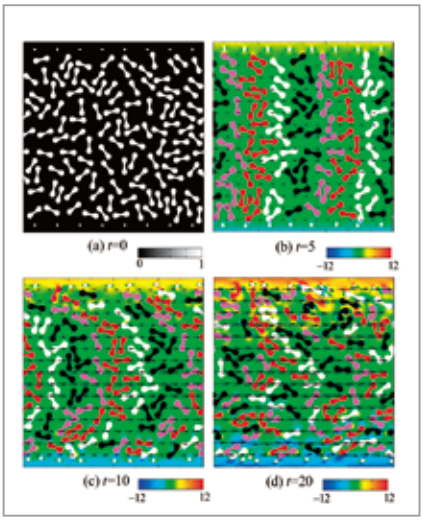


Figure 2: A distribution of many quasi-red blood cell particles in a flow through a 2-D vessel[2]. The direction of flow is left to right. (a): distribution of the solid volume ratio at time t = 0. (b), (c), (d): distribution of particles, velocity (arrow) and vorticity (background color) at times t = 5, t = 10 and t = 20, respectively. Different colors in particles are used to show the change in the position of each particle with time.

Report on Research

A Large-scale Simulation Model of Cortical Microcircuits: CMDN (Cortical Microcircuit Developed on NEST)

Brain and Neural System Team

Jun IGARASHI

The neocortex (part of the cerebral cortex), which covers a large area in the brains of mammals, is involved in higher functions such as sensory perception, voluntary movement, learning and cognition. Areas in the neocortex each respond to a specific sensory stimulus such as vision, hearing, smell and somatic sensation, with a majority of the areas consisting of six layers from the surface to the inner part of the brain. The characteristics of neurons are unique to each layer, while the synaptic connections between neurons in and outside of the layers are extremely complex, yet orderly. Although it is clear that the neurons in all neocortical layers work closely together to process information, the exact rules of processing remain unknown.

The key phenomena in understanding the mechanism of neocortical information processing are patterns of brain waves in specific frequency ranges, called gamma and theta waves. It is reported that, when an animal performs certain cognitive activities, brain waves within a specific frequency range, such as gamma waves, are generated in the neocortex, in addition to an increased correlation of brain waves between the linked neocortical areas. Brain waves are oscillatory phenomena in the local field potentials, believed to be caused mainly by the synaptic current generated when the synchronization of populations of neurons occurs. Consequently, this synchronization is likely to have some role in neocortical information processing, although the details are not yet understood. Understanding the neural mechanisms behind the synchronization in each cortical layer and its influence on interlayer correlation is essential in uncovering the information processing structure of cortical microcircuits. A simulation model should allow direct comparison to electrophysiological data as a means to navigate through the complexity of cortical microcircuits.

Building on the cortical microcircuit model developed by our co-researchers, T. Potjans and M. Diesmann, we are currently conducting simulation research with the aim of uncovering the role of the synchronization activities in the gamma-frequency range within the neocortical layer structure (Figure 1). The simulation model achieved a detailed recreation of synaptic connection patterns in neurons in each neocortical layer by applying a vast range of findings obtained from anatomical experiments. While a relatively-simple neuron model was previously employed in order to reduce the amount of calculation, the model that we currently use is a conductance-based inhibitory interneuron model as our research necessitates the recreation of neuron behaviors, especially in relation to the synchronization activities in the gamma-frequency range (Figure 2). The conductance-based model recreates more authentic electrophysiological characteristics of neurons by taking into account ion channels in the neuronal membrane. As it is reported that physiological experiments showed that certain types of inhibitory interneurons are linked to generation of subthreshold oscillations and the induction of synchronization activities in the gamma-frequency range, the ability to recreate the characteristics in detail is crucial in understanding the nature of such activities.

