Next-Generation Integrated Simulation of Living Matter
Strategic Programs for Innovative Research Field 1 Supercomputational Life Science

BioSupercomputing Newsletter

2013.3 Vol.8

4th Biosupercomputing Symposium
December 3 -5 , 2012 Tokyo International Forum

CONTENTS

Looking back on Six and a Half Years of the "Next-Generation Integrated Simulation of Living Matter"

Grand Challenge opens the way to the future of life science through innovative approach

Program Director, Computational Science Research Program, RIKEN
Koji Kaya

Six-and-a-half-year support of the "K computer" from the viewpoints of hardware and software

When RIKEN took the central role in the development of today's "K computer" and worked with many public research institutions and universities in its development, and made environmental arrangements for its usage, we carried out many discussions at RIKEN Science Council which I chaired at that time. Although studies utilizing supercomputers had progressed across a wide range of areas including material science, high end computation did not play an important role in the life science field. In these circumstances, RIKEN ended up making the proposal that supercomputer-based studies would become very important in the life science field, and RIKEN should take a proactive stance. I think administration officers including the Ministry of Education, Culture, Sports, Science & Technology which took the initiative in the project probably held active discussions. However, I believe that our proposal contributed to appointment of RIKEN as the core agency.

Along with developing a supercomputer, the R&D of Grand Challenge Application software to allow maximum utilization of the new computer became one of the major goals of the project. Regarding "Next-Generation Integrated Nanoscience Simulation Software," a supercomputer had already been introduced in the early 2000s when I was Director of the Institute for Molecular Science, National Institutes of Natural Sciences. Then, theorists involved in nano-technology gathered to start joint study of computational science. Therefore, the Institute for Molecular Science served as a center to start the development of Grand Challenge simulation software for understanding and predicting the characteristics and phenomena of nanomaterials through electronic-, atomic- or molecular-level sophisticated, large-scale computation. However, regarding the life science field, the computational science approach had hardly been explored. Still, Dr. Ryutaro Himeno and Dr. Makoto Taiji et al insisted that research and development should be promoted in this field, and played a driving force for Grand Challenge "Next-Generation Integrated Simulation of Living Matter Software." As their insight was brilliant and the introduction of computational science in life science was a must, 13 institutes (ultimately 15 institutes) gathered to establish the R&D system. It officially started in September 2006 after its application was approved by the Ministry of Education, Culture, Sports, Science & Technology in June 2006. Since I did not major in life science, I honestly hesitated for a moment when I was asked to assume the position of Program Director. However as I was once involved in computational science in the field of nano-technology, I thought I could be of some help in organizational control and decided to assume the position. From 2008, I have been concurrently serving as Deputy Director of the RIKEN Next-Generation Supercomputer R&D Center. As a result, I have watched over the development of the "K computer" from the viewpoints of both hardware and software.

Tangible results produced by Grand Challenge

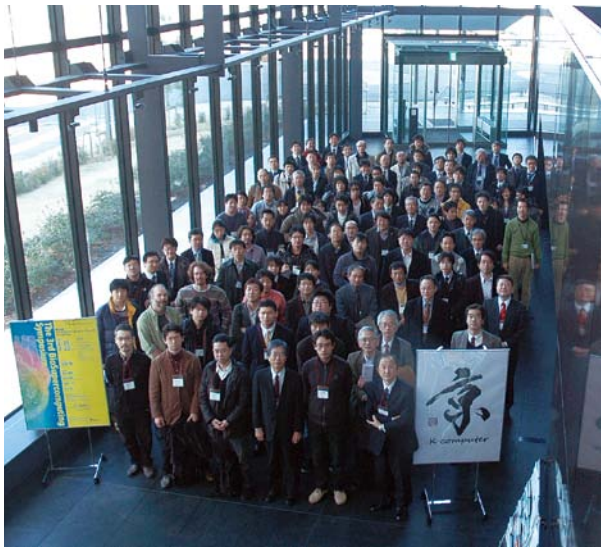
While no one in the material science field had doubts about the research and development of simulation software, it was initially very difficult to gain understanding from those involved in the life science field. After all, we had to address a considerably broad spectrum ranging from the molecular level to whole-body scale. What we had to do was a multi-scale simulation. In addition, we had hardly any established theory for each hierarchy. Regarding cells for example, RIKEN of course had been studying them by experiments. However, cells are not simple, anyway. The cell contains various substances including proteins, whose concentrations in the cell are incredibly high. It is almost unbelievable that proteins can be dissolved in water at such high concentrations. Everyone doubted whether they could replicate it correctly and calculate cell functions. Cell simulation was such a very difficult theme. However, as a result, research has progressed in those 6 years so that we can see some daylight. Of course it is also a fact that actual replication of the cell is far beyond the computational capability of the "K computer."

After I had just assumed the position of Program Director of Grand Challenge, our group including Dr. Himeno met many leading professors in the life science field. Many of them gave us rather dismissive answers. They neither showed active interest in nor extended support to our attempt. Only 6 to 7 years ago, it was a trend in the life science field. Now our research and development in life science has advanced by leaps and bounds. Simulation studies have advanced understanding at a cellular level, and also at the molecular level we can now replicate various protein functions. Brain research has advanced such that the "K computer" would in a real sense replicate functions like those of a real brain on a computer for the first time in the world. The Organ and Body Scale WG has enabled very excellent heart simulation. In blood flow simulation, a radically new fluid-structure interaction analysis method was developed for replication of the thrombogenic process. Sonic simulation studies which may lead to the development of ultrasonic therapy apparatus also show successful results. These are very important achievements also in terms of the

Building of Advanced Institute for Computational Science under construction (April, 2009)

Supercomputer room under construction was shown to the press (February, 2010)

Commemorative picture for completion of "K computer"



The 3rd BioSupercomputing Symposium (March, 2011)

contribution to medical care and medical engineering. In the field of data analysis fusion, some are initially doubtful about processing bioinformatics data by a petaflops-class supercomputer. When the environment necessary for easier genome analysis is put in place, the era of personalized medicine which requires the processing of enormous amounts of data will arrive. In this context, the supercomputer with high calculation performance was proven to be useful in various processing tasks.

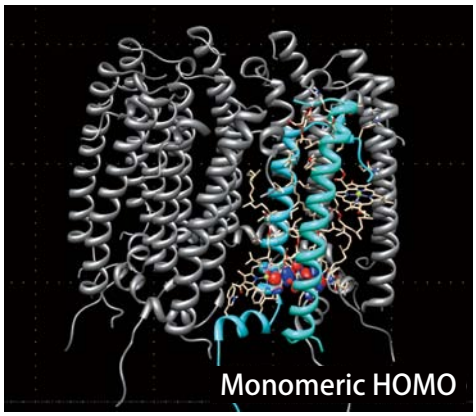
Initially, many people doubted whether a computational science-based approach can provide adequate results in the complex and broad-ranging life-science field. In spite of this background, there is no doubt that we got some positive response and discovered many possibilities during those 6 to 7 years. Until we reached this stage, the High-performance Computing Team led by Dr. Taiji took a key role in the development of software technologies and supported this project. I cannot forget their important role. Of course our accomplishments are still a far cry from clarification of complex life phenomena and mechanisms. However, it is a fact that tangible results were obtained. I think Grand Challenge made a powerful impact. I do feel that high end computation will make a strong positive impact on the biology and medical care in the 21st century.

Hoping for young people's contribution to development of future life science

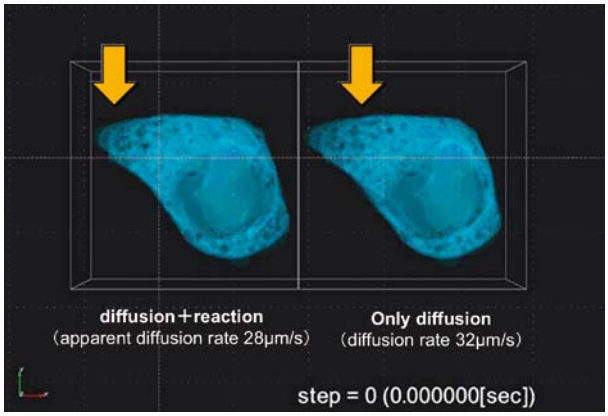
Another achievement of Grand Challenge is that as simulation science grew in the field of life science, many young researchers became interested and joined the field, and we can now collaborate with people from a wide range of research fields. I believe those young scientists will become a very important human resource which supports future research. However, the problem is that the new computational science, which those young researchers will be engaged in, has not yet won enough recognition in Japan. In the USA where computer science enjoys greater recognition, reasonable posts are offered to researchers engaged in computational science, but the Japanese life science field doesn't even go that far. Still today, vacant posts tend to be allocated to experimental scientists. Simulation studies will definitely have a key role in future life science. However, I cannot help feeling that professors who have been leading life-science researchers do not seem so sure about that. In addition, we have to make the best of today's tight research budget. I think this is one of reasons why it is not easy to increase the number of posts for the computational bio-science.

The Grand Challenge Program ends in March 2013. While producing excellent results, there still remain many courses of research and development which have to be challenged again by the collaborations among colleagues in the Grand Challenge Program. I believe how to inherit and expand it will be a major challenge from now on. The life science field had a very limited interaction and collaboration with other research fields. However, in Grand Challenge, researchers across various fields gathered to create a new life science. I hope more and more young researchers overcome the barrier, absorb the knowledge and information in various fields, and pave the way for a new world.

In any age, it is always hard to open up a new field, and we may face many difficulties in doing so. However, one thing that I can say without fear is that life science is facing an era of revolutionary change. More and more innovations must be produced from now. I believe Grand Challenge is a runway for that. It is my fervent hope that as many young researchers as possible take off from this runway, and continue the effort to blaze a new path to the future.



