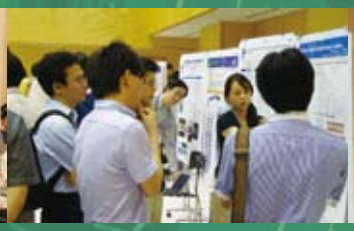
ISLiM
Integrated Simulation of Living Matter

BioSupercomputing Newsletter

Vol.1 2009.10

Next-Generation Integrated Simulation of Living Matter



The Second Computational Science Joint Workshop held on July 9 and 10, 2009

CONTENTS

●INTRODUCTION	Computational Science Research Program Program Director Koji KAYA	2
●SPECIAL INTERVIEW	○Innovative Approach for Understanding Phenomena of Life Exploring New Possibilities with Bio-supercomputing Computational Science Research Program Deputy Program Director Ryutaro HIMENO	2-3
●A Message from the Team Leader	○Simulations to Understand the Functions of the Biopolymers that Play Fundamental Roles in Life Molecular Scale Team Team Leader Akinori KIDERA	4-5
	○Develop a 3-D Model of the Entire Human Body and Understand In Vivo Phenomena to Utilize for Medical Purposes Organ and Body Scale Team Team Leader Shu TAKAGI	6-7
	○The Fourth Methodology (Data Analysis Fusion): Transforming Biology into a Predictable Science Data Analysis Fusion Team Team Leader Satoru MIYANO	8-9
●Report on Research	○Prediction of Transmembrane Dimer Structure of Amyloid Precursor Protein using Replica-Exchange Molecular Dynamics Simulations Molecular Scale Team Naoyuki MIYASHITA ／ RIKEN Advanced Science Institute (Molecular Scale WG) Yuji SUGITA	10
	○Simulation for Charged Particle Therapy Organ and Body Scale Team Kenichi L. ISHIKAWA	11
	○Prospects of Prognostic Prediction Based on Genome-wide Association Study and Genetic/Non-genetic Factors Riken Center for Genomic Medicine (Data Analysis Fusion WG) Naoyuki KAMATANI	12
	○Key Technology Supporting Petascale Computing High-performance Computing Team Kenji ONO ／ Satoshi ITO ／ Daisuke WATANABE	13
●ISLiM Participating Institutions ／ Administration		14
●Joint Workshop with VPH		15
●About Our Logo ／ Event Information ／ About the Cover Photo		16

BioSupercomputing Newsletter

introduction

Computational Science Research Program
Program Director
Koji KAYA

The 21st century has been called the “era of prediction”, and the importance of computational science in making this a reality is recognized worldwide. Computational science is theory-based, but the additional grounding of concrete experimental data makes it as vital to the future advancement of science and technology as pure theory and experiment.

Research and Development of the Next-Generation Integrated Simulaton of Living Matter is the name of a new simulation software project launched in October 2006. RIKEN is collaborating with a number of other institutions to develop petaflop-scale simulation software that will make full use of the next-generation 10-petaflop supercomputer scheduled to be completed in 2012. Two approaches are being taken in the development of this software—an analytical approach to understanding natural phenomena using basic principles, and a data-driven approach to discovering unknown pathways and laws using large-scale experimental data—to integrate and organize a variety of micro- to macro-scale research studies and data. We seek, through such efforts, to unlock the mystery of life and bring about breakthroughs that can be applied to the development of medicines and medical equipment. And in the process of doing this, we hope to establish computational science as a new methodology for the life sciences.

This newsletter has been launched to make our research activities and achievements more widely known. Inside you will find news on the latest hot topics and updates on our various research projects.

SPECIAL INTERVIEW

Bio-supercomputing: A Critical Component in the Future of Life Science Research

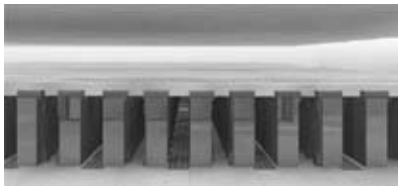
Innovative Approach for Understanding Phenomena of Life Exploring New Possibilities with Bio-supercomputing

Computational Science Research Program
Deputy Program Director
Ryutaro HIMENO

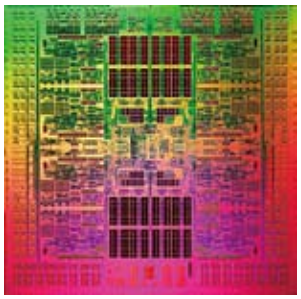
Concurrent with the research and development of a world-leading, maximum performance, next-generation supercomputer, steady progress is also being made with the application software that will take full advantage of such a computer’ s superior capabilities. In collaboration with a number of researchers, RIKEN, as the research and development base for the “Grand Challenge” application in the life sciences, which is part of the Next-Generation Supercomputer Project, is undertaking the “Research and Development of Next-Generation Integrated Life-Science Simulation” program and advocates the science of “bio-supercomputing,” a new academic field. We asked Dr. Himeno, Deputy Program Director, to share his views as to how the world of life science will change and evolve with the emergence of bio-supercomputing.

Until recently, supercomputers were used only in limited areas in the life sciences. Today, however, supercomputers enabling highly-sophisticated simulations are becoming a reality in new areas such as studies of various structure changes of protein or blood circulation. In the meantime, against the backdrop of a significant performance improvement in testing equipment over the last decade, the need for bigger computing resources and superior software capable of analyzing a vast amount of experimental data is increasingly being recognized. The advanced computing capabilities of a next-generation supercomputer will most definitely benefit areas of study that require such mass data analysis. Furthermore, we consider the "Next-Generation Supercomputer" project a great opportunity to sow the seeds to expand the use of supercomputers in other areas in the life sciences that have so far not required them.

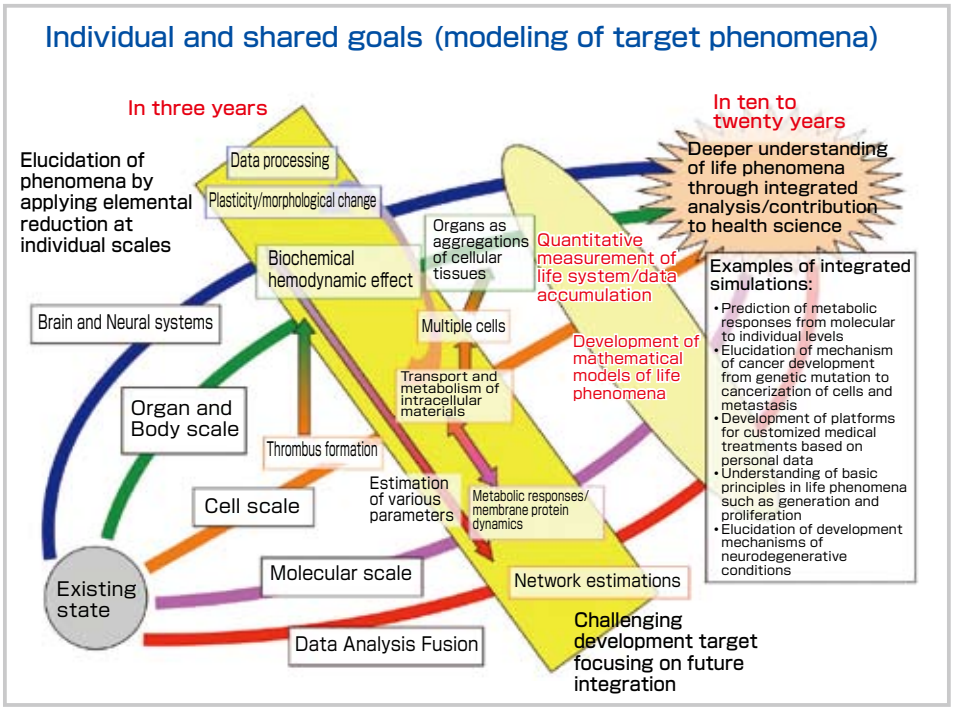
Image of the next-generation supercomputer facility



An image of the next-generation supercomputer in computing room
(source: Fujitsu Ltd.)



The CPU in the next-generation supercomputer
(SPARC64™ VIIIfx, source: Fujitsu Ltd.)



Generally, researchers tend to narrow in on a single phenomenon in a sophisticated, detailed and keen manner. Likewise, studies in the life sciences have also been moving toward unraveling mechanisms by focusing on a single element of natural phenomena, such as proteins or genes. Such an approach alone, however, cannot explain a phenomenon in its entirety. The motion of atoms and molecules, for example, basically follows the laws of physics. This behavior could be described perfectly using equations from quantum chemistry and molecular dynamics. It is not so straightforward, however, when it comes to a phenomenon within a cell, as it involves highly-complex interactions between many molecules. Furthermore, different reactions can occur simultaneously depending on locations within the cell. What we call natural phenomena are not small individual reactions or elements. Rather, these are phenomena that perform essential functions such as the creation of tissues and organs through cellular division. That is what life is. One cannot complete a picture of life by simply putting together the pieces of a jigsaw puzzle. What I believe we need is an approach that connects the pieces by integrating elements from all hierarchies, while remaining focused on the entire picture. Our focus has now shifted from analyzing elements to gaining a comprehensive understanding of life, by shifting from an empirically-defined model to one based on invariable principles.

I would like this project to be a forum for bringing researchers together to connect the individual pieces that each researcher has worked very hard to uncover. Opportunities for a molecular biophysicist and a medical practitioner to meet and have discussions have to date been quite rare. The project can provide a roundtable for people who have been isolated in various research specialties in their individual quests for answers to use as a starting point for new discoveries and the creation of new methodologies. This roundtable is what we call "bio-supercomputing." Drawing in diverse individuals under the identity of "bio-supercomputing" would cultivate a deeper awareness of being in a shared place and together we can explore and develop a new field of science. Needless to say, it would be difficult to achieve such goals with researchers in the life sciences alone. Collaboration with computer science researchers is vital for success. Furthermore, opening participation to an even broader scope of specialists will be an amazing opportunity that can leads us in new directions. Scientists in astronomy and elementary particles, for example, will be able to interconnect hierarchies in the phenomena of life at the mathematical level.

