Next-Generation Integrated Simulation of Living Matter
HPCI Strategic Program Field 1 Supercomputational Life Science

BioSupercomputing Newsletter

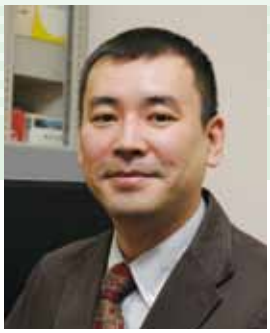
2012.3 Vol.6

ISLiM Interim Accomplishment Meeting held in December 2011 (see page 14)

CONTENTS

SPECIAL INTERVIEW	○Development of New Fluid-structure Interaction Analysis (ZZ-EFSI) Resulting in Rapid Achievement of High Operation Performance Research Associate Professor, School of Engineering, The University of Tokyo Kazuyasu Sugiyama	2-3
	○Interview with High-performance Computing Team Members: Continued Efforts in Tuning to Harness the Potentials and the High Capability of the K Computer Group Director of Computational Molecular Design Group, Quantitative Biology Center, RIKEN Makoto Taiji Senior Researcher of High-performance Computing Team, Integrated Simulation of Living Matter Group, Computational Science Research Program, RIKEN Yousuke Ohno Senior Researcher of High Performance Computing Development Team, High Performance Computing Development Group, RIKEN HPCI Program for Computational Life Sciences Hiroshi Koyama Researcher of High-performance Computing Team, Integrated Simulation of Living Matter Group, Computational Science Research Program, RIKEN Gen Masumoto Research Associate of High-performance Computing Team, Integrated Simulation of Living Matter Group, Computational Science Research Program, RIKEN Aki Hasegawa	4-5
Report on Research	○ Functional Analysis of Multidrug Efflux Transporter AcrB by All-Atom Molecular Dynamics Simulation Graduate School of Nanobioscience, Yokohama City University Tsutomu Yamane, Mitsunori Ikeguchi (Molecular Scale WG)	6
	○Multi-scale Modeling of the Human Cardiovascular System Computational Science Research Program, RIKEN Liang Fuyou (Organ and Body Scale WG)	7
	○ Toward a spiking neuron-level model of the early saccade visuomotor system Kyoto University Jan Morén Nara Institute of Science and Technology Tomohiro Shibata Okinawa Institute of Science and Technology Kenji Doya (Brain and Neural Systems WG)	8
	○Developing the MD Core Program for Large Scale Parallel Computing Computational Science Research Program, RIKEN Yousuke Ohno (High-performance Computing Team)	9
SPECIAL INTERVIEW	○Pioneering the Future of Computational Life Science toward Understanding and Prediction of Complex Life Phenomena Program Director of RIKEN HPCI Program for Computational Life Sciences Toshio Yanagida Deputy- Program Director of RIKEN HPCI Program for Computational Life Sciences Akinori Kidera Deputy- Program Director of RIKEN HPCI Program for Computational Life Sciences Yukihiro Eguchi	10-11
Report on Research	○Free Energy Profile Calculations for Changes in Nucleosome Positioning with All-Atom Model Simulations Quantum Beam Science Directorate, Japan Atomic Energy Agency Hidetoshi Kono, Hisashi Ishida, Yoshiteru Yonetani, Kimiyoshi Ikebe (Field 1- Program 1)	12
	○ Estimation of Skeletal Muscle Activity and Neural Model of Spinal Cord Reflex Information Science and Technology, The University of Tokyo Yoshihiko Nakamura (Field1 - Program 3)	13
Report	○ISLiM Interim Accomplishment Meeting in 2011 Computational Science Research Program, RIKEN Eietsu Tamura	14
	○Computational Life Sciences Classes Held in High Schools HPCI Program for Computational Life Sciences, RIKEN Chisa Kamada, Yasuhiro Fujihara, Yukihiro Eguchi	15
Event information		16

Development of New Fluid-structure Interaction Analysis (ZZ-EFSI) Resulting in Rapid Achievement of High Operation Performance



Research Associate Professor, School of Engineering, The University of Tokyo
Kazuyasu Sugiyama

Reviewing algorithms for establishing a new analytical method

The development of fluid-structure interaction analysis has been promoted mainly in the manufacturing field. Unlike machine parts, the living body is not created according to an original blueprint. The starting point of development of the application was thus the desire to establish a unique fluid-structure interaction analysis that was fully compatible with the medical images created by CT scanning and MRI, and allowed simultaneous handling of fluids and solids that were different in mathematical expression of stress characteristics. By taking this approach, we attempted to develop an application that can be easily used in the healthcare setting and contributes to understanding of the very essence of life, clarification of disease mechanisms, and drug discovery.

For this purpose, we made efforts to develop a new analysis technique based on the Eulerian method that does not require the process of mesh generation and reconstruction. We took the formulation approach so that all the physical quantities can be updated on the fixed mesh by utilizing voxel data. In this manner, we attempted to easily simulate problems with complicated geometry and those containing many dispersed objects. Because this method facilitates expansion of computation scale, it is suitable for massively parallel computation. We also re-created a time integration algorithm. Recently, researchers have widely reviewed the pseudo-compressibility method that enables completely explicit time integration. We introduced dynamic parameters into this method, and perform optimization processing for minimization of velocity divergence to achieve higher effective performance and parallel scalability with numerical stability. The intention behind these developments was that they would like to achieve uniformity of operation quantity per node as much as possible, and make the most of the performance of the K computer by avoiding iterative processing.

Because the computational technique developed is substantially different from conventional interaction analysis, its validity needed to be verified. A high level of reproducibility was confirmed by making a comparison with the computational results of previous fluid-structure interaction problems and fluid-membrane interaction problems that had been sufficiently verified. When we attempt to numerically handle a multimedia system in which the physical quantity jumps at the interface, we often encounter collapse of the conservation law. However, we examined our problem focusing on the budget of kinetic energy transport, and confirmed that its conservation requirement was sufficiently met. Moreover, it is important to make a comparison with actual experimental observation. For example, in the examination of the behavior of red blood cells in a fine vessel, a phenomenon of axial accumulation occurs. This phenomenon leads deformed red blood cell groups to the central axis of the vessel. We confirmed that our simulation results also clearly reflected such behavior of red blood cell groups (cell free layer near the vessel wall) and their shape (slipper-shaped red blood cell).

Currently, we are engaged in the development for practical application and extension of this new method to thrombus simulation. First, we attempted to numerically predict the process of platelet thrombus, or how a platelet attaches to the injured vessel wall. We have already developed part of a blood flow analysis simulation at the continuum level. When we try to figure out platelet attachment, we should consider the bond formation between proteins, in other words, a molecular phenomenon with fluctuation that is definitely different in scale. Because of the extremely large difference in scale, we do not directly approach the phenomenon on the molecular scale, but adopt a method that enables stochastic handling of the effect of fluctuation when observed on the continuum scale. Specifically, whether a platelet attaches or not can be expressed by the ligand receptor binding. We use a model statistically reflecting this molecular scale effect. We have been engaged in development of analytical methods based on such multi-scale multi-physics. Since we have nearly completed preparations for the new project, we would like to expand the scale and actually operate the K computer to analyze the process through which platelet attachment occurs in a vessel filled with many red blood cells.

Achievement of about 46% of the peak computation performance

We started to use the K computer, which is being increasingly used for practical computation, this year. Now, a scalar-type super computer is said to be unsuitable for fluid analysis. For example, it is said that with a fluid application for which the effective performance using the "Earth Simulator" of a vector-type super computer is about 70%, only about one-tenth of the performance can be realized with a scalar-type super computer. For this reason, our first goal was 10% performance. However, we made efforts to effectively utilize the functions of the K computer and elicit its high computation performance. Consequently, we succeeded in achieving a peak computation performance of about 46% in the K computer environment. We are satisfied with this relatively reasonable result. Although the K computer is a scalar-type super computer, it incorporates the hardware idea that is common to a vector-type super computer. Therefore, this result was predictable, but we can honestly say that all our efforts have been rewarded.

One of the advantages of the K computer is its high communication speed. The hardware reduction processing is a particularly important strength of the K computer. Communication may be adjacent or global communication. When we attempt to estimate the residual error in the entire field, we need the reduction processing for adding up the error at each node. In this case, global communication is realized. The present

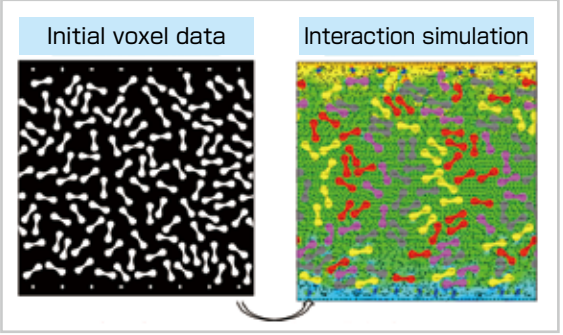
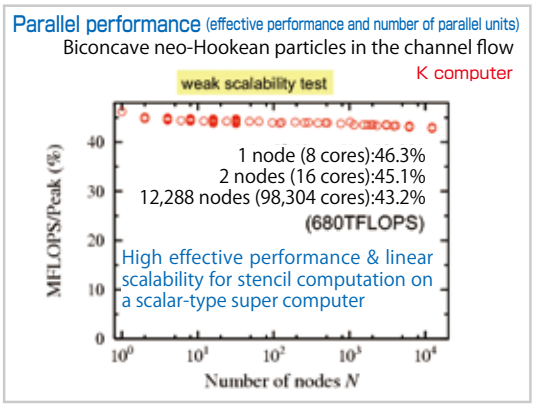


Fig.: Characteristics of Euler method
All the physical quantities are updated on the fixed mesh. The fluid-structure interaction problems that have complicated boundary shapes and many dispersed objects can be simulated easily. Generation and reconstruction of a boundary-fitted mesh is unnecessary.

The K computer adopts the innovative network structure called a six-dimensional mesh/torus network interconnect (Tofu). Its adoption has contributed toward saving internode communication time. The photo shows the conceptual model of Tofu.



pseudo-compressibility method also requires reduction processings. The percentage of computation time needed for global communication increases with the increase in the number of nodes. At the beginning we were not able to predict the exact rate of increase, and speculated that this point would be a bottleneck for our code. In the practical setting, however, we succeeded in high-speed processing. Global communication actually accounts for less than 1% of computation time at the current operation level of about 100,000 cores. Adjacent communication accounts for 3-4% of computation time, thus showing that the transmission speed was higher than expected. We expect that the communication will be within the range of about 8% even if we use 80,000 nodes and 640,000 cores.