Protein DF - Protein All Electron Calculation



Ca+ kinetics simulation (HepG2)

Looking back on Six and a Half Years of the "Next-Generation Integrated Simulation of Living Matter"

A Landmark Project that Brought on an Innovation in the Field of Life Science

Deputy-Program Director, Computational Science Research Program, RIKEN
Ryutaro Himeno

Phenomenal Findings by the Project

The "Next-Generation Integrated Simulation of Living Matter," a research and development project which we have committed to since October of 2006, essentially aimed to develop a "show-case model" of simulation software in the life science field as part of the "Next-Generation Super Computer Project". This project was set up to prove that "new and astounding studies could be performed" with a supercomputer with a performance at the 10 PFLOPS level.

Looking back on the last six and a half years, there were many things we could and could not accomplish. But in the perspective of research and development of simulation software, I believe we were able to produce a large number of software streamlined for the "K computer", which were able to display a higher performance than the theoretical performance that we originally expected.

Not only did the "MD core program for large-scale parallel computers (cppmd)", the molecular dynamics software streamlined for the "K computer", result in a performance value close to 40% as expected, the "Fluid-structure interaction analysis program for whole body voxel simulation (ZZ-EFSI)" also attained a remarkable performance value of 43%. Further, the "multi-scale multi-physics heart simulation (UT-Heart)" also attained a performance value close to 30%, and "non-invasive treatment simulation (HIFU)" also attained a performance value higher than 20%. Originally, our goal was to develop simulation software that exceeded a performance of 1 PFLOPS using the entire "K computer", and we were hoping to possibly develop two software. But these four software will definitely result in performance values higher than 20% if they are actually run on the entire "K computer". Therefore, it might sound a little overstated, but I feel we can say we surpassed our initial goal. Regardless of how we say it, simulation software that yields high performance values has been created. There are thirteen software including these four that were able to produce solid results on ten thousand nodes of the "K computer". The fact that one third of approximately thirty software programs in total attained honorable records is a success that surpasses our original expectations.

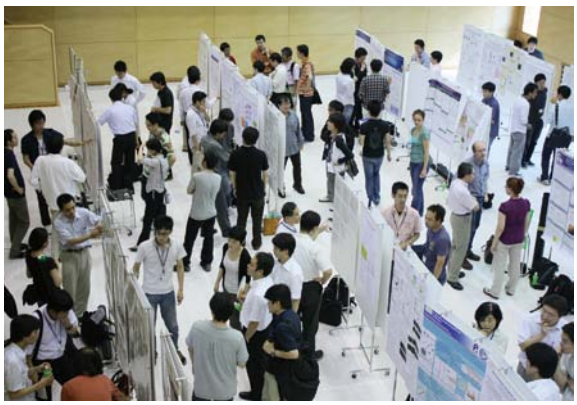
You may think, "Isn't that a little over dramatic for 30%?", but in the year 2006 when the project started, there was much speculation and criticism regarding development of simulation software in the life science field for the supercomputer. In those terms, to be able to develop thirteen usable software is a wonderfully remarkable achievement.

Looking back, it was good that we were able to produce respectable findings every year. People have gradually started to accredit our accomplishments through these yearly achievements. Most of all, the actual researchers who are developing the software understand that their efforts for improving parallel performance have been fed back to them as achievements, and are at a stage where they say, "We can't operate without the "K computer", actually even the "K computer" isn't enough". It is now hard to believe that there were worries like, "Can we fully utilize a 10 PFLOPS machine?" when we first started.

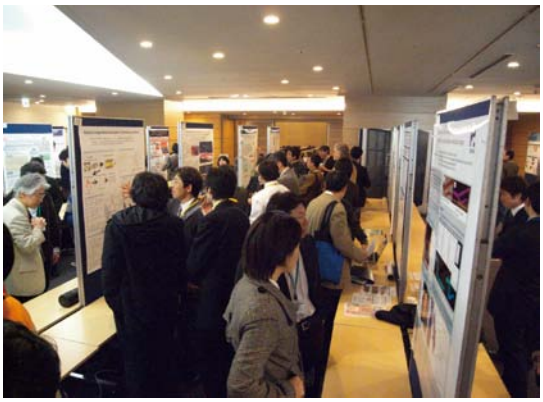
Especially in the year 2012, we only had a period of two months starting in July to be granted priority to use 10 PFLOPS together with the five fields of the Strategic Programs for Innovative Research. We had a difficult time on how to handle everyone's concerns regarding the lack of calculation time. Reflecting on our initial worries, we can say it is a gratifying miscalculation, but it is very disappointing that some were not able to produce sufficient scientific findings due to lack of time to use on a priority basis. For example, the simulation software "Neural Simulation Tool (NEST)", which replicates input and output of the entire brain in order to reproduce the growth and learning capabilities of the brain, has the largest calculation in the field of brain science in the world when it uses the entire "K computer", but we could not get enough calculation time. In that context, I believe a lot more could be accomplished if we had another year.

Likewise, there are many topics that are left uncompleted, but I feel that the project was able to secure respectable achievements overall. Without this project, I believe there still would have been many researchers that would say "the "K computer" is not necessary for the field of life sciences".

Ryutaro Himeno, Deputy-Program Director, who gave his lecture at "Next-Generation Integrated Simulation of Living Matter Symposium 2007"



Poster Session held at the Joint Computational Science Workshop 2009 (July, 2009)



Poster Session held at the 2nd Biosupercomputing Symposium (March, 2010)

Integration Born from Summer School

The "K computer" has a relatively fast network and is constructed with highly distinguished hardware. The work of the High-performance Computing Team whose researchers found bugs in compiler before others use it and made various efforts to bring out full performance values cannot be ignored. Their work is wonderful. Through their steady efforts, the "K computer" has been built as a machine that can be used with ease, and I believe the existence of this team has led to the development of a large number of software that attain high efficiency. For improving computation speed, it is critical to share knowhow among all software development researchers. The role they played in sharing a precise pool of information for generating higher performance was exceptional.

The word "sharing" brings to mind the early stage of this project. Personnel with research backgrounds of various scales from many research institutions and universities joined the project, but since they spoke their own technical language and abbreviations, it was difficult to understand each other when we initially held symposiums (laughing). Although explanations became more explicit and understandings deepened, it did not lead to integration. In the midst of this, the summer school in which we collected younger researchers was effective. While each researcher worked on his/her new model, people from varying research fields used the computers in a similar manner but with a different approach, and this observation directly stimulated everybody. Through summer school, researchers became to understand what researchers from other fields were pursuing, which led to the overcoming of research field and technical language barriers, and eventually progressed to integration. You can say that the simulation of blood clotting through the fluid-structure interaction analysis method was also born out of this new kind of integration.

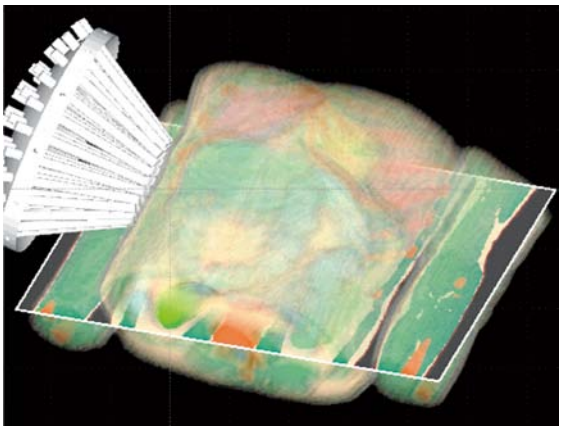
I am very pleased that this project stimulated growth in each field due to the new integration. This itself cannot be explained in a research paper, and may not be perceived as an achievement. Even so, I hope there will be a day when people will say that in the long run this project became the turning point for generating new connections, and led to wide-ranging progress in research and development.

To Expand on the Accomplishments of the Project

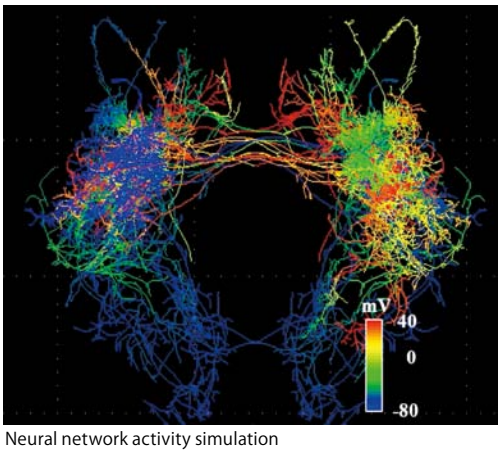
To simulate the phenomenon of life which is a multi-body and multi-layered system, it is necessary to explore new phases through new integrations. In this project, new integrations have been formed in each research and development team, and interesting findings are coming to light. The main focus of the project was software development, but at this point interests are shifting towards the research findings generated from the new software which is an exciting result. You can also say we were science-driven and software had been developed as seedlings for the researchers' interests, but some could possibly be put to practical use for actual drug discovery and medical care.

The project will end in the end of fiscal 2012, and each research and development team will be dissolved, but those that have the potential for future needs will be incorporated into the Strategic Programs for Innovative Research (field 1). Other software that should be inherited and extended, rather than ending up as they are, will be undertaken by RIKEN's Advanced Center for Computing and Communication for further development, updates, and support. We hope they will be actually used by the research centers at RIKEN. We will need to maintain a running environment, or else efforts will end with the development stage. By continuing the efforts in RIKEN's research topics, I hope many will realize and appreciate the value of operating this project at RIKEN.

I mentioned earlier, the young researchers whose research attitudes have changed through the summer school. We have founded a bio-supercomputing research group in order to support them, and hope to continue to help this research group evolve. It will be a wonderful place to continue to cultivate these budding young researchers that possess new insights.