There are two internationally-acknowledged keywords in the life sciences today: "multi-scale" and "integration." The multi-scale method relates micro-scale to macro-scale. It is a universally-pursued approach in which models of different scale are integrated through the creation of new models that employ and interlock these models. In brain science, for example, simulation studies have been initiated to apply multiple scales comprehensively at one part of the human body. Our "Research and Development of Next-Generation Integrated Life-Science Simulation" program is a unique in the world, however, as it aims to integrate every phenomenon within living organisms ranging from the molecular level to the entire system.

This is an extremely difficult undertaking, as life is all too complex and diverse. Truth be told, conducting a project that focuses on a single phenomenon of life would be more likely to produce a result with a bigger scientific impact. The goal of this particular program, however, is not merely to produce short-term achievements. As development projects utilizing next-generation supercomputers continue to make further progress, studies in the life sciences are also ongoing and will be carried through into the future. Creating a specialized model designed solely around a specific aspect means the loss of universality with no possibility for further expansion. Our six teams, which focus on different themes such as molecular and cellular scales, are currently tackling several software applications and issues in parallel efforts. Although they may appear to be heading toward different goals, this is because we are pursuing subjects in the order in which the data required for model development and simulation testing are sufficiently available. In the near future, when all the other data are available, individual achievements by the research and development teams for various scales will be applied in a collaborative effort. I envision that is how the project will eventually lead us to a comprehensive understanding of the phenomena of life. For the success of the project, it is important that, in partnerships reaching beyond Japan's borders, we source software that can be standardized globally and accumulate our knowledge as a shared asset of all humanity.

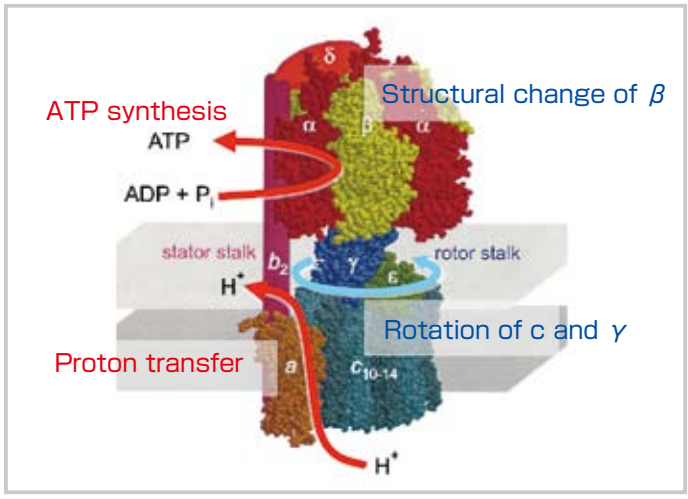
A Message from the Team Leader

Molecular Scale Team

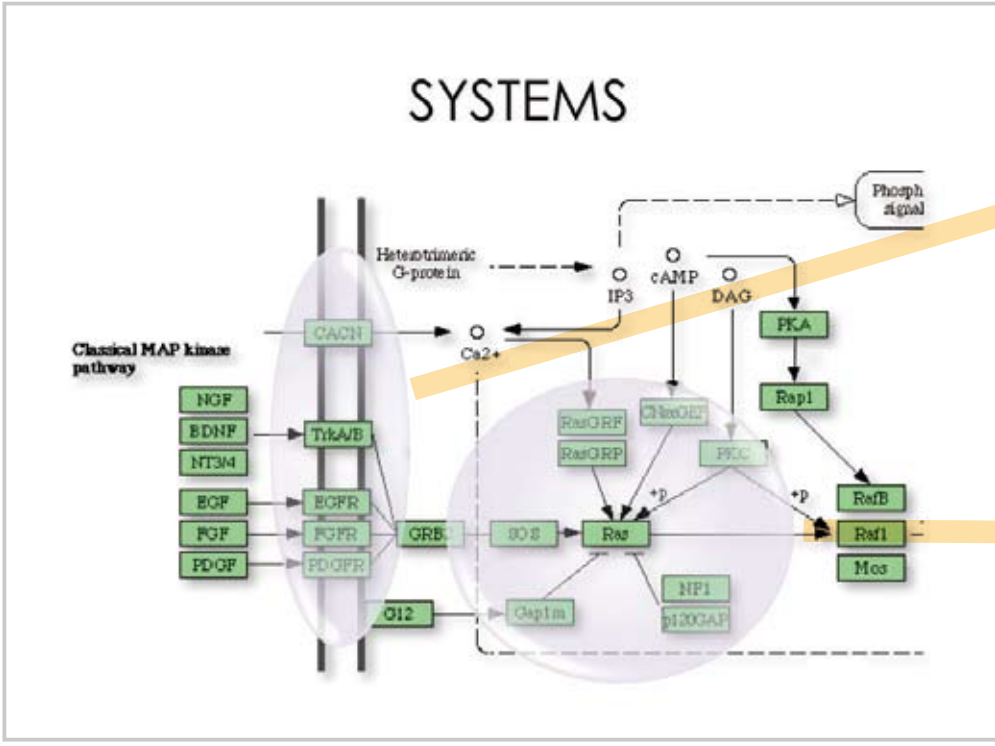
Simulations to Understand the Functions of the Biopolymers that Play Fundamental Roles in Life

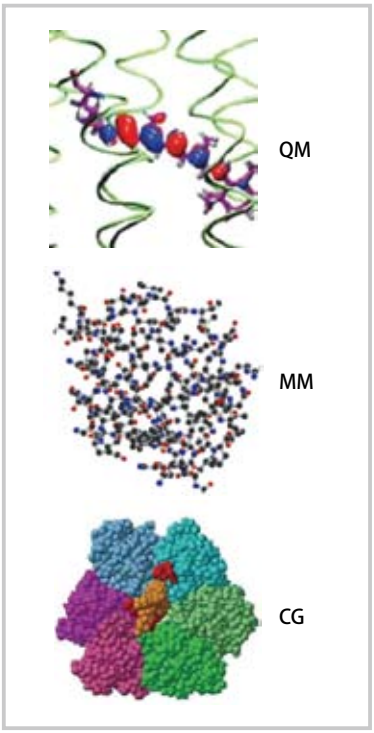
Molecular Scale Team
Team Leader
Akinori KIDERA

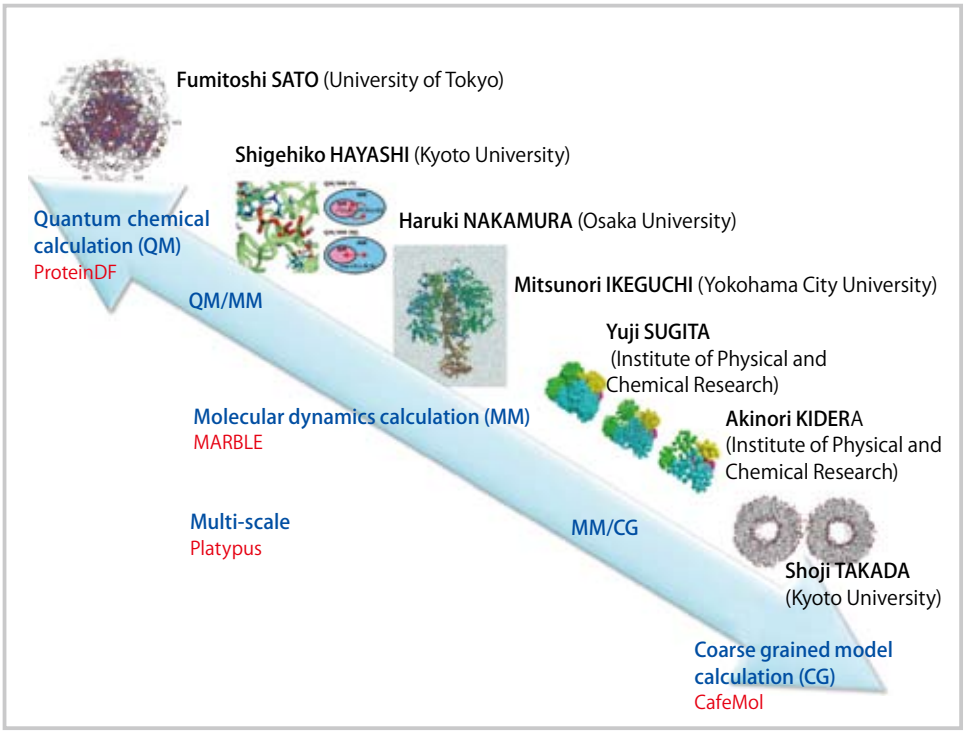
Genes can be found on the bottom level of biological hierarchy, while individual living organisms are at the top. The Molecular Scale Team has been researching the world of molecules, which come closest to the bottom level. Genes can be compared to hard disks in which an amount of information is accumulated. In order for genes to actually perform the biological activities, they need to be transformed into something functional; more specifically, proteins. Molecules include DNA or genes, in other words, as well as the RNA that is generated in the process of transcription and a variety of relevant components, but proteins are the main players among molecules. Proteins come together to perform a variety of functions. Cells can be found at the top of the bottom level, forming tissues and organs and ultimately forming individual living organisms. Our main challenge is to clarify what kind of roles proteins play in individual living organisms. Our aim is to understand the world of the molecules that form the fundamental basis for the activities of life at the most basic level as is possible by virtually activating the molecules in the supercomputer with fully utilizing the molecular simulation methods.