The cortical microcircuit model consists of approximately 80,000 neurons and five hundred million synaptic connections, making the amount of calculation for simulation far from small. Still, the size of the model translates into the equivalent of a mere 1mm^2 of the brain surface, only a fraction of the size required for the simulation of the entire neocortex. Current computers, however, do not have sufficient calculation capacity to handle the large-scale simulation model we would need next. To this end, our team is now working on the development of a large-scale simulation model for the cortical microcircuit using the next-generation super computer. The large-scale simulation model will allow us to study the interactions that occur between neocortical columns.

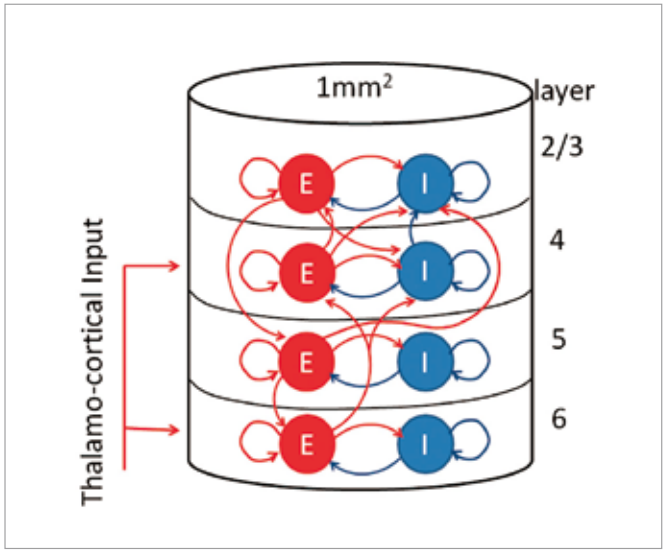


Figure 1: Diagram of local cortical neural network model
The cortical microcircuit model consists of four layers and each has two types of neurons, both excitatory (red) and inhibitory (blue). The arrows indicate major synaptic connections between neurons.

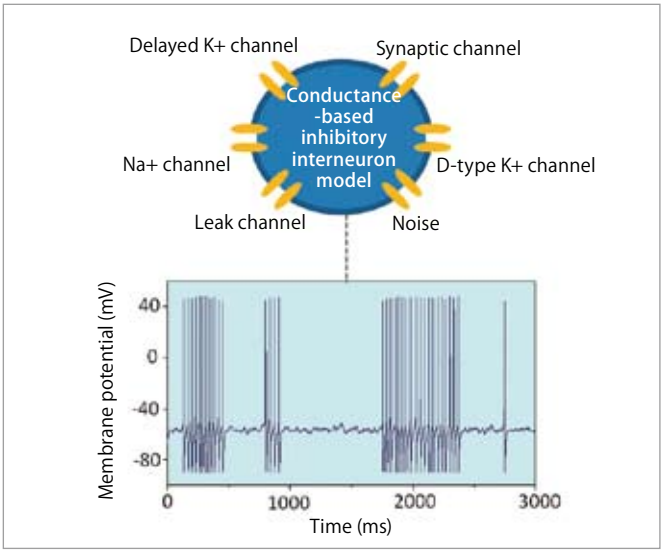


Figure 2: Diagram of conductance-based inhibitory interneuron model and firing patterns
The interneuron model has several ion channels in the cell membrane and exhibits subthreshold oscillations in the gamma-frequency range in response to noise.

Development of Next-generation Molecular Dynamics Simulation Programs

High-performance Computing Team

(From top) Hiroshi KOYAMA, Yosuke OHNO, Gen MASUMOTO, Aki HASEGAWA, Gentaro MORIMOTO

The High-performance Computing team has been currently developing molecular dynamics simulation programs as basic applications for the next-generation supercomputer. Molecular dynamics is a calculation technique used to explain the motion of molecules, such as proteins, by solving Newton's laws of motion in classical mechanics. Of course, the behavior of molecules would follow quantum mechanics in principle. The calculation of many-body quantum mechanics, however, requires a large number of calculations, even using the next-generation supercomputer. In addition, it is easier for us to understand the dynamic characteristics if we replace the motion with classical mechanics. This is why molecular dynamics is still an effective method. Molecular dynamics simulation can be one of the basic, classical research methods in the life sciences.

