

BioSupercomputing Newsletter

2015.9 Vol.13

CONTENTS

Open Up SPECIAL TALK Discussion meeting

Accurately predict the absolute binding free energy using K computer to support the drug design of next-generation
Takefumi Yamashita / Keiko Shinoda The University of Tokyo 2

SCLS Eyes S-cruise From the field of software development
Realize molecular dynamics simulations of huge biological systems with K computer
Yuji Sugita / Jaewoon Jung / Takaharu Mori RIKEN 6

ZOOM IN Close-up on SCLS Research

Theme 1 Simulations of biomolecules in cellular environments
Whole bacteria cytoplasm modeling and all-atom molecular dynamics simulations
Isseki Yu RIKEN 8

Theme 3 Hierarchical integrated simulation for predictive medicine
Modeling human for understanding human
Akihiko Murai National Institute of Advanced Industrial Science and Technology 9

Theme 3 Hierarchical integrated simulation for predictive medicine
Neural control of human quiet standing and biped walking
Movement fluctuation and its degeneration caused by neurological disease
Taishin Nomura / Chunjiang Fu Osaka University 10

SCLS Gotcha!

“On Interdisciplinarity studies - the integration of biology, physics, chemistry and mathematics” 11
Third Advisory Committee Meeting 11

NEXT STAGE • Challenging the priority issues on post-K computer

Innovative drug discovery infrastructure through functional control of biomolecular systems
Yasushi Okuno RIKEN 12

Integrated computational life science to support personalized and preventive medicine
Satoru Miyano The University of Tokyo 14





Theme 2 Simulation Applicable to Drug Design

Accurately predict the absolute binding free energy the drug design of next-generation

Clarify biomolecular behavior by the simulation science to support exploration of small molecular drug and antibody drug candidates

Theme 2 Simulation Applicable to Drug Design (Representative: The University of Tokyo, Professor Hideaki Fujitani) searches for candidate compounds that strongly bind to target proteins of diseases using the absolute binding free energy computation method based on the originally developed molecular dynamics (MD) method. Furthermore, they analyze interaction between proteins aiming at renovation of the drug design process by the computational science. This time, we interviewed two researchers who are involved with the studies on new chemical compounds which suppress functions of target proteins of diseases and design of antibody drugs utilizing the K computer.

Encounter with the biosimulation research

● First, I'd like to hear about how you two started working on the present research.

Yamashita : I have worked on the computational science since I was attracted to describing physical phenomena in mathematics in my high school days and worked on the computational science. I was interested in quantum mechanics and studied on proton. Having found a lot of phenomena greatly affected by the proton in the field, I was engaged in the

biological field such as the molecular biology in the US and studied on proteins. That is my background. I was interested in drug design using protein simulation. I thought the simulation would be applied more widely than I expected, so I decided to challenge this theme. The research directly connected to medical service is related to priority issues in the society, so that is worth doing. I am really happy if I can contribute to the society. I joined the University of Tokyo in January,

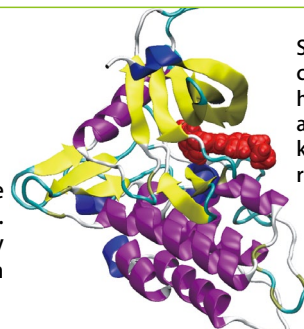
2011. Since then I have studied on biosimulation and the molecular dynamics calculation that targets proteins related to drugs.

Shinoda : I was very attracted to the phenomenon called "phase transition" in physics in my college days. So, I worked in an experiment laboratory first, and was involved with studies mainly with X-ray diffraction experiments of alloy using synchrotron radiation. I had a chance to observe phenomena that could rarely be seen in the experiments

Takefumi Yamashita

Associate Professor, Ph.D
 Research Center for Advanced Science and Technology
 Systems Biology and Medicine
 The University of Tokyo

In the simulation research with the K computer we should not be afraid of any failure. That is important. Through those experiences like trial and errors we surely obtain reliable results. Further, we need to find a vision that leads to the next step from such experiences.