As an example of molecular simulation methods, I'd like to explain ATP synthase. ATP (adenosine triphosphate) is the most important energy resource for living organisms. Energy extracted from organic matter is stored in the form of the energy compound ATP and energy generated at the time of ATP breakdown, which occurs on an as-needed basis, is utilized for a variety of the biological activities. ATP synthase is involved in the process of synthesizing ATP. A synthase is a form of molecular machine made of proteins and features sophisticated functions. The synthase remains stuck in the cell membrane. The transfer of hydrogen ions from the outside toward the inside (proton transfer) generates a rotating force on the synthase, which is transformed to the energy needed for protein conformational change. This energy is used for synthesizing ATP. These functions have been confirmed through experiments and we intend to explain the principles of the functions through molecular simulations. There are two main methods for understanding these functions. One is the quantum chemistry calculation (QM), which can handle first principle calculations of chemical reactions. The other is the classical level molecular dynamics calculation (MM), which can calculate the movements. In respect to ATP synthase, proton transfer, for example, is a reaction that should be handled by quantum chemistry. The rotations of macromolecules are dynamic movements and can be understood through molecular dynamics calculations. The conformational changes can also be understood through molecular dynamics, and the subsequent chemical reaction between phosphoric acid and ADP should be through quantum chemical calculations. By utilizing these two methods, the principles of the ATP synthase functions can be understood. Computer experiments are carried out by constructing protein complexes, and activating them to initiate the reactions , such as ATP synthesis. Through these procedures, we can clarify reactions under various conditions, including the responses of protein complexes to certain







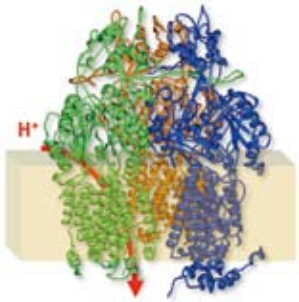
external influences, resulting in new discoveries and theories that cannot be achieved through experimentation.

The problem here is computer resources. In order to carry out the aforementioned simulations on a computer, a huge amount of calculations are required. Even if we use a supercomputer, it would be impossible to calculate everything starting from quantum chemical calculation or even from molecular dynamics simulation. To overcome this difficulty, we have been developing another method – coarse grained (CG) model calculation – in order to see the overall picture of the functional expression of proteins. For instance, the properties of each of the amino acids that comprise proteins are modeled, replacing each of the amino acids with a sphere specified with certain interaction parameters. We are trying to express biomolecular systems in three levels, specifically QM, MM and CG. The CG model calculation is a new method in the field of molecular simulation. The major issues for our team include the establishment of coarse graining methodologies, the development of new software and the development of methodologies for multi-scale simulations that are carried out by combining the three methods of QM, MM and CG (coupled simulation connecting different levels of biological hierarchy). Our main and ultimate goal is to lead our molecular level efforts into an understanding of cells, one layer above the molecular level. In other words, we are trying to explain phenomena in cells from the molecular level.

1. Membrane proteins that transport the information and substances that play the main roles in the network

Operating principle of molecular machines - Mechanistic

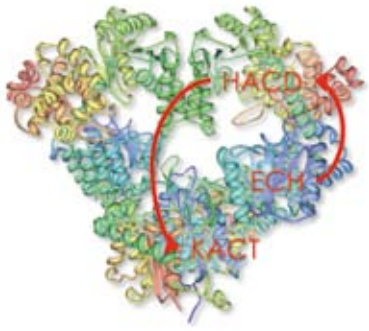
Multidrug efflux transporter AcrB



2. Multi-enzyme complex that expresses the actual condition of a network connection

Mutual interaction and dynamics - Thermodynamic

Multi-enzyme complex for fatty acid metabolism



A Message from the Team Leader

Organ and Body Scale Team

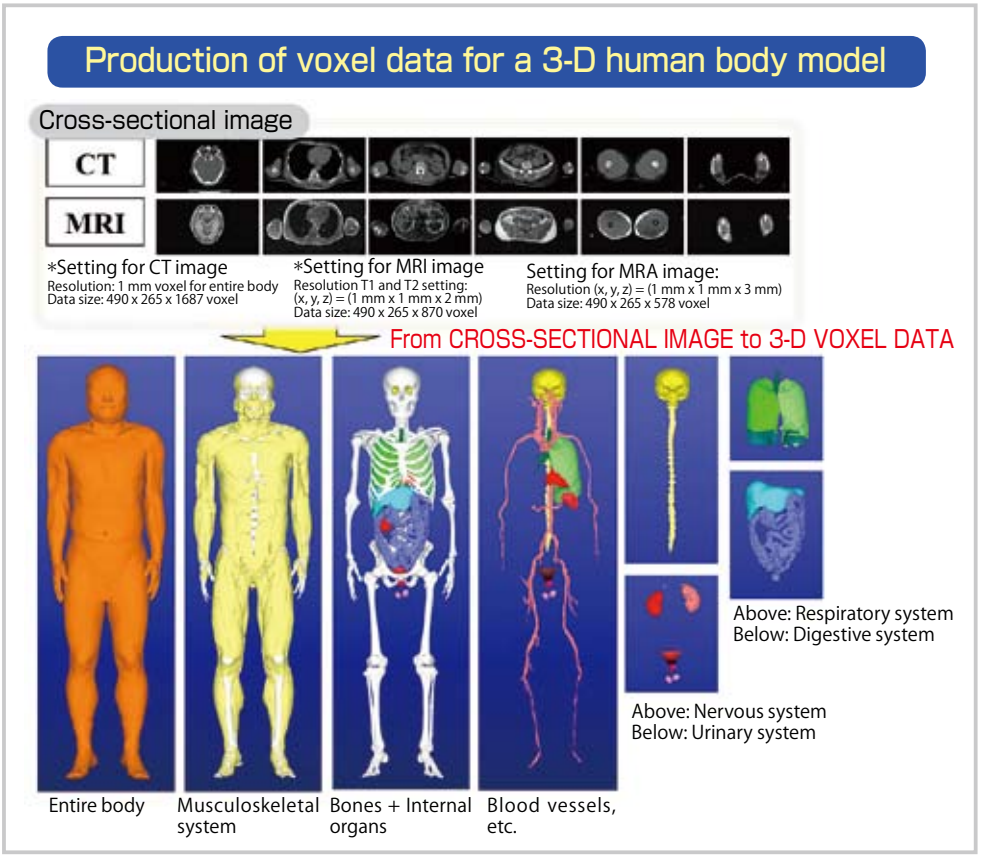
Develop a 3-D Model of the Entire Human Body and Understand In Vivo Phenomena to Utilize for Medical Purposes

Organ and Body Scale Team
Team Leader
Shu TAKAGI

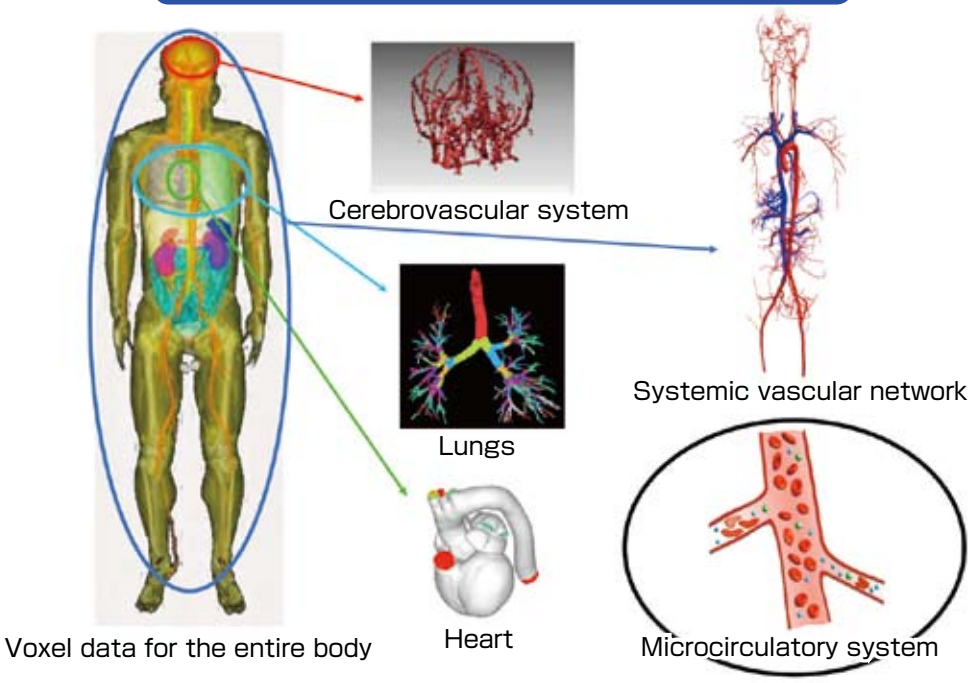
The general goals of our Organ and Body Scale Team are to develop a computer 3-D model of the entire human body with all its organs and systems, including the circulatory system, the respiratory system, the musculoskeletal system and the nervous system, based on image data obtained through MRI, CT and ultrasonography, etc. and to perform various simulations with this virtual human body for the purposes of acquiring a further understanding of biological phenomena, demonstrating the mechanisms of diseases, predicting pathological paths and ultimately making the simulations a useful tool for diagnosis and treatment in clinical practice. In other words, our goals are to bring the "static information" of medical image data to life and to revive it as dynamic information. Therefore, we think that it is a priority to develop and adapt new analytical methods such as the Eulerian fluid-structure coupling algorithm. The Lagrangian method is generally used in the analysis of solids and related equations are written based on the trajectory of a material point. This method requires the generation of a mesh in accordance with organ distortion and has the problem of poor affinity for voxel data of the entire body. On the other hand, the Eulerian method is commonly used in the analysis of fluids and related equations are written based on a fixed position in space through which a material moves. This method demands greater mathematical complexity, but becomes a very effective means of dealing with medical image data.