Then, why do we need to develop a new program? One reason is a serious requirement for scalability. The following might describe what scalability is in a simple way; "computation time is reduced by half when we use two computers." This may sound obvious, but it is essential for large-scale computing not to slow down performance, even if the number of computers is increased to a hundred or a thousand and so on. This means that we cannot create a supercomputer by just connecting cheap PCs. More importantly, scalability depends on not only the level of perfection of the hardware and software, but ultimately also the computational algorithm, since the purpose of scientific computing is to find a solution.

Figure 1 shows the run time of simulations using our molecular dynamics program for three different scales. This figure shows that computation time decreases in inverse proportion to the number of parallel computers, that is, it is scalable. On the other hand, every line has a plateau, while the number of parallel computers increases. This occurs when the communication time exceeds the computation time. In the graph, these plateaus occur at approximately 100 atoms per CPU core, regardless of the scale. This limit indicates the balance between the computing performance of a CPU and the performance of the data communication network. The number of parallel computers for the next-generation supercomputer is 100 times more than these simulations, as indicated by the arrow in Figure 1. To compute 100 times faster, we need to increase the communication speed 100 times, reduce the data amount to 1/100 or make the scale of matter 100 times larger. Since the communication speed depends on the hardware specifications, the improvement of software has a limitation. As the reduction of amount of data is a mathematical issue through an algorithm rather than a programming issue, we need to seek a new method. While this is hard to achieve, it is also interesting and challenging.

What can we do if large-scale molecular dynamics simulations become possible through the next-generation supercomputer? One application is drug discovery. Drugs are substances (keys) that bond closely to the surface of target proteins (keyholes) and develop chemical activities. Searching for and choosing these candidates is called "drug-discovery screening." The compound data used by conventional screening contains data on approximately eight million compounds in general. It has been pointed out that discovering new probable structures is already difficult. According to the theoretical estimation that assumes that approximately 10^{63} compounds satisfy drug likeness, however, screening has only been performed on a small percentage of these compounds. If we can make a virtual compound library consisting of one billion compounds that are likely to result in new structures, the possibility of discovering new drugs dramatically increases. The next-generation supercomputer and next-generation molecular dynamics programs will make this possible by increasing the speed of this task. The new library needs to calculate the assay indices for diversity, newness, drug likeness and assay availability by comparing the calculated indices to the indices in the conventional library. To achieve this, we are currently creating a prototype on the scale of 10 million to 100 million compounds (Figure 2).

Finally, we would like to thank the Advanced Center for Computing and Communication at RIKEN for letting us use the RIKEN Integrated Cluster of Clusters (RICC) system to execute the numerical simulations in this report.