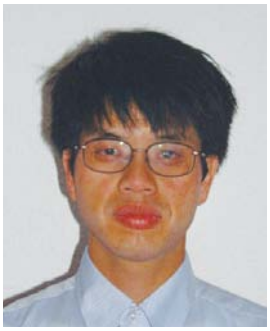
Ultrasound treatment simulator: ZZ-HIFU



Neural network activity simulation

Multi-scale, multi-physics heart simulator UT-Heart

Graduate School of Frontier Sciences, the University of Tokyo
(From the above) Toshiaki Hisada, Seiryo Sugiura, Takumi Washio, Jun-ichi Okada, Akihito Takahashi
(Organ and Body Scale WG)



The human heart comprises multi-physics problems that widely cover various electric (ion current, excitation propagation, electrocardiogram, etc.), chemical (mass transport, reaction, energy conversion, etc.), and mechanical (myocardial tension, blood pressure, blood flow, etc.) phenomena originated from the biochemical reaction by ATP (adenosine triphosphate) hydrolysis as an energy source. In terms of spatial scale, it comprises multi-scale problems that range from protein molecules (up to 10 nm) to organs (up to cm) via cells (up to 100μm) and tissues (up to mm), and finally ejection of blood is achieved.

ATP is consumed as energy for the relative sliding movement between the numerous actin filaments and myosin filaments that comprise 50 to 60 myofibrils present in a single cardiac muscle cell. It is known that the relative sliding of the filaments results from the tractive force that is produced by the myosin molecule heads extending from the myosin filaments combining with the actin filaments. Hypotheses for the specific mechanism involved are however diverse, and still remain under discussion. Figure 1 (right) shows a hypothesis of the rotation of myosin heads. In the meantime, cardiac hypertrophy is generally interpreted as a phenomenon where the wall thickness (myocardial cross section) increases to cope with a situation of high blood pressure such as valvular diseases and hypertension. In a disease called hypertrophic cardiomyopathy, however, the heart wall is thickened for no reason at normal blood pressure. In an initial study, the thickening had been interpreted as occurring to compensate for degrading cardiac function due to point mutation of myosin molecules (represented by mutation of the 403rd amino acid at which the myosin molecule binds with the actin molecule). In a later study, however, it was reported that the mutated myosin is rather hyperactive[1]. Then, what happens to the heart by admixture with the hyperactive myosin?

We have developed cardiac multi-scale analysis technologies on massively parallel computers so far[2], and are earnestly addressing this issue using the "K computer". More specifically, we had inevitably tried to solve the probabilistic binding state of actin molecules and myosin molecules by conceiving an average molecule, and applying it to an equation expressing the state transition in the past, due to limitations of computer power[3]. However, we came to be able to seamlessly solve the tissue and organ movements through cell contraction by starting with a concrete simulation of the probabilistic movements of the individual myosin heads with intramolecular elastic elements modeled on a spring as shown in Fig. 1, by taking advantage of the characteristics of massively parallel computers. The myosin head movement has a characteristic, known as "Cooperativity," whereby binding of a certain myosin head with an actin filament encourages its neighboring myosin heads to bind with the actin. This computation model enables the introduction of such a characteristic as it is, and thus the analysis of heartbeat in the case of admixture with the previously-mentioned hyperactive myosin. The following specific procedures will be undertaken:

(1) Simulate the probabilistic micro-scale behavior of myosin molecules with cooperativity directly by the Monte Carlo method;

(2) Embed the above-mentioned Monte Carlo model in the muscle fiber portion of the mesoscale cell structural element model including the intercellular space and so on to properly address the dynamic interactions between scales physically; and

(3) Bind the movements of the above secondary structural element model and the macro-scale organ model by the homogenization method.

The realization of the above analytical method enables the simulation-based analysis of such issues as what effects are produced by the molecular level state change law and cellular tissue level structural elements on heartbeat performance and energy efficiency and, conversely, what effects are produced by the feedback of macromuscular contraction and relaxation on the molecular level state change. Such a simulation platform may also serve as a means for molecular biologists to verify their own hypotheses from a new perspective in future.

The research and development and validation of the UT-Heart are jointly implemented by the Fujitsu Limited Next Generation Technical Computing Unit (a group led by Yoshimasa Kadooka, General Manager and Head, Application Research and Development Division), and the University of Tokyo Hospital (a group led by Visiting Professor Ryozo Nagai). The clinical study was approved by the Institutional Review Board at the University of Tokyo.

[References]

[1] Belus A, Piroddi N, Scellini B, Tesi C, Amati GD, Girolami F, Yacoub M, Cecchi F, Olivotto I, Poggesi C, The familial hypertrophic cardiomyopathy-associated myosin mutation R403Q accelerates tension generation and relaxation of human cardiac myofibrils, J Physiol, 586(Pt 15), 3639-3644, 2008

[2] Hosoi A, Washio T, Okada J, Kadooka Y, Nakajima K, Hisada T, A Multi-Scale Heart Simulation on Massively Parallel Computers, SC10, NewOrleans, proceedings 1-11, 2010

[3] Washio T, Okada J, Sugiura S, Hisada T, Approximation for Cooperative Interactions of a Spatially-detailed Cardiac Sarcomere Model, Cellular and Molecular Bioengineering, 5, 113-126, 2012

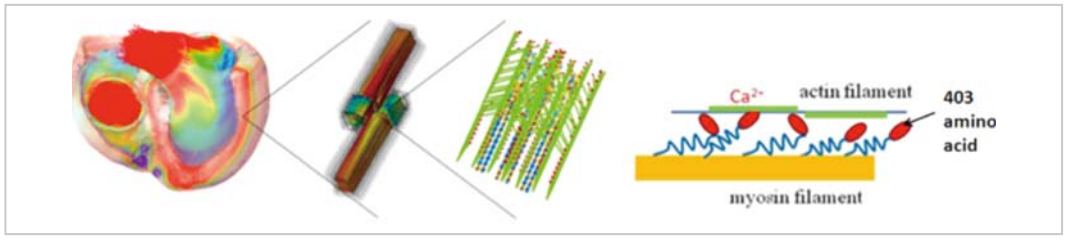


Fig. 1: K computer-based Multi-scale Heartbeat Simulation from Sarcomere Kinetics

Simulation Model for Insulin Granule Kinetics in Pancreatic Beta Cells

Graduate School of System Informatics, Kobe University
Hisashi Tamaki
(Cell Scale WG)

Diabetic patients have been increasing dramatically all over the world today, especially among Asians. Most Asian diabetic patients suffer from Type 2 diabetes caused by inadequate insulin secretion. Healthy people produce insulin granules in the Golgi body of pancreatic cells. Then the granules are transported to the cell membrane for secretion. Type 2 Diabetic patients have insulin granules in the cell though their secretion is inhibited by some factor. For revealing the factor and dealing with it properly, it is necessary to elucidate the insulin secretion mechanism in pancreatic cells.

Since many studies have been carried out on the insulin secretion mechanism of the cell membrane, considerable progress has been made. However, it is also a fact that the intracellular mechanism of granule kinetics has still hardly been clarified. Insulin granule kinetics after glucose stimulation has recently become observable. Based on the result of this partial observation and biological evidence, we have prepared a simulation model for insulin granule kinetics in pancreatic cells. With this simulation model, we are trying an approach to acquire new insight into insulin granule kinetics by reproducing typical phenomena found in insulin granule secretion.

In insulin granule kinetics or the transportation process of insulin granules produced in pancreatic cells to the cell membrane for secretion, the granules are believed to move along the cell skeleton. For constructing the simulation model, we divided the model into four stages based on the skeleton of a pancreatic cell and the principle of granule kinetics, i.e., granule producing site, interior layer, exterior layer and granule secretory part (Fig. 1).

Regarding the phenomena we aim to reproduce, we considered that the insulin granule secretion process consists of three major phases, (1) Phase I: Steep increase in secretion occurring in connection with glucose stimulation, (2) Phase II: Moderate increase in secretion after glucose stimulation, and (3) Phase III: Small amount of continuous secretion called basal secretion. Then, we focused on the typical secretion patterns found in healthy subjects and diabetic patients, and after drug administration, respectively. (a) Healthy subjects: Phase I, II and III are observed successively. (b) Diabetic patients: Phase I does not appear. (c) At drug administration: Both Phase I and II appear strongly. We aim to reproduce the secretion processes with those typical patterns.

As an example simulation, Fig. 2 shows the results for two cases simulating a healthy subject and a diabetic patient, respectively. Although the difference in parameter setup between Case 1 and Case 2 is only the number of secreted granules at the granule secretion site under glucose stimulation, it was confirmed that biphasic insulin secretion and insulin secretion lacking Phase I, both of which are found in real observation, were able to be reproduced in Case 1 and Case 2, respectively.

Currently we are searching for a parameter setup for reproducing insulin granule secretion after drug administration. Additionally, we think it is also vital to conduct a more detailed study on the impact of the individual parameter setup on kinetics and secretion by improving simulation results which can reproduce typical secretion patterns. Further, we would like to perform studies for setting parameters based on available data of granule kinetics in pancreatic beta cells (which would be considerably limited and fragmentary), creating a satisfactory simulation scenario and refining the simulation through coordination with metabolic system simulation.

