Strategic Programs for Innovative Research Field 1 Supercomputational Life Science

BioSupercomputing Newsletter

2016.3 Vol.14

CONTENTS

SCLS SUPERCOMPUTATIONAL LIFE SCIENCE

www.scls.riken.jp/

Strategic Programs for Innovative Research Field 1, special discussion meeting

Reflecting on the five-year activities of "Supercomputational Life Science"

Computational life science carving out a new era of life science

The Strategic Programs for Innovative Research Field 1 "Supercomputational Life Science" (SCLS), which aim to produce the world's leading research products in the field of life science by utilizing Japanese High-Performance Computing Infrastructures (HPCI) centered on the K computer and got into full swing in FY2011, will come to an end at the end of FY2015. We interviewed Dr. Toshio Yanagida, Dr. Akinori Kidera and Mr. Yukihiro Eguchi, who have significantly contributed to the promotion of the project, to find out the achievements they attained in the course of research activities for the purpose of exploring the potentials of computational life science for forecasting and controlling biological phenomena, and applying such achievements to medical care and drug development, as well as the perspectives they obtained for future development of computational life science.

Is computational science useful after all?

— How did you feel when SCLS started?

Yanagida : Actually, when I was appointed as Director out of the blue, I felt a little embarrassed, asking myself why I was appointed as Director in spite of my career as a specialist of single-molecule measurement, and the fact that I had never been professional in computers. I remember that, in my first speech, I said: "I doubt if computational science will ever be so beneficial to life science" (laugh).

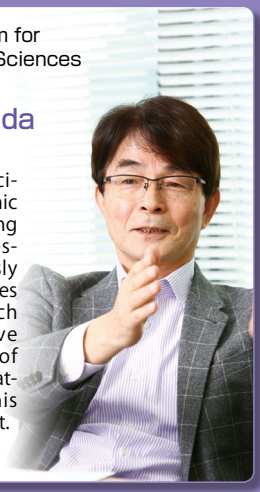
Kidera : Yes, I remember that (laugh).

Yanagida : I think I made such a speech because I had an impression that something had been achieved in the field of life science without any deep knowledge, and I felt no need to know complex processes. However, as a matter of common sense, life scientists were empirically thinking that it was time for a change, and obliged to ask themselves what they should do to overcome such a situation. Meanwhile, measurement technologies such as genetic analyses did not stop advancing, and the complexity of data was increasing beyond our understanding. Thus, as the momentum for recognizing the importance of computing increased, I gradually started to gain an understanding.

But when we started, we recognized the sheer lack of computer power. In spite of that, when the members started to make efforts to use the K computer, we recognized the potential of the K computer for giving birth to a science which is different in kind, and the Director changed his mind and got earnest about promoting our efforts. All this happened thanks to the efforts made by those in computational science.

Kidera : The first speech of Dr. Yanagida was so impressive and striking. If I remember correctly, I said: "Please

RIKEN HPCI Program for Computational Life Sciences
Director
Toshio Yanagida
(RIKEN)

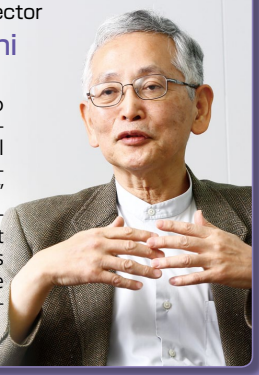


Computational life science is a new academic field. To keep it growing and maturing, it is essential to continuously expand the major bases and conduct outreach activities. And above all, the enthusiasm of researchers for cultivating and fostering this academic field is a must.

RIKEN HPCI Program for Computational Life Sciences
Deputy-Program Director
Akinori Kidera
(RIKEN, Yokohama City University)

With expansion of computational scope, one future challenge is to provide interfaces which can be readily used by many life scientists, rather than developing computers requiring a wider range of expertise.

RIKEN HPCI Program for Computational Life Sciences
Deputy-Program Director
Yukihiro Eguchi
(RIKEN)



I think that the key to broadening the horizon of computational life science is "a successful experience." Once "a success experience" like the heart simulator "UT-Heart" is created, studies in the field will be accelerated significantly.

don't say something which would discourage the researchers" after his speech (laugh). His statement was so striking that I couldn't help giving him a warning. Why? Because his statement was right on the mark. In a word, at that time, computational science did not offer any results which would truly contribute to life science. SCLS was preceded by the "Next-Generation Integrated Simulation of Living Matter (ISLiM)," in which we basically conducted software development. Although its achievement forms the basis of this project, we were engaged only in software development but not allowed to use the K computer, so we did not get to attain any achievements which would convince life scientists. Besides, no clear or specific attention was directed toward the kinds of targets, calculation methods or desired achievements. So, as to the achievement of ISLiM, we could only say "We succeeded in developing general-purpose software. It's amazing, isn't it?" As we knew that, we did not want Dr. Yanagida to make a statement which would reiterate that. However, once we started to use the K computer, we marveled at its potential, and found our specific targets and what we should see, while the research system to achieve them was prepared. As researchers knew how they should use the K computer to do computations, they became experienced to some extent and confident, and finally were no longer afraid of what Dr. Yanagida says. That was when the project was halfway finished. At last, the image of life science based on the supercomputer was emerging, and achievements were being made little by little. After that, research activities progressed more rapidly by gaining more momentum. We have had five years so far.

Yanagida： In that sense, I succeeded in my first speech (laugh).

Kidera： Let alone succeeding… Anyway, the strongest impact was your first speech and various organizational changes at the intermediate stage. That is, you stressed the importance of thinking how to collaborate with life scientists who are outside the field of computational science. I believe that the project got rolling smoothly only because of such efforts. Before then, though researchers were not withdrawn into the shell of computational science, they succumbed to a vague and indefinable sense of understanding by creating an excellent piece of software without definitive products. But now, it is different.

Yanagida： No, actually, I said more gently: "Computers are interesting, but they will be more interesting if you apply them to understand the mechanism of life together with life scientists" (laugh). In this way, a new science was born, which was different in kind.

Eguchi： As Dr. Yanagida said at the beginning of the project, unlike other fields, only little-known researchers use computers in the field of life science. In a word, this is still a minor field quantitatively as well as qualitatively. The researchers were not sure about what they should do in this situation. That has been true. But, I wonder if it is still true even now. What should we do under such circumstances? In my opinion, the most important thing is to connect those researchers involved in research activities using the K computer with the research sites of universities and private companies. I always worked with the hope that those universities and industries approach and see the researchers who use the K computer and discover its usability, potential for drug development or feasibility in medical settings, and they consider to use it by seeing it working. I started like that. Specifically, one successful project used "MP-CAFEE" developed by Dr. Fujitani. At first I consulted Dr. Kidera, and he answered: "I think that 'MP-CAFEE' would be friendly to everyone. The others are too professional and unfriendly to laypeople." Consequently, a group was formed for industries based on "MP-CAFEE." At first, it was joined only by two pharmaceutical companies. But now, it has been joined by more than 20 companies. Maybe, such a "connection" between university researchers is very important, and I believe it is essential for outreach activities.

Excellent achievements in each of four Themes

— I would like to ask what kinds of achievements have been made during the past five years with regard to the four Themes of SCLS.

Kidera：In each of the four Themes, achievements have been made, which are slightly different from each other. In the ISLiM, the need for software development was first advocated. However, what was required actually resulted in achievements which would be significant to life science. The most significant theme at that time was the interconnection of the hierarchical layers of life. Life consists of spatially small to large components, that is, molecules to cells, organs and the entire body, and temporally short to long phenomena on a tremendously wide scale. We were required to make their interconnections understandable in detail. I truly felt it would be impossible, as the K computer was still unavailable. But the only way was to keep trying. After the start of SCLS, calculations on the K computer level were made possible for the first time, and the way to connect the hierarchy of life began to open up, although it was only in its beginning. For example, for the molecular-level research of Theme 1, efforts were made to handle the biological phenomena of cell scales containing a lot of proteins and biomolecules, and a simulation, which may be the first cell model accurately reproducing the behaviors of proteins and so forth in cells, is about to be developed.

Theme 3 is a challenge to the hierarchical connection on a larger scale. In this Theme, the heart simulator called "UT-Heart" was developed by miraculously establishing a connection from the molecular scale to the organ, heart. This is truly a marvelous achievement. Needless to say, if molecules are handled at the same level as Theme 1, even the K computer's capacity cannot cope with them. However, by adopting the same spirit, a model representing the behaviors of upper cardiac muscle cells was successfully developed, resulting in a very accurate and sophisticated whole heart model. Furthermore, Theme 3 tackled a more challenging theme, that is, to connect the brain, nerves, muscles and bones, and built an entire body model, which is now about to be put into practice. For Theme 3, I believe that not only the hierarchical connection was realized based on a simulator consisting of diverse algorithms on small to large scales, but also an integrated hierarchical simulation was realized. By making full use of the K computer, the answer to ISLiM's issue, that is, hierarchical connection, is sufficient. As a result, we are finally getting ready to respond to various questions concerning life science.