In our team, research is vigorously conducted by five sub-groups based on the target phenomena or study approaches. Specifically, these are "The sophistication of voxel-data production for the entire body and the development of a model with respect to systemic dynamics among organs and their systems," "The development of a simulator for the in vivo propagation of ultrasound and heavy particle beams as a supportive tool for minimally invasive therapies," "The implementation of a heart simulator in next-generation supercomputers," "The development of a vascular network model and the integrated simulation of blood circulation" and "The integrated simulation of pulmonary respiration and pulmonary circulation." One of the ultimate goals of these projects is to find an answer to the question "What is life?" To us, the body is not simply represented by an assembly of separate organs. It functions as a whole as a result of organs that interact with each other. As such, an understanding of the entire body is our primary attitude towards our research. For example, vascular networks connect organs with each other and enable their different functions to work together organically as a whole. We therefore consider vascular networks to be very important and using voxel data of the entire body, we are currently aiming to develop an integrated multi-scale circulatory-system simulator by establishing a dynamic model of blood circulation involving the heart, the arterial/venous vascular system and even the capillaries. Numerical analyses of the circulatory system are also performed in terms of dynamics in blood circulation, including the agglutination of platelets and red blood cells and adhesion to vessel walls in a condition of thrombosis or microcirculatory disturbances. In particular, studies on thrombosis are important because thrombosis, as a circulatory disorder, can cause serious harm to the body. When considering only blood vessels in the heart or brain being blocked with clots, we tend to take it as a local phenomenon, but blood clots can actually be formed as a result of the accumulation of the platelets activated in damaged vessels in any part of the body. Thrombosis, therefore, can affect every part of the circulatory system.

Blood plays an important role as a medium for transporting substances throughout the body. Extremely important, in particular, is the transportation of oxygen and carbon dioxide. In this sense, studies on the functions of the lungs become indispensable for research and we are working on the simulation of the process of gas exchange in the lungs. The heart, of course, is a crucial organ for life and a highly-advanced simulator for the heart, the multi-scale and multi-physics heart simulator, developed by the University of Tokyo, has been developed. If this simulator is implemented in next-generation supercomputers, it will even be possible to simulate everything from the movements of each cardiac muscle cell to the movements of the heart as a whole, thus making a major contribution



Simulation of circulatory system

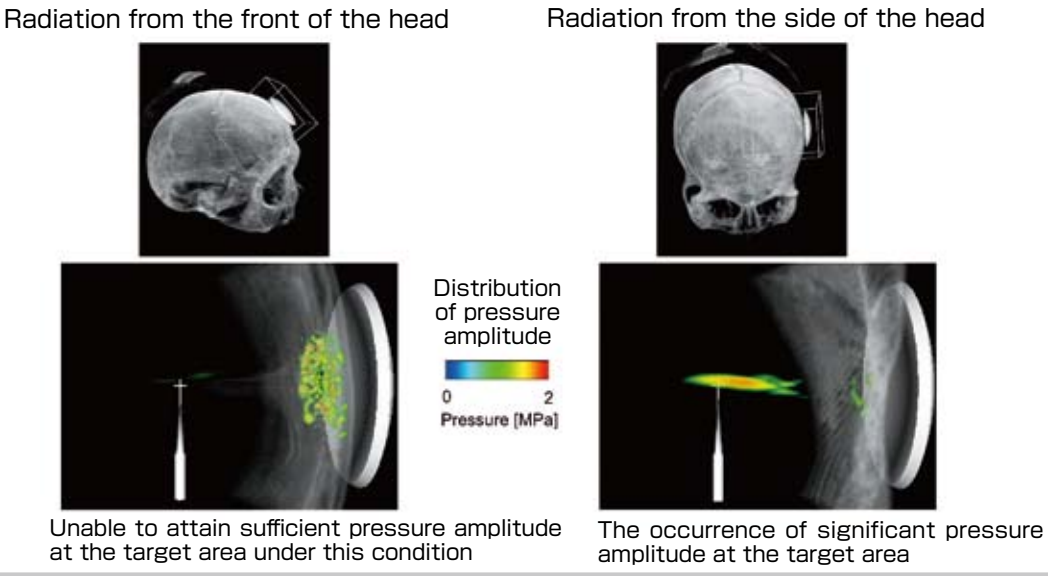


toward advanced medical treatment for cardiac disease.

For medical applications, the research project we are committing to the most is the development of a simulator for less invasive radiation therapy using ultrasound or heavy particles. The cauterization of tumors found in prostate or breast cancer, with focused ultrasound, has been performed in actual clinical practice. Radiation can be very difficult, however, when the target organ is surrounded with bones or located deep inside the body, because the body consists of various parts that may have totally different acoustic wave propagation properties. Determining the intensity and focusing area for irradiation depending on the characteristics of individual bodies will become very useful in clinical practice. In the future, even the cauterization of brain tumors through the skull will become possible.

The key to simulating a living body, we believe, is a clear definition of where uncertainty is involved. For example, medical image data used to produce voxel data for the entire body itself has some degree of uncertainty, because it cannot illustrate the precise position of a surface. In my field of basic research on fluids, many researchers discuss how to reduce this uncertainty by a few percentage points, but in many cases, such strict accuracy is unnecessary in clinical practice. It is considered fundamental that "the human body is too complicated for the mechanisms of the body to be described simply and clearly." In the process of developing a model for something as complicated as the body, the occurrence of some uncertainty cannot be avoided. This does never mean, however, that uncertainty can be allowed in numerical calculation methods. Calculation methods have to be developed with mathematical consistency and with less uncertainty being involved. In each calculation method, it is essential to first have a clear understanding of where uncertainty is involved and where, consequently, accuracy is reduced and then to use the results of simulation with a living body model, which has some degree of uncertainty.

Calculation for radiation of ultrasound through the skull



A Message from the Team Leader

Data Analysis Fusion Team

The Fourth Methodology (Data Analysis Fusion): Transforming Biology into a Predictable Science

Data Analysis Fusion Team
Team Leader
Satoru MIYANO

I may face scorn for saying this, but biology really is the only field of science that does not have a "scientific" language. Practically speaking, no discoveries of what can be called a principle have been made and researchers continue to basically present facts similar to theorems in Mathematics. For example, we may prove with a molecular biology method that a certain gene controls a particular gene cluster and, as a result, can be identified as the cause of a particular phenotype. We write about facts that support such findings. However, while this may be fascinating in its own right, we are basically just letting the world know we have discovered this or that. With this approach, even after a thousand years, life science would not evolve to be a "predictable science."

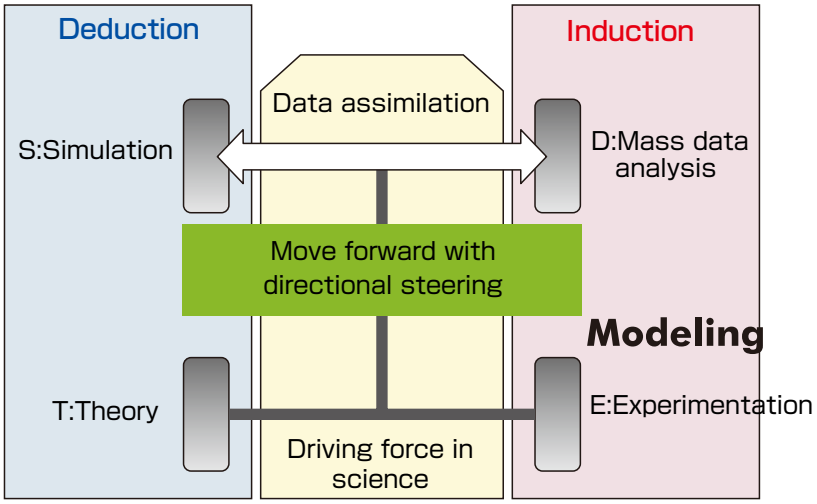
In 2003, the National Institute of Health (NIH) in the U.S. launched a roadmap for biomedical research to fully benefit from the completion of the human genome project. It includes the following sentence: "All of these techniques generate large amounts of data and biology is fast changing into a science of information management." At the time of publication, biologists and medical researchers were bewildered by the message. Now, however, everyone strongly recognizes this to be a reality. That is, although data generation continues to be important, the focus of biology as a science is shifting toward how to analyze and interpret vast amounts of data. Concurrent with efforts to effectively deal with huge quantities of facts, there is also a need to change biology into a "predictable science." An acknowledgement of the importance of computational science in the life sciences reflects this need. Therefore, it is important to pursue the development of life science software that uses next-generation supercomputers, which will lead to the establishment of an information infrastructure to make this a reality.

Our Data Analysis Fusion Team is divided into four groups, each with its own theme: "estimations and applications of large-scale genetic networks," "development of new algorithms that relate to large-scale genome polymorphism data and phenotypic data and a review of the validity and usability of these algorithms," "estimations and applications of large-scale protein networks" and "development of data assimilation technologies for the simulation of living matter." The groups work synergistically to establish more sophisticated modeling technologies by analyzing networks of molecular interactions based on gene expression data, etc., then combining the extracted models with estimated dynamics.