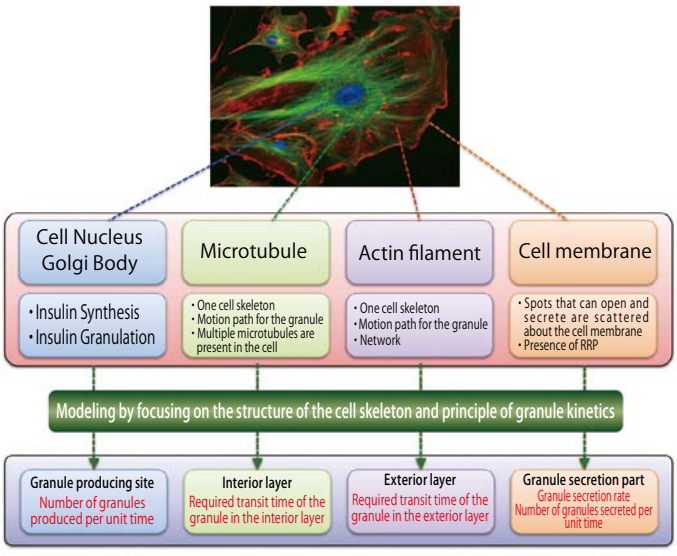


Fig. 1 : Outline of insulin granule kinetics model

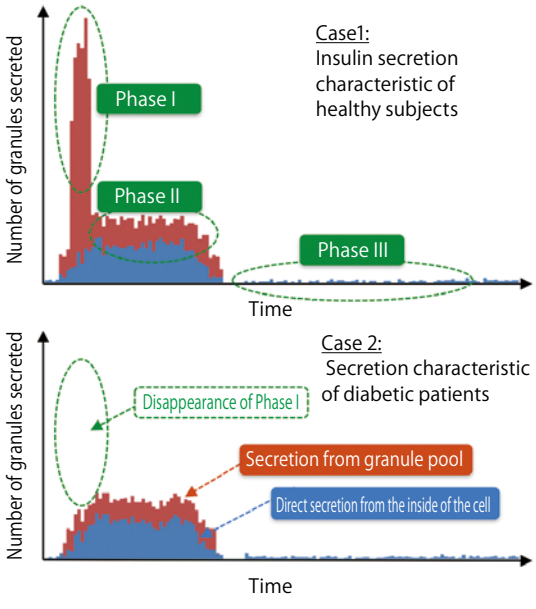


Fig. 2 : Simulation Results

The road to brain-scale simulations on K

Brain Science Institute, RIKEN
Institute of Neuroscience and Medicine (INM-6), Juelich Research Center
Medical Faculty, RWTH Aachen University
Markus Diesmann
(Brain and Neural Systems WG)

This is the story of the endeavor of the Brain and Neural Systems Team (BNT) to make the computational power of K available to the field of computational neuroscience. An extended version of this report can be found at [1].

The human brain comprises about 10^{11} neurons, each connected to 10000 others. In computational neuroscience, the bottom-up approach often starts from a mathematical description of neurons and their interactions in order to investigate network dynamics [2]. The NEST simulator [3] is tailored to this resolution. Neurons are represented as small systems of differential equations, which interact by δ-impulses [4,5] to form networks of natural size and complexity.

The top-down approach starts from an abstract description of a particular brain function and investigates how this function is implemented at the neuronal level (e.g. [6]). The functional circuits of the brain typically involve several areas. Due to the demand of computer memory [7], however, simulations were previously constrained to the local microcircuit (e.g. [8,9]) or non-spiking macroscopic models. With supercomputers like K, the multi-scale connectivity of the brain is within reach.

At the time of the 1st BNT meeting (Nov 2008), NEST (2nd generation kernel, 2g) scaled to 1024 processors for 10^5 neurons. At the 3rd meeting (Oct 2009) the JUGENE computer at Juelich reached 10^6 neurons. By Feb 2010 we had set milestones for K with network size increasing from 10^6 to 10^8 and above. NEST compiled Nov 2010 on a prototype, May 2011 on the test system, and in Sep on K as reported on meetings 6 and 7.

In the 3g kernel, the data structures account for the sparseness of neurons with local targets and enable the efficient storage of information about non-local neurons [7]. At the 8th BNT meeting in Mar 2012, the second milestone came into sight, and was reached in May [10] utilizing just 12288 nodes of K. The upcoming NEST 2.2 release is based on this kernel.

The contacts between neurons exhibit different types of dynamics and plasticity [11,12]. In the limit where a neuron has either none or only a single synapse on a given compute core, however, this heterogeneity collapses to a single type. The 4g kernel presented at the 9th meeting in Sep 2012 exploits this uniqueness. The figure shows the maximum network size (triangles) for a given number of nodes and the runtime (dots). The 4g kernel enables simulations of 10^9 neurons on K. Neither kernel compromises on generality; functionality and user interface of NEST remain the same [13,14].

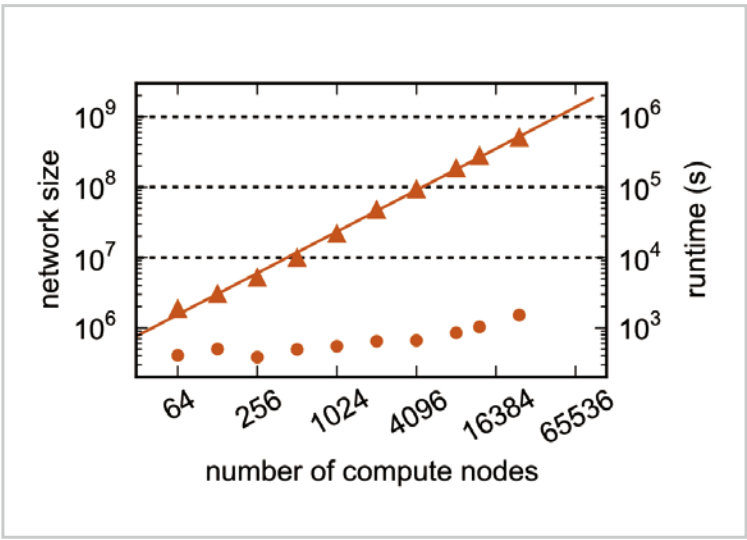
We hope that the simulation technology now available will help neuroscientists to integrate the data on the anatomy and dynamics of the brain into models to gain insights into brain function.

Our challenges ahead are to make the 4g technology available as a NEST release and to develop software concepts for the exa-scale machines now in preparation.

www.csn.fz-juelich.de
www.nest-initiative.org

Acknowledgements

Partly supported by early access to K at RIKEN AICS, Next-Generation Supercomputer Project of MEXT, Helmholtz Association: HASB and portfolio theme SMHB, JARA and EU Grant 269921 (BrainScaleS).



[References]
[1] Diesmann (2012) p83-85 In Proc 4th BioSupercomputing Symposium, Tokyo
[2] Lansner and Diesmann (2012) Chap 10 In Computational Systems Biology, Le Novére ed., Springer
[3] Gewaltig and Diesmann (2007) Scholarpedia 2(4):1430
[4] Morrison et al. (2008) Biol Cybern 98:459-78
[5] Hanuschkin et al. (2010) Front Neuroinform 4:113
[6] Potjans et al. (2011) PLoS Comput Biol 7(5): e1001133
[7] Kunkel et al. (2012) Front Neuroinform 5:35
[8] Potjans et al. (2012) Cerebral Cortex doi:10.1093/cercor/bhs358
[9] Wagatsuma et al. (2011) Front Comput Neurosci 5:31
[10] Helias et al. (2012) Front.Neuroinform 6:26
[11] Potjans et al. (2010) Front Comput Neurosci 4:141
[12] Pfeil et al. (2012) Front Neuromorph Eng 6:90
[13] Eppler et al. (2009) Front Neuroinform 2:12
[14] Gewaltig et al. (2012) Chap 18 In Computational Systems Biology, Le Novére ed., Springer

MD Core Program Development for Large-scale Parallelization

Computational Science Research Program, RIKEN
Yousuke Ohno
(High-performance Computing Team)

Our team has been developing a MD core program for large-scale parallelization in order to provide a high speed core program library for the "K computer", and to attain and collect application speedup technology knowhow on the "K computer".

Molecular dynamics (MD) simulation is a method for estimating the movement of molecules and changes in structure through calculation. In life sciences, biomolecules such as protein which are the foundation of life phenomena are targets of calculation. In the simulations, all movements of atoms are calculated by calculation of the all forces between atoms and numerically integration of the equation of motion. The amount of calculation for the numerical integration of the equations of motion and forces of covalent bond approximated by classical is proportional to the number of atoms. On the other hand, Van der Waals forces (intermolecular forces) and Coulomb forces incur a calculation load that is proportional to the number of pairs of atoms, which means the square of the number of atoms. Van der Waals forces are very small after a distance of 1.4nm, therefore pairs of atoms that are distant are ignored which is called a cutoff method. However, Coulomb force is inversely proportional to the square of the distance, and cannot be blindly ignored. Therefore the PME (Particle Mesh Ewald) method, which reduces the calculation load of long distance Coulomb force as low as $O(N \log N)$ by using FFT, is a popular way to calculate Coulomb force. Yet, FFT requires a high amount of communication when parallelized, which leads to lower performance in large-scale parallelization. Therefore, methods like the FMM (Fast Multiple Method) which demands less communication compared to FFT are starting to be employed. In addition, the PME method is very strict under periodic boundary conditions, and intracellular targets which do not have periodic cycles may have problems in using a pseudo-period. In that case, a cutoff method that corrects electrical charge and dipole moments under neutral conditions may be applied [1, 2]. The cutoff method requires communication between adjacent nodes, which is suitable for large-scale parallelization.

In our research, optimizing a cutoff method on the "K computer" became the center of our study because some form of cutoff calculation is the major part of calculation time regardless of the methods used. We attained a calculation efficiency of 60% for the cutoff calculation portion on a single CPU using a cutoff method with the best conditions. In the all node evaluations of the "K computer" (Note 1) in October 2012, we attained 4 PFlops effective calculation speed and 38% efficiency on 5.4 hundred million atoms with a cutoff distance of 2.8nm and once per four-step potential calculation. When potential is calculated in each step, effective calculation speed was 4.4 PFlops and 43% efficiency for a calculation of a 79872 node with 5.2 hundred million atoms.