Results of practical use of K computer

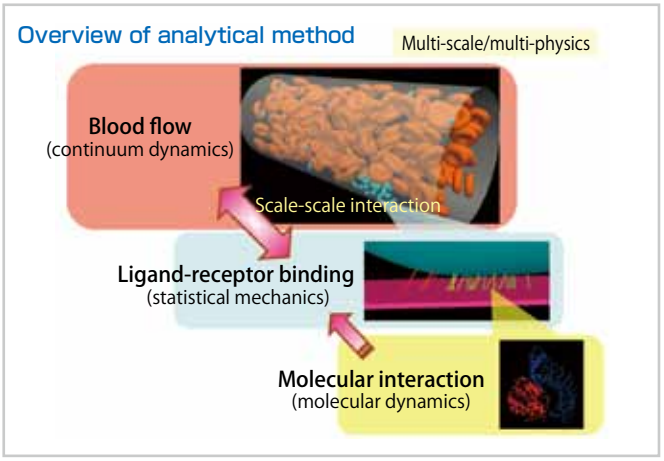
We are engaged in continuum dynamics characterized by relatively simple principles and concepts. The parts to be described in codes in continuum dynamics are also quite simple. As the principles and concepts would be extremely complicated in some research fields, our results with the K computer and our attempt to generate higher performance might not be widely applicable, but some of our experiences would be informative.

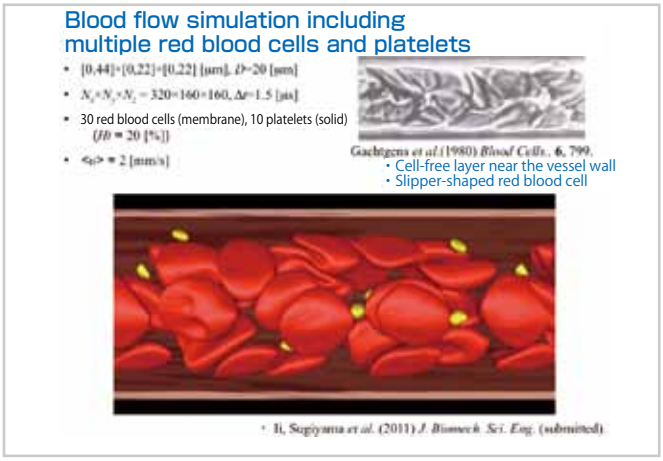
We first directed our attention to the hardware monitor information by the profiler so that we could identify the problems and improve performance. This information enables display of effective performance by loop and the speed of communication with the memory. In this manner, we can identify where the problem (hotspot) exists. The manual always instructs users to "identify the hotspot and cope with it," and we followed this common formula and started to improve the hotspot. We adopted this restricted approach to each loop on the one hand, and a comprehensive approach to the entire program flow on the other. Thus, we reviewed the program by integrating the processing divided into several parts, or alternatively dividing the processing into several parts. It seems that a developer familiar with these code flows is the only person who could complete this task. The work may result in remarkable improvement in performance, and actually, the performance was increased in a very effective manner. A different member of our group wrote the code from the beginning to ensure cross-check. Although this approach appears to be a roundabout way, it solves the problem efficiently, and effectively promotes rapid development. As there are various types of code in the tuning stage, this resulted in many advantages. For example, we can reveal invisible problems and recognize merits by making a comparison between profiler results. As a result of continuously taking such a conservative approach, we were able to successfully enhance the application performance.

We are using Fortran, and we believe that we made the right choice. While some researchers say Fortran is out of date, the HPC compiler that a manufacturer first procures is generally Fortran. C, which would be more capable than Fortran, requires more effort for completion of the compiler. On the other hand, Fortran, within its limited ability, can easily complete a high-performance compiler. In the early stage of introduction of the K computer, writing codes with Fortran allows a higher degree of optimization and acquisition of more information than writing codes with C.

Logical thinking is of prime importance in increasing performance. A user of the K computer needs to be familiar with a hardware composition with a stratified structure. However, we can not find the best answer if we take the logical approach alone. We occasionally use logic and speculate that "this action should result in a successful outcome," but encounter a trade-off, resulting in failure. In this case, we should try to progress only through trial and error. It is unfavorable to aim at creating a high-quality finished program in the first trial. I think more attention should be directed to an approach that enables writing of programs in a manner that allows trial and error.

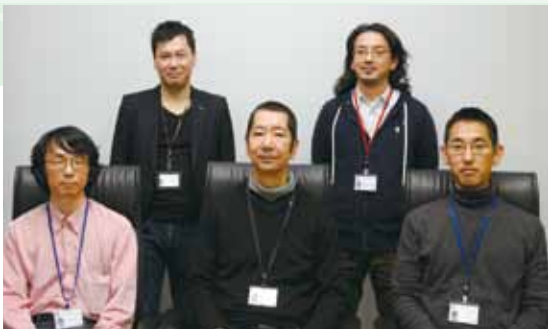
Note: The K computer is currently in the process of development. The numerical data included in this article are those obtained to date.
Acknowledgment: The data on computation with the K computer reflect the results of its experimental use.





Interview with High-performance Computing Team Members Continued Efforts in Tuning to Harness the Potentials and the High Capability of the K Computer

Group Director of Computational Molecular Design Group, Quantitative Biology Center, RIKEN — Makoto Taiji
Senior Researcher of High-performance Computing Team, Integrated Simulation of Living Matter Group, Computational Science Research Program, RIKEN — Yousuke Ohno
Senior Researcher of High Performance Computing Development Team, High Performance Computing Development Group, RIKEN HPCI Program for Computational Life Sciences — Hiroshi Koyama
Researcher of High-performance Computing Team, Integrated Simulation of Living Matter Group, Computational Science Research Program, RIKEN — Gen Masumoto
Research Associate of High-performance Computing Team, Integrated Simulation of Living Matter Group, Computational Science Research Program, RIKEN — Aki Hasegawa



Yousuke Ohno, Makoto Taiji, Gen Masumoto (from the left, front row), Hiroshi Koyama and Aki Hasegawa (from the left, back row)

The K computer proved to be more stable than expected.

As the development of the supercomputer K continues, we are taking up the great challenge of running application programs on it, and performing tuning and tentative computation although on a trial basis. The "MD core program for large-scale parallel computers (cppmd)" you developed has already recorded 1.3 PFLOPS*, and development of other application programs has continued with a view to achieving more than 1 PFLOPS. Today, members of the High-performance Computing Team who are providing various support for researchers talk about their actual impression of the K computer along with other problems found in their support work.

TAIJI (Honorifics omitted)**:** Speaking of the hardware aspects, I have the impression that the system was quite stable from the word "go". I have hardly ever heard of any sudden unexpected shutdown, and the system managed to operate on a rather large scale right from the beginning. At the outset, we expected much trouble due to malfunctions, but the system became very stable by last spring when we got involved in it.

OHNO: Admittedly, it was really surprising that we had so little hardware-related trouble. Rather, it was software-related problems that we had difficulty in coping with. In particular, since the compiler was under development, C or C++ based compilers did not keep up with the development of the Fortran compiler. We had compilation failure and erroneous results although the program as such is not defective. In any case, with regard to software-related issues, we are fully prepared to cooperate in finding bugs, because we encourage and benefit from earlier delivery…(laugh).

Fortran user achieved higher performance

Does it mean that you obtained better results from Fortran?

OHNO: In the beginning, Fortran was also faster in terms of performance. In addition, I tried to write such codes that allow optimization using the compiler under development. In the meantime, we are getting improvements from other compilers these days.

KOYAMA: This is not because I am a Fortran user, and I have no intention to value specific manufacturers over others, but automatic parallelization of the Fortran compiler is very complete. Regardless of the relatively limited number of users, it will be easy to use by those who have already been accustomed to Fortran for many years. On the other hand, we have to pay attention to the problem of memory bandwidth. I first thought that Fortran users who have used vector computers such as earth simulators can readily adapt to porting, but there are two distinct cases, i.e., successfully exploiting performance within the range of available memory bandwidth, or failing to do so. The demarcation is very clear. If that's not possible, you will have to make changes to the algorithm itself to exploit higher performance. Reluctantly, we have to accept the constraints as they are, although that's of course challenging…

Certainly, when compared with the earth simulators, we need to accept the memory bandwidth on an as-is basis.

KOYAMA: I'd agree, but the memory bandwidth issue affects the real consequences, I mean, some people succeed, and others do not succeed. That's hard.

In that context, it seems that ZZ-EFSI proved that it works successfully.

OHNO: Certainly, loop statements written in Fortran leads to favorable performance in automatic parallelization.

KOYAMA: It's of course a result of our continued development efforts, but in a sense we were very lucky (laugh). However, many K computer users want sufficient compiler optimization for programs using C and C++. I want manufacturers to respond their needs.

MASUMOTO: People having experience in fluid dynamics ought to have experience in vector supercomputers. On the other hand, quite a few people in biotechnology have not used supercomputers. They should have made necessary preparations taking the many C users and C++ users into consideration.

KOYAMA: At any rate, that applies to the current stage. The situation will be quite different when the system is in full-scale operation.

OHNO: Being sophisticated in terms of language specification makes it more difficult to construct a compiler. It's possible that construction of the compiler may be intentionally postponed. Also, there must have been some pressure to realize high performance from the very beginning.

MASUMOTO: So getting started with relatively easy-to-handle Fortran, that's not unreasonable.

TAIJI: The K computer incorporates a parallel computation model called VISIMPACT for effective massively parallel computing. It facilitates high-speed core synchronization. That function facilitates vectorization-oriented parallelization. For that reason, usually, parallelization between cores is done at an upstream location, but advantageously, parallelization in the K computer takes place at a downstream location and then the loop is decomposed. Probably, it depends on whether or not automatic parallelization can take place as to whether automatic parallelization as a whole is successful.

KOYAMA: People using many loop structures and a lot of data in them are getting good and faster performance. Although there is the memory bandwidth issue, I think that, in terms of optimization, people accustomed to vector types are can readily adapt to automatic parallelization.

MASUMOTO: The application programs that I was involved in are all written using C and C++, and I had trouble with compilation at the beginning. Some application programs have to demonstrate performance immediately, and Mr. Ohno cooperated with me to do as what can be done in the current situation, so recently I have had much better results. With regard to other application programs, I am wondering whether we had better wait for maturity of the compiler or modify codes in accordance with the currently available compiler. If time allows, we should refrain from doing something tricky and wait for completion of the compiler…

So you have to solve the problem of how to deal with computers under development.