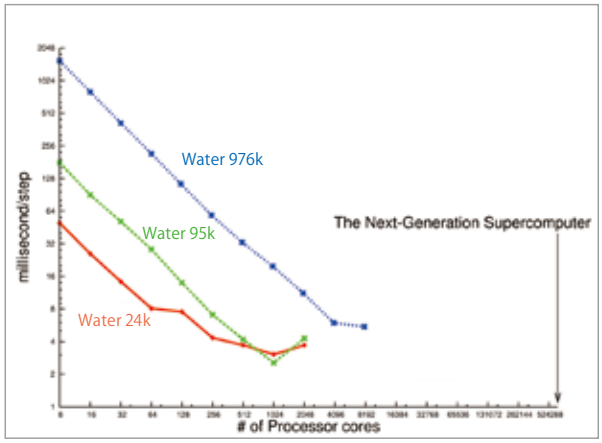
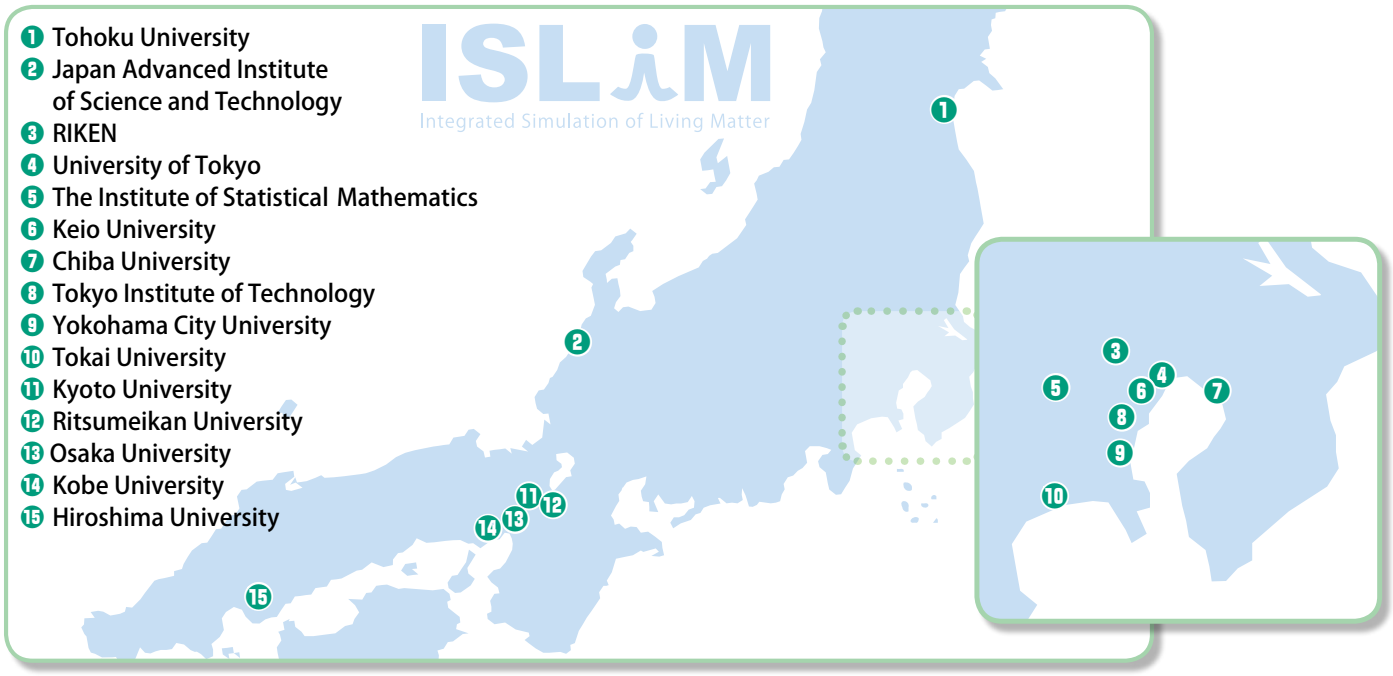


Figure 1: Water benchmark (see text)

Figure 2: Virtual compound library and drug-discovery screening

ISLiM Participating Institutions



Administration (as of September 1, 2009)

RIKEN

Headquarters/Centers
- Next-Generation Supercomputer R&D Center
- Advanced Center for Computing and Communication, etc.

Close collaborations

Close collaborations

Computational Science Research Program

Koji KAYA, Program Director
- Steering Committee
- Advisory Committee
- Research Review Committee

Ryutaro HIMENO, Deputy Program Director

Toshikazu TAKADA, Akihiko NAKANO, Atsushi MOCHIZUKI, Coordinators

Ryutaro HIMENO, Director of Integrated Simulation of Living Matter Group
- Akinori KIDERA, Leader of Molecular Scale Team
- Hideo YOKOTA, Leader of Cell Scale Team
- Shu TAKAGI, Leader of Organ and Body Scale Team
- Satoru MIYANO, Leader of Data Analysis Fusion Team
- Shin ISHII, Leader of Brain and Neural Systems Team
- Makoto TAIJI, Leader of High-performance Computing Team

In Collaboration with

Molecular Scale WG
Akinori KIDERA, Yuji SUGITA (RIKEN)
Mitsunori IKEGUCHI (Yokohama City Univ.)
Shoji TAKADA, Shigehiko HAYASHI (Kyoto Univ.)
Fumitoshi SATO (Univ. of Tokyo)
Haruki NAKAMURA (Osaka Univ.)