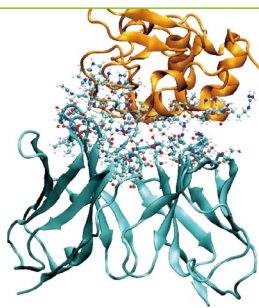


Structure of kinase combined with inhibitor: The ribbon and red balls indicate kinase and inhibitor, respectively

Keiko Shinoda

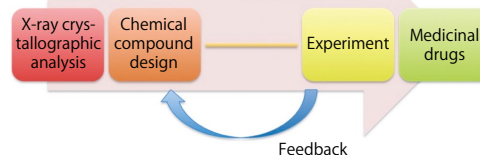
Assistant Professor Dr. Sci.
 Research Center for Advanced Science and Technology Systems Biology and Medicine
 The University of Tokyo

I think that it is important to continue making researches with simulation methods or systems to make use of the high-performance by the K computer in the light of using the K computer. Performing the computation in large quantities enables us to obtain statistical interpretation even in the case of the computation, whose scale is not big respectively.



Antigen (orange) and an antibody (blue). The important residues which interact in the antigen and antibody boundary surface are shown as atoms.

Conventional computer-assisted drug design method



New computational drug design with the supercomputer

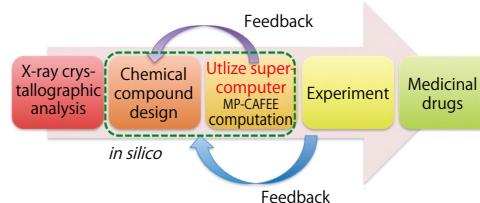


Figure: Possibility of new computational drug design method

using K computer to support

and I was very excited with it. But, at the same time I hoped to dig into model researches and theory to find how and why such phenomena occur. I took a master course and a doctoral course of different graduate schools. I studied on phase transition of diluted ferromagnet in the master's course and on quantum liquid based on an integral equation theory in the doctoral course. I was involved with a research with the MD simulation for the first time when I was a post-doctoral researcher of National Institute of Advanced Industrial Science and Technology. After that, I

had worked on various themes such as material development or quantum chemical calculation as a post-doctoral researcher of enterprises. Meanwhile, I had been in half-hearted attitude all the time. I could take time to hypothesize and carefully by myself. I could not dig into my research, which made me frustrated. I was employed by the University of Tokyo in December, 2010, a month before Dr. Yamashita was. With MD simulations of biomolecules using the supercomputer and examination of their structures and dynamics, I researched to develop drugs. Because

the researches including examination of their structures and dynamics to support drug design have its clear purposes and direct relation to the output. Such researches really interested me and were worthwhile doing. The professors of the wet team working with me are eager to develop drugs with the results of computational science. The results of simulation research are directly applied for drug design, which gave me a great pressure on me; however, I will make an effort to do it.

Promote research that contributes to new drug design

● What researches are you two working on for Theme 2?

Yamashita : What I first dealt with Theme 2 was a research on molecular target drugs that inhibit kinase. Many drugs have an effect by suppressing functions of the proteins that cause the disease. In

order to discover a drug, it is necessary to find out chemical compounds which strongly interact with the target protein in vivo. Therefore, we attempt to evaluate absolute binding free energy of target proteins and chemical compounds by simulations with a

supercomputer so as to find chemical compounds that can be used as drugs. The absolute binding free energy is the physical quantity that indicates how much molecules can recognize target proteins. We predict the activity of new candidate compounds with high

precision by the method of the absolute binding free energy computation, "MP-CAFEE", developed by Prof. Fujitani and his colleagues. We design and search for chemical compounds with high inhibitory activity. The MD program, the engine of "MP-CAFEE", is "GROMACS". It has been developed as an open source program. We tuned this up for the K computer in cooperation with developers of Stockholm University and others. The computation improved approximately twice as efficiently as the primary version. This speeding up allows the absolute binding free energy computation to be performed more efficiently. In this way, a technical base to find out promising candidate compounds has been established. Large-scale data began to be accumulated since last year. It enables us to write papers that examine possibility of new drug design. Now our important mission is to analyze the computation results and provide outcomes as science.