As Dr. Yanagida told us at the beginning, in the field of traditional life science, connections including the hierarchical one were forcibly built in many ways. The genotype and phenotype are good examples. In Theme 4, large-scale computations were carried out by using the K computer for analyzing a huge quantity of data to thoroughly research them and find out their relationships, focusing on the fact that the phenotype is more complicated. Such large-scale life data analysis is now about to enable a connection between the genotype and the phenotype in a real sense, and clarification of their complicated system. Not only has the data analysis become large-scaled. The phenotype is information on diseases from clinical sites, while the genotype is information on patients of those diseases. Efforts are being made to build a system by gaining the trust of medical doctors on site in that diseases might be understood by connecting the genotype and the phenotype, and to enable their connection in the true sense. This is truly the most marvelous thing. It is highly valuable that an infrastructure was built on which information obtained by the next-generation sequencer in clinical practice is being utilized as truly useful information.

Theme 2 was built around Dr. Hideaki Fujitani, who was invited to respond to direct social demand: "What is the supercomputer used for? Achieve results!" I was surprised by him at first, who responded to the question "Can the K computer develop a drug?" by saying "It's worth trying." I was secretly wondering if he was really sure. However, contrary to my expectation, he ultimately achieved three outcomes which led to the preclinical stage including an antibody. Of course, I believe he gained much support from various parties, but he successfully demonstrated the fact that such a variety of results can be obtained by use of the K computer for a limited time. This is a great achievement. His attitude in pursuit of maximum efficiency also greatly stimulated this strategic field. The researchers involved in the project including us had much to learn from him.

Life science led by computational science

Yanagida：Above all, I would like to stress the fact that the computer power of the K computer made the achievement of such outcomes possible. Some may think that, even by using a low-performance computer, such outcomes could have been achieved with more time. That is not true. It is evident that every Theme has tackled issues which could not have been solved without the computer power of the K computer. As I said earlier, the project demonstrated that computational science was changing the quality of life science. I think that is the important point.

Kidera：For the hierarchical connection, for example, we actually want an infinite computer capacity. However, the capacity is limited. Researchers are always thinking what to do within such limitations. The computation method depends on the upper limit. However, with the advent of the K computer, computational flexibility, as well as the range of potentials to be pursued, increased. I think this

is truly significant. In future, we will have much more opportunities. The K computer is, so to speak, a prototype which will bring such opportunities.

Eguchi : I have a feeling that these four research groups are fundamentally changing the concept of life science. In recent years, we have witnessed astounding progress in measurement technologies, and we can now obtain a huge amount of data at one time. As a consequence, there is growing demand for a method to describe a phenomenon in an integral manner based on a large amount of data, as in the case of Theme 4, instead of conventional life science in which attempts are made to understand a phenomenon by focusing on a certain part of the data obtained. The K computer responds to such demand. I am confident that unprecedented research styles which were not realized before will emerge one after another.

Yanagida : Quite a few life scientists still believe that computers are a kind of tool which support and demonstrate their experimental results and tentative theories. However, computational science is not just to support life science. I hope they understand that computational science is pioneering life science. It is in a position to lead data-driven life science.

To this end, I hope that those with potential for understanding the most advanced software programs tackle life science, and make full use of the K computer to explore new data-driven life science fields. Computational science is already a part of life science.

Toward an era where life scientists use computers perfectly

— How do you plan to build on the achievements obtained in the past five years?

Kidera : I think that research activities and their outcome using the K computer will sprout forth buds when computer resources increase further in future. We successfully built a solid model for conducting research activities, that is, how to accept a large amount of information, how to handle it, and what outcome should be achieved. I hope that more researchers will participate in this field.

Yanagida : What was achieved by the K computer is not necessarily sufficient. If a computer equipped with a computer power 100 times greater than the K computer is developed, we will want to do the computations to connect hierarchies further, and to solve more general problems. For example, we may want to understand how the heart works as part of the entire physical system, or how it is coordinated with mental status. I was a university physiology professor, and the most important area for medicine is physiology. For example, we want to know the physiological conditions of the entire body, such as how the mental status controls the immune response. I am confident that computational life science will pave the way for the understanding of this aspect, too. This will revolutionize the fields of life science and medicine.

Eguchi : I think it is measurement technologies and computers that are going to change the field of life science. However, even if a computer increases its capacity 100 times, high-precision calculation would not be possible without the development of measurement technologies. Measurement technologies also have to evolve in response to the development of computational science.

Yanagida : You are right. Modern measurement technologies do not measure a complex thing as is. Under this hypothesis, a target phenomenon is focused on, simplified, and then measured. So, around 99% of data is discarded, as only the data satisfying the hypothesis is left. A measurer will provide 100 times more data than at present provided that any measured parameter of the entire body can be evaluated by the computer. Also, with such a system, a measurer will be able to provide required data accordingly and enter it into the computer for subsequent computations. It will also become possible to survey the reaction after the introduction of a drug in detail by focusing on the target through computation and measurement. Measurement methods will change, and the quality of life science will also change, which will drive a paradigm shift.

Kidera : In my opinion, data assimilation, which enhances simulation accuracy by taking experimental data into a model, will be recognized increasingly as a key concept in the field of life science. For this purpose, at the front of life science, researchers are now required to use computers. That is to say, there's no need to open up the field of computational life science. It can be integrated into the field of life science. It would be rather ideal.

Eguchi : You mean that computers and software programs are becoming like conventional test tubes and reagents. If the time comes when we can see computers in a life science laboratory as ordinary items, we will no longer need to use the words "computational life science," right?

Kidera : I think it was the first project to draw words like "We cannot do anything without a supercomputer any more" from some researchers at the site. I hope that the number of such researchers will increase.

Yanagida : I am sure that the time will come when life scientists make full use of computers like test tubes.

Theme 1 Simulations of biomolecules in cellular environments

RIKEN
Yuji Sugita

Molecular dynamics simulation is commonly used as a method to find out the relationship between the molecular movement (dynamics) of macromolecules such as proteins and nucleic acids, and their functional structures. While conventional computations were targeted only at simulations under water or in lipid bilayer membranes, the objective of this Theme was to study molecular dynamics in a realistic cellular environment (biomolecules in cellular environments). At the initial stage of the study, we received criticism for ambiguity of specific computation targets. So we closely re-examined our research targets, and decided to focus on two research subjects: "modeling of signaling pathways considering cellular environments", and "multi-scale modeling of chromatin and nucleosomes". A research method termed "multi-scale simulation" played a fundamental role in this Theme. With the participation of specialists in quantum mechanics/molecular mechanics (QM/MM), all-atom molecular dynamics, coarse-grained molecular dynamics and single-particle simulation, biological phenomena from molecular to cellular scales were successfully analyzed by using molecular models with different resolutions. In particular, for the former, unprecedented large-scale simulations, such as all-atom molecular dynamics for bacteria cell cytoplasm, single-particle simulation for EGF signal pathways, and QM/MM computations taking the cellular environment into consideration, were conducted through the use of the K computer. For the latter, we succeeded in performing free-energy calculations on nucleosomes using a number of replicas, modeling of a tri-nucleosome through coordination between Small Angle X-ray Scattering (SAXS) and molecular dynamics (MD/SAXS and CGMD/SAXS), large-scale chromatin modeling based on coarse-grained dynamics, and so forth. At this point, finally, the image of our initial objective, "biomolecules in cellular environments", assumed a tangible form. I will continue to engage in my research activities by strengthening close cooperative relationships with other experimental researchers.