MASUMOTO: Sure, but that's a problem specific to C++. C is much better than that. As regards Mr. Koyama's comment, I am very much surprised by the fact that Fortran is much, much better… (laugh)

KOYAMA: It was certainly easy, and I have a very good impression of Fortran (laugh).

Real performance to be assessed in future

HASEGAWA: It was not so complicated or time-consuming to port the data analysis application programs that I am in charge of. However, there is much room for expanding capability through hybrid parallelization. I'm going to direct my efforts toward that end. In addition, many I/O components are still pending because the system itself has not entered the full-scale operation phase, so tuning will be done somewhat later.

MASUMOTO: When compared with other application programs, I feel that I/O-related components create a bottleneck.

HASEGAWA: It may sound misleading to say that I don't have much trouble with it (laugh). I do have a lot of trouble. The point is that we don't have many options we can choose at this stage. Further development of the I/O-related components will offer us more options in due course.

KOYAMA: For example, in the computation regarding fluids, the computation results do not differ so much from computations using dummy data, and they are meaningfully predictable, so the I/O configuration can be defined afterward. In contrast, in biotechnology, data should be given first. This means that the computation could lack significance in the absence of a predefined benchmark affecting the hit ratio. An operation using dummy data would give you no hits, and you have to use more or less fully-fledged data to tune the system and thereby get meaningful results. Accordingly, in some cases, I/O features have to be available at an early stage. That is where some of our efforts are directed.

—— The K computer incorporates many advanced techniques that have drawn much attention as early as in its development phase, including the 6D Mesh/Torus Interconnect (Tofu Network), and the sector cache feature in which specific data can be maintained in the cache memory. What is your impression of these after actually using them?

OHNO: In the MD computation in the cppmd, the amount of computation is relatively large compared with the amount of data, and the computation does not cause much memory bottleneck. Since the cache capacity is, actually, sufficient in most cases, we were successful in delivering performance without using the sector cache functionality. In the application programs in life science files that I now handle, I have not had to use the sector cache thus far. With regard to networks, in the case of MD, the communications ratio is relatively low, and favorable performance is achieved even when tuning is left to be done to the maximum extent possible. Either of them is not used full-scale, and it is a long way to go until it proves its true performance.

Challenges specific to development phase

—— When the K computer was under development, Mr. Taiji told me that one can't tell what will happen until a computer this large is actually used. Is there something you did not expect or predict, unexpected findings, or something surprising?

TAIJI: I still hold to the idea "you can't tell unless you try," but what was the most surprising was that, as I said previously, we didn't have much trouble in the hardware aspects and the system had more stability than expected (laugh).

—— So it was an unexpected benefit, wasn't it? And…

TAIJI: Yes. The downside to this, as we were discussing, the compiler is still Fortran- based. We of course understand that it is in the development phase, but I want a C++ compiler that really works at this stage.

KOYAMA: What is most challenging is that some features are not implemented in the development phase, and we have to decide on the extent to which the application should be modified to avoid bugs. For example, version upgrades could lead to improved performance even when major modifications are not made at this point so that the performance is improved. We need to select the most promising out of the available approaches…

MASUMOTO: We have oftentimes had similar situations up to now (laugh). However, such efforts might not be meaningful to the application programs, but they help us build our skills, and it is somewhat useful for us to understand the characteristics of the K computer.

KOYAMA: I understand that gaining experience is very important. However, it would be too much burden on those involved in development of individual application programs to yield their own research results, for that increases their workload.

MASUMOTO: Even in that case, it will never happen that you get splendid results from the original codes without making modifications. You need to try many things. Of course, having tried many things often ends up going back to the original state (laugh).

—— What are the questions frequently asked by users who develop application programs?

MASUMOTO: I had many inquiries before, saying, "I cannot compile successfully" (laugh).

KOYAMA: In some cases we had trouble in cross-compiling.

—— Cross-compiling?

OHNO: Since the K computer is a machine in which the front-end CPU and the computation node CPU are different, they may fail to run in different environments. The operation is successful now, but at the beginning I had much trouble.

KOYAMA: As the K computer is under development and there are many functional limitations, some people report that system development is affected by such functional limitations specific to development phases, although they suffer from their own bugs, of course (laugh).

MASUMOTO: Since there are still many beginners, they often ask me elementary questions. Meanwhile, the manuals are very complete. They seem more complete than expected in the first version.

OHNO: Rather, the machines have to be upgraded to be in line with the integrity of the manuals (laugh).

Efforts by application developers are required.

—— I think many people would like to use the K computer. Please give them some advice from a professional viewpoint.

KOYAMA: There are many specific topics, but it is difficult to offer general advice.

OHNO: When performing large-scale computation using the K computer, it will be necessary to provide hybrid parallelization where two parallelization schemes are used. I have observed some cases where people find difficulty in this hybrid parallelization.

MASUMOTO: The Headquarters also recommend hybrid parallelization.

KOYAMA: Not only in the K computer but also in other large-scale machines, the number of cores per node tends to increase, so now we will have to rely on hybrid parallelization to exploit the capability of CPUs, given the current trends of high-performance computing (HPC).

OHNO: As we go back to basics, we need much effort by those who write the application programs. In fact, their efforts are the key to harnessing HPC potentials.

—— Nevertheless, researchers developing application programs use computers to output their research results, so it would be rather demanding for them.

OHNO: In that sense, a High-performance Computing Team like us can provide support for them. Complete division of labor is an extreme example, where there are two distinct groups, i.e., members dedicated to code tuning and members dedicated to research using the tuned codes. In that case, revising the algorithm itself will have to ensure that researchers' computation targets are left intact. Even when you do not have a clear idea of computation targets, you have to tune codes with computation methods, the presence of computation targets, and specific objects of computation taken into account. Complete division of labor could cause communication problems. I think that people who can handle code tuning while being involved to some extent in the research field as such are required.

—— The K computer achieved first place in the TOP500 on the LINPACK benchmarks in two consecutive years, and research on silicon nanowire materials using the K computer won the Gordon Bell Prize, which further raised expectations for the research results that will follow these two achievements. Your endeavor will become most important for the K computer to meet expectations. I am looking forward to you getting superb achievements. Thank you for making time in your busy schedule.

* Results obtained by special use for 2011 Gordon Bell Prize

Functional Analysis of Multidrug Efflux Transporter AcrB by All-Atom Molecular Dynamics Simulation

Graduate School of Nanobioscience, Yokohama City University
(From the above) Tsutomu Yamane
Mitsunori Ikeguchi
(Molecular Scale WG)

Drug resistance is a phenomenon where drugs become ineffective against pathogens or cancer cells, which leads to a major social problem. There are some known mechanisms for drug resistance. One of the causes is where membrane proteins, located in the cell membrane of the bacteria and called multidrug efflux transporters, play an important role. A multidrug efflux transporter is a membrane protein which works to discharge various cytotoxic chemicals actively to the outside of the cell.

Among RND-type multidrug efflux transporters typically found in multidrug resistant Pseudomonas aeruginosa which is the main cause of nosocomial infection, the atomic-level crystalline structure of Escherichia coli-derived AcrB was determined in 2006 by Dr. Satoshi Murakami (presently professor of Tokyo Institute of Technology), et al. According to the results, AcrB functions as an assembly of three giant proteins (homotrimer), each of which consists of approximately 1,000 amino acids. The respective proteins have three different 3-D structures that act for access, binding and extrusion of the drug, respectively (Fig. 1A). With the drive force obtained by transfer of protons into cells using the difference in proton concentration (pH) of periplasm and cytoplasm, each protein switches between the 3-D structures in turn to perform the drug-efflux function (functionally rotating mechanism, Fig. 1B). In addition, three charged amino acids located at the membrane spanning region (Asp407, Asp408, Lys940) contribute to proton transfer into cells, and the side chain structure of Lys940 is different only in the extrusion state (Fig. 1C). Through this, a change in the protonated state of these charged amino acids due to proton transfer is believed to trigger the functionally rotating mechanism.

AcrB is the common research target of our Molecular Scale Team, and its molecular simulation has been performed by various methods. Since a more realistic environment is considered in the all-atom molecular dynamics simulation we use, our computation covers a system including membrane lipids, water molecules, ions and even hydrogen atoms (Fig. 2, center). Therefore, our computations have to be performed on a very large system of which the total particle number is approximately 470,000, but we can take detailed views of the various interactions in the system. Such an all-atom molecular dynamics simulation revealed the following.

(1) When Asp407 and Asp408 were both in the deprotonated state in the extrusion protomer, it was found that the side chain of Lys940 changes to have the structure found in the binding state and the access state (Fig. 2, right panel). Meanwhile, when only Asp408 is in the protonated state, the extrusion-type conformation was found to be stably maintained.

(2) In the simulation where both Asp407 and Asp408 of the extrusion protomer were deprotonated concurrently to change the structure of the Lys940 side chain, the drug entry point which is usually closed in the extrusion state opens, and a structural change toward the access state was observed (Fig. 2, left panel).

Currently we are studying to reveal the mechanisms of how a structural change of the extrusion-type Lys940 side chain leads to a structural change for opening the drug entry point.