I have been doing a basic research, and my greatest motivation is in there while I am attracted to mysterious aspects of basic phenomena and things not clarified when doing the applied research for drug design. While application progresses steadily toward a practical use, there actually are a lot of thing remaining insufficient in basic

parts. Not all are completed as perfect methods. As long as we belong to Academia, we need to continue the research further. In order to achieve it, we perform actual computation widely and obtain results, continuing researches for such aspects.

Shinoda : I have studied molecular dynamics for antibody drug development against cancer. The use of antibodies that bind specifically to antigens has achieved considerable success in cancer therapy in recent years. Antibodies are applied to the methods such as pre-targeting method, which involve separating the targeting antibody from the subsequent delivery of therapeutic agent that binds to the tumor-localized antibody. Now we try to improve the affinity of current treated antibodies and investigate basic mechanism: a specific way of binding antigens - antibodies by simulations. This is because the antibody treated by us is very unique. There is a proline in the region where antigens are recognized, and its conformation changes from cis to trans by binding to antigen (cis-trans isomerization). This transition requires great amount of energy and it is said that it assumes a switching role for various biological processes. It takes too much time to see the mechanism of this transition by the MD simulation

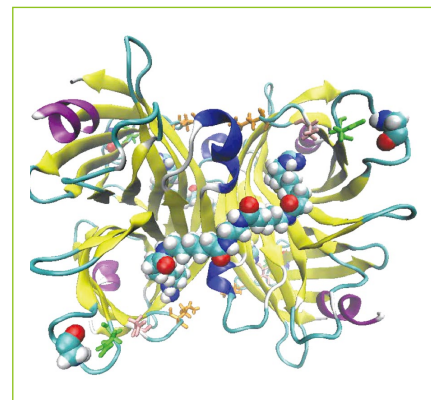


Figure : Binding structure of "Cupid" and "Psyche" used for the pre-targeting therapy. The antibody connected to "Cupid" (indicated as ribbon) binds to the target protein, and "Psyche" (indicated as van der Waals in the middle) that carries drugs specifically binds to "Cupid".

directly; however, if we can obtain a clue to understand the mechanism, it can be applied to the design of the part for recognizing antigens, we believe.

In order to clarify the binding processes and structures, we performed nearly 1,000 MD simulations using the K computer, and obtained various trajectories of antigen binding to antibody. In other words, we intend to estimate the mechanism observing various cases by performing the extensive simulations. It is biologically essential to clarify such an underlying mechanism, which leads to drug design in the future.

Building up research vision for the next generation is also a mission

● Going into the final year, how do you wish to advance your research based on the results obtained so far?

Yamashita : As I have just mentioned, a method for high speed computation being established so we are challenging a large-scale computation of over 200 chemical compounds for one theme. In the future, it will be important to brush up the method to elucidate "Why" through more detailed analyses. I'm sure drugs such as the kinase inhibitor will be designed and explored efficiently. Further, the experiment is important for applying simulation results to drug design, I believe. No matter how good the simulation result is, what simulation

results can show is only imitation of the reality and nothing but prediction. In order to get close to the real phenomena, it is necessary to conduct proper experiments and verifications at various phases, so as to verify consistency between the simulation and the reality. I believe that repeating such processes will lead to realization of highly accurate simulations and also be helpful for improving the speed and efficiency of drug design process.

About the relation between experiments and simulations, I suggest that young researchers should understand that simulations are not performed to agree with experimental results. Models themselves are based on physics and

built up purely from the theory. My remarks "experiments are important" doesn't mean that simulations results are to be adjusted to make them consistent with experimental results conveniently but means that we find some parts of the reality that the model can't reproduce. Experiments are, so to speak, something like "checking the answers" and are important to find out why the simulation results do not agree with experimental results, which is very important and the researchers must be



done for in the simulation research.