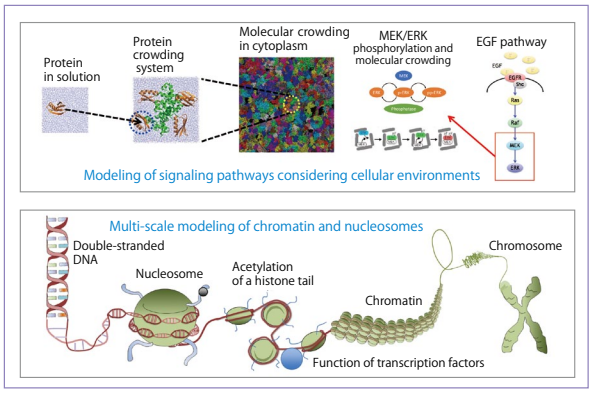


Fig. : **"Modeling of signaling pathways considering cellular environments", and "multi-scale modeling of chromatin and nucleosomes".**

Theme 2 Simulation applicable to drug design

Research Center for Advanced Science and Technology, The University of Tokyo
Hideaki Fujitani

Life activities are based on the atomic interaction and molecular reactions within the lives, whose behaviors are ruled by the physical principles such as quantum mechanics and thermo-dynamic statistics. Molecular force field well describes the room temperature behaviors of bio molecules such as water, small molecules, proteins, nucleic acids. We developed FUJI force field for accurate and unified descriptions of bio molecule taking advantage of high level quantum molecular orbital calculations. Since the simulation should be checked by quantitative comparison with experiments, we developed the binding free energy calculation method of MP-CAFEE utilizing Jarzynski equality in nonequibrium thermodynamic statics to compare with binding constants commonly measured in the drug development. In the collaboration with Professor Lindahl at the Stockholm University, we developed *SIMD kernels of GROMACS molecular dynamics suite, which are highly optimized and doubly speed up the K computer performance.

The project target proteins were decided after careful investigations, IT company members engaged in the fragment based de novo design of drugs, University researchers performed MP- CAFEE on the K computer, and pharmaceutical companies performed synthesis and assay of the designed compounds. We performed about 300 MP-CAFEE calculations for the cancer target protein and the most promising compounds are now under investigational new drug enabling study (animal test).

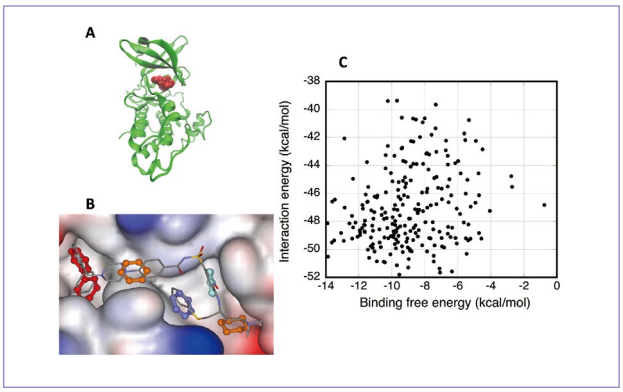


Fig. : **Cancer target protein and drug (A) Fragment based de novo design (B), MP-CAFEE result by the K computer (C)**

*Molecular dynamics computation program accelerated at the assembler level through use of the vector processing function

Theme 3 Hierarchical integrated simulation for predictive medicine

School of Engineering,
The University of Tokyo
Shu Takagi

The objectives of this Theme are to reproduce the functions of tissues and organs as an assembly of various cells at a level never explored before and, to predict clinical conditions and contribute to medical treatments. For instance, for the brain-nervous and musculo-skeletal system, we have conducted our research with the aim of reproducing the differences in tremor and rigidity, and the postural maintenance disorder in Parkinson's disease by integrating the NEST, which succeeded in a brain-nervous system simulation consisting of the world's largest number of neuron cells, the multi-scale skeletal muscle simulator Hi-MUSCLE which reproduces the behaviors of an entire muscle as an assembly of muscle fibers, and the generalized musculo-skeletal simulator K-Body. As to the achievements up to now, we succeeded in reproducing the β-band (approx. 15Hz) tremors in the basal ganglia which are caused by the lack of dopamine and observed it in animal tests using a monkey, and simulated the behavior of the signal that reduces its frequency by half in the thalamus and transmits it from the cerebral cortex to muscle fibers through the spinal cord, resulting in the tremors of hands (see article "Zoom in" on page 9 of this issue).

For simulation of the circulatory system, the achievements made by the multi-scale and multi-physics heart simulator, UT-Heart, are significant. It succeeded in realizing a 3-hierarchy simulation integrating sarcomere, cardiac muscle cells and the entire heart for the first time in the world, making a significant impact to the field of computational science. A video created based on those achievements got the Best Visualization or Simulation Award from the International Conference SIGGRAPH, an authority in the field of CG, in 2015. Up to now, this video has received over 250,000 views on YouTube (English version) (see Fig. below), and this heart simulator is now renowned globally rather than in Japan. Furthermore, the simulator has already entered the stage of predicting clinical conditions using clinical data, and a comparison with clinical results demonstrated its feasibility in accurate postoperative prediction of pediatric congenital heart disease. For simulation of thrombosis, through this project, we proposed new analysis methods to evaluate drug efficacy based on an integrated analysis of simulations and flow chamber experimental results, and obtained new findings concerning the functions of antiplatelet drugs.

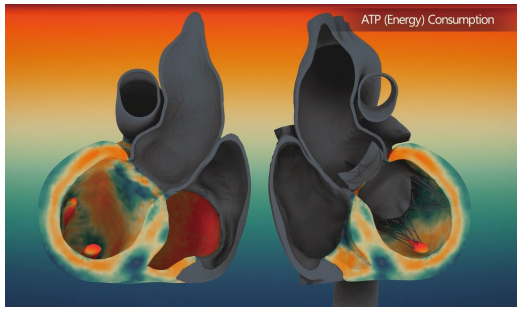


Fig. : **Image from a video of UT-Heart, which received over 250,000 views on YouTube**

Theme 4 Large-scale analysis of life data

The institute of
Medical Science,
The University of Tokyo
Satoru Miyano

The K computer has made a "warp" of medical and life sciences into a new dimension.

In some cases, cancer is caused by a virus infection. Japan is the major endemic area of human T-cell leukemia virus type I (HTLV-v). After infection during infancy, a highly malignant blood cancer called adult T-cell leukemia-lymphoma (ATL) develops after a latent period of a few decades. A large-scale omics data analysis that was conducted using the world's largest number of ATL cases (400 cases or more) revealed the full details of genetic abnormalities of ATL, and successfully discovered the targets for the development of new therapeutic drugs (Nature Genetics 2015). Also, a large-scale data analysis clarified the comprehensive histology of colon cancer which evolves spatiotemporally, obtains heterogeneity and develops metastasis to the liver, and its simulations were conducted (PLoS Genetics 2016). In the prediction of effects on anticancer drugs in individual patients, the world's largest-scale gene network analysis was conducted based on gene expression data of more than 600 cancer cell lines, and data for sensitivity and resistance against more than 100 kinds of drugs, to build an infrastructure for individualized anticancer drug administration with the world's highest accuracy and precision (PLoS One 2014).

The body temperature of a baby is very high, because his/her brown fat cells generate heat. However, as he/she grows up, those cells disappear. Meanwhile, in case of a lab mouse at 4°C, its fat cells are known to be converted into beige anti-metabolic cells which are different from brown fat cells, and generate heat 100 times stronger than that generated by skeletal muscles. By performing a large-scale gene network analysis, the relationship between physiologically active substances involved in the inflammation pathway called IL-1β and the heat generation mechanism was discovered for the first time, and the molecular mechanism of conversion into anti-metabolic cells was completely unraveled (Fig.) (Cytokine 2016).

With the advent of GHOST-MP which carries out massive parallelization and high-speed search of homologous sequences, the analysis of the metagenome of human feces can be completed within 10 minutes, and the identification of resident bacteria which can stimulate cross-antigenicity against cholera is advancing dramatically in cooperation with immunology researchers. There is a high expectation for vaccine development through the use of the K computer.

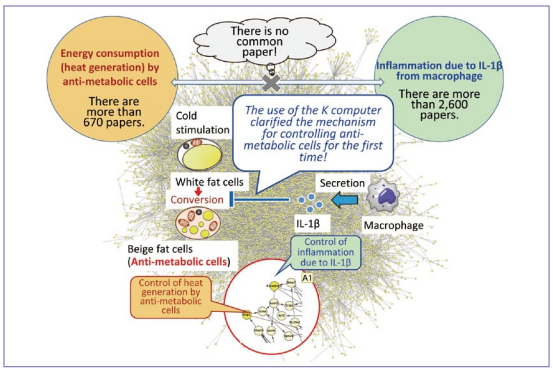


Fig. : **A large-scale network analysis made possible only by the K computer clarified the mechanism of conversion into anti-metabolic cells.**

See the URL on the right for detailed information on this report http://www.scls.riken.jp/eng/newsletter/Vol.14/openup.html

Visualizing the real molecular pictures, which are not identified through experimentation, through the use of computing power – Searching the structures of histone tails –

Molecular Modeling and Simulation Group, Japan Atomic Energy Agency Jinzen Ikebe

For sustaining life, all living beings including human beings are undergoing diverse chemical reactions (i.e. metabolism) in their bodies. Metabolism is produced by transcription proteins which are synthesized (or expressed) based on the genetic information recorded in DNA. By expressing as many of those proteins as necessary when required, living beings strictly regulate their metabolism. Abnormalities in the mechanism for regulating the genetic expression constitute a factor for causing cancer or genetic disorders. To elucidate this mechanism, we are conducting research by using a molecular dynamics (MD) simulation which traces the structural change in proteins on a computer.