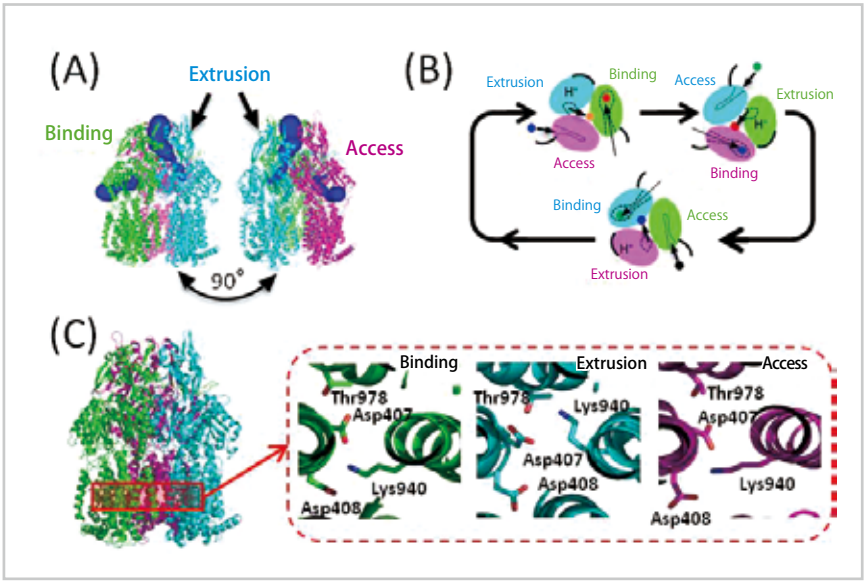


Fig.1 :Structure and drug efflux mechanism of multidrug efflux transporter AcrB
(A) Structures of respective monomers, and the route of the drug (blue)
(B) Drug efflux mechanism (functionally rotating mechanism) In the figure, Proton: H^+, Drug: circle
(C) Amino acids in membrane spanning region involved in intracellular incorporation of protons

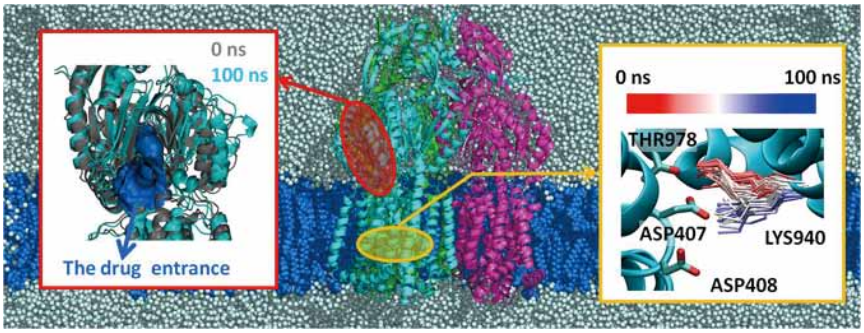


Fig.2 : System used for all-atom molecular simulation (center: a water molecule is indicated by a gray sphere, and a lipid by a blue sphere in the figure), and the results (left and right panels)

Multi-scale Modeling of the Human Cardiovascular System

Computational Science Research Program, RIKEN
Liang Fuyou
(Organ and Body Scale WG)

The cardiovascular system plays an important role in maintaining homeostasis; on the other hand, it is also highly vulnerable to diseases. Cardiovascular diseases are currently the number one killer of human beings. Research and technological development aimed to improve or support the diagnosis and treatment of circulatory diseases has been widely performed since decades ago. In particular, the correlations between vascular diseases and hemodynamic factors have been one of the most important issues.

The combination of medical imaging and fluid dynamics computing techniques makes it possible to noninvasively visualize in-vivo 3D-blood flows in a patient-specific manner, thus having been considered to carry some potential for clinical use. However, there are several critical problems with the application of the method. For example, the measurement of blood flow/pressure at the boundaries of a 3-D model is subject to various limitations related to measurement accuracy or vascular location suitable for the implementation of measurement. Moreover, the measurement can generally be done only when patients are in the resting state; as a consequence, the results of most hemodynamic computations reflect only the resting hemodynamic characteristics. As a matter of fact, hemodynamic characteristics are always changing in response to variations in physiological conditions.

Besides local blood flow patterns, hemodynamic phenomena viewed from a larger scale, such as systemic hemodynamic behavior, pulse wave propagation, carry medical value as well. For example, it is known that arterial pulse waves carry some characteristics that reflect the functional state of the heart and the morphological and mechanical properties of the arterial system. In fact, arterial pulse waves have long been used as a critical indicator in the diagnosis of cardiovascular diseases.

The complexity of the cardiovascular system raises another problem. For example, the cardiovascular system of a human is comprised by hundreds of arteries (veins), tens of millions of arterioles (venules), and billions of capillaries. These vessels are connected to form an extremely complex network structure. The complexity determines that it is impossible to construct a computational model that involves all the blood vessels since the data (such as those on vessel geometry and mechanical properties) for each vessel are not available, not to speak of the unaffordable computer resource required by such a model.

To handle the dynamic characteristics and the complexity of the cardiovascular system, we have developed a multi-scale modeling method. First, the microcirculation (including the capillaries, arterioles and venules), the arterial system and local large artery segments, are described respectively by a lumped-parameter (0-D), 1-D and 3-D model. Then, the models are coupled with each other to yield an integrated model of the entire cardiovascular system (Fig 1). The integrated model contains sub-models of different scales and hence is usually called a multi-scale model. An advantage of this modeling method is that it significantly reduces the computational cost for the whole model by representing the most complex microcirculation with a 0-D model that demands the least computer resource. From the point of view of hemodynamic analysis, the present model is preferable as well in that it allows computation of 3-D blood flows under various physiological/pathological conditions by coupling 3-D models of local blood flows with models of the remaining system.

The model has been applied to investigate various cardiovascular physiological phenomena related to hemodynamics. For instance, the model-based simulations successfully reproduced the physiological phenomenon that the peak value increase and waveform variation of arterial pressure wave along the aorta are pronounced in young subjects but disappear with aging. More importantly, our model study indicated that such changes in arterial pressure wave propagation are caused by aging-associated arterial stiffening. We also coupled the model with an upper-arm cuff model and succeeded in evaluating the measurement accuracy of an automated sphygmomanometer and examining the validity of a device that assesses arterial stiffness based on the information obtained from an oscillometric cuff. Presently, we are developing methods to incorporate patient-specific clinical data (such as medical images for blood vessels, blood flow rate) into the model, aiming to preoperatively predict the postoperative outcomes or the risk associated with a surgery. In the future, by using a supercomputer, we will perform blood flow analysis that accounts for 3-D hemodynamic phenomena in a wider arterial region and incorporates more patient-specific information and physiological mechanisms.

[References]
[1] Liang F.Y., Takagi S., Himeno R., Liu H., Biomechanical characterization of ventricular-arterial coupling during aging: a multi-scale model study, J Biomech, 42:692-704, (2009)
[2] Liang F.Y., Takagi S., Himeno R., Liu H., A computational model of the cardiovascular system coupled with an upper-arm oscillometric cuff and its application to studying the suprasystolic cuff oscillation wave concerning its value in assessing arterial stiffness. Comput Methods Biomech Biomed Engin. (in press)

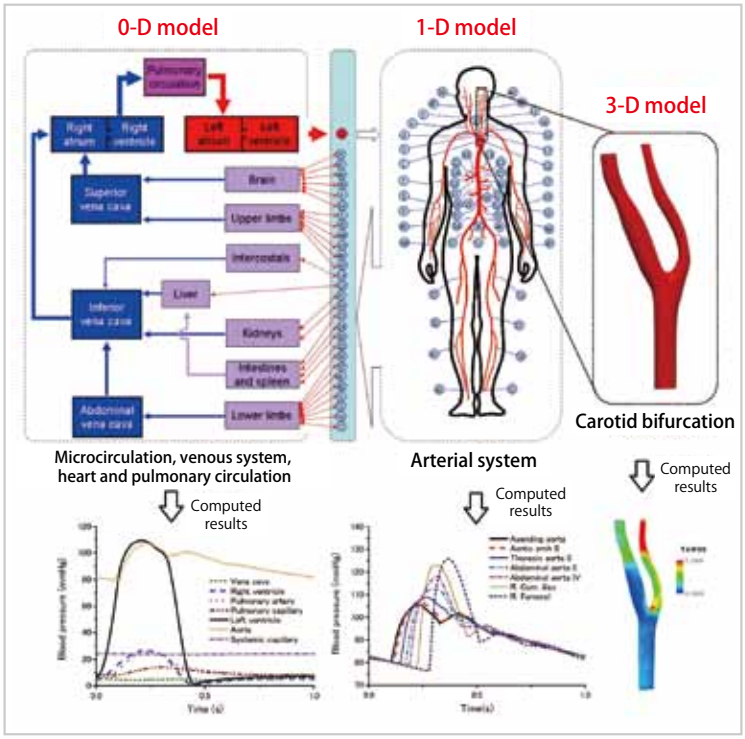


Fig 1 : Multi-scale modeling of the human cardiovascular system
(Upper panel: A schematic description of the model, Lower panel: Examples of the results computed with each sub-model)

Toward a spiking neuron-level model of the early saccade visuomotor system

Jan Morén (photo) ①
Tomohiro Shibata ②
Kenji Doya ③
(Brain and Neural Systems WG)

①Kyoto University
②Nara Institute of Science and Technology
③Okinawa Institute of Science and Technology

Macroscopic visuomotor behaviour depends on high level interaction among system-scale structures in the brain. The functional mechanisms of such structures are however rooted in the low-level neurophysiology. To fully understand organic behaviour we need to look at these structures at both scales. Multiple description levels rapidly increase both the structural complexity and computational demands of the resulting model.

Our ultimate goal is to create the entire perception-action loop, from retina to the eye muscle motor systems, by ourselves and by collaboration with other groups, and we currently aim at neuron-level modeling for generating saccadic eye movements (saccades) with a particular emphasis on superior colliculus (SC), as saccades are a good target in that they are highly repeatable and stereotypical.

We use the NEST simulation tool to create a large-scale spiking-neuron level model of the intermediate SC (SGI) (Fig. 1), based on the present best knowledge on neurophysiology. We use the conductance-based Adaptive Exponential IaF neuron model augmented with an NMDA synaptic input. The number of neurons in the model is on the order of 100k, with burst, buildup, quasivisual, deep inhibitory and inhibitory interneurons, as well as supporting units for input generation and activity integration. The SGI is retinitopic and forms the first part of the oculomotor system. The retina projects to the cortical vision systems but also to the superficial SC. It in turn selects locally salient areas and activates the corresponding point in the intermediate division directly and through indirect pathways. Whereas the superficial division is a sensory area, the intermediate division is part of the motor system; they make up the sensor-motor convergence point in this perception-action loop.

A saccade is initiated in the SGI with buildup neuron activation spreading around the salient point. The lateral SGI is normally inhibited by neurons in the brainstem and by other areas, simplified here to a single inhibitory input. With the release of external inhibition, burst neurons generate an NMDA synapse-mediated burst of spikes that activates burst neuron systems in the pons and midbrain. They in turn move the eye horizontally and vertically, regulated by the cerebellum.