Thinking of those things, I'm going to reflect the major theme that I have been in charge of in order to research outcome, adding the simulations to respond to new questions while brushing up methods and techniques. The simulation results are given to the researchers for the next experiments but the beneficial of the results is in my hands. So, if we can make use of them for the improvement of the future analysis techniques, or acquire hints to understand interesting characteristics of proteins from them, the research will be even more meaningful and interesting. Not only using the methods that have already been established, but also making them improving with our knowledge is fantastic. In some cases, the absolute binding free energy is not the only way and there may be other techniques by which better results are obtained in combination with the absolute binding free energy. Algorithms can also be improved based on a new viewpoint. We have mainly focused on speed-up but I believe there must be other approaches that we can

attempt. For instance, the variation of effective candidate compounds can be widened by changing the approaches for the activity prediction. They may be worthless in the immediate future but I believe it is also our important mission to propose the ideas and theories that can be utilized in the next generation or next challenge.

Shinoda : Last year, we performed approximately 1,000 MD simulations in an antigen- antibody system containing water with different initial velocity, and analyzed binding process and structure, water dynamics, and so on. We continue the analysis now, and our job for this year is to summarize the analysis results, I suppose. Dr. Yamashita has just mentioned that "GROMACS" has come to run in the K computer. I actually have received its benefit, which made my job much easier. However, as the MD computation was tough work, we paid great attention to it. In particular, the K computer in the early stage was not always capable of giving 100% of the expected data. It caused some problems such as half-way system freezing and data crashed by errors. At

last conducting nearly 1,200 simulations, we could acquire approximately 1,000 data out of them. Further, data being obtained successfully, we needed to check if

they were complete being. Anyway, as it was my very first time to perform such a massive computation at the beginning, we had no idea which part to check in the data, so as to draw out data sets bringing outcome. Finally, we managed to finish checking, and then at least approximately 1,000 of them were all right. The analysis still going on, we've found that the time of binding, duration and frequency of the binding vary in each case and that some do not even bind. Thus we have observed various behaviors. I presume that such a massive simulation had never been performed for a binding process of antigens and antibodies so far. In this sense, we carefully examine the dynamics of the binding process from these results.



Wish to make use of the research results for medical service

● Look forward five and ten years later, what researches do you wish to work on in the future?

Yamashita : Thinking of the cycle of a research, it usually takes about five years obtain answer of the research. So, ten years are not a long period. As I said at the beginning, being interested in physical mechanisms, I am not going to continue only biology researches. However, because molecular behavior in proteins and cells have been clarified one after another and the biosimulation is making great progress, I find this field very interesting. In this sense, in the future we are going to focus not only on one single protein but perform simulations for a protein complex and focus on proteins playing special roles. I continue exploring interesting physical phenomena based on the results of such

researches. Of course, as long as we treat biomolecules, we cannot ignore actual medical service so I try to perform a simulation research to support it. A research to clarify the cause of a disease is a good example. After all, the clues seem to be hidden in various diseases. I think slight difference in proteins causes unexpected serious symptoms in some cases. I am sure that it will be more important in the future to reveal what small errors in design or assembly should cause at molecular or atomic levels. I want to contribute to that.

Shinoda : At present I have only treated only one kind of antibody. I want to is a kind of be a workman, or a specialist who knows all about simulations of antibodies in five years. Experts of the quantum chemistry would find errors intuitively looking

at the molecular structure, and said to me "there's something wrong with it. I feel something strange in this result". Surprisingly I asked him why he knew that, and he said "you'd see it putting yourself in the molecule's position".

I wish I would say, "Make it little moved, and we'll get an antibody to bind better" only looking at visualized images of molecular simulation. Further, through the research of antibodies, I do my best to clarify the mechanism of important phenomena, which contributes to the medical service. Of course I work on researches on the computational science with the computer simulation technique, however at the same time I have a feeling to face the biology or other approaches such as statistical dynamics, which is still a vague idea.