DNA is folded into compact structures in a cell nucleus. The structural unit formed by DNA wrapped around the histone proteins is called a nucleosome, and a fiber of packed nucleosomes called "chromatin". When the chromatin structure is condensed, transcription proteins cannot access DNA to decode the genetic information, and the gene expression is disabled. Meanwhile,if acetyl groups are added to histone tails (or terminals of histone proteins) (i.e. acetylation), the chromatin condensation is decreased, activating the gene expression (Fig. 1). To understand the mechanism for switching gene expression on and off, it is necessary to examine how the structures of histone tails change due to acetylation. However, as histone tails assume a number of structures called “intrinsically disordered states”, it is impossible to survey the molecular structures in detail using experimental methods such as crystal structure analysis or NMR spectroscopy. Also, as histone tails carrying positive electrical charges strongly cling to DNA carrying negative electrical charges through electrical forces, conventional MD simulation has not identified the diverse structures of histone tails sufficiently.

As part of a project for the K computer, we succeeded in developing a new MD simulation method called adaptive lambda square dynamics (ALSD). This method allows us to examine diverse structures of histone tails by intentionally scaling their electrical charges in the simulation to eliminate the clinging to DNA. As a result, by using the computational power of the supercomputer, detailed pictures of histone tails, which had not been revealed through conventional experiments or simulated computations, emerged for the first time, such as the facts that the acetylation reduces the structural size of histone tails, leading DNA wrapped around histone proteins to relax easily and thereby allowing structural changes in chromatin to be promoted. (Fig. 2), and that acetyl groups are actively exposed over the surface of nucleosomes as a marker to induce transcription proteins. We will further clarify the gene expression mechanism and embark on research activities aimed at application such as the development of new drugs through the ALSD simulation method.

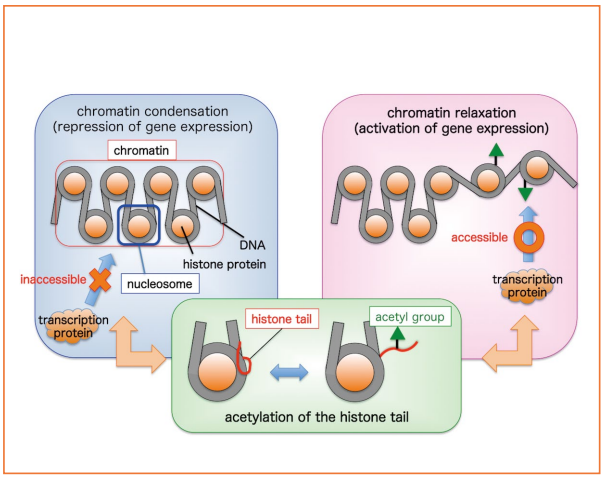


Fig. 1 : Gene expression is mainly activated through acetylation of histone tails.

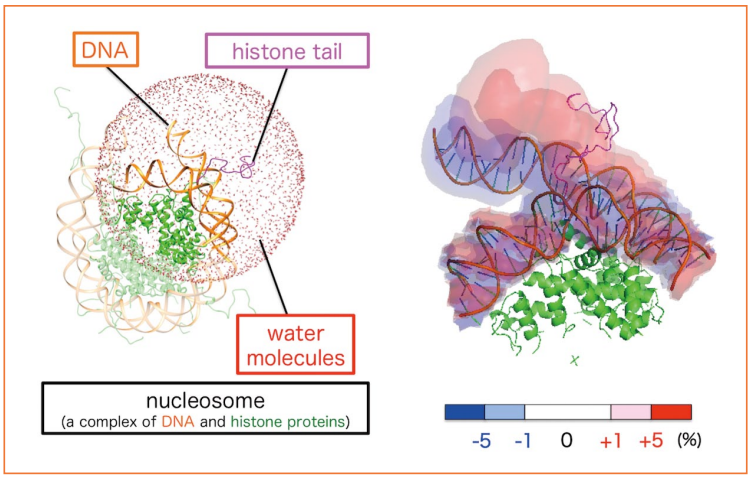
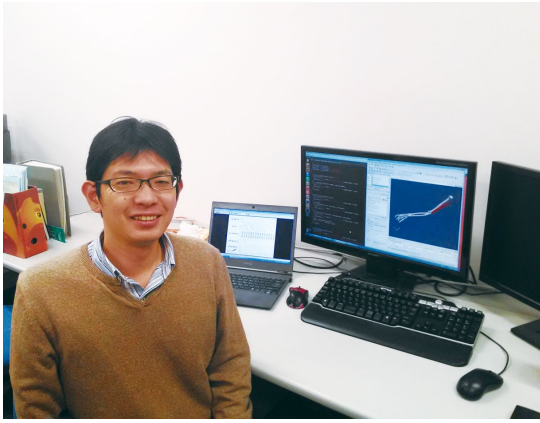


Fig. 2 : (Left) Atomic model around a histone tail. (Right) Changes in the DNA spatial distribution through acetylation. DNA tends to emerge in the red area, and does not in the blue area, compared to when the acetylation process is not carried out.

See the URL on the right for detailed information on this report and profile of the author.
http://www.scls.riken.jp/eng/newsletter/Vol.14/zoomin01.html

Integrated simulations of the nervous system and musculo-skeletal system for reproducing Parkinson's disease symptoms

School of Engineering, The University of Tokyo Naoto Yamamura



When you intend to move your body voluntary, motor commands generated inside your brain are transmitted from the motor cortex to the spinal cord through nerve fibers. In the spinal cord, the motor commands are integrated and coordinated with the feedback signals from muscles, skin and so forth, and sent to the motor neurons which are connected to the muscle fibers. When motor neurons are activated, all of the muscle fibers innervated by the neuron contract and produce force that causes joint movement.

Any trouble in the process from the cerebral nervous system to the musculo-skeletal system may develop motor dysfunction. Parkinson's disease is a neurological brain disorder that results from degeneration of neurons in the basal ganglia which controls movement. The degeneration of neurons creates a lack of the neurotransmitter called "dopamine", causing the movement dysfunctions. The motor symptoms of Parkinson's disease include resting tremor, rigidity, bradykinesia and postural instability, but the mechanism whereby such symptoms are generated remains unexplained. We are developing an integrated simulator of the central nervous and musculo-skeletal systems of the entire human body with an aim to provide a useful tool for investigating the mechanisms underlying motor dysfunctions, and exploring effective therapeutic approaches through the use of computer simulations.

Our simulator is an integration of a simulator of the cerebral nervous system consisting of the basal ganglia, thalamus and motor cortex circuits, a simulator of the cerebrospinal nervous system computing the activity of motor neurons, and a simulator of the musculo-skeletal system computing behaviors of the musculo-skeletal system from the activity of motor neurons, which were developed by the Doya Research Group (Okinawa Institute of Science and Technology Graduate University), the Nakamura Research Group (The University of Tokyo), and the Takagi Research Group (The University of Tokyo) respectively, so as to expand them over the entire human musculo-skeletal system (Fig. 1).

Fig. 2 shows an example of an integrated simulation based on a central nervous system model and a human upper arm musculo-skeletal model in Parkinson's disease. The neural activities in the motor cortex, thalamus, basal ganglia in the brain, the activity of motor neurons in the spinal cord and the joint angle of an elbow joint are shown through time. The upper side of each neuron is projected to the biceps muscle, while its lower side is projected to the triceps muscle. In the basal ganglia, abnormal neural activity specific to patients with Parkinson's disease are reproduced. For neurons in the thalamus, alternate neural activities are reproduced with a frequency similar to that of Parkinsonian tremors (4 to 6Hz) in the biceps and triceps. The output from the simulator of the cerebral nervous system is sent to the spinal cord as the pyramidal tract activity of the motor cortex, and motor neuron activity is computed by the simulator of the cerebrospinal nervous system. The simulator of the musculo-skeletal system computes the force of contraction based on motor neuron activity, and tremor-like movements are observed in the elbow joint.

The development of this integrated simulator enables the simulation of human body movements originating from activity of the central nervous system. Using this simulator, we will proceed further with our academic studies with an aim to reproduce motor symptoms such as Parkinsonian tremor and rigidity, understand the mechanism generating such symptoms, and eventually develop treatment methods.