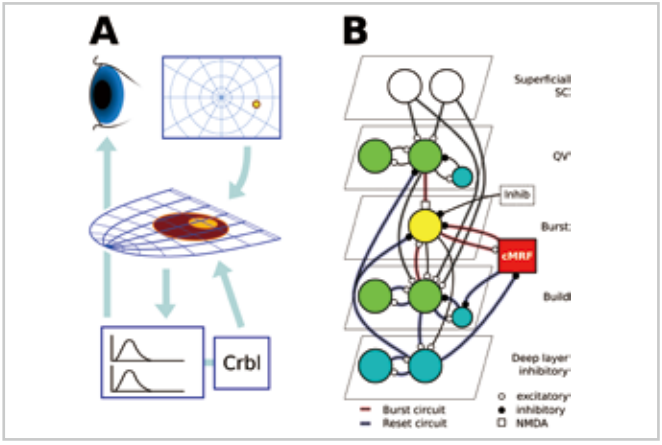


Fig. 1 : A) The main components and flow of the subcortical saccadic system. Retinal output activates the retinitopic SC, which in turn activates horizontal and vertical eye movement systems in the pons and the midbrain, and the cerebellum acting as a regulator.

B) The intermediate SC (SGI) model circuit. Red connections form the burst generation circuit, and blue connections the spreading activation and saccade system reset. Grey connections are common to both.

The SGI is fairly well understood but there are still a number of questions surrounding the nature of the activity in this area. The bursts may be a retinotopic target indicator, a movement vector or a velocity command. The significance of the spreading activation among buildup neurons is also not clear. We attempt to account for recent results on the spiking neuron activity and the spreading activation among buildup neurons, while keeping with the known neurophysiology of this area (Fig. 2).

Burst neuron activity indicates a saccade distance-dependent eye movement profile over time. The bursting profile is created through NMDA synapse-generated bursts with linear inhibitory feedback from a spike accumulator in the mesencephalic reticular formation. The resulting activity follows experimentally determined bursting profiles. Buildup neuron activity spread is accomplished by local intraconnections and inhibitory interneuons. We propose that the amount, but not the position, of the activity tracks saccade progess over time. This triggers inhibitory neurons in the deep SC that in turn inhibit burst neuron activity directly, and indirectly restores control to the rostral pole. This improves saccade end timing accuracy.

We aim to create a superficial SC and to pair our model to similarly detailed models of the retina and brainstem systems. We are also exploring embodiment by connecting the model to physical systems, enabling us to study computationally heavy model behavior in more naturalistic settings.

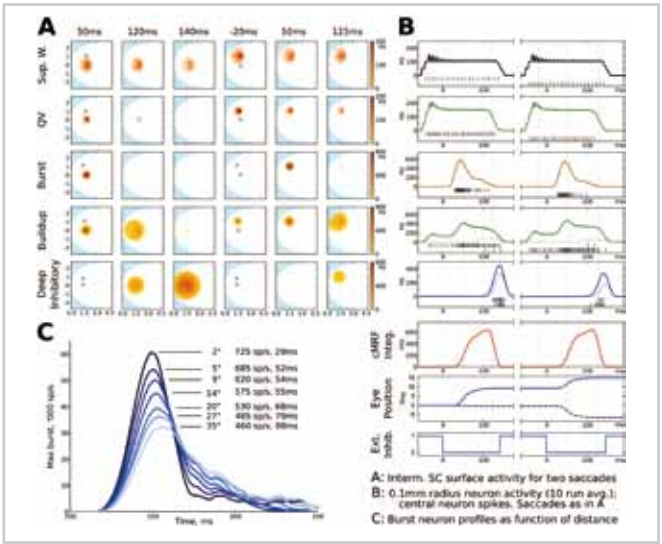


Fig. 2 : A) SGI surface activity over time for a 9° radius, 0° angle, then a 8° radius, 45° angle saccade. Each point is a 10ms activity average of one model neuron. Times relative to saccade deinhibition. From the top, superficial wide-field neurons; quasivisual neurons; burst neurons; buildup neurons and deep inhibitory neurons. From left, burst peak, buildup activity peak, deep inhibitory reset; preparatory activity for second saccade, second burst peak and second buildup peak.

B) 10-trial averaged activity in a 0.1mm radius around the central point for each saccade in A), in the same row order. black bars are spikes of the centermost neuron during a single trial, upper row for first saccade, lower row for second. Dotted lines for each saccade shows activity at the point of the other saccade. Below, total cMRF integrator activity; estimated horizontal (solid) and vertical (hashmarks) eye position; and external inhibition.

C) SGI burst neuron activity over time as a function of saccade distance. Longer saccades have a lower activity peak but a longer duration.

Developing the MD Core Program for Large Scale Parallel Computing

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

Our team has been developing a MD core program for large-scale parallel computing to deliver high-speed core program libraries designed especially for a "K computer", and to understand and develop high-speed technologies for applications using "K computer ".

Molecular Dynamics (MD) simulation is a method for representing molecular activities and changes in structure by calculating the effects and forces acting on the atoms that make up molecules.

In life sciences, this is used to explain the nature of the biomolecule such as protein, and is the foundation of living phenomena. In biomolecule simulations for proteins, covalent bonding, Van der Waals, and Coulombs forces act on the atoms. For covalent bonding, there is a simple classical force field that acts like a harmonic oscillator, in which case sequences of up to four connected atoms are used, and the computational effort is proportional to the number of atoms in the covalent bond. But for Van der Waals and Coulombs, forces act between any atoms, meaning an interaction could exist between every pair of atoms. The computational effort in this case will be proportional to the number of atoms squared. The calculation effort for Van der Waals and Coulomb forces will take up most of the processing time if there are many atoms. The simplest way for speeding up the calculation is to ignore the forces between distant atoms and cut off calculation. Van der Waals forces decrease in inverse proportion to the power of 7 and 13 of the distance between atoms, and ignoring distances of 1.4nm and more will hardly affect calculation precision. Coulomb forces also decrease in inverse proportion, but with the power of 2, and a simple cut-off calculation will affect calculation precision. Therefore, a fast and common way for calculating Coulomb forces between distant atoms is the PME (Particle Mesh Ewald) method which uses FFT. However, FFT requires a large range of communication, and leads to a rise in communication time if the level of parallels is high. Therefore, we are considering using a calculation method that does not use FFT such as the fast multipole method (FFM).

As a large-scale test on cut-off calculation, we performed simulations on how "molecular crowding" affects protein dynamics. In an intracellular environment, proteins exist in a higher density than normal experiments, and it has been verified that the crowding of proteins affects their structure and interaction. To simulate protein crowding, it is necessary to perform simulations of a large number of proteins. Therefore, we have high hopes for "K computer" which is capable of large-scale computation. In a calculation test, we compared the structure of the TTHA1718 protein which is made of approximately 1,000 atoms. This was performed in an underwater environment, and in an environment of egg whites which are 30% ovalbumin proteins (consist of nearly 6,000 atoms). Fig.1 shows the fluctuation in positions of the amino acids that make up TTHA. Although we have not retested due to shortage of calculation time, we were able to confirm differences in TTHA activity in an underwater environment (red line vit1, blue line vit2), and an environment with ovalbumin (green line viv). Fig. 2 is a visualization of TTHA surrounded by ovalbumin.

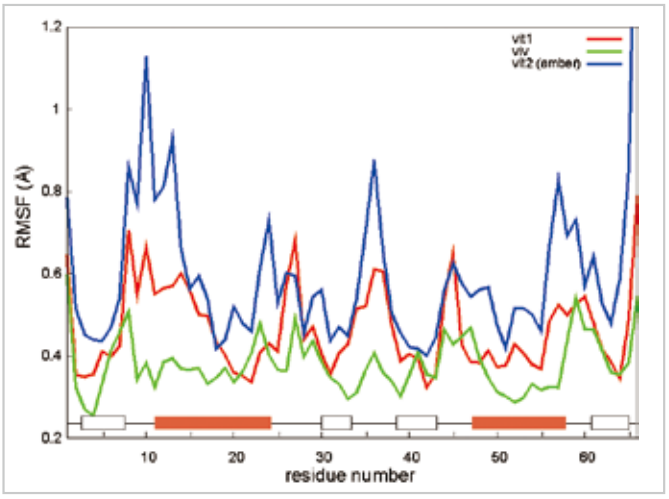


Fig 1 : Fluctuation in TTHA structure. Red and blue are underwater environments, and green is an environment with ovalbumin. The horizontal axis is the number of amino acids, and the vertical axis is the size of fluctuation.

In the performance evaluation of "K computer" (note), we were able to obtain an efficiency increase of 30%, an actual calculation speed of 1.3 PFlops using 1.8 hundred million atoms, and a cut-off calculation of 2.8nm. An efficiency increase of 40% or more has been achieved for the cut-off calculation part.

As of this point, we have implemented a cut-off calculation, the PME method, and FMM. Detailed algorithms of the PME method and FMMs have been refined for optimization. We have implemented a communication method that is designed for "K computer"'s mesh or torus network structure, but in cases where the number of atoms per node is small, the communication time exceeds the calculation time. Therefore, we are continuing improvements for reducing communication time.

Our research is a joint research project with the High-performance Computing Team whose members are Hiroshi Koyama, Gen Masumoto, Aki Hasegawa, Gentaro Morimoto, Noriaki Okimoto, Hidenori Hirano, and Naoyuki Miyashita from the Life System Research Center.

Some of the results of this research have been collected by using the RIEKN Integrated Cluster of Clusters (RICC) system.

(Note) Performance evaluation of a next-generation supercomputer, the petascale computer "K computer", by RIKEN's next-generation supercomputer development implementation division

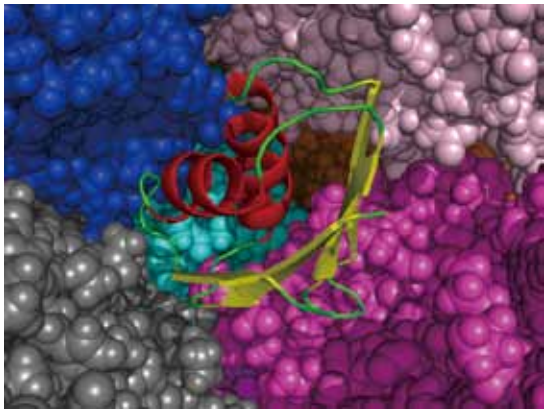


Fig 2 : TTHA surrounded by ovalbumin. The spiral in the center and arrows represent TTHA. The surrounding spherical shapes are ovalbumin atoms, where each different color represents one ovalbumin.