Here is a simple overview on our research and development. We first predict a large-scale molecular network by computing data such as gene expression data. Using the results as a map, we then look for molecular interaction or genes that may be relevant to drugs and diseases. In the field of geophysics, for example, exploration by artificial satellites made global earth maps readily available and allowed researchers to keep track of temperature and humidity data. Likewise, in life science, DNA chips allowed us to identify the gene expression of a particular cell at a glance. With such information, we can ascertain the correlation between certain molecules. In other words, we

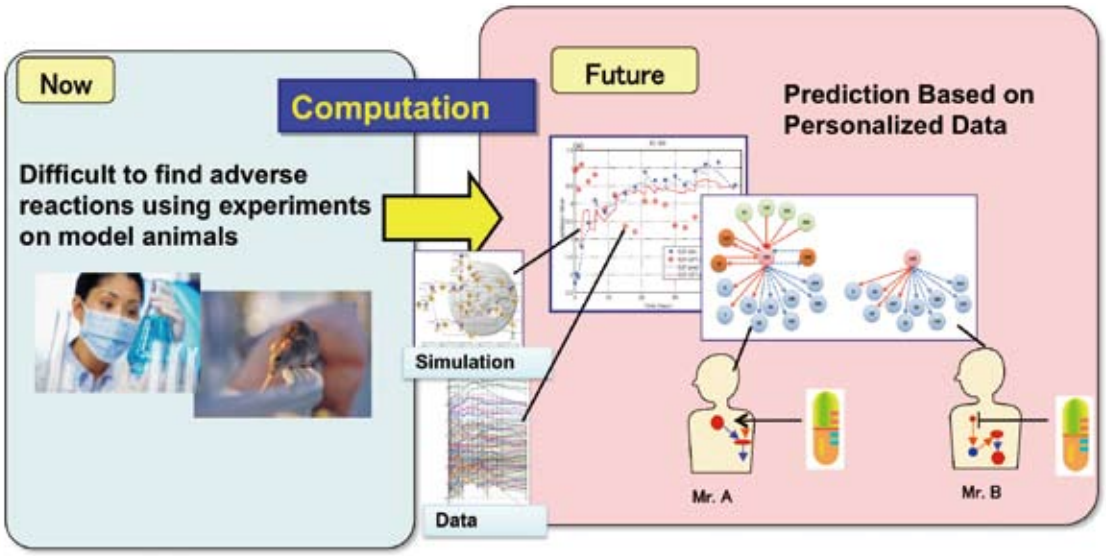
Data Analysis Fusion Team

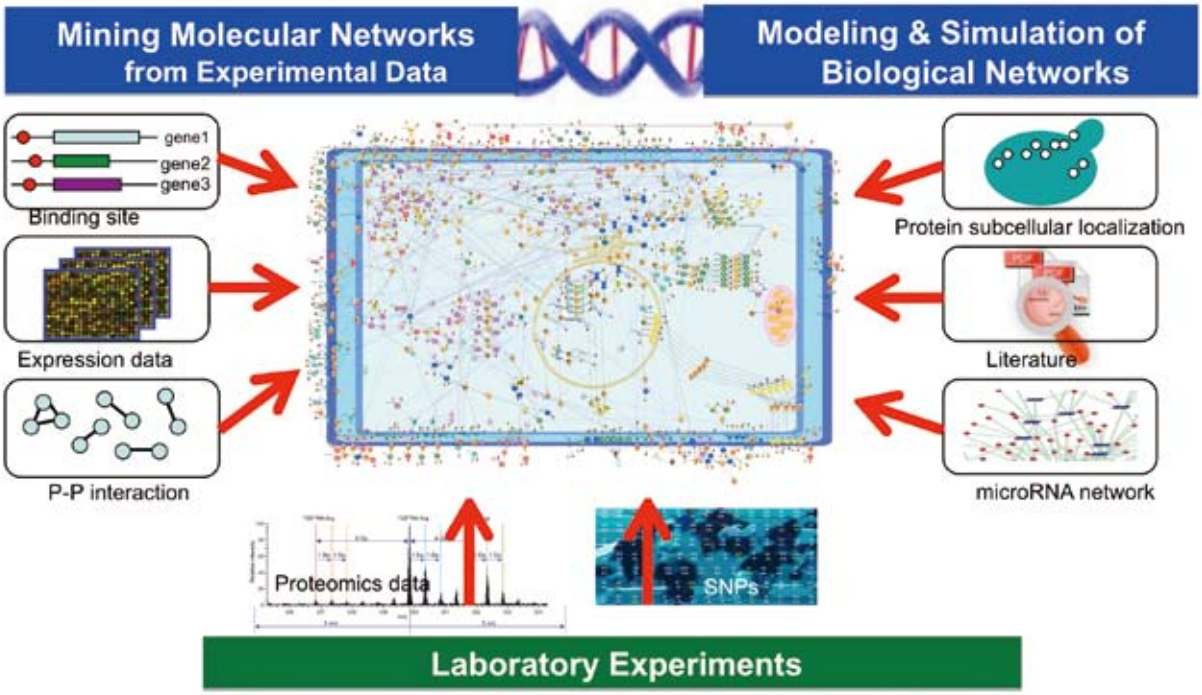
Data Analysis Fusion: the fourth science, or the fourth methodology, following theory, experimentation and simulation



The team works on the development of applied technologies such as algorithms for petascale analysis of ever-expanding quantities of genomic and genetic data, as well as modeling technologies that fuse data and simulation models through data assimilation. Our future goals include the creation of medical information technologies that can be applied in drug target searches and individually-designed treatments. Such technologies will make drug target searches from the entire human gene structure a reality.

With strong synergies both within and outside of the team, we are currently working on establishing networks within living organisms, estimations of dynamic simulation models and a new technology that can create an individual model from a standard model using data assimilation, under the theme of "lung cancer and related drugs."





are developing a map of a dynamic network, which is to be further refined by identifying where the gene responsible for a certain disease and its molecular interaction are located on the map, in order to develop an even more sophisticated model. Due to the number of variations of phenomena in an actual human body, however, a standard model is not always applicable. We could overcome such issues by employing data assimilation techniques. This approach will allow us to create a personalized map in which the individual's information is incorporated into a standard model based on molecular-biological facts along with networks and dynamics from a vast amount of data.

What's important here is that as long as a biologically-sound standard model exists, it is possible to then build a specific model for an individual by applying personal data, even with the existence of many variables. Building a tailor-made, personal model from scratch could take an untold number of years. If we can make such a model with the use of data assimilation, it will allow us to conduct simulations to identify which drugs are needed for each patient and open up new possibilities in therapy.

Our new approach, "data analysis fusion," integrates deduction (i.e., simulations) and induction (i.e., mass data processing) through data assimilation. It is the fourth science, or the fourth methodology, a new addition to the three pillars of science: theory, experimentation and simulation. In addition, under the theme of "lung cancer and related drugs," we are currently developing a new technology that can create an individual model from a standard model using the state-space modeling method.

Prediction of Transmembrane Dimer Structure of Amyloid Precursor Protein using Replica-Exchange Molecular Dynamics Simulations

Molecular Scale Team
Naoyuki MIYASHITA (left)

RIKEN Advanced Science Institute
(Molecular Scale WG)
Yuji SUGITA (right)

Alzheimer's disease is a neurodegenerative disorder and its main symptom is to disturb cognitive functions. In the progressive process of this disease, the death of nerve cells in the brain (neurons) occurs. The most widely accepted hypothesis for the cause is the amyloid hypothesis, in which Alzheimer's disease results from the aggregation and accumulation of amyloid β(Aβ) peptide in the brain. Therefore, it is vitally important to acquire better knowledge of the formation of Aβ peptides in understanding Alzheimer's disease.

Amyloid precursor protein (APP) is composed of approximately 700 amino-acid residues in the membrane of neurons. Aβ denotes a series of approximately 40 residues in APP and is derived from the sequential cleavage at both ends of these consecutive residues (upper and lower) by β - and γ -secretases, respectively. A$\beta_{1\text{-}40}$ consisting of 40 amino-acid residues is a primary isoform, whereas a varying γ -cleavage site can produce A$\beta_{1\text{-}42}$, which is a major component of senile plaques. How can these different lengths of Aβ peptides be produced from APP? The transmembrane structure of APP has not been revealed by X-ray crystallography, presumably due to the large fluctuations of APP. We attempt to predict atomic structure of APP in the membrane, using molecular dynamics simulations and compare the results with recent biochemical experiments.

In this study, we focus on A$\beta_{23\text{-}55}$, in the transmembrane (TM) and juxtamembrane regions , because contains key residues in APP. A$\beta_{23\text{-}55}$ includes five consecutive glycine (Gly) residues and is considered to form a pair (dimer) in the membrane. In fact, recent experiments have demonstrated dimerization of wild-type APP (having a normal amino-acid sequence) in the intramembrane. Interestingly, it has been reported that when two Gly residues (Gly_{29} and Gly_{33}) out of five are replaced with hydrophobic* leucine (Leu) residues, this mutant also forms a dimer as in wild-type APP but cannot be cleaved by γ -secretase (which results in no production of Aβ peptides)[1]. How can the mutations of only two residues in the amino-acid sequence affect the structure and function of APP ?

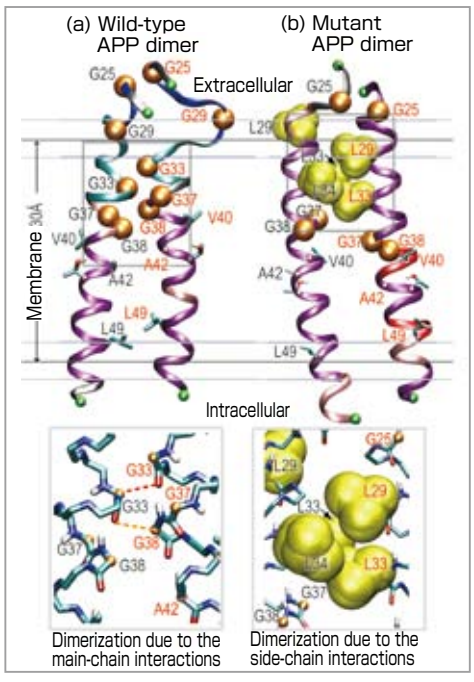


Figure1 : Predicted structures

To answer this question, it is usually effective to perform molecular dynamics simulations that can deal with molecular interactions in protein. If atomic structure of the relevant protein is not determined by X-ray crystallography, a thermodynamically stable conformation selected from various possible conformations by computer simulations can be the basis for discussion. Generalized-ensemble algorithms, including the replica-exchange molecular dynamics method[2] used in this study, are especially suitable for this purpose and many new methods have been developed by researchers in Japan, such as Professor Yuko Okamoto in Nagoya University. Our results[3], as shown in Figure 1, indicate the differences in dimer formation between wild-type and mutant APPs. In the mutant APP, the increase in the number of hydrophobic amino-acid residues causes APP to tilt toward the membrane and thus the site for cleavage in APP no longer matches the active site of γ -secretase [Figure 2]. The calculation results are compatible with the recent experimental data and moreover the molecular mechanism suggested by us has attracted the interest of many experimental researchers.