Fig. 1 shows the parallel performance. The green dotted line represents parallelism when the calculation size (number of atoms) per node is constant (weak scaling). The number of atoms per node is 52338, 6542, 818, 102 in descending order. Calculation time is fixed in every case and a high level of parallelism is attained. The red line represents parallelism when the entire calculation size is fixed (strong scaling) and the number of atoms is 418707, 3349656, 7536726, 26797248, 214377984 from left to right. Parallelism depends on the number of atoms per node, and performance improved proportionally to the number of nodes when there were 6000 and more atoms. But when the numbers of nodes were multiples of 8 or 64 and the numbers of atoms were 800 or 100 per node, parallel performance fell to 50% or 20%. Ten milliseconds for calculating one step on a node of 100 or so atoms seems to be the margin for practical use. Regarding the test evaluation of long distance Coulomb calculation using FMM, parallel performance did not weaken as much compared to cutoff calculations. Further, we confirmed that performance levels did not weaken much on a scale that is equivalent to a scope using all nodes of the "K computer".

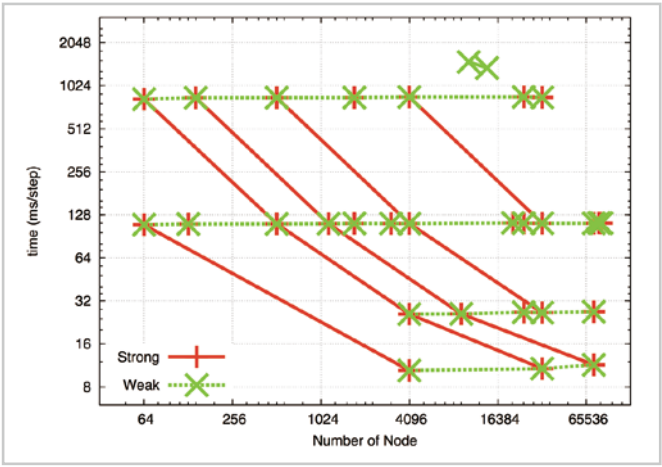


Fig. 1 : Parallel performance of the cutoff method. Cutoff distance of 2.8nm. Horizontal axis represents number of nodes, vertical axis represents calculation time per step (milliseconds). The green dotted line is a fixed number of atoms per node (weak scaling), the red line is when the number of all atoms is fixed (strong scaling).

Table 1: Calculation performance of the cutoff method on a scale worth all nodes of the "K computer". Cutoff distance of 2.8nm.

Node configuration	48x52x32	48x52x32	48x48x36*
Number of nodes	79872	79872	82944
Number of atoms	522546336	522546336	542644272
Calculation time per 1000 steps (sec)	116.357	112.414	112.085
Frequency of potential calculation	Every time step	Once per four steps	Once per four steps
Effective calculation speed (PFlops)	4.387	3.871	4.031
Efficiency of calculation	0.429	0.379	0.38

* Note: It differs from the actual network configuration of the "K computer" which is 48x54x32, but this does not greatly affect the results.

The optimization on the "K computer" attained in our research has been applied in the development of applications and optimizing by other teams. Further, this MD core program will be released either as an example and a reference code for the optimization on the "K computer", or as an optimized MD code that can be reused.

Our research is joint research project with the High-performance Computing Team whose members are Hiroshi Koyama*, Gen Masumoto, Aki Hasegewa, Gentaro Morimoto, Noriaki Okimoto, and Hidenori Hirano.

* Hiroshi Koyama is now a researcher at National Institute for Materials Science (NIMS).

Note 1
Use of RIKEN's super computer "K computer" as a research program using the HPCI system (program number:hp120068)

[References]

[1] Wolf, D. and Keblinski, P. and Phillpot, SR and Eggebrecht, J., J. Chem. Phys. 110 8254 (1999)
[2] Fukuda, I. and Yonezawa, Y. and Nakamura, H., J. Chem. Phys. 134, 164107 (2011)

Aiming to realize hierarchical integrated simulation in the circulatory organ system and the musculoskeletal / cerebral nervous systems

Professor, Department of Mechanical Engineering
and Department of Bioengineering The University of Tokyo
Shu Takagi
(Theme3 GL)

Elucidate life systems in top-down approach

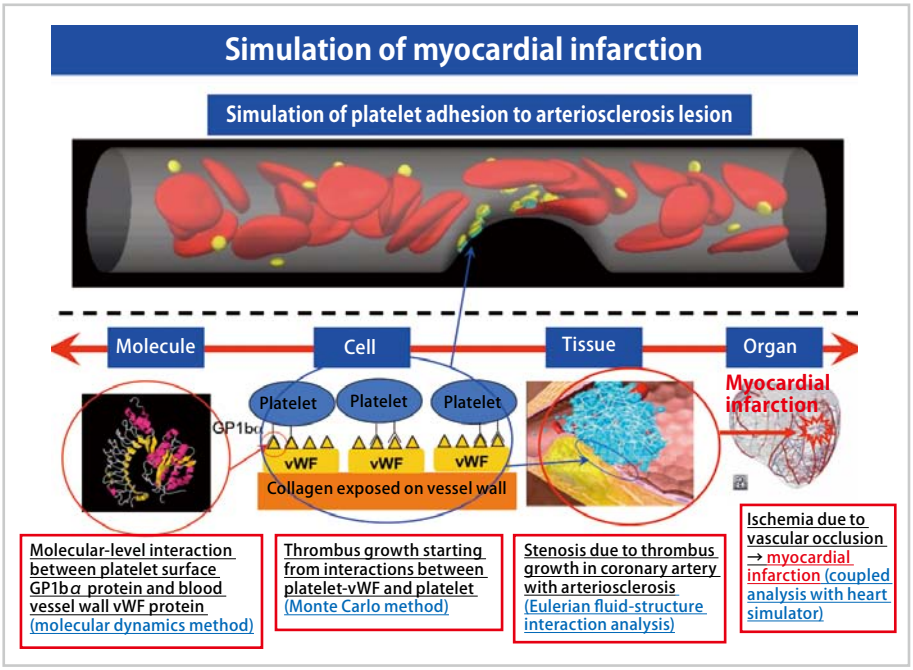
The ultimate goal of Theme 3 is to elucidate the meaning of "being alive." To this end, we will work on hierarchical integrated simulation, which forms a hierarchical human body model, with the aim to understand life as an integrated system.

Our lives are sustained by ingesting nutrition from the mouth, which is transported throughout the body via the blood and by breathing air to obtain oxygen in the lungs, which is transported throughout the human body again via the blood. Through these processes, the cells constituting the human body play their respective functions with the nutrition and oxygen delivered to them. After this process, a direction for further nutrient intake will be transmitted to our brain and our brain will feel hungry and encourage us to eat more. There are naturally various smaller processes, which are required for the sustainability of life, with a system that transfers heat, mass, and information in various forms. In plain words, Theme 3 aims to elucidate life systems in such an approach as explaining what is happening in the body with the above phenomena as a starting point.

For example, when a disease was cured by an effective medicine, it is correct to explain, on the molecular level, that "the disease was cured as a result of normal cellular functioning with a certain function of protein inhibited by the medicine." If the disease was not cured, however, there are many cases where the reason cannot be explained, only locally on the molecular level, why the disease was not cured. In fact, there are many medicines that are not effective although they should take effect on the molecular level. When symptoms of a disease disappear, your brain will feel that "the disease has been cured." In that case, "the cure of your disease" is felt in the brain finally via the nervous system after your cells normally function with the protein functions changed, and then your tissues and organs are normally activated as populations of such cells rather than the individual functioning of proteins or cells. In other words, outcomes are mostly not determined only by local events. Accordingly, if a medicine is not effective, you should think of other various factors involved. In order to understand "the meaning of being alive," various factors need to be grasped from the molecular level to the systemic level. It is difficult, however, to elaborately simulate all factors from the molecular level to the systemic level and there are inevitable limitations. What is important is the methodology as to how the hierarchies should be integrated.

There are two approaches to the understanding of life phenomena where various events are inter-related in a complicated manner both in terms of scale and time. One is a bottom-up approach that builds up small factors to lead to the understanding of larger phenomena, and the other is a top-down approach that starts with a condition, where the whole body is well-maintained, and goes down to smaller factors to explore how the small factors function and what is effective in maintaining the whole body. The Theme 3 Team aims to elucidate life phenomena with the latter approach.

Our body is equipped with a function called homeostasis that tries to restore to a normal condition. Thanks to homeostasis, a slightly disturbed condition will not aggravate immediately. The condition will drastically become pathological, however, when it exceeds a certain level. Although it may be difficult to grasp quantitatively what significant changes will occur at which stages due to what effective functions, we believe that a top-down approach is better to find what significant changes will occur at what levels of which factors. On the contrary, it is very difficult in a bottom-up approach to find what factor remains last. Hierarchical integration is coarse graining in the sense that it tries to focus on different hierarchies by cutting off something. It should be noted that simple averaging will erase the functions and features that may appear in the next hierarchies. What should be left depends on the functions and features that will appear in upper hierarchies in the end. Which factors should be linked to what parts and how they can



We will aim to reproduce the scenario leading to myocardial infarction by integrating the thrombosis simulator and the heart simulator. We will reproduce vascular occlusion in a contracting/dilating coronary artery through the coupled analysis of heartbeat by further expanding the multi-scale thrombus simulator, incorporating the previous simulation ranging from protein-molecular-level interactions to fluid-dynamics-level blood flow phenomena.

be linked together to leave cardiac functions for the heart and leave blood flow functions for blood flow properly based on the theories of the lower hierarchies—it is no exaggeration to say that most of the study subjects, which we have dealt with so far, focused on the handling of this point

Exploit the results of the Grand Challenge Program for further development

In the Grand Challenge Program, which has been carried out for the past five years, various living body simulators with excellent performance have been developed such as the thrombus simulator, the heart simulator, the musculoskeletal simulator, and the cerebral nervous system simulator. In Theme 3, we wish to improve the infrastructure tool for reproducing complicated processes relating to various diseases by integrating these simulators, which have been developed with the "K computer" on an individual basis so far, by positively exploiting the previous results with an eye to developing the efforts into construction of frames for integration into a hierarchical human whole-body simulation in future. In addition, we will work on applying disease-related simulation results to the prediction of pathological conditions, and support for treatment.