Pioneering the Future of Computational Life Science toward Understanding and Prediction of Complex Life Phenomena


Dr. Kidera, Dr. Yanagida and Mr. Eguchi (from the left)

Program Director of RIKEN HPCI Program for Computational Life Sciences Toshio Yanagida
Deputy- Program Director of RIKEN HPCI Program for Computational Life Sciences Akinori Kidera
Deputy- Program Director of RIKEN HPCI Program for Computational Life Sciences Yukihiro Eguchi

The HPCI Strategic Program, which aims to produce the world's best research products in a wide variety of fields by utilizing Japanese High-Performance Computing Infrastructure (HPCI) centered on K computer and returning their fruits to society, got into full swing this year. Field 1, "Supercomputational Life Science," has four R&D themes. What are the details of the projects? How will the researchers proceed with the themes? Following the preceding issue, we had Director Toshio Yanagida, Deputy Program Director Akinori Kidera, and Deputy Program Director Yukihiro Eguchi talk about that.

Four R&D projects

YANAGIDA (Honorifics omitted)**:** Then, let us start by introducing themes 1 to 4 to explain what kinds of themes we are going to work with in Field 1.

KIDERA: Our goal is to use the maximum performance of the K computer and produce results that contribute to life science. For this purpose, we don't start from scratch but in essence utilize the established "workable computation approaches." In that sense, the four themes are roughly divided into three categories. In Theme 1, "Simulations of biomolecules under cellular environments" and Theme 2 "Simulation applicable to drug design," we do biomolecular simulations utilizing molecular simulation techniques. In Theme 3, "Hierarchical integration simulation for predictive medicine," fluid-structure interaction analysis incorporating fluid dynamics calculation into structure analysis is performed. In Theme 4, "Large-scale analysis for life data," analysis of life data centered on the genome is performed. By taking a more detailed look at each theme, it is seen that the themes are not limited to those mentioned above. For example, Theme 3 includes simulation of the cerebral nervous system. In any case, we use methods that have been developed so far. For example, in biomolecule simulation for Theme 1 and 2, we use existing methods such as quantum chemical calculation, molecular dynamics calculation and coarse grained model calculation, which are scaled up with the K computer.

The reason why molecular simulations are divided into two themes is, as suggested by "Application in Drug Discovery and Medical Development" clearly stated in the title of Field 1, because we want to clearly express that we attach importance to contribution to drug discovery and medical development as a result of the simulation studies. In order to promote drug discovery approaches intensively as an application of molecular simulations, we defined Theme 2. Of course, we proceed not only with Theme 2 but also all themes with contribution to drug discovery and medical development in mind.

EGUCHI: A certain level of model has already been established for Theme 2. Meanwhile, not only scaling up of traditional and conventional models, but also contriving a new model, is required for Theme 1, I think. Don't you think such a challenge is imposed on Theme 1?

KIDERA: Exactly. Theme 2 has established models. A significant result is expected through efficient use of the models. However, it is not the sole method. There is a possibility for Theme 2 to develop into something superb. In this regard, Theme 2 is still a challenge.

EGUCHI: What do you think about Theme 1?

KIDERA: What we have to do in Theme 1 is make maximum use of the K computer at the existing level of molecular dynamics models to improve the level of existing simulations. Then, perform the calculation of a large system for a longer time to replicate molecular-level biological events that have not been observed so far or have finally been elucidated. We have to start off from there. Through such work, we should search for models for the next-generation simulation. That would be the order of things in Theme 1.

YANAGIDA: You can say life systems are a world which is made up of hyper-degrees of freedom. Until now, we have been trying to understand life systems by limiting the degree of freedom to a controllable range with various hypotheses. Of course, it is not possible that you only have to make the K computer calculate a system with hyper- degrees of freedom to get instant results. Nevertheless, we have to construct models by trial and error. When we return the model to hyper-degrees of freedom, we perform simulation with a computer and get predictions. Therefore, simulation should be used not only for explaining the phenomena observed so far but also for leading to the next prediction. That is why we have to do modeling. For this reason, it is important to combine experiments, model and simulation.

KIDERA: Actually, it is very difficult to combine modeling and prediction. The model must be predictable because we cannot move forward only by explaining previous data. For that, we should do ab initio (non-empirical) simulation in which as high a degree of freedom as possible is built in. In the results obtained, we should find phenomena that can be expressed with a new low degree of freedom and improve the model. Of course, we can use experiment itself to improve the model.

Exploratory challenges focusing on the final goal are also important.

EGUCHI: Theme 3 is also very challenging.

KIDERA: Theme 3 is a project focusing on interaction and integration. For example, in traditional blood flow simulations, blood vessels and the

heart are viewed as structure, and blood flow as fluid. In our simulation, structure and fluid are coupled together. In addition, we are going to combine heart and systemic vasculature simulators, which have been simulated separately so far. On top of that, we introduce totally different simulators such as systemic skeletal musculature and the cerebral nervous system, and link them together for doing human whole-body simulation. We also strongly emphasize medical support, and clearly set forth simulations that contribute to prediction of disease state, review of treatment method and evaluation of drug efficacy as our goal.

EGUCHI: Some researchers may doubt that integration can be achieved successfully.

YANAGIDA: However, from the standpoint of clinicians, this is one type of research they definitely want us to do. To tell the truth, I was a professor of physiology at Osaka University. In physiology, that research was so far done in a classical manner. If that research is combined with computer science and progresses further, I think it would have a great impact on medical science. Although it will be very challenging, I believe biology must aim at this. It is of course important to get results with the K computer, but it is also important to consider what our final goal is.

KIDERA: It is a fair argument that we have to define a clear-cut, achievable goal. However, there is another concept. Are we allowed to target only what would be visible by scaling up with the K computer? We want to carry out challenging trials having more heuristic elements with the K computer. We want to start research towards a real goal with the K computer. Such a desire is also very important. I believe the attempt to conduct an integrated whole-body simulation with multi-scale multi-physics simulation in Theme 3 is crucial in the sense of exploratory challenges that precisely cultivate such budding ideas for the future.

Contribution to medical treatment by life data analysis

EGUCHI: Themes 1 to 3 partially overlap one another, but Theme 4 is relatively distinct from the others as it handles data analysis.

YANAGIDA: Theme 4 aims to understand life programs and their diversity through large-scale analysis of life data based on the genome, and utilizing them for individualized medicine and industrial use of genome information.

KIDERA: Various new experimental technologies produce a huge amount of life information. By analyzing it, we have to understand what it means and step into the field of prediction. Theme 4 is a project using the K computer to analyze such life information. What is the most important part and one we have focused on ahead of the others is the large-scale genome information produced by the next-generation sequencer. We have to analyze life system data which is becoming larger and rapidly more refined. However, now, it is too much to handle with conventional computer resources. Therefore, we are seeking a basis to analyze life system data with the K computer. This theme is directly linked to experimental data, and we believe we can realize a data-analytical science as part of new life sciences with the K computer.

EGUCHI: Prof. Satoru Miyano, who is head of Theme 4, emphasizes that what can be explained by first-principle equations is very limited in life science, and most phenomena do not have a clear-cut first principle. I believe he thinks this way: in order to view life science as science, we have to let the data speak. That is, we have to strive to find some regularity in the huge amount of data produced. He said he wanted to focus on preparing a system that can analyze data obtained with a next-generation sequencer efficiently on the K computer for the first couple of years.

KIDERA: There is an extremely strong demand for that at least. This is one reason. In addition, if such a system does not work on the K computer sufficiently well, we would not be able to perform further large-scale life data analysis on a full scale. With the development of a sequencer, we will certainly have an increasing number of data for genome analysis.

YANAGIDA: First, we have to organize the data. Then, from this enormous amount of data, we extract meaningful information, for example information leading to prediction of a disease. For this purpose, larger computer resources are required. We can say we have to do data mining.

EGUCHI: On the other hand, in the study on cancer metastasis Prof. Miyano released recently, he found a gene suspected to be involved in metastasis through gene network analysis, and was able to demonstrate its involvement through experiments. Those approaches which are not limited to prediction will become important.

We aim at heuristic simulations

KIDERA: We have viewed the four R&D themes one by one. Now we have to conceive the impact of K computer on whole projects. As an orthodox research and development concept in the HPCI Strategic Program, research results are required which can be obtained only by computation-scale expansion with the K computer. If we only perform very large calculations and it does not lead to the fruits of life science that were our original target, it is not the goal or result we are seeking.

YANAGIDA: Yes. If we have to continue symbolic research just to upkeep the K computer, it is exactly like putting the cart before the horse.

KIDERA: It is still difficult to say that computer science has become widespread in the field of life science. I'm sorry to say it, but there is no prototype of life science research using the K computer which is a high-end computer resource. In circumstances such as these, we should not use the K computer simply as we are told, and get results anyway. Eventually this project should be evaluated by how much genuine scientific results can be obtained. I think that is what is important.

YANAGIDA: As we said before, we have been trying to understand life systems with a hyper-degree of freedom by limiting the degree of freedom to a controllable level. It might be a good idea to put higher degrees of freedom into the K computer anyway. By doing so, the key degree of freedom may surface.

KIDERA: The K computer may make it possible to jump from a simulation without any conflicts to the area of "heuristic simulation" in which something that has not been expected is discovered in the results, right?

YANAGIDA: I like the phrase, "heuristic simulation." We aim at heuristic simulation in life systems with a hyper-degree of freedom. We can summarize our discussion on that note.

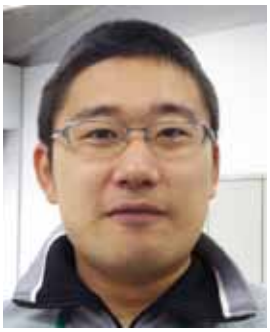
KIDERA: We also have to remember that we are trying to achieve it with the K computer. Something very positive is going to happen in biology. We should be thankful that we are lucky enough to witness it, and keep trying our best to reach our goal.

Free Energy Profile Calculations for Changes in Nucleosome Positioning with All-Atom Model Simulations

Quantum Beam Science Directorate, Japan Atomic Energy Agency

(From above) Hidetoshi Kono, Hisashi Ishida, Yoshiteru Yonetani, Kimiyoshi Ikebe

(Field 1 - Program 1)



Human genes are encoded by a genome DNA which consists of approximately 3billion base pairs and approximately 2meters long. The genome DNA of eukaryotes including that of humans, forms chromatin structures and is stored compactly in the cell nucleus whose diameter is approximately several μmeters long. The fundamental, repeated structure unit of chromatin is called nucleosome, in which about 150 base-pair DNA is wrapped nearly twice around a protein core called a histone. Recently it has been found that structural changes in the genomic DNA play key roles in the regulation of gene expression. While our genes are protected from being damaged and are stabilized by forming the nucleosome structure, during DNA metabolism which is the basis of vital functions such as transcription, duplication, repair and recombination, disruption of the nucleosome structure and changes in nucleosome positioning occur so that the DNA can directly interplay with regulatory proteins and RNA polymerases (Fig.1). The nucleosome structure has such seemingly-contradictory functions. Our interest is understanding the mechanism which brings about the disruption of the nucleosome structure and changes in nucleosome positioning.