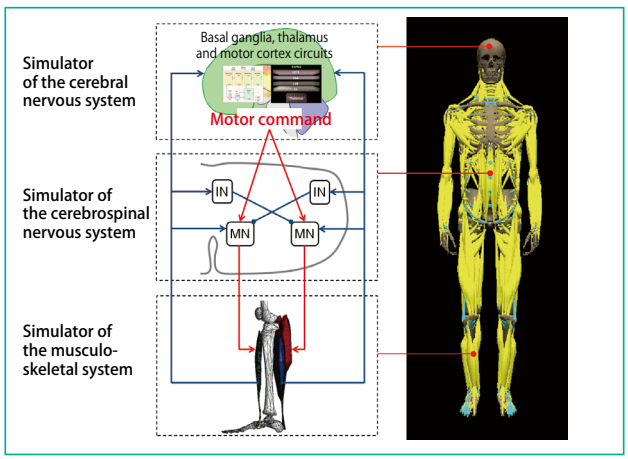


Fig.1 : Simulator integrating the musculo-skeletal and nervous system. "MN" means motor neuron, and "IN" means interneuron.

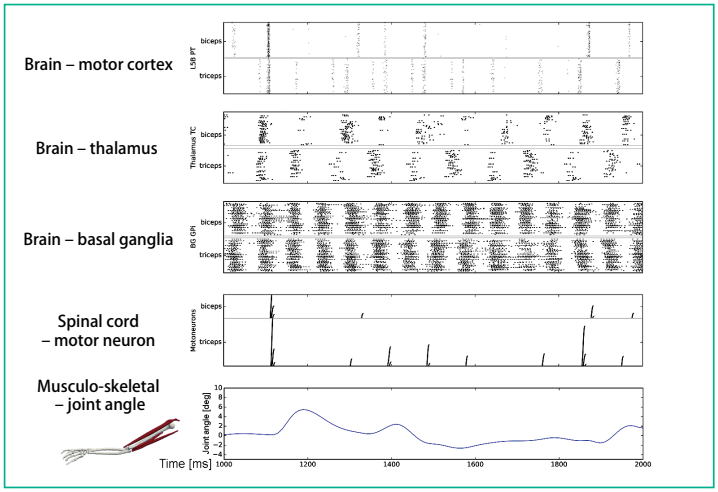


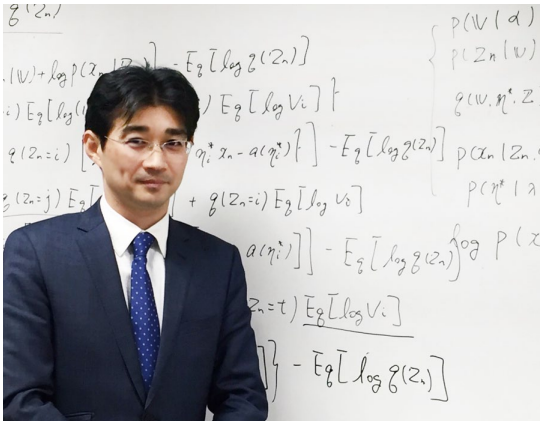
Fig.2 : Integrated simulation of the cerebral nervous, cerebrospinal nervous and musculo-skeletal systems in Parkinson's disease. The black dots represent neural activity.

See the URL on the right for detailed information on this report and profile of the author.

http://www.scls.riken.jp/eng/newsletter/Vol.14/zoomin02.html

From supercomputing analysis of big cancer genome data to life science and medical care

Health Intelligence Center, The Institute of Medical Science, The University of Tokyo Seiya Imoto



We use a supercomputer to analyze cancer genome data for identifying abnormalities of systems in cancer, and finally aim to create prediction methods for drug efficacy and side effects, and novel treatment and prevention methods. Genes form a complex communication network depending on the phosphorylation of proteins they produce, the interaction between proteins, and the control of expression of other genes. This system generates the proteins required to maintain our lives. Meanwhile, cancer is believed to occur due to accumulation of various acquired mutations in genomes. Such mutations may cause some genes to produce abnormal proteins and then deactivate them, abnormally activate them, or even make them acquire new functions. Such "failures" of genes affect the network of genes mentioned above. If the system falls into an uncontrollable state, "cancer" is said to have occurred.

Thanks to the breakthrough of the next-generation sequencers, the cost required for determining the entire DNA sequence continues to decrease at a pace faster than Moore's law. Nowadays, it is as low as 1,000 dollars. This technology also enables measurement of the amount of messenger RNAs produced by each gene. We are now in an age when we can readily obtain individual genome data and genome-related data. Nevertheless, the amount of human genome data or RNA data is overwhelming. The volume of whole genome sequencing data is around 200GB in terms of a contemporary measurement. Consequently, supercomputing capacity is indispensable for data analysis.

Whereas we simply refer to "cancer" as a type of disorder, it bears diverse genomic variations and mutations. They differ from patient to patient, and even one patient may have multiple cancer-cell populations whose genomes are damaged in different ways. Research activities to understand genomic mutations which determine "cancerous characteristics," such as high malignancy and drug resistance, are currently in progress.

We successfully established a new data analysis method called "NetworkProfiler" by using statistical theory. This method allows us to find differences in the gene network which determines the characteristics of cancer cells, such as drug resistance, based on data obtained by measuring the amount of RNAs produced by each gene (i.e., RNA expression data). By using this technique, we extracted the gene network relating to the effects of approx. 100 kinds of anticancer drugs based on data for more than 600 cancer cell lines. This was made possible only by using the K computer. Currently, we are engaged in the development of a data analysis technique which can more accurately forecast the effects of anticancer drugs for each cell by combining data from this gene network with a great amount of data on genomic variations from the next-generation sequencer, so as to perform a comprehensive analysis. The results of such an analysis will lead to individualized medical treatments based on the genomic data.

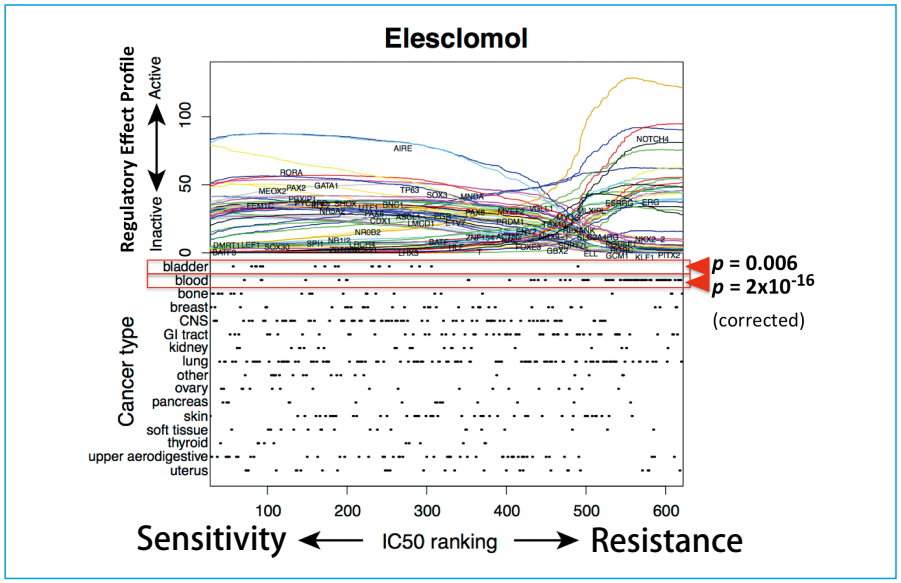


Fig. : Sensitivity to anticancer drug Elesclomol (horizontal axis) and transcription factor activity prediction value (vertical axis of the upper panel).
It shows that, for example, while AIRE controls strongly the controlled genes in cancer cell lines highly sensitive to Elesclomol, its influence is limited in resistant cancer cell lines.
The lower panel shows the biases depending on cancer types. A slight bias toward sensitivity is observed in bladder, while a bias toward resistance is observed in blood. However, no significant biases are observed in other cancer types. The resistance and sensitivity mechanism of Elesclomol is expected to emerge beyond cancer types.

See the URL on the right for detailed information on this report and profile of the author.
http://www.scls.riken.jp/eng/newsletter/Vol.14/zoomin03.html

Report on workshop for media personnel, titled "2nd Forefront of life sciences led by K computer!"

Planning and Coordination Group
RIKEN HPCI Program for Computational Life Sciences

The workshop for media personnel, "2nd Forefront of life sciences led by K computer!" was held on Wednesday, September 30, 2015. The purpose of this workshop was to widely inform press personnel of research achievements delivered by integrating this cutting edge supercomputer with life science, as well as of new science dreams arising out of such achievements. Four representatives tackling research subjects under Strategic Programs for Innovative Research Field 1 Supercomputational Life Science (SCLS), and two representatives tackling priority issues for the post-K computer, delivered lectures from the Tokyo Liaison Office of RIKEN and RIKEN Advanced Institute for Computational Science (via teleconference).