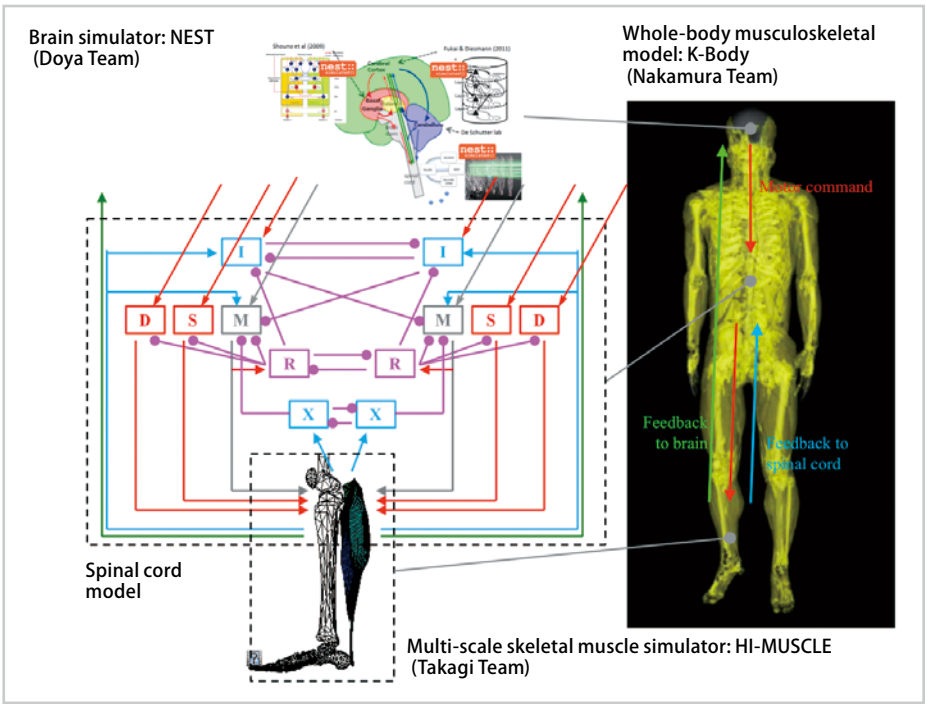
We selected myocardial infarction and Parkinson disease as two specific simulation targets. As mentioned above, the circulatory organ system is a mass transfer network and the nervous system is a signal transduction network, which are vital in the human body for transfer of heat, mass, and information for the sustainability of life, and function as what may be called human body trunk line networks. The circulatory organ system and the nervous system are also very important for future construction of a systemic integrated simulator. With such importance taken into consideration, we selected myocardial infarction and Parkinson disease as targets.

In the simulation of myocardial infarction, we aim to reproduce not only a primary thrombus (platelet thrombus), where platelet adhesion and aggregation occurs in the initial stage of thrombus formation, but also the scenario, in which the growth of a secondary thrombus (fibrin thrombus) initiated by blood coagulation function develops into vascular occlusion, by expanding the thrombus simulator developed in the Grand Challenge Program. In addition, we will perform coupled analysis of the heart simulator developed by Toshiaki Hisada (The University of Tokyo) et al., to simulate the thrombus growth process from thrombotic adhesion to arteriosclerotic lesions in the coronary artery surrounding the heart to myocardial infarction with the effects of the heart simulator incorporated into the simulation result. We also wish to provide important findings useful for the development of new drugs by evaluating the effects of thrombus-related drugs.

We also aim to reproduce the simulation of Parkinson disease as an integral part of the musculoskeletal simulator and the cerebral nervous system simulator. Parkinson disease is considered to cause limb tremor and physical rigidity even in a healthy musculoskeletal body due to abnormal signals transmitted from the brain. We want to contribute to the pathological prediction and treatment support for Parkinson disease, which is one of the functional motility disorders resulting from a cranial nerve disease, by reproducing a pathology called tremor that spike signals from the brain transmit to the muscle fibers via motor neurons and cause tremors, and a pathology called rigidity that stiff muscles cause a rigid posture.

Promote studies with an eye to the construction of human whole-body integrated simulator

In addition to the ongoing simulation of myocardial infarction and Parkinson disease, we also intend to work on the construction of infrastructure software for an integrated simulator that can be applied to a wide variety of diseases by integrating musculoskeletal and circulatory organ simulators via the nervous system. For the heart part, various diseases are expected to be simulated to study therapeutic methods and evaluate efficacy including abnormal cardiac functions due to myocardial infarction, angina pectoris, dilated cardiomyopathy, hypertrophic cardiomyopathy, and microvascular damage, and influence of stress and excitement. For Parkinson disease, we wish to develop the simulation into prediction of an optimum prescription such as concomitant use of electric stimulation with dosing based on data of individual patients on symptoms and motor tests, for example, which are different from the conventional symptomatic treatment with electric stimulation, by elucidating the mechanism and evaluating therapeutic methods through a whole-body integrated simulation including the cerebral nervous system, with consideration given to the effects of mass and signal transfer on the neuron level and the effects of signal transduction. In addition, we believe that the use of such a simulator will enable the analysis of a wide range of phenomena such as prediction of changes in body balance and damage in the event of overturning, and the difference in the effects of muscular or brain fatigue in the future.



We will aim to reproduce Parkinson`s disease, which is one of the motor dysfunction resulting from a cranial nerve disease, by integrating the brain nervous system simulator and the musculoskeletal simulator.

Leading-edge large-scale sequence data analysis with K computer in order to promote the understanding of life programs and their diversity

Professor, Human Genome Center, The Institute of Medical Science
The University of Tokyo
Satoru Miyano
(Theme4 GL)

Elucidation of life systems through large-scale data analysis

In Theme 4 "Large-scale analysis of life data", after establishing an infrastructure for leading-edge large-scale sequence data analysis optimized for High Performance Computing Infrastructure (HPCI) centering on the K computer, we are proceeding with research, in which we consider life programs found in cancers and cell differentiation processes as a system and understand their complexity and diversity based on the genomes. We also conduct biomolecule network analysis research centered on the data based on genomes. Through this research, we aim to contribute to practical applications such as better predictions of drug efficacy and adverse reactions, projections of causes of toxicities, personalized medicine, and prediction of survival.

"Life data" in the title of course includes the genome. It also covers a massive amount of data based on the genomes such as the epigenome (the entire set of DNA modifications) and transcriptome (all transcription products). They are key data for understanding life programs.

As an example to understand their importance, let's consider why we get cancer. Cancer is a kind of disorder in life systems induced by the complex combination of genetic factors inherited from parents, DNA modification (epigenome) by environmental factors and genetic mutation accumulated in tumor cells. Systems disorder may result in malfunction of a self-destruction system for the prevention of abnormality (apoptosis) or self-ordering for growth due to malfunction of external order for arresting growth. In addition, cancer cells join hands with normal cells such as vascular endothelial cells, immunocytes and inflammatory cells to acquire drug resistance, or spread anywhere through infiltration and metastases. Due to this, cancer is a complex and abnormal cell group that evolves spatiotemporally. It is the genome that regulates the malignancy grading of cancer, responsiveness to treatment and likeliness to have adverse reactions. In order to elucidate systems disorder, we have to examine the genome, epigenome, trascriptome as well as proteome (the entire set of proteins), metabolome (all metabolic products), and interactome (the entire set of interactions). However, we cannot understand a treatment approach or state of disease with these alone. For elucidating the diversity, complexity and dynamism of cancer, it is crucial to understand the system by large-scale data analysis and mathematical modeling taking advantage of mathematics and supercomputers.

Breakthrough for understanding life systems through mathematical analysis by use of the K computer

In 2003, the 13-year "Human Genome Project", which cost 100 billion yen, finally succeeded in decoding the human genome. This year, the National Institutes of Health (NIH) provided a roadmap after decoding the human genome, which says "Biology is changing fast into a Science of Information Management." In 2003, it surprised many Japanese researchers, who thought it was an exaggerated claim. Now, 10 years since then, everyone keenly feels that the claim is coming to be realized. From 2003, the International HapMap Project to elucidate individual differences on a DNA level was put into practice mainly in the USA, and a basis for efficiently finding genes involved in human diseases and drug responsiveness was established. The International Cancer Genome Consortium started to prepare a genomic aberration catalogue of major cancers from 2008, and has analyzed whole genome information on 50 thousand human genomes, consisting of cancer samples and normal cells of 25 thousand humans. Meanwhile, Barack Obama who assumed the presidency in 2008 introduced "The Genomics and Personalized Medicine Act" when he was a senator. In 2009, "the Genetic Information Nondiscrimination Act" was enacted. Consequently, the USA steadily implemented medical and healthcare strategies based on genomic information. NIH started to allocate its budgets to the $1,000 Genome (a project to drop the cost of full genome sequencing per human to roughly $1,000) from as early as 2004. Thanks to this, development and practical application of sequencing technology was promoted, and now Biology is becoming a science to analyze and interpret massive amounts of genome data.