Cell Scale WG
Hideo YOKOTA (RIKEN)
Makoto SUEMATSU (Keio Univ.)
Yoshihisa KURACHI (Osaka Univ.)
Susumu SEINO (Kobe Univ.)
Shinya GOTO (Tokai Univ.)

Organ and Body Scale WG
Shu TAKAGI (RIKEN)
Yoichiro MATSUMOTO, Mari OHSHIMA, Toshiaki HISADA (Univ. of Tokyo)
Takami YAMAGUCHI (Tohoku Univ.)
Teruo MATSUZAWA (Japan Advanced Institute of Science and Technology)
Taishin NOMURA, Shigeo WADA (Osaka Univ.)
Hiroshi RYU (Chiba Univ.)
Shigenobu OKAZAWA (Hiroshima Univ.)
Akira AMANO (Ritsumeikan Univ.)

Data Analysis Fusion WG
Satoru MIYANO (RIKEN)
Tomoyuki HIGUCHI (Inst. of Statistical Mathmatics)
Yutaka AKIYAMA (Tokyo Inst. of Technology)
Naoyuki KAMATANI (RIKEN)

Brain and Neural WG
Shin ISHII (RIKEN/Kyoto Univ.)
Ryohei KANZAKI, Shinya KURODA (Univ. of Tokyo)
Tomoki FUKAI, Shiro USUI, Markus Diesmann (RIKEN)

Towards Multidisciplinary Research, Birth of BioSuperComputing Research Community

What is the BioSuperComputing Research Community?

Recently, rapid advances are being made both in Japan and abroad in computational bioscience, in which individual phenomena of living matters are comprehensively studied and predicted by means of computer simulation and large-scale data analysis. In the midst of such a trend, like-minded researchers joined and launched a research community called the "BioSuperComputing Research Community (BSCRC)," aiming at developing this research area, which can be notably developed only through the use of supercomputers with speeds in excess of one PetaFlops.

The "BioSuperComputing Research Community (BSCRC)" was first proposed by the BSCRC establishment preparation committee in December 2008 and launched by the founding members with a large membership of 81 (as of May 2009) on July 1, 2009.

BSCRC is interested in a wide area of computational bioscience research at various levels such as individuals, organs, tissues, cells and molecules, as well as comprehensive research integrating the above research areas. It aims at creating and fostering opportunities to get together to share experiences and ideas among researchers from various fields such as medicine, cellular biology, molecular biology, biochemistry, biophysics, pharmacology, chemistry, physics, engineering and informatics from industry, government and academia. The promotion of international collaboration and the provision of information is another aim. The community already boasts approximately 100 participating members.

While envisaging the progress of the "Next-Generation Integrated Simulation of Living Matter," BSCRC aims at undertaking activities to assist the academic culture of new bioscience that utilizes supercomputers to flourish from a broader and longer-term standpoint. It plans to contribute to the joining forces of specialists, especially young researchers, at universities, research institutions and private enterprises in various fields of study. For more information about the BSCRC, see the BSCRC web page (http://www.bscrc.jp).

The first general meeting of the BioSuperComputing Research Community was held on October 8, 2009, two and half months after the foundation.
In spite of stopped traffic in Tokyo by a typhoon, a large number of members attended.
The following five persons out of the establishment preparation committee were elected as the first BSCRC board members.