In this study, because of the limitation of computational resources available at present, we do not explicitly include solvent and phospholipid molecules, but rather use a model in which the influence of solvent and membrane are included implicitly, considering the effect of the excluded volume. In the next-generation supercomputer, we can include solvent and phospholipid molecules in the simulations of membrane proteins and expect to obtain more reliable simulation results. The replica-exchange molecular dynamics simulation is a promising methodology in the next-generation supercomputer, because it can effectively simulate substrate binding or conformational changes in large proteins as the number of CPUs available increases. In addition to Alzheimer's disease, phenomena involved in many other diseases are also associated with the behavior of membrane proteins. We hope that our studies will lead to a better understanding of the biological activities associated with various diseases and can be utilized for their treatments.

* Note : Twenty kinds of amino acids, which constitute proteins, share a common structure called the main chain, in which a hydrogen atom, an amide group (NH_2) and a carbonyl group (CO) are bound to a carbon atom (C_α) in the center. Each amino acid is characterized by its side chain that is also bound to the C_α atom. Twenty kinds of side chains can be categorized into two groups: hydrophilic, with a property of attracting water and hydrophobic, with a property of repelling water.

References
[1] P. Kienen-Campard et al., J. Biol. Chem. 283, 7733-7744, (2008)
[2] Y. Sugita and Y. Okamoto, Chem. Phys. Lett., 314, 141-151, (1999)
[3] N. Miyashita, J.E. Straub, D. Thirumalai and Y. Sugita, J. Am. Chem. Soc., 131, 3438-3439, (2009)

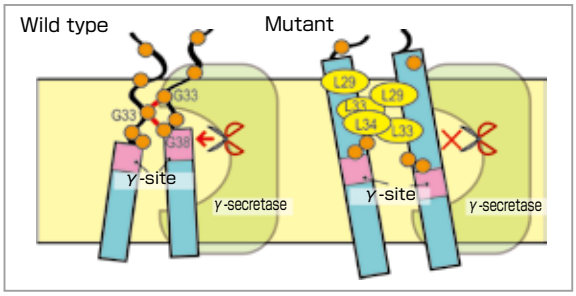


Figure 2 : Relationship between γ-secretase and γ-site

Simulation for Charged Particle Therapy

Organ and Body Scale Team

Kenichi L. ISHIKAWA

Charged particle cancer therapy is an advanced radiotherapy to cure cancer noninvasively (without surgery) by intensively irradiating the tumor with a charged particle beam (a beam generated by accelerating ions such as carbon ions). X-rays used in conventional radiotherapy do not stop at the tumor and go through the body, thereby uniformly affecting the tissue along their paths. In contrast, a charged particle beam deposits maximum energy immediately before it stops, as shown in Figure 1. Thus, a charged particle beam with an appropriately chosen energy stops exactly at the cancer focus to attack the tumor in a pinpoint manner. Charged particle therapy is presently approved as a highly advanced medical technology and is attracting increasing attention.

When performing a surgical operation, a doctor can visually confirm the area for resection, but during charged particle therapy, it is impossible to monitor which area is being irradiated. Thus, it is required beforehand to calculate the desired condition of the beam irradiation so that the radiation dose (the amount and the effect of the radiation) can be efficiently concentrated on the cancer focus. Hence, the quality of therapy can be further improved by sophisticated simulation.

A charged particle beam collides with atoms in the human body and causes various kinds of phenomena, such as ionization and nuclear reaction. In a human body, composed of different kinds of tissues, how and where the beam reacts is complex and stochastic. Based on the Monte Carlo simulation, which uses random numbers to trace the transport and reaction of numerous incident particles in its entirety as much as possible, we are endeavoring to develop a method to calculate the dose of charged particle radiation more accurately.

In the current therapy, calculations usually cover only the effect on the cancer focus and its periphery. Since a charged particle beam is a type of radiation, however, it can cause secondary cancer years or decades after recovery from the initial cancer. It is an irony that the therapy has become so effective that the treatment planning now has to take account of the risk of secondary cancer that the therapy itself may cause several decades later (how greedy humans are!). For that purpose, we are conducting research on how to calculate the dose distribution of charged particle beams in the entire human body. Figure 2 illustrates an example of the results of the whole-body dose calculation using a voxel phantom, which represents the body as a group of small cubes (voxels). Through the calculation mentioned above, our simulator can determine the type of reaction or particle that contributes to effectively delivering a dose to areas far from the cancer focus. Thus, the simulator is expected to provide safer therapy with a lower risk.

The human body moves during respiration, but even so, if CTs are taken repeatedly, the amount of X-ray exposure may be increased. To overcome this problem, we are researching the effect of respiration by simulating the movement of the lung by computers. By utilizing the simulation coupled with the model of lung motion (the spring network model) developed by Prof. Shigeo Wada's group at Osaka University, we calculated how a dose of charged particle radiation is affected by respiration. The calculation results are shown in Figure 3. The dose distribution at inhalation differs from that at exhalation. It is expected that more precise therapy is realized if such effects are taken into consideration.

The simulation we are working on requires a huge amount of calculation. We are dreaming to develop a system capable of elaborate treatment planning on the entirely computational basis from a single CT scanning, with account of the movement of organs and the risk of secondary cancer, using the next-generation supercomputer.

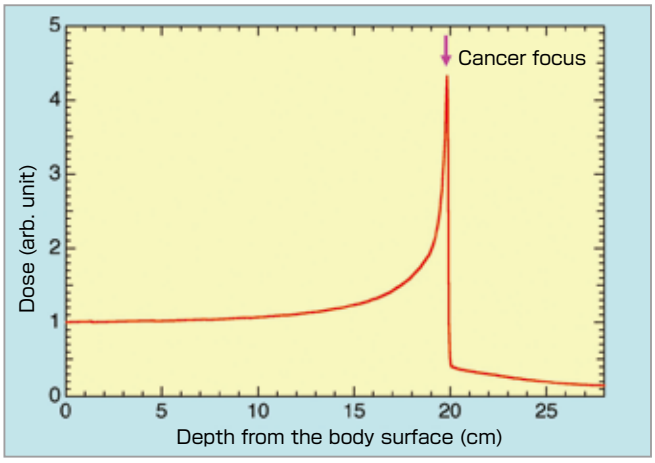


Figure 1 : Relationship between the depth and dose of charged particle radiation

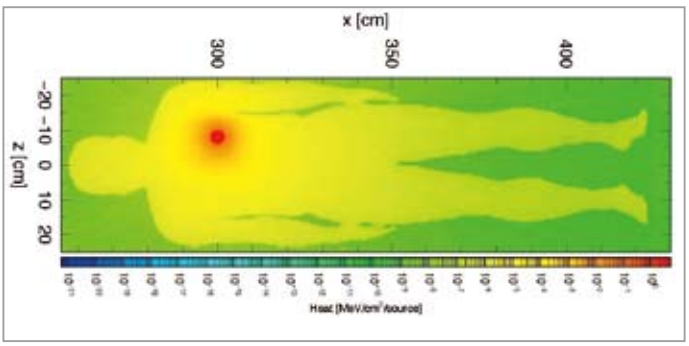


Figure 2 : Whole-body dose distribution calculated with a human body voxel phantom. (for the case of carbon ion beam incidence to the lung with an energy per nucleon of 140 MeV/u)

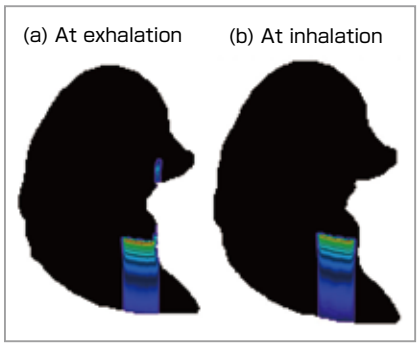


Figure 3 : Calculated dose distribution on a horizontal plane at exhalation (a) and inhalation (b) when a carbon ion beam with an energy per nucleon of 70 MeV/u and a diameter of 2cm is incident to the lung.

Prospects of Prognostic Prediction Based on Genome-wide Association Study and Genetic/Non-genetic Factors

Riken Center for Genomic Medicine
(Data Analysis Fusion WG)
Naoyuki KAMATANI

Studies on the relationship between personal difference in genome sequences and traits have been rapidly promoted since the elucidation of the human genome in 2003. Traits refer to attributes that vary from person to person, for example, "being disease or non disease" or "responsiveness to a certain drug." In respect to Mendelian disorders, a method called "linkage analysis" has been established that can almost certainly identify causative genes if sufficient genealogic information is provided. Linkage analysis was first proposed by Fisher in 1922 with the use of the "maximum likelihood method," a mathematical approach developed by Fisher himself. Along with the development of numerous markers for the human genome and the improvement of computer performance, linkage analysis was quickly applied to the elucidation of genetic diseases.