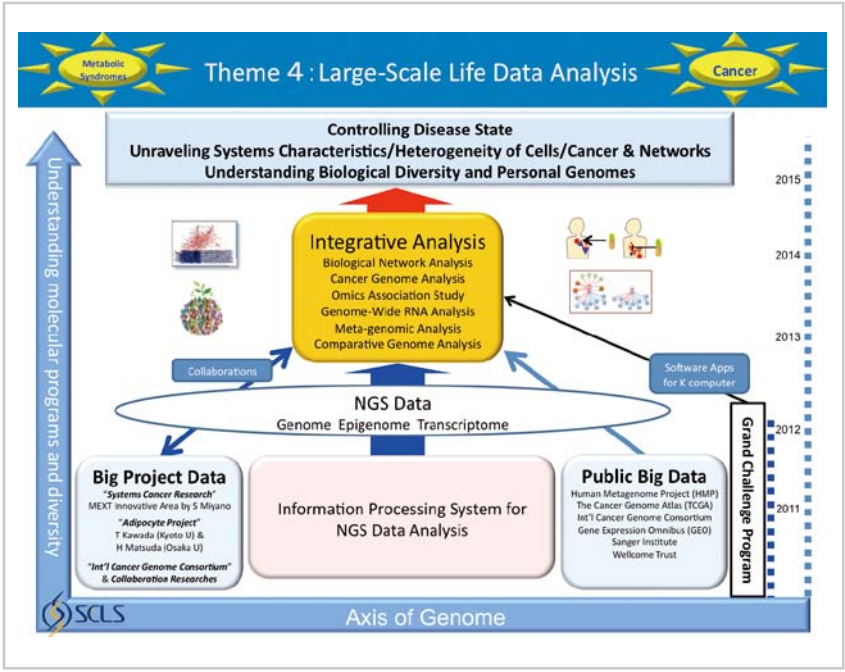
These days, sequencing technology is advancing rapidly. The next-generation sequencer using conventional optical sequencing technology is an expensive instrument, and also has some problems such as need of extravagant fluorescent reagents and limited lead length. However, today, cutting-edge sequencers such as a semiconductor sequencer using low-cost silicon chips have been developed, and an instrument which can make personal DNA sequence information available "at $1,000 or JPY 100 thousand within several hours" will be put to practical use in 2013. It will give steady service a couple of years later. Moreover, sequencers providing personal genome information "at $100 within an hour" are said to be "around the corner". With the spread of super-low price, high-speed, high-precision sequencers, personalized medicine, in which disease diagnosis and selection of therapeutic strategy and type and dosage of the drug are decided based on clinical sequence to check your own DNA, will rapidly become popular. I believe the day is not far off when everyone has his or her own DNA sequence. When the personal genome era begins in earnest, an enormous amount of genome data will be produced, and new breakthroughs will be applied to the understanding of life systems.

The first results have already been produced. It is the "discovery of causative genes for chronic myelodysplastic syndrome (MDS)." Many cases of MDS transform to acute myelocytic leukemia since bone marrow cannot produce normal blood. Although its cause had been unknown, the team of Seishi Ogawa (Cancer Board, University of Tokyo Hospital) and his colleagues fully employed the supercomputer system at Human Genome Center (The Institute of Medical Science, The University of Tokyo) and discovered from the thorough analysis of patient samples' exomes (remnant after an RNA splicing reaction) using next-generation sequencers that the cause was mutations in 4 genes in the RNA splicing pathway. This is said to be a historic result in cancer research, since it is not only the first discovery of the causative genes for this disease, but has also shown that abnormal RNA splicing plays a role in the development of cancer for the first time in the world. This can moreover be said to be the result of a successful combination of sequencing technology, supercomputer and a statistical mathematics analysis team. Also in the "Large-scale analysis of life data" of Strategic Programs for Innovative Research Field 1, in which we are now engaged, we would also like to achieve such breakthroughs.

Content of the research Theme 4 address

The major objective of Theme 4, "Large-scale analysis of life data," is the understanding of the life program and its diversity. The first thing required to achieve this goal is the high-speed processing of massive amounts of next-generation sequencer data. For this purpose, we made great efforts to establish the basis for high-speed analysis of massive amounts of next-generation sequencer data optimized for "the K computer" from 2011 to 2012. In this analysis system, we are trying to realize high-speed analysis of massive genome sequence information by achieving the performance of 10 million reads per hour and performing the world`s deepest high-speed sequence searching which can detect more challenging smaller homology. In this way, we are aiming to create a system to ensure a 10-to-1 advantage over the rest of the world.

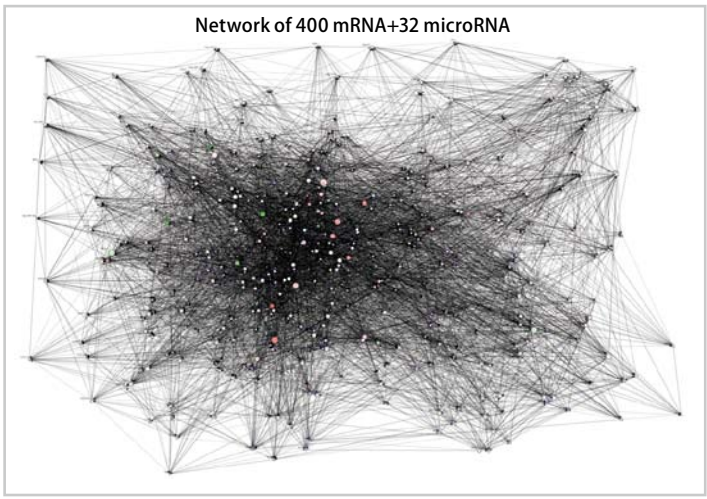
The basic data for the analysis includes not only public data but also the Cancer TCGA Project and the Meta-genome Project data, and we collaborate with the International Cancer Genome Consortium which researchers of Theme 4 participate in, the Innovative Areas "Systems Cancer Research" and "Adipocyte Project." Validation by experiment is also carried out in those projects.



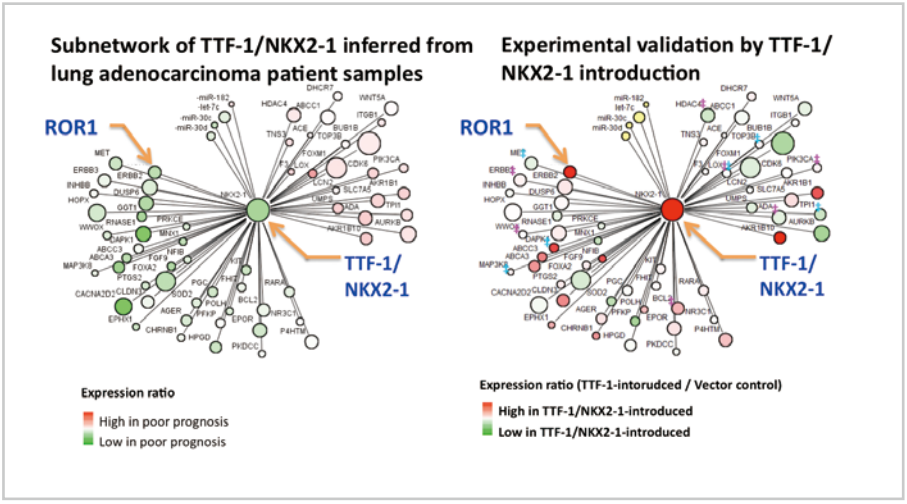
Research advancement plan of Theme 4

The specific content of the research is roughly grouped into "Understanding of the lineage of cancer and its diversity," "Deepening of personal genome understanding," "Personal medical intervention prediction," "Cell differentiation network," "Drug response network" and "Personal caner network." Although the research will be extensive, it has 3 goals. One is "Deepening the understanding of the biological diversity and personal genomes." We especially focus on the metagenome and the diversity of cancer. In addition, by advancing the understanding of personal genomes, we believe we can deepen understanding of the biological diversity. The second is "Unraveling systems characteristics/heterogeneity of cells/cancer and networks." The third is "Control of disease state" by leading those findings to prediction of medical intervention. We are going to advance our own research to achieve those three goals.

Simulation studies are being vigorously performed in Life Science. However, the life system has not yet been sufficiently clarified to simulate even *E. coli* fully. Therefore, although understanding of the human life system is advancing, it is probably still very difficult nowadays to draw new biological findings or develop a drug effective for everyone by constructing a mathematical model for simulation. In Theme 4, we start from data for data-driven science. It is our approach to develop methods to seek a clue to the cause for disease of each individual by analyzing data in an integrated way, and identifying something visible from that.



MicroRNA/mRNA gene network of lung adenocarcinoma. Overall view of the causal relations among genes (by courtesy of Professor Takashi Takahashi (School of Medicine, Nagoya University)).



Introduction of the gene TTF-1/NKX2-1 changed the gene expression of surrounding network that shows a significant relationship to survival. It is seen that the switch to determine whether or not lung adenocarcinoma survives has been altered (by courtesy of Professor Takashi Takahashi (School of Medicine, Nagoya University)).

4th Biosupercomputing Symposium Report

Computational Science Research Program, RIKEN
Eietsu Tamura

"Next-Generation Integrated Simulation of Living Matter (ISLiM)" has been comprehensively working on the Grand Challenge for Life Science, which is research and development of software in the field of life science targeting the "K computer" since the start of the project in October 2006. This project, which is organized by six research and development teams for Molecular Scale, Cell Scale, Organ and Body Scale, Brain and Neural Systems, Data Analysis Fusion, and High Performance Computing, will be completed at the end of the fiscal 2012.

At the final phase of the program, the "ISLiM International Symposium: 4th Biosupercomputing Symposium" was held from December 3 to 5, 2012, at Tokyo International Forum (Hall D) by inviting the following seven renowned overseas researchers in the field of modeling and simulations for expanding international visibility and collaborations:

Pharmacueticals – Sandra R. B. Allerheiligen (Merck)

Brain Science – Sten Grillner (Karolinska Institutet)

Cell – Ion I. Moraru (University of Connecticut Health Center)

Medicine – Grace Peng (National Institute of Health)

Organ – Aleksander Popel (Johns Hopkins University)

Molecules – Karissa Sanbonmatsu (Los Alamos NationalLaboratory)

Molecules – Ruhong Zhou (IBM T.J. Watson Research Center)

Koji Kaya, Program Director, gave an opening speech, and two keynote speakers, seven invited speakers and 14 ISLiM speakers presented highlights on the research and development that have been conducted in this program.