Six lecturers gave stimulating presentation of the research objectives, the achievements exclusively attained by the K computer, and the future application and evolution in industrial areas under the four research subjects: "Hierarchical integrated simulation for predictive medicine," "Analysis of the molecular dynamics of proteins and nucleic acids in cellular environments," "Innovative drug development enabled by the supercomputer" and "Large-scale analysis of life data," peppered with their own future initiatives and dreams. Also covered were the future prospects for the potential breakthrough achievements soon to be witnessed in the field of life sciences thanks to the projects of the post-K, which is expected to succeed the K computer, under the titles, "Integrated computational life science for individualization and preventive medicine" and "Drug development innovation enabled by the post-K."

According to the results of the questionnaire survey conducted afterward, what impressed the participants were the research activities directed at understanding the diversity of abnormalities in cancer, the results of ATL's large-scale analysis, and the cranial nerve system and musculo-skeletal system, indicating their wide-ranging interests in diverse research fields. The workshop sessions were held in Tokyo and Kobe in the presence of a number of press personnel. We appreciate their attendance and strong interest. For a summary of the lectures and non-classified presentation documents, please visit the website of SCLS (http://www.scls.riken.jp/information/material.html).

A lecture held at the venue in Tokyo

-The K computer creates the future of science and society - Supercomputational Life Science 2015

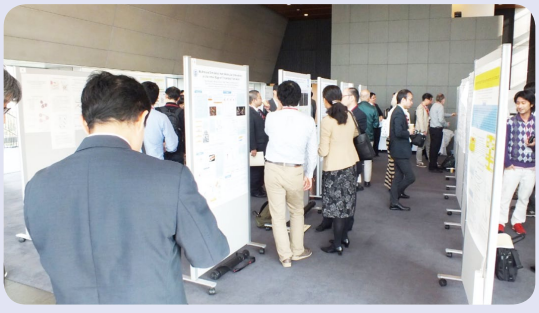
Planning and Coordination Group
RIKEN HPCI Program for Computational Life Sciences

SCLS comes to an end in March 2016. To present its final achievements, a symposium titled "-The K computer creates the future of science and society- Supercomputational Life Science 2015 (SCLS2015)" was held in the Takeda Hall on the Asano Campus of The University of Tokyo, on Tuesday, October 20 and Wednesday, October 21, 2015. This two-day symposium was attended by 143 persons including those from educational institutions, research institutions, private companies and governmental entities, as well as general citizens.

At the international workshop held on the first day, four researchers from overseas laboratories and four researchers from SCLS delivered lectures to present the research activities being conducted at the forefront of computational life sciences. At the achievement report session held on the second day, 15 SCLS researchers reported the latest research achievements attained through use of the K computer, such as analysis of intercellular molecular dynamics based on molecular dynamics simulation, prediction of binding of drug development candidate compounds based on high-precision binding free energy calculation, hierarchical integrated simulation, and exhaustive analysis of abnormalities in cancer, and received a great response from the attendees. The poster session was attended by 29 persons, and offered an opportunity for participants in the symposium to exchange opinions and information in a lively and active way.

With positive comments such as "I realized that unprecedented achievements are being made at a pace which was unthinkable only five years ago," and "I believe that some of the software programs and approaches developed by them have sufficient potential for leading the world," we are pleased that this symposium provided a precious opportunity to think about the future of computational life sciences, whose importance is increasing as part of the further development of life sciences. Taking this occasion, we would like to extend our cordial thanks to all the lecturers and participants.

Lecture session



Poster session

NEXT STAGE

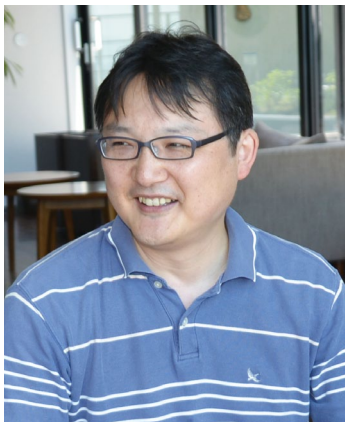
We interviewed four Subtheme leaders of priority issues 1 and 2 in "Flagship 2020" Project (post-K computer project for the development

Priority issue (1) : Innovative drug discovery infrastructure through functional

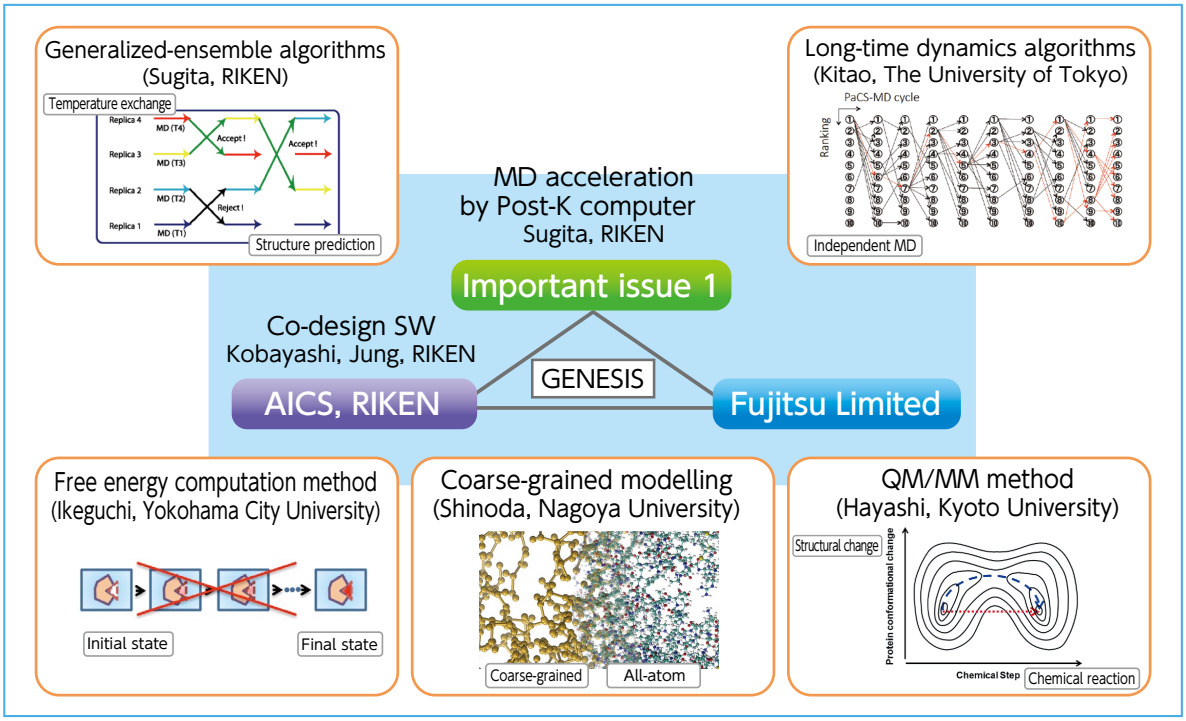
Subtheme A

Advancement of MD and associated algorithms by post-K computer

Priority issue 1, Subtheme A, Supervisor
Yuji Sugita
RIKEN



Under this issue, we engage in acceleration and algorithm development required for maximizing efficiency of the post-K molecular dynamics simulation. For drug design application, molecular dynamics simulation covering various target systems is required. In addition to simulation forecasting of protein and ligand binding conditions based on a small system consisting of several tens or hundreds of thousands of atoms, and the dynamics of membrane proteins or protein complexes consisting of a maximum of one million atoms, it is necessary to compute a large-scale system consisting of 10 million atoms or more in the case of a biological system like a virus or cellular environment. To allow the post-K to handle such biological systems, design work is carried out by focusing on both hardware and software to optimize overall computing efficiency, that is, by adopting a co-design methodology. Together with priority issue 1, RIKEN Advanced Institute for Computational Science and Fujitsu Limited are co-designing a molecular dynamics software called GENESIS developed under the initiative of RIKEN. We plan to allow small- and medium-scale systems to implement a number of simulations with different parameters and ligands simultaneously using a post-K, and large-scale systems to challenge complex and large computation limits by using all the nodes of the post-K. By adopting a new molecular dynamics algorithm for drug design application under priority issue 1 in the sophisticated GENESIS, we will attempt to simplify and accelerate the computation of molecular dynamics and the analysis of free energy necessary for drug design through use of the post-K. We will also publish the resulting software and computation so that general users can use them at the same time as the release of the post-K to encourage their use.



Under Subtheme A, "Post-K MD acceleration and algorithm deepening," we will make the molecular dynamics software program called "GENESIS" more sophisticated with a view to the post-K concept through the co-design process, and adopt advanced MD algorithms sequentially.

of a machine that will succeed the K computer) promoted by the Ministry of Education, Culture, Sports, Science and Technology.