In addition, recent studies have revealed that the changes in nucleosome structure affect the cell differentiation mechanism. The individual cells that constitute our body essentially have the same genetic information (genome DNA). However, once a cell becomes differentiated, it usually only generates cells of the same type; a skin cell generates only skin cells and a liver cell only liver cells. Furthermore, recent research has revealed that the mechanisms of DNA metabolism and the cell differentiation are closely related to chemical modification of histone proteins that are a constituent of the nucleosome structure. In this way, changes in gene expression or cellular phenotype can be inherited as a kind of memory without accompanying any sequence change in the genome DNA, which is called epigenetics. This is one of the most active research areas in molecular biology nowadays.

In our study, we elucidate the mechanism of changes in nucleosome positioning from the aspect of the free energy profile by use of computer simulation. By examining how the difference in DNA base sequence, chemical modification of DNA, histone proteins and their variants affect the free energy profile, we pursue research on epigenetics. We have developed a computer program called SCUBA that is suitable for large-scale molecular dynamics simulations. We are among the first to introduce an algorithm in SCUBA which can divide the target system in space and enables us to efficiently compute it in parallel. With SCUBA, we have executed a dynamics analysis of Holliday Junction branch migration, which is observed in DNA homologous recombination, and a ribosome molecule in a system composed of over 2million atoms. As shown in Fig.2, it is known that nucleosomes spontaneously undergo slow conformational fluctuations whereby a partial unwrapping of DNA which was wrapped around the histone protein (opening) occurs due to thermal fluctuation, as if they were breathing. In our project, by using "K computer" we are aiming to clarify how such a dynamics and structural stability of nucleosomes are changed by chemical modification of the DNA and the histone proteins, compositional changes in histone variants, and differences in the base sequence of wrapped DNA. Since such a dynamics occurs in the order of sub-seconds, even the world's fastest "K computer" cannot simulate for long enough to track the dynamics by simply waiting for thermal fluctuation phenomena to occur during computation. Therefore, we set an appropriate reaction coordinate, and induce a structural change along the coordinate to compute the free energy profile for the structural change. Through this molecular simulation approach, we intend to physic-chemically understand the basic molecular mechanism of living processes, or how gene expression is regulated.

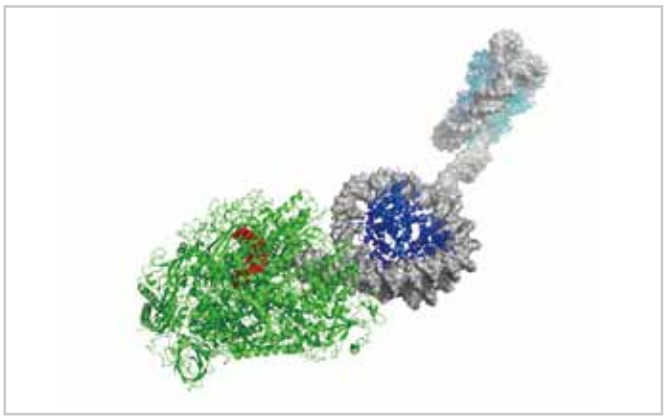


Fig 1 : Structural view of mRNA synthesis by RNA polymerase. How does RNA polymerase carry out transcription process against genome DNA that forms a nucleosome structure? We approach the problem through molecular simulation. RNA pol II: green; synthesized RNA: red; DNA: white; histone: blue and cyan.

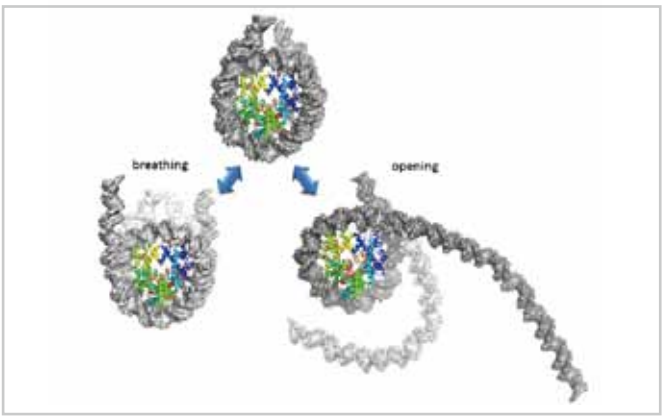


Fig 2 : We perform an analysis of the free energy profile for the structural change of nucleosomes shown in the figure through molecular simulation. It is experimentally suggested that partial unwrapping and rewrapping of DNA occurs spontaneously as if it were breathing.

Estimation of Skeletal Muscle Activity and Neural Model of Spinal Cord Reflex

Information Science and Technology, The University of Tokyo
Yoshihiko Nakamura
(field1- Program3)

We have been developing an "Inverse system" [1] , which is a technology for estimating information in the human body from dynamics and feedback [2][3] based on an efficient computation of a rigid body system that was developed in the field of robotics. That is, we are trying to take a look at the information in the body by extraneous measurement and computation.

We conducted a project known as CREST, "Creating the Brain" (Research Supervisor: Prof. Shunichi Amari), from 1998 to 2003, and started research to generate and control a body movement pattern of humanoid robots by attracting symbolic nonlinear dynamics. Meanwhile, we developed a parallel computing method called Assembly and Disassembly Algorithm, in which O (log N) is usable for calculation, as a direct dynamics computing method [4], and we took a cue from that and started to compute human systemic skeletal musculature as a physics model. In the model, 997 wires are used to link a systemic skeletal system with 155 degrees of freedom [5]. It became one of the major projects through two Grants-in-Aid for Scientific Research (S): "Deployment of Dynamics-Based Information Processing Model of Intelligence" (2003 to 2007), and "Establishing Human-Machine Communication through Kinesiology and Linguistic Integration" (2008 to 2012).

Estimation starts with calculating the force applied to the skeleton with mass (the entire mass of the human body is allocated to the skeleton) from geometric information about motion, which is measured by motion capture, by a calculation scheme of inverse dynamics. The next computation is back calculation of systemic muscular tension expressed in the wire model without mass. There is an inequality constraint that the wire tension shall be positive or zero. The solution of a problem having high redundancy, $155 < 997$, is explored as a secondary programming problem. In order to take the activity of antagonist muscles into account, wireless electromyography is attached to the body during motion capture, and muscle activities of up to 32 channels are measured. In the optimization problem, the secondary programming problem is solved using examples from the activity (Fig. 1).

Then, our interest was directed towards the deeper part of the body and construction of an inverse model for estimating nervous activity from muscle activity, and we are now studying the motor portion of the nervous system from the spinal cord toward the periphery. The spinal nerve consists of total 31 pairs of peripheral nerves emerging from the spinal cord (8 pairs of cervical nerves, 12 pairs of thoracic spinal nerves, 5 pairs of lumbar nerves, 5 pairs of sacral nerves and a pair of coccygeal nerves). As each nerve branches into an anterior branch and dorsal branch, the spinal nerve branches into 124 nerve fascicles. The function of skeletal muscle, to which each nerve fascicle projects, was clarified. The 124 nerve fascicles are supposed to represent the same number of signal pathways for estimating their activities. We placed a totally-coupled mathematical neural network in the spinal cord, and made it learn by activity of the nerve fascicles and voluntary signals (in which various hypotheses are used) descending from the brain. The spinal neural network obtained this way is regard as an optimized nerve system specialized in the anatomical body model and movement pattern typically found in humans, and represents the reflex arc by the spinal cord (Fig. 2).

Strategic Program, Program 3: In a hierarchical integrated simulation for predictive medicine (represented by Shu Takagi), a high-dimensional systemic neuromusculoskeletal system and its parallel computation algorithms are being developed by creating a model, in which the mass is dispersed among muscles and viscera, and computing the stress distribution of contraction and contact associated with elastic deformation. By integrating the multi-scale model from the muscle fibers of skeletal muscle studied by the Takagi Group and the basal ganglion model by the Doya Group, we are now aiming, for example, to duplicate a mechanism for altered dopamine metabolism in basal ganglions in Parkinson disease as altered motion in walking and reaching. In this way, we are going to establish basic computational technologies for predictive medicine helpful in drug discovery, diagnosis and treatment.

Fig. 1: Visualization of muscle activity

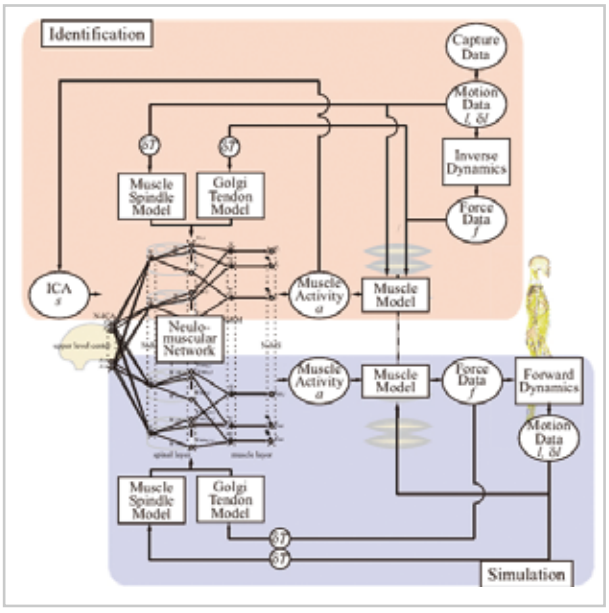


Fig 2 : Composition of spinal cord reflex model

[References]
[1] M. Kaneko and T. Kodama, Inverse system – Disclosure of the mechanisms of market and life, Iwanami Shinsho (New red cover 875), 2004
[2] A. Murai, K Yamane, and Y. Nakamura. Modeling and identification of human neuromusculoskeletal network based on biomechanical property of muscle. Proceedings of the 30th IEEE EMBS Annual International Conference, Vancouver, Canada, 2008.
[3] A. Murai, K Yamane, and Y. Nakamura. Effects of nerve signal transmission delay in somatosensory reflex modeling based on inverse dynamics and optimization. Proceedings of the 32nd IEEE International Conference on Robotics and Automation, pp. 5076–5083, 2010.
[4] K. Yamane and Y. Nakamura. "Efficient Parallel Dynamics Computation of Human Figures". Proceedings of IEEE International Conference on Robotics and Automation (ICRA2002), 1:530–537, Washington D.C., May, 2002.
[5] Y. Nakamura, K. Yamane, Y. Fujita, and I. Suzuki. Somatosensory computation for man-machine interface from motion capture data and musculoskeletal human model. IEEE Transactions on Robotics, 21:58–66, 2005.