Chairman	Haruki Nakamura	Professor and Director of Research Center of Structural and Functional Proteomics, Laboratory of Protein Informatics, Osaka University
Vice Chairman	Ryutaro Himeno	Director of Advanced Center for Computing and Communication, Group Director of R&D group of Next Generation Supercomputer R&D Center, Deputy Program Director of Computational Science Research Program, RIKEN
Executive Board Members	Yutaka Akiyama	Professor of Graduate School of Information Science and Engineering, Tokyo Institute of Technology
	Akinori Kidera	Professor of Department of Supramolecular Biology, Yokohama City University
	Makoto Suematsu	Professor and Dean of School of Medicine, Keio University

At the first general meeting, the following proposal made by the executive board was approved, which is that annual member fee is waived for a certain time since the community is very young, and some fees are set for participating in lectures and seminars held by the community. Meanwhile, there was a discussion about the importance of collecting annual fees for stabilizing the operation of BSCRC and enhancing its research programs at the first general meeting. The executive board follows the discussions.

BioSuperComputing Committee (URL: http://www.bscrc.jp)

Scenes from the General Meeting

Chairman Nakamura talks about the future policies and plans of BSCRC

Executive board members introducing themselves
From the left: Vice Chairman Ryutaro Himeno acting as the moderator, Board Member Makoto Suematsu, Board Member Akinori Kidera and Board Member Yutaka Akiyama

In spite of the stopped traffic by typhoon, many members attended the first general meeting.

Event Report

Lecture on the next-generation supercomputer

On December 24, 2009, researchers participating in the Next-Generation Integrated Simulation of Living Matter (hereinafter ISLiM) attended a lecture held at RIKEN's Next-Generation Supercomputer R&D Center. Topics of the seminar included details of the next-generation supercomputer currently under development and key points in application development for obtaining optimal performance results.

ISLiM is developing software that uses the next-generation supercomputer's superior capabilities to its fullest. However, as the next-generation supercomputer itself is still in the developmental stage, an environment for the verification of such software is not presently available. As a result, our software development relies on information obtained from the next-generation supercomputer being developed.

The lecture was held twice due to the turnout of approximately seventy participants, which exceeded the capacity of the room. Lecturers from the R&D Center described the configurations, characteristics and specifications of the next-generation supercomputer, followed by key programming data and focus areas in order to ensure optimal performance.

Topics during the Q&A session ranged from a streamlined library to programming language options and compilers, as well as to the development of support software, enabling the participants to have a tangible image to reference with regard to software development. A proposal made by the ISLiM researchers in respect to applicable programming languages will be reviewed by the hardware developers.

The seminar provided an opportunity for software and hardware developers to exchange views on a wide range of topics. Through the promotion of this kind of partnership, we are continuing our efforts in developing software that enables the optimal use of the capabilities of the next-generation supercomputer.

Event Information

The 2nd BioSupercomputing Symposium

In addition to research updates, the event features presentations by leading researchers in various fields invited from around the world.

- Keynote lecture

  High performance and distributed computing in biomedical research

  Prof. Peter Coveney, Centre for Computational Science, Department of Chemistry
  University College London

Date : Mar 18 (Thu) and 19 (Fri), 2010 from 10:00 AM to 6:00 PM (Reception: March 18 from 6:30 to 8:00 PM)

Location : My Plaza Hall (Marunouchi, Chiyoda-ku, Tokyo)

- For program and registration, go to **http://www.csrp.riken.jp/2009/2ndBSCS_j.html**

About Our Logo

ISLiM
Integrated Simulation of Living Matter

The three main logos of the Next-Generation Integrated Simulation of Living Matter are essentially the same with only minor differences such as the position and size of the program's abbreviated name, as you can see in the designs on the left and on the cover page (there are actually six variations, to be precise).

They were designed to express the goal of our project, the analysis of the human body, by combining an abstracted image of a genome and the shape of the body. The small letter 'i' in "ISLiM," the abbreviated version of the project's name, is also shaped like the body.

BioSupercomputing Newsletter Vol.2 2010.3

Issued : March 15, 2010

RIKEN Computational Science Research Program

2-1 Hirosawa, Wako, Saitama 351-0198 Japan TEL: +81-48-462-1488 FAX: +81-48-462-1220 http://www.csrp.riken.jp/