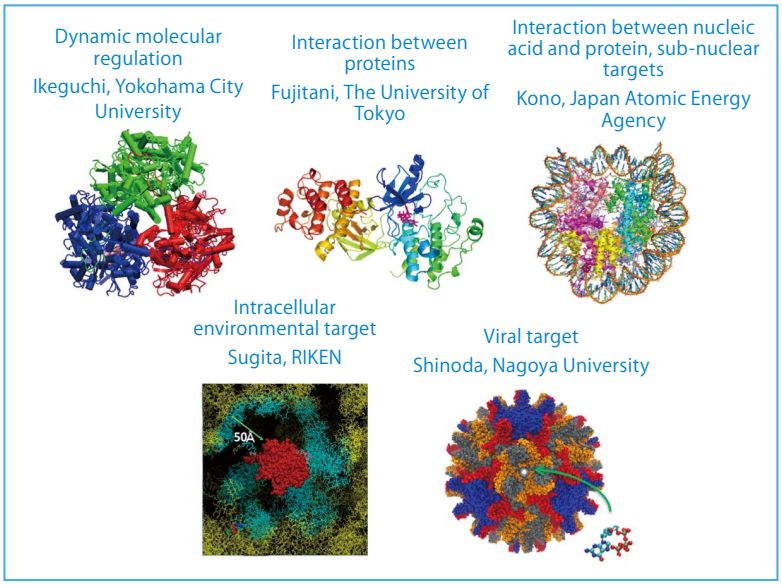
control of biomolecular systems

Subtheme B

Development of next-generation drug design computational techniques

Priority issue 1, Subtheme B, Supervisor
Mitsunori Ikeguchi
Graduate School of Medical Life Sciences, Yokohama City University

One of the nine priority issues designated under the concept of the post-K computer, "Innovative drug discovery infrastructure through functional control of biomolecular systems," is now going to start in full swing. In this priority issue, I was appointed as a supervisor for Subtheme B, "Development of next-generation drug design computational techniques." In Subtheme B, we will work unwaveringly on computing methods for drug design and structural biology, which are not attainable with conventional computing capabilities, by using the sophisticated molecular simulation software and algorithm optimized for the post-K to be developed under Subtheme A, so as to pave the way for the integrated computational drug design system to be developed under Subtheme C. Specifically, first of all, we set a goal to control dynamic molecular functions by focusing on motions and structural changes of drug design target proteins. Computing methods for conventional computational drug design often treat the structures of target proteins as firm and immovable, in accordance with the lock-and-key model. However, actually, the structures of proteins are very soft and often change due to the binding of compounds such as pharmaceuticals. In this issue, we try to understand and control the motions and structural changes of target proteins by capitalizing on the superior computing capabilities of the post-K. The next issue is to control the interactions between proteins, and those between nucleic acids and proteins. In recent years, bio-pharmaceuticals containing proteins as medicinal chemicals such as antibody drugs also have been developed. Such proteins and nucleic acids are much more complex than low-molecular weight compounds, and their steric structures are flexible. Therefore, it is necessary to conduct large-scale computation using the post-K in order to compute the interactions between molecules in detail. Our issues also include the entire computation of virus capsids and computation targeted at the cellular environment where a plural number of biological supermolecules are intricately entwined. Actually, these targets are supergiant systems, which could be computed only by the world's leading supercomputers. This would be an unprecedented challenge. As described so far, this is already an attractive research theme only in terms of computation. However, its value will be increased by closely coordinating with experimental system measurements. In this issue, we plan to conduct research activities in cooperation with the most advanced structural biology experiment facility SPring8 and NMR facilities. We sincerely hope that we can implement unprecedented computations for drug design using the post-K.



Outline of Subtheme B "Development of next-generation drug design computational techniques." It consists of the subjects "dynamic molecular regulation," "control of interaction between proteins," "control of interaction between nucleic acid and protein," "sub-nuclear environmental target," "intracellular environmental target," and "viral target."

See the URL on the right for detailed information on this report. ***http://www.scls.riken.jp/eng/newsletter/Vol.14/nextstage01.html***

Priority issue (2) : Integrated computational life science to support personal

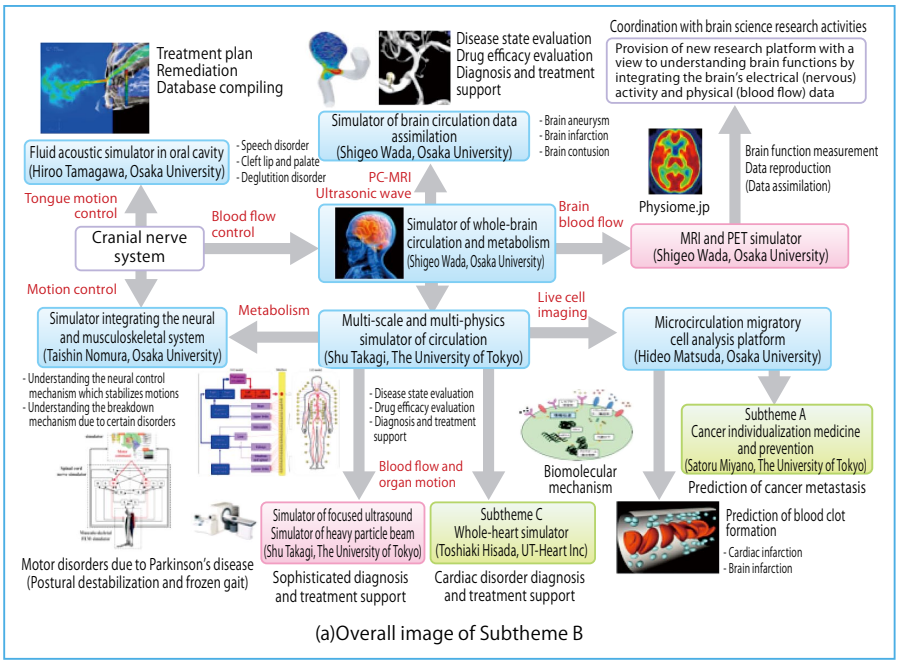
Subtheme B

Supporting personalized medicine by data assimilation based biological simulation

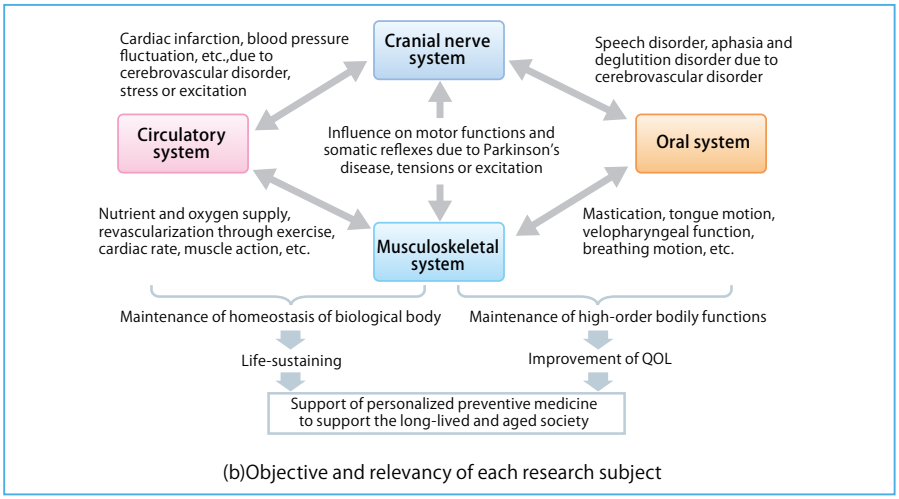
Priority issue 2, Subtheme B, Supervisor
Shigeo Wada
Graduate School of Engineering Science, Osaka University

The development of measuring techniques which enable non-invasive observation in the body such as MRI, X-ray CT and ultrasonic echo has contributed significantly to the advancement of medicine. Meanwhile, the development of computing science enabled dynamic phenomena to be analyzed at diverse dynamic layers from biological molecules and cells to tissues, organs and individuals, resulting in the reduction of the distances between biology, life sciences and physics. One of the roles of bioengineering is to apply understanding of biological phenomena from such physical aspects to medicine. However, in the field of evidence-centered medicine, much importance is attached to measurements, and technical analyses based on mathematical models are not utilized sufficiently in clinical practice. As the amount of data available because of progress in measuring devices increases and the target phenomena become more complicated, sophisticated diagnosis and treatment require deep understanding of the biological phenomena observed, and analysis of a huge quantity of measurement results. However, no medical engineering techniques to support such tasks have so far been developed.