Next, researchers became interested in multi-factorial traits. Multi-factorial traits do not have the Mendelian genetic form, but rather a complex genetic form and are presumed to be influenced by multiple genes and the environment. Traits can be classified into quantitative traits and qualitative traits, many of which have two phenotypes. The types of influences on qualitative and quantitative traits induced by multiple genes and the environment were formulated by Fisher in 1918 as an additive polygene model. Based on this model, linkage analysis using data on numerous (500,000-1,000,000) markers for the human genome is currently under way. This approach is known as the "Genome-wide Association Study (GWAS)" and it is a prominent method to understand the genetic factors of multi-factorial traits. GWAS was successfully adopted for the first time anywhere in the world by the Riken Center for Genomic Medicine (then called the SNP Research Center) in 2002.

The most important task of GWAS is data cleaning. Since several hundred thousand pieces of information for each individual are obtained from hundreds or thousands of people, it takes a lot of work to clean the data involved. The second most important task is to perform an assay, i.e., to investigate whether there is a relationship between traits and genomic diversification. Here, the problem relating to multiple comparisons arises because the assay has to be conducted several hundred thousand times. The normal statistical significance of $P < 0.05$ is insufficient and the P value must be at the level of 10^{-7}-10^{-8}. We have developed and proposed an algorithm for performing linkage analysis with the use of numerous markers while taking into account linkage disequilibrium (Figure 1). Furthermore, an analysis of population structuring is also important, because it may lead to false positives. Based on a principal component analysis, we reported that Japanese are classified into two distinct clusters (mainland and Ryukyu clusters) and that people in the mainland cluster also have significant genetic differences depending on the areas where they live (Figure 2). The third most important task is the estimation of various parameters and the interpretation of the results. Finally, we develop algorithms that use data from various analyses to predict disease susceptibility and drug responsiveness and then evaluate these algorithms.

Each of the above steps is important, but many of them require a great deal of time for calculation. Moreover, longer calculation times are required as the number of samples or control markers increases. Normally, calculations are performed on the assumption that the influence of each gene or each environmental factor is independent, but the calculation time becomes even longer when interaction is factored in. Recently, the amount of genomic data that can be obtained from one person is increasing dramatically with the introduction of ultrafast sequencers. In short, although we have the data, we are unable to accomplish the task because the calculations take too long. It is certain that the genetic causes of diseases, which are presently unidentified, will be identified when petaflop computers become available.

Despite the huge amount of data obtained as described above, prediction accuracy is usually not as high as one might expect due to the unstable probability. Probability in the laws of genetic inheritance is quite stable, however, ensuring the accuracy of prognostic prediction based on genomic data.

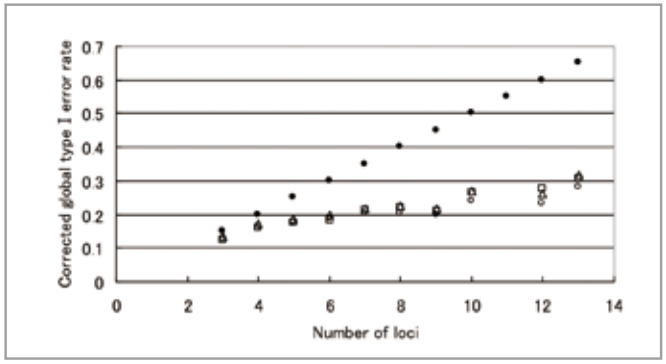


Figure 1: Assay to investigate the relationship between multiple SNPs and qualitative traits factoring in linkage disequilibrium. Assays require a significant amount of time.

Linkage analysis was conducted using genotypes of multiple SNP loci with linkage disequilibrium. The analysis was then followed either by the conventional Bonferroni's correction of multiple comparisons or one of our newly proposed methods. According to Bonferroni's correction, the rate of type I errors increases linearly with an increase in the number of loci, whereas with our methods, which use allele frequencies, the dominant mode and the recessive mode, respectively, are all capable of reducing the rate of type I errors. This resulted in a non-conservative assay.

○: Allele frequency; △: Dominant mode; □: Recessive mode; ●: Bonferroni's correction

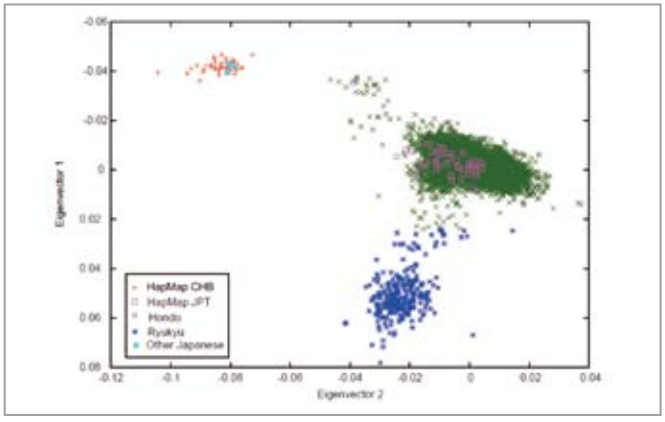


Figure 2: Clustering of approximately 7,000 Japanese, as well as Chinese, based on principal component analysis using around 140,000 markers

For the approximately 7,000 Japanese and the Chinese listed in the HapMap database, 140,000 pieces of SNP genotype information were obtained for each person. The data were analyzed using principal component analysis and the individuals were two-dimensionally plotted based on the first component (Eigenvector 1) and the second component (Eigenvector 2). As a result, the samples were classified into three distinctive clusters, namely, Chinese, Mainland Japanese and Okinawan Japanese.

HapMap CHB: Han Chinese in HapMap; HapMap JPT: Japanese in HapMap; Hondo: Mainland Japanese cluster; Ryukyu: Ryukyu Islands Japanese cluster; Other Japanese: Japanese other than the above

Key Technology Supporting Petascale Computing

High-performance Computing Team
Kenji ONO (left)
Satoshi ITO (middle)
Daisuke WATANABE (right)



The High-performance Computing team has been researching and developing the elemental technologies and software development frameworks for the development of high-performance applications. These technologies have contributed to the "Research and Development of Next-Generation Integrated Life-Science Simulation Software," a project to significantly improve the performance of developed applications and the efficiency of software development. We are now developing a visualization system that efficiently and effectively visualizes the calculation results of large-scale simulations and helps present information on these results.

All current supercomputers are parallel computers consisting of multiple CPUs/cores. In order to achieve high performance on such computers, it is inevitable to parallelize (the distribution of input data and message passing) and optimize applications. Parallelization purely depends on programming and puts a strain on application developers. Without optimization, an application can only utilize a small percentage of the parallel computer' s performance and as a result, optimization is required to derive the full potential of parallel computers. Optimization procedures, however, differ depending on an architecture, which significantly decreases portability. To alleviate such problems in the development of a large-scale simulation system, we are developing application middleware known as SPHERE (Figure 1). SPHERE supports both the development and operation of applications. It helps with efficient development, providing various functions including data input and output, algebraic operations such as inner product calculations and boundary condition operations, etc. Every function is parallelized so that the developed application can run on a parallel computer without modification. Since each function is optimized for various computers, users can expect high performance. Aiming for higher performance, we are planning to implement auto-tuning technology to enable the system to automatically select the most appropriate method and parameters during execution. For operation, SPHERE uses XML files to describe the parameters and manage execution of each application. As a result, the same format is applied to multiple applications (for specifying analysis conditions and input/output file names, etc.) so that the operational workload decreases significantly. In addition, SPHERE provides various utilities (XML file creation support, definition of boundary conditions, domain decomposition) to improve the operational efficiency of an overall analysis. These functions are designed to improve development efficiency for industrial-use applications and aim to handle several thousand parallel processes.

Next-generation supercomputers can perform extremely large numerical simulations and output the results as a large file group. To efficiently understand and analyze such data, we are researching and developing a functional data management and visualization system that can handle large-scale data expected to reach several hundred TBs. This visualization system aims to provide an interactive visualization environment for the handling of large-scale data (Figure 2). Users can interactively specify the desired observation areas and visualization parameters and repeat visual exploration of data in real-time. This helps users have a better understanding of phenomena. The visualization system is designed as a server-client system and performs parallel distributed processing. If each node can use a GPU to create images, the visualization is faster and of better quality. To reduce the cost of loading and transferring large-scale data, we are developing the following technologies in addition to simple area selection and downsampling: Out-of-Core technology that loads only the required part of the data as necessary and fast data compression technology that achieves high compression rate. Furthermore, we provide functions for versatile use, including local visualization with a standalone PC and batch processing that visualizes data according to a pre-defined scenario. This visualization system will extract new physicochemical phenomena and useful information from large-scale simulations and contribute to scientific discoveries.

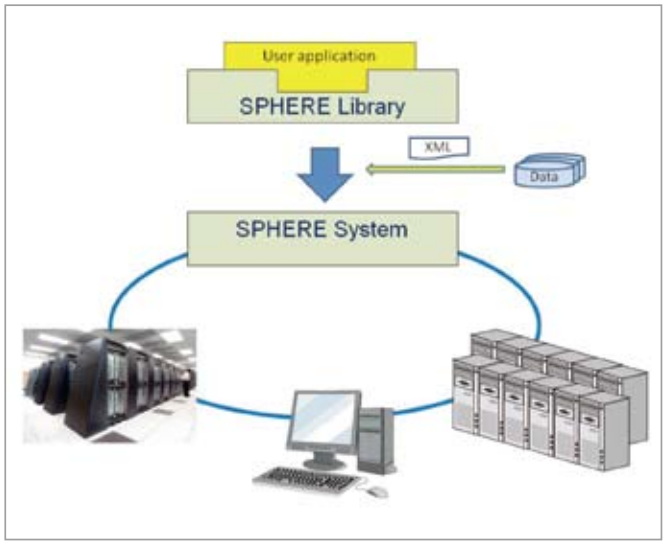


Figure 1: SPHERE Framework
Users develop applications using the functions provided by SPHERE. Linked to the system installed on the given computer, the developed application is optimized and executed.