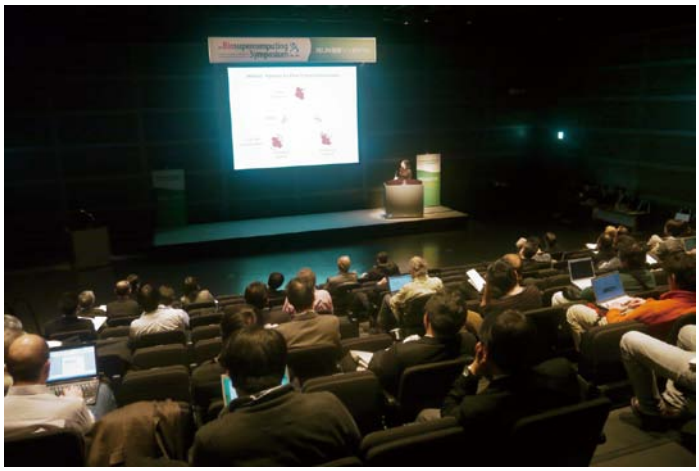
There were 145 participants successfully joining the symposium.

The following are comments of some invited speakers from abroad.

- The symposium was very broad and I have expertise in only some and not all areas, but it is a very stimulating symposium and learned a lot.
- The speeches of Program Director KAYA and Deputy Program Director HIMENO were both helpful in presenting a great picture of the ISLiM Project, and all the other speeches were also well-designed.
- The ISLiM Project is evidently a success story. It is extremely important to let other countries know about the software researched and developed in the ISLiM project and its experience.

During three days of symposiums, invited speakers and ISLiM team members communicated actively together. This was a good opportunity to present all the work of ISLiM and its results to major overseas researchers in the field, and gain their international visibilities.

The proceedings of this symposium is available in the Web site (URL: http://www.csrp.riken.jp/4thbscs/program.html).



"K Computer" Compatible Computer: Installation of SCLS Computer System

HPCI Program for Computational Life Sciences, RIKEN
Yoshiyuki Kido

Background and Mission

The HPCI innovative research field 1 is proceeding with research activities using supercomputers that are centered on the "K computer", while taking responsibility for the establishment of the research systems for computational science such as the dissemination of research findings, cultivating human resources, and creation of human networks. In that effort, we have decided to install a SCLS computer system which is compatible with the "K computer". Fig. 1 shows the field site for SCLS computer system installation. The purpose of this system is as follows:

1. To promote the use of supercomputers such as the "K Computer" in life science research
2. To provide a test bed environment for research findings from innovative research field 1 and the ISLiM project
3. To create a community for developers and users of the "K computer", and "K computer compatible computers".

Fig. 1 : SCLS computer system installation field site

Therefore, we are widely calling for participation of life scientists to use the SCLS computer system [1]. The specifications of the SCLS computer system which is to be installed are given in Table 1. The specifications of the "K computer" are also provided as a reference for comparison.

Promoting the Use of Supercomputers such as the "K Computer" in Life Science Research

In the field of drug discovery, virtual screening and docking simulations are performed using computers to narrow down candidate compounds for improving efficiency. Furthermore, it is becoming realistic to perform precise molecular dynamic simulations through calculations of intermolecular free bonding energy. To execute precise molecular dynamic simulations, the calculation load rises exponentially when factors increase, and the calculation time would be unrealistic without the use of supercomputers. However, it is difficult for small and medium-sized enterprises and research institutions to introduce supercomputers because they are highly expensive. Therefore, we have developed the SCLS computer system as a super computing environment that can be used casually on a small scale for life scientists. We have prepared an open application system for using the SCLS computer system, and programs that qualify and pass the open application may use the SCLS computer system free of charge. Although the number of nodes of the SCLS computer system is small, it is compatible with the "K computer", and we hope the know-how attained through the programs on the SCLS computer system will lead to applications for using HPCI centered on the "K computer".

Table 1 : Comparison of SCLS computer system and K computer specifications

		SCLS computer system	K computer
CPU	Name	SPARC64™ IXfx	SPARC64™ VIIIfx
	Logical performance	211GFLOPS(1.65GHz)	128 GFLOPS（2GHz）
	Number of cores	16	8
Whole system	Number of nodes	48	88,128
	Logical performance	10.1 TFLOPS	11.28 PFLOPS
	Memory per node	32GB (whole system 1.5TB)	16GB (whole system 1.5PB)

Providing a Test Bed Environment for Research Findings

In innovative research field 1 and the ISLiM project, applications that run on super parallel computer machines such as the "K computer" are also being developed. As a means for disseminating research findings, the SCLS computer system will be developed as a test bed environment and will promote the usage of these applications. Applications developed in the ISLiM project range widely over six categories, resulting in development of a total of 31 applications [2]. These applications will be installed on the SCLS computer system, and the SCLS computer system will be available as an environment for their use and evaluation.

Creation of a Developer and User Community

Many applications that run on the "K computer" and "K compatible computers" are developed by large projects, and maintenance of these applications will be difficult once the project ends. Such a situation would be a large loss, therefore we aim to grow a developer community to allow continuous maintenance of the applications running on the "K computer" and "K compatible computers". While providing computational resources, the SCLS computer system will also provide a developer community with the opportunity for lively discussions. Specifically, a Web-based system for developers to freely read and write (see Fig. 2), and a mailing list, will be provided. It is planned to operate the Web-based system as a place to contain various information such as system troubleshooting and knowhow for parallel development, and a system operation engineer will be included in the editorial staff.

Conclusion and Future Challenges

Life science, medical care, and drug discovery involve massive data and computational load that cannot be processed by humans. However, with the progress of computers, this problem can be solved. In other words, you can say that computers are absolutely necessary in the field of life sciences. The "K computer" has taken first place on the 37th and the 38th TOP500 list in 2011[3], and remains a top class computer in the world. In order to use the "K computer," it is necessary to provide supercomputer use records such as application scalability, which requires knowledge of supercomputers. It is too difficult for life scientists who do not have this expertise to use large scale computers like the "K computer". Therefore, we believe it is necessary to provide a supercomputer which is small-scale and can be used with ease by such life scientists. We hope that the SCLS computer system will become a catalyst in promoting the use of supercomputers.

[References]
[1] HPCI Innovative Research Field 1, http://www.kobe.riken.jp/stpr1-life/en/
[2] Next-Generation Integrated Simulation of Living Matter, http://www.csrp.riken.jp/index_e.html
[3] TOP 500, http://www.top500.org

Fig. 2 : SCLS computer system portal site under test operation
(Our English portal site will be constructed in the near future.)

Event information and news

● Announcement of an open symposium

Open Symposium on the "High-performance Computing Infrastructure (HPCI)" by the Ministry of Education, Culture, Sports, Science and Technology and the Strategic Programs for Innovative Research "Grand Challenge application software development project"

- Date : 10 : 00 - 18 : 20 on March 11 (Mon.) (A social gathering will be held after the symposium)
- Place : Sanjo Conference Hall, The University of Tokyo (Bunkyo-ku, Tokyo)
- Participation fee : Free (Social gathering participation fees will be charged)

For more details, visit our website (http://www.csrp.riken.jp/).

● Updated information on SCLS including events and research results is posted on our Twitter and Facebook at the earliest possible time. Our twitter account is "@HPCI_Senryaku1" and our official Facebook page can be searched by the account "HPCI Senryaku1". We hope you will follow us on our Twitter and Facebook, and click the "Like" button with your comments.

● Information on acceptance of applications from the public to use the SCLS computer system (K-computer-compatible supercomputer system) is available on the SCLS website (http://www.kobe.riken.jp/stpr1-life/).
Applications will be accepted in April, July and October in 2013. We look forward to receiving your application.

● In the Computational Biology Research Center (CBRC) of the National Institute of Advanced Industrial Science and Technology, "Strategic Programs for Innovative Research Field 1 Human Resources Fostering Program" is being implemented. The program consists of a wide range of activities, from training for novices in bioinformatics, to cultivation of human resources who can make full use of the supercomputer, "K computer", by providing seminars, workshops, tutorial training, and e-learning to the public, and by combining computational science with life science in a sophisticated way. The details of activities for FY 2013 will be announced on the website (http://hpci.cbrc.jp) as soon as they are decided.

The Development and Use of the Next-Generation Supercomputer Project of the Ministry of Education, Culture, Sports, Science and Technology (MEXT)

Next-Generation Integrated Simulation of Living Matter

ISLiM

The "Next-Generation Integrated Simulation of Living Matter" is a project sponsored by the Ministry of Education, Culture, Sports, Science and Technology(MEXT), in which research and development of simulation software to understand various phenomena occurring in the biological systems, including molecules and the human body, have been undertaken to realize a petascale simulation by making full use of the performance of supercomputer "K computer".

Strategic Programs for Innovative Research Field 1

Supercomputational Life Science

SCLS

SPIRE (Strategic Programs for Innovative Research) is a MEXT program aiming to produce ground-breaking results in computer science technology by maximizing the benefits of the HPCI (High Performance Computing Infrastructure) centered on the supercomputer "K computer", and encouraging developments in five research fields that need to be strategically addressed.
"Supercomputational Life Science" has conducted research with the mission of understanding and predicting life phenomena based on large-scale simulation and advanced data analysis, and to apply these results to design and implement medicine and medical care with its research.

BioSupercomputing Newsletter Vol.8 2013.3

Issued: March 2013

RIKEN

Research Cluster for Innovation
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

HPCI Program for Computational Life Sciences
7-1-26 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan
Tel : +81-78-940-5692 Fax : +81-78-304-8785
http://www.kobe.riken.jp/stpr1-life

RIKEN 2013-028