The purpose of this Subtheme is to develop a computational simulator acceptable to the medical field where actual measurement data are important by assimilating and integrating large-scale physical simulations of the biological body through use of the post-K computer with biometric data supplied in various forms, and build fundamental technologies to support personalized medical care based on the medical and biological data of individual patients. For a large-scale simulator which can only be created with the post-K, we will develop a whole-brain circulation and metabolism simulator which connects cranial nerve activity with brain circulation. By providing the medical field with physical data obtained from the data assimilation based biological simulation, we will evaluate the biological functions in the human body correctly and quantitatively and build a database, and create new medical engineering technologies built on computational science in order to ensure sophisticated diagnosis, treatment, prevention and forecast methods which support a healthy, long-lived society.



(a)Overall image of Subtheme B



(b)Objective and relevancy of each research subject

Fig. 1: In Subtheme B, we will extend the simulation technologies developed by the K computer, and develop the biological body's physical simulator based on each patient's medical and biological data for the blood circulating system that plays the major role in life sustenance, the neural-musculoskeletal system responsible for body movements indispensable for the maintenance of QOL in the aging society, and the oral system responsible for speech.

ized and preventive medicine

Subtheme C

Bridging basic medicine and clinical medicine by fusing heart simulation and molecular simulation

Priority issue 2, Subtheme C, Supervisor
Toshiaki Hisada
UT-Heart Inc.

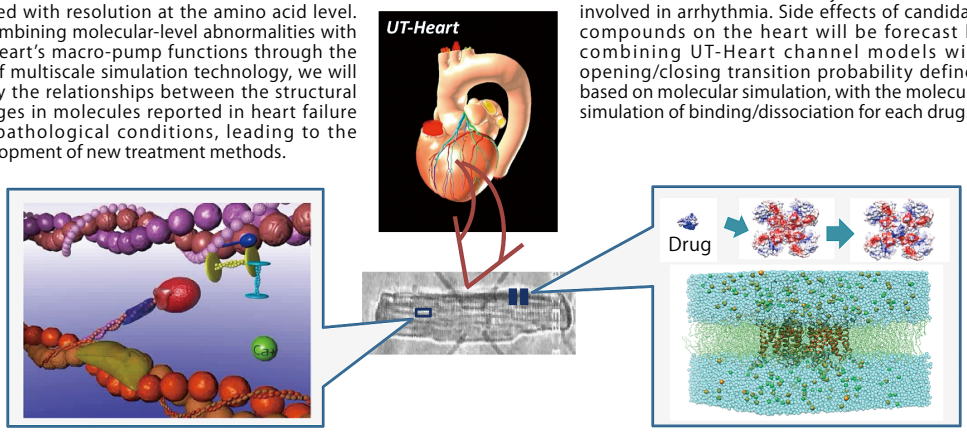
Currently in the middle of transition from the K computer to the post-K computer, the "importance of planning research activities by forecasting progress in computers and the necessity for developing technologies for implementing this" are further increasing. We are urged to think outside the box and steadily deepen our academic understanding.

I would like to briefly present what we are going to achieve under the post-K project over the coming 10 years. The idea of our research activities is reflected in the subtheme, "Bridging basic medicine and clinical medicine by fusing heart simulation and molecular simulation." In development activities by UT-Heart so far through the use of the K computer, multiscale simulation technology connecting micro contraction proteins which constitute the sarcomere with macro heartbeats was successfully developed. In this technology, a mathematical model, in which ATP is bound to the myosin head whose arm is expressed as a spring and undergoes hydrolysis, causing stochastic oscillations to/from actin filaments, is defined in accordance with statistical mechanics, and the resultant contraction force of the sarcomere is computed. For the post-K project, we will take this technology further to develop a sarcomere model consisting of coarse-grained molecular simulation models, and coordinate with the heart model. We consider that this will enable us to, for example, understand how the load observed and measured at the macro level is conveyed to the molecular level and activates the intracellular signal transducing system, and clarify the mechanism for causing pathological conditions. Moreover, by using the molecular model, we will build the ion channel which controls the input/output into/from the cells of various ion currents which contain calcium ions that modulate the motion of contraction proteins, and simulate binding/dissociation to/from drugs, to evaluate the risk of arrhythmia caused by drugs to the heart. This will enable swift screening of candidate compounds which produce side effects on the heart and make the process of drug design more efficient. This study will be conducted in tandem with Priority issue 1.

Through good luck, we witnessed milestones, that is, the emergence of the K computer and subsequent evolution to the concept of the post-K. However, time passes so fast. Our achievements must be good enough to make constructive discussions during the next 10years. We would like to create a new milestone in the history of computing science by keeping our spirits high and concentration intense.

The molecular model within the sarcomere is created with resolution at the amino acid level. By combining molecular-level abnormalities with the heart's macro-pump functions through the use of multiscale simulation technology, we will clarify the relationships between the structural changes in molecules reported in heart failure and pathological conditions, leading to the development of new treatment methods.

There are four to six major ion channels involved in arrhythmia. Side effects of candidate compounds on the heart will be forecast by combining UT-Heart channel models with opening/closing transition probability defined based on molecular simulation, with the molecular simulation of binding/dissociation for each drug.



CafeMol sarcomere molecule model
Shoji Takada, Kyoto University

μ^2lib ion channel molecule model
Tohru Terada, The University of Tokyo

Research concept of Subtheme C "Bridging basic medicine and clinical medicine by fusing heart simulation and molecular simulation"

See the URL on the right for detailed information on this report. http://www.scls.riken.jp/eng/newsletter/Vol.14/nextstage02.html

News & Events

Reflecting on remote interactive lecture titled "Foundations of Computational Life Sciences"

Since FY2014, we have offered remote lectures on computational life sciences, which is a cross-disciplinary research field of life sciences, science and engineering, for university students, postgraduate students, postdoctoral researchers, university teachers, laboratory researchers and company researchers. Thanks to the cooperation of a number of researchers, we created curricula so that they can learn the most advanced fields of computational life sciences in a structured manner, and we contributed to the building of standard curricula of computational life sciences in the future. We had much positive feedback from approx. 450 trainees from Japan and overseas, such as, "I was impressed because the lecture covered the most advanced knowledge in a wide range of fields with explanation of basic knowledge," "I was satisfied as the lecture provided me with knowledge about computational tools and their concepts," and "I was inspired because application examples were brand new and unprecedented."
The cutting-edge the K computer is located in the Kobe Port Island area, where the Education Center on Computational Science and Engineering, Kobe University, and SCLS are headquartered. The area also enjoys beneficial geographic conditions as the Kobe Biomedical Innovation Cluster (KBIC). We strongly hope that such a structured educational program of computational life sciences will be instrumental in developing human resources for research and development in the field of computational life sciences, creating innovative technologies, and contributing to the advance of computational life sciences and the prosperity of KBIC.

Foundations of Computational Life Sciences II

Integration of life sciences, science and engineering for understanding of life and application to health and medicine" (15 sessions in total)

- Date : October 7 (Wed) in 2015 – February 3 (Wed) in 2016 Every Wednesday 5:00 p.m. - 6:30 p.m.
- Location : 1F the school building of Faculty of Engineering, Kobe University C3-101 (Creative Design Studio on Technology 2)

The 39th Annual Meeting of Japanese Society of Biorheology

- Date：June 18 (Sat) – 19 (Sun) in 2016
- Location：The Tokai University Club (Chiyoda-ku, Tokyo)
- Theme : Biorheology based on medicine-engineering collaboration on thrombotic events, blood flow and arteriosclerosis
- Annual President: Shinya Goto (Cardiovascular Medicine, Department of Medicine (Cardiology), Tokai University School of Medicine)

Editor's note

This will be the final issue of the "BioSupercomputing Newsletter" since the first issue published in 2011.
We started to issue the Newsletter with the desire to launch a reading material which would be handy and plain for every age bracket from high school students to general adults, and encourage them to get interested in cutting-edge research fields of computational life sciences. In issuing each volume, as I came to know the dedication of researchers, I increased my desire to convey their ardent beliefs and dedication to you all. I believe that this has driven me to creative activities, incited enthusiasm for my job, and deepened my interest in life sciences. As a spokesperson, I am so honored to have had such great opportunities.
We would like to extend our cordial appreciation to our readers and those involved in its issue for their long-time support. We would be very grateful to receive any comments or opinions regarding the Newsletter. We solemnly ask for your continued support for the field of computational life sciences. (E. Jinnai)

Strategic Programs for Innovative Research Field 1

Supercomputational Life Science

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.14 2016.3

RIKEN HPCI Program for Computational Life Sciences
7-1-26 Minatojima-minamimachi, Chuo-ku, Kobe Hyogo 650-0047, Japan
Tel:+81-78-940-5835 Fax:+81-78-304-8785

Issued: March 2016
RIKEN 2016-014