Realize molecular dynamics simulations of huge biological systems with K computer

Highly parallel molecular dynamics simulation software "GENESIS"

RIKEN Theoretical Molecular Science Laboratory, Chief Scientist,
RIKEN Advanced Institute for Computational Science, Computational Biophysics Research Team, Team Leader
Yuji Sugita

RIKEN Advanced Institute for Computational Science, Computational Biophysics Research Team, Research Scientist
Jaewoon Jung

RIKEN Theoretical Molecular Science Laboratory, Research Scientist
Takaharu Mori



A joint research team consisting of Yuji Sugita, the team leader of the Computational Biophysics Research Team of RIKEN Advanced Institute for Computational Science, and his colleagues, has released the molecular dynamics simulation software, "GENESIS", for massively parallel supercomputers. It was released as an open-source software on May 8, 2015. "GENESIS" has successfully realized high parallel efficiency by introducing their original computational algorithms suited to the architecture of the K computer. It has also enabled us to perform a fast molecular dynamics simulation for a system in cellular environments, consisting of 100 million atoms, which could hardly have been simulated so far. Moreover, it allows simulations of biomolecules like proteins, cell membranes, glycans, and nucleic acids at the molecular level. Therefore, it greatly attracted researchers' attention. It is expected to be utilized widely from basic studies to drug discovery in the future. We interviewed the developers about the process of development of GENESIS and key features.

The molecular dynamics software which makes full use of the maximum performance of K computer

—When did you start the development of "GENESIS"?

Sugita We started it around 2010, before the emergence of the K computer. The computational method to simulate motions of molecules according to the laws of physics is called the "Molecular Dynamics (MD) method". A number of MD simulation software has been developed all over the world so far. If you try to apply conventional computational algorithms to a large-scale molecular assembly system with a large number of CPUs like the K computer, the computation speed is easily saturated due to the significant communication between CPUs. Therefore, we started to develop a new MD simulation program that is capable of fully utilizing the maximum performance of the K computer. However, to be honest, it was obviously not easy to develop such software, and barriers

to start the development were really high. Moreover, in order to create software that can be used for research, various functions should be introduced. I often felt it was a hopeless task during the development. I was afraid that Dr. Mori and Dr. Jung, who played a key role in the development, had to continue their work without clear prospects. We had no idea when we could reach the goal.

Mori First, we made efforts to reproduce the same energy and force values as existing software. If we cannot obtain the same results, it means "GENESIS" has errors somewhere. After that, simulation methods such as temperature and pressure control algorithms were introduced. The algorithm that I introduced first was actually wrong. The temperature control was okay, but there were a lot of troubles in pressure control. Anyway, we solved

the problems one by one while keeping studying. Because such basic algorithms are prerequisites for research, we planned to do basic work prior to developing new methods.

Jung I had been involved in quantum mechanical/molecular mechanical (QM/MM) calculations before I joined this project. So I didn't know much about MD simulation. I had to learn about the theories of MD one by one, and wrote the source codes from scratch. It was really attractive for me to use the K computer, one of the most advanced computers in the world. This was my motivation to develop "GENESIS". I was sure that I would have a chance to experience new things I had never done. I expected to make new achievements, and thought that the new research topic was a good challenge for me.

ATDYN and SPDYN

—When did you overcome the barriers, and when did the development work with great progress?

Sugita I think it was since the development of SPDYN got off the ground. "GENESIS" consists of two MD simulators named

ATDYN and SPDYN, and analysis tools. In ATDYN, where the atomic decomposition method is introduced, interactions to be calculated are distributed over processors for parallelization. On the other hand, SPDYN uses an algorithm that divides the

entire space into small subdomains and cells by the space decomposition method. In SPDYN, each subdomain is assigned to a processor, and interactions are computed in each subdomain. In other words, ATDYN divides interactions simply by the number