ISLiM Interim Accomplishment Meeting in 2011

Computational Science Research Program, RIKEN
Eietsu Tamura

The Next-Generation Integrated Simulation of Living Matter (ISLiM) program held an Interim Accomplishment Meeting in 2011 at Takeda Hall in the University of Tokyo on December 21 (Wed) and 22 (Thu), 2011.

ISLiM program was launched in October 2006 along with the development of the K computer, with the aim to understand life phenomena more deeply and to contribute to drug discovery and healthcare in particular. It remains one year to go before completion of this program. The researchers of 14 universities and institutes have been making collaborative efforts to develop ISLiM software and optimize them for the K computer. ISLIM develops 33 homegrown application programs, and challenges the most efficient, large-scale, and highly accurate simulation with K computer. A total of 141 researchers specializing in life sciences and healthcare from industry and academia attended the ISLiM Interim Accomplishment Meeting in 2011, where ISLiM software developers reported their latest results for two days.

First, Koji Kaya, Director of the Computational Science Research Program, gave an opening speech. Then, Ryutaro Himeno, Deputy Program Director of the program, explained the current status for development of application software, each of which has been optimized by using the K computer in trial operation from March or April 2011, and showed interim performance data. Himeno said, "Our efforts will be focused on optimization of ISLiM software until November 2012 when K computer becomes generally available. The performances for some software products (cppmd, ZZ-EFSI) have reached to 20-30% of the peak performance of the K computer. Further efforts will be directed to improvement of performance and to dissemination of the software."

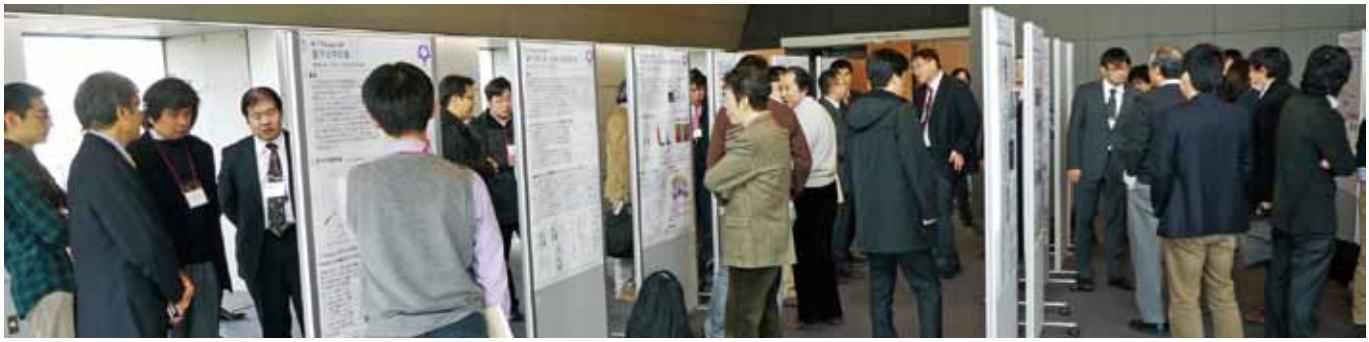
Afterwards, the leaders of the six ISLiM teams (Molecular Scale Team, Cell Scale Team, Organ and Body Scale Team, Data Analysis Fusion Team, Brain and Neural Systems Team, and High-performance Computing Team) and ISLiM's researchers in charge of development of ISLiM software reported their latest achievements. The details of their reports (mostly in Japanese) can be downloaded from the website of ISLiM (URL: http://www.csrp.riken.jp/2011/islim-houkokukai2011_j.html).

Because two days are too short to present all the ISLiM developed software in depth, four-poster sessions were held in the foyer at Takeda Hall. A total of 38 posters were exhibited there, and the software developers gave explanations to audiences from industry, etc. Some of them had deep discussions in front of the posters.

Participants from industry appreciated the presentations in the meeting, saying that they understood how widely and deeply ISLiM researchers were involved in research and development of ISLiM software, even in vivo study. But, such comments suggest that ISLiM software has been little recognized by people in our target area outside the project. Then, we will organize an ISLiM Accomplish Meeting in the fall of 2012, and foster programs that can improve visibility of ISLiM software as well as dissemination.

Acknowledgment

The performance data with the K computer are obtained through the experimental use since April 2011 or its special operation in March 2011 before general availability in November 2012.



Computational Life Sciences Classes Held in High Schools

HPCI Program for Computational Life Sciences, RIKEN
Chisa Kamada, Yasuhiro Fujihara, Yukihiro Eguchi

Last December, we held a class named "What is the K computer? Will it change the face of biology?" at Kobe High School and Nishiwaki High School in Hyogo prefecture. We worked on it hoping that many young high school students would participate in computational life sciences and contribute to further development of computational life sciences even after the limited five years of this program. At the opening of the class, we provided three main points that we wanted to talk about in the class:

1) The K computer has become the fastest computer in the world. The key technology that makes it the fastest computer is parallel computation.

2) You also need mathematical principles (mathematics, physics, and chemistry) to study biology.

3) The K computer is a catalyst for integrating biology with mathematical principles.

In May, when we asked the teachers at Kobe High School about the class, they said "please don't give us a lecture using only slides." So, we tried hard to design a class with two practical training sessions. A student commented, "I was glad I had some practical training and I really enjoyed it", which made all our efforts worthwhile.

About 60 students from Kobe High School and about 90 students from Nishiwaki High School joined our classes. First we explained the details of the K computer, and introduced five challenges we are working on in our strategic field. Then, we had practical training sessions for binding and editing a base sequence and for forming peptide bonds, to introduce HPCI Strategic Program Field 1 "Computational Life Science and Application in Drug Discovery and Medical Development" We provided students with eleven pairs of fragments that have about fifteen base pairs, and let them actually try binding and editing. Some students finished in about five minutes, while others were not able to finish even after ten minutes, creating a lively atmosphere in the class. One student commented, "I and my partner tried to bind a base sequence. Despite the small number of base pairs, we had difficulty in finding the correct answer. It's amazing that a computer can bind a much longer base sequence in a moment of time without making any mistakes." We provided all students with molecular models of alanine and the phenylalanine, and let them try to form a peptide bond using dehydration synthesis. This time, the students worked quietly and seriously, and they seemed to be very intent on what they were doing. After we visually showed them the result of a molecular dynamics simulation of a protein following the practical training session for forming a peptide bond, one student said with emotion: "I got a creepy feeling when I watched the moving protein, but I thought it was great when I knew this phenomenon actually occurs in my body." We were surprised and grateful that the students took an interest in this type of analogous practical training, because we thought they were only interested in mobile games, etc.

The HPCI Strategic Program involves the research and development team, as well as a team for promoting supercomputational science and technology. The mission of the team for promoting the technology is to "support the research and development team on achieving their research and development target." The support includes enhancement of massive parallel computation, as well as establishing a framework that allows the research and development team to be supported by society as a whole. The class for high school students is a part of the latter activity. We, as those who were involved with the class, are very happy to receive comments from students who joined the class, such as: "I thought the K computer was just a high performance computer. Now, in this class, I understand the K computer is utilized in various fields, such as medical services, disaster prevention, next-generation manufacturing, and creation of new material and energy. I want to be involved with this kind of work in the future," or "Until now, I thought the K computer was just a computer for performing calculations. I was surprised to hear it relates to biology and physics, and so I now have a greater interest." We would like to hold this kind of class again in the future, as we were requested.

Finally, we would like to thank Dr. Takano of Osaka University for kindly accepting to hold the class in Kobe High School, HGS Hinomoto Plastics Co., Ltd. for kindly accepting to manufacture a special molecular model for this class, and Ms. Namba for preparing the required number of base sequence puzzles for the students who joined the class. We would also like to express our deep appreciation to Mr. Nagasaka of Kobe High School and to Mr. Fujiwara of Nishiwaki High School for giving us these great opportunities.

Event Information

Nano-Life Public Symposium
Next Generation Integrated Nanoscience Simulation Software (Nano)
Next-Generation Integrated Simulation of Living Matter Software (Life)

- Date : March 5 (Mon), 2012, 10:00 - 18:20 (joint sessions and get-together)
  March 6 (Tue), 2012, 9:00 - 17:55 (parallel sessions)
- Location : Nichii Gakkan Kobe Port Island Center (Kobe City) 7-1-5 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047
- Conference Fee : No charge (except get-together fee)

Program (TBA)

Day One / March 5 (Mon): 10:00 - 18:20 Integrated Nano-Life Joint Sessions (18:30 Get-Together)

Life Molecular Scale Team / Integrated Nanoscience Biomaterial Joint Session
Special Lecture Hiroki Ito (Fujitsu)
Joint Panel Discussion "The K Computer and What Lies Ahead (TBD)"

Day Two / March 6 (Tue): 9:00 - 17:55 Parallel Sessions and Poster Session

Nano Parallel Session

9:00 - 9:55	Integrated Nanoscience Main Report
10:10 -12:10	Next-Generation Functional Nanomaterials for Information Technology
13:40 -15:40	Joint Poster Session
15:50 -17:50	Next-Generation Energy

Life Parallel Session

10:00 -12:00	Panel Discussion "Life Revealed by HPC"
13:40 -15:40	Joint Poster Session
15:50 -17:50	Panel Discussion "ISLiM's Contribution to Drug Discovery and Medical Development"

For details, please refer to the web page: http://nanogc.ims.ac.jp/nanogc/sympo2012

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

Ministry of Education, Culture, Sports, Science and Technology Project for high-degree application of high performance general-purpose computer

HPCI Strategic Program Field 1

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, 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 a ten quadrillion speed computer known as "K computer".

Supercomputational Life Science

The HPCI Strategic Program is a program of the Ministry of Education, Culture, Sports, Science and Technology aiming to produce ground-breaking results in computer science technology by maximizing the benefits of the HPCI (High Performance Computing Infrastructure) centered on the K computer, and encouraging developments in five research fields that need to be strategically addressed.
"Supercomputational Life Science" as a representative organization of RIKEN, 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.6 2012.3

Issued: March 2012

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