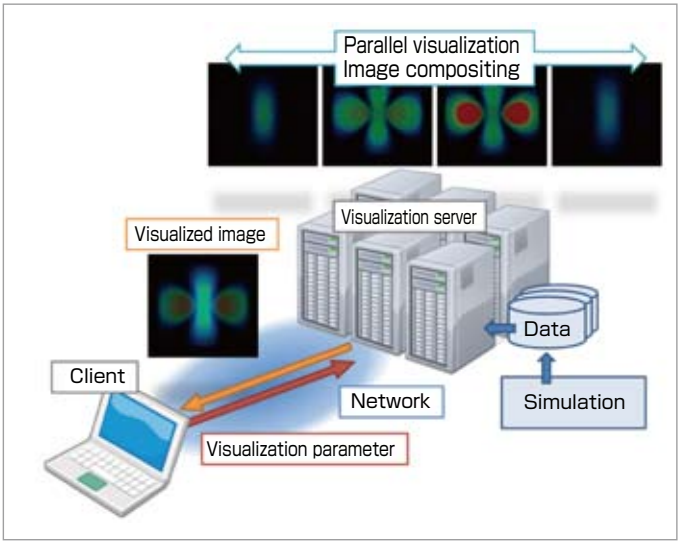
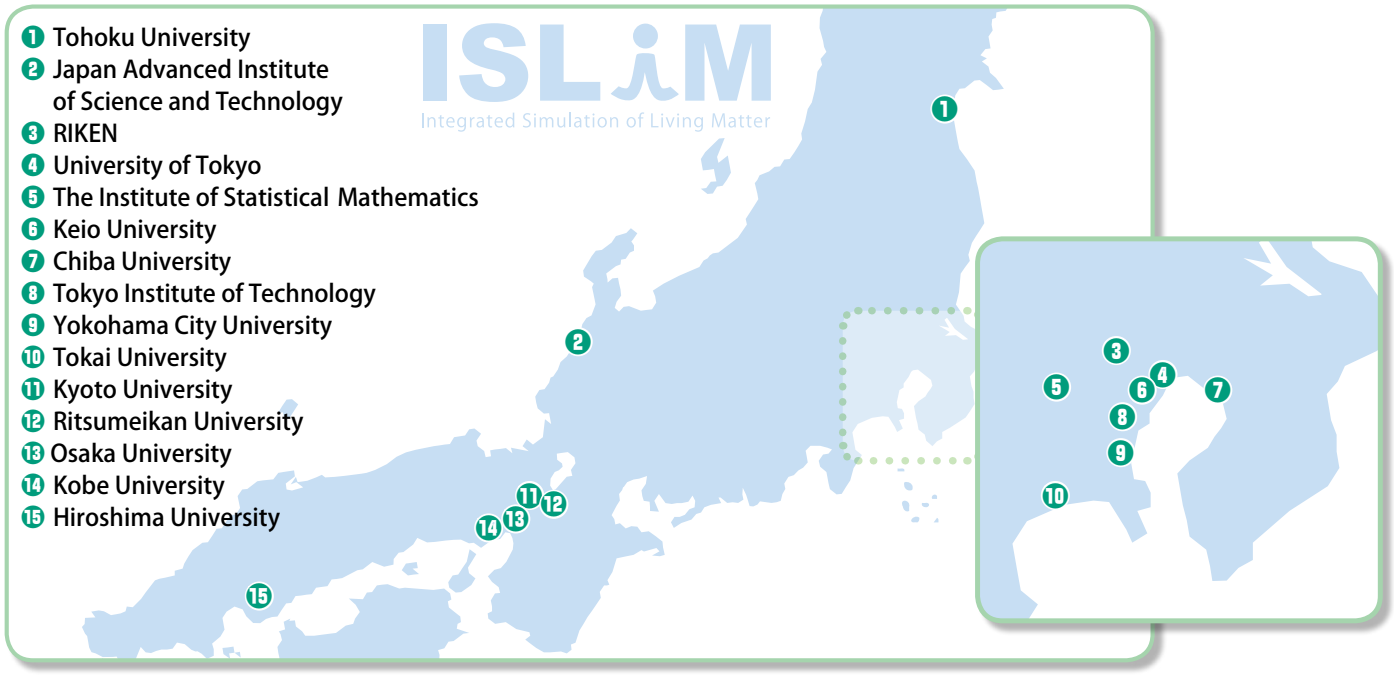


Figure 2: Remote Interactive Visualization Environment
Users interactively send the visualization parameters and receive the visualized images via a network. The visualization server visualizes large-scale data in parallel.

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)

Joint Workshop with VPH

In Europe, the EU's Virtual Physiological Human (VPH) Project is currently underway, led by Prof. Peter Kohl and Prof. Peter Covney. Prof. Kohl's project group visited Kyoto to attend the International Congress of Physiological Sciences held in July and on July 31, RIKEN's Wako Institute hosted a workshop to exchange information and discuss possibilities for future collaboration between his team and our Integrated Simulation of Living Matter Group. Prof. Kohl's group, consisting of researchers mostly from Oxford University in VPH, aims to contribute to medical science by modeling and simulating the entire heart.

The members who visited RIKEN for the workshop were Professors Kohl, Smith, Masindova and Quinn, as well as Professor Iribe of Okayama University, who until recently was a researcher at Oxford. The visiting group was shown RIKEN's testing equipment, which included a 3D internal structure microscope* and a tensile tester, as well as our new supercomputer RICC and 4D visualization system. The visitors showed particular interest in the 3D ISM and the supercomputer. The tour was followed by a meeting at which Dr. Kaya, Program Director, gave an overview of RIKEN and introduced projects run by both parties, while facilitating further understanding of each project through active discussion. As the discussion became too diffuse, covering a wide range of topics, we agreed that a subsequent meeting would be held in the U.K. (or a TV conference session) sometime around October and for interim communication to continue via e-mail to facilitate further review of each other's proposals. During the workshop, we explored areas for possible collaboration projects and long-term co-research themes. The participants from our group were Dr. Kaya, Program Director, Dr. Takagi, Leader of the Organ/Body Scale Team and its members, Prof. Matsuzawa of JAIST, Dr. Yokota, Leader of the Cell Scale Team and its members, staff from RIKEN's Advanced Center for Computing and Communication and the reporter, Ryutaro Himeno.

*The 3D-ISM takes continuous, cross-sectional pictures of a frozen biological sample in progressive slices, eventually creating a three-dimensional image of the entire sample piece. By combining a microscope and a laser, it is capable of taking images at ten micron resolution.

Timetable

Time	Subject	Speaker
11:00-11:50	Laboratory & RICC/4D Theatre Tour	
11:50-12:10	Brief Introduction to the VPH Initiative	P. Kohl
12:10-12:30	eu-Heart - Integrated Cardiac Care Using Patient-specific Cardiovascular Modeling	N. Smith
12:30-13:20	Lunch and Welcome Address	K. KAYA
13:20-13:40	preDiCT - Computational Prediction of Drug Cardiac Toxicity	A. Masindova
13:40-14:00	Integrated Cardiac Investigation into Mechano-Electrical Interactions: Macro to Micro	A. Quinn
14:00-14:20	Integrated Cardiac Investigation into Mechano-Electrical Interactions: Micro to Nano	G. IRIBE
14:20-14:50	Q&A, Discussion	
14:50-15:05	Break	
15:05-15:25	Overview of Japan's Life Science Grand Challenge	R. HIMENO
15:25-15:55	Research Activities of Organs and Whole Body Scale Research Team	S. TAKAGI
15:55-16:25	Research Activities of Cell Scale Research Team	H. YOKOTA
16:25-16:55	Q&A, Discussion	
16:55-17:10	Break	
17:10-18:10	Discussion on Collaboration	

Dr. IRIBE and Prof. Kohl

Workshop

Members of Oxford team

Post-workshop party

Members of Japanese team

About Our Logo

Our new logo for Research and Development for Next-Generation Integrated Simulation of Living Matter was designed to express the goal of our project, 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. The logo will be used widely in our future publications and Web pages.

Event Information

■ Next-Generation Supercomputing Symposium 2009

Date : October 7 (Wednesday)-8 (Thursday), 2009
Location : Conference room at My Plaza Hall (Marunouchi, Chiyoda-ku, Tokyo)
Fee : No charge *Fee required for reception (optional)

＊For details, go to http://www.nsc.riken.jp/sympo2009/09/symposium2009.html

■ Biosupercomputing Research Community: First General Meeting

Date : October 8 (Thursday), 2009 12:00-13:00 (pending)
Location : My Plaza Hall (Marunouchi, Chiyoda-ku, Tokyo)

＊For details, go to http://bscrc.riken.jp/event.html

■ The 2nd Biosupercomputing Symposium

Date : March 18 (Thursday)-19 (Friday), 2010
Location : In Tokyo (TBA)
Fee : No charge *Fee required for reception (optional)

＊Details TBA at http://www.csrp.riken.jp/

About the Cover Photo

It was the second event of the workshop, co-hosted by the Integrated Simulation of Living Matter program and the Advanced Computational Sciences Department at the Advanced Science Institute. As the promotion of the simulation of living matter requires collaborative efforts that involve wide-ranging research areas and methods, the workshop aims to offer opportunities for researchers from a variety of specialties to come together, present their findings and exchange ideas.

- Number of participants: 143
- Number of posters: 91
- Number of lecturers: 13

BioSupercomputing Newsletter Vol.1 2009.10

Issued : October 1, 2009

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/