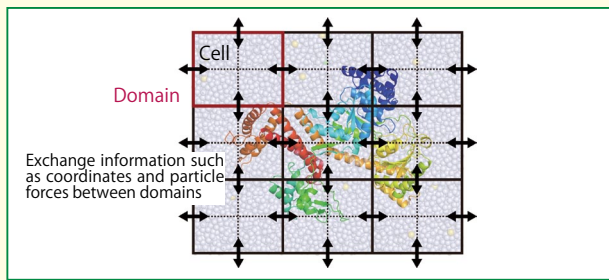


Fig. 1: Domain decomposition method in SPDYN

In the domain decomposition method in SPDYN, the whole system (box) of the simulation is divided into subdomains (solid line frames) and cells (dotted line frame). In parallel computation by the MPI protocol, each MPI process is assigned to each domain, and information on particle like forces and coordinates and particles moving across the domains are communicated (Arrow).

of CPUs, while SPDYN divides data according to the space. Since, data communication is performed mainly between neighboring domains in SPDYN, communication costs are greatly reduced. Therefore, MD simulation could be performed faster for larger systems with SPDYN. We obtained efficient data partitioning by excluding unnecessary data. To develop SPDYN, I guess that there had been a lot of problems, but Dr. Jung did a really great job.

— **Another important function of "GENESIS" is the Replica-Exchange Molecular Dynamics method (REMD), isn't it?**

To establish a higher-performance MD method

— **"GENESIS" was released as open-source software from May 8, 2015.**

Sugita Because GENESIS has been updated frequently, we have to choose the best timing of release carefully. GENESIS has the top benchmark performance in the world, and we published a paper about it. At that time, we thought it was the best timing for the first release. Our future goal is to improve "GENESIS" up to the global standards of MD programs, in other words, the representative MD simulation software. In our laboratory, we expect that individual researchers add new functions to "GENESIS", and utilize them in their own research. Eventually, we will update "GENESIS" by incorporating those into the future release version.

Mori I am now doing application research on membrane proteins using "GENESIS", and also developing "GENESIS"

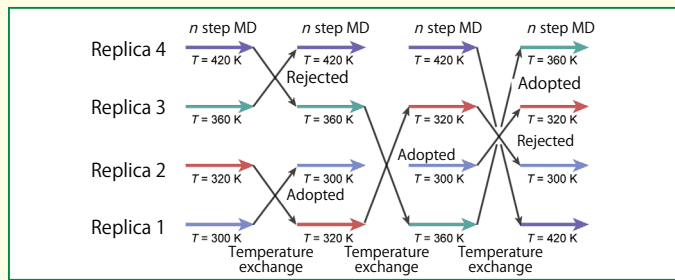


Fig. 2: Replica exchange molecular dynamics (REMD) method

In the REMD method, copies (replicas) of the target system are prepared, and different temperatures are assigned to each replica. MD simulations are performed in parallel, and temperature exchange is attempted at certain intervals. If exchange is accepted, temperature is exchanged, and if rejected, the computation is continued without the temperature being exchanged.

Sugita In the REMD method, simulations of different conditions are mixed by preparing a number of copies of the target system. For example, simulations at different temperatures are performed in parallel, and temperatures are exchanged between replicas at a certain frequency. Exchanging temperatures enable us to search various structures of proteins in the simulations. Of course, the conventional MD may be able to do a similar structure sampling, but it requires much longer computational time. Presumably, it is difficult even with the K computer. In this respect, the REMD method enables us

to obtain a lot of structures in the same computational time. In the conventional MD, proteins often stay in states close to the initial structure at low temperature, while REMD helps us to search protein conformations in different states by mixing high temperature simulations. In addition, "GENESIS" is capable of changing various parameters such as pressure, surface tension and biasing forces as well as temperature. Dr. Mori worked on the development of the REMD method with great efforts.

on the side. To do my application research, I have to introduce new algorithms into "GENESIS" that are not available currently. I think "GENESIS" should be developed gradually by doing this.

Jung As for me, I'm developing an even faster and user-friendly "GENESIS" that can be used not only for the K computer but also for usual PC clusters. For example, the development of "GENESIS GPGPU version", where graphic processing units (GPU) are used for MD simulation, is going to be finished soon. It makes the MD simulation much faster. Furthermore, a project for the post-K computer has already started. "GENESIS" has been selected as one of the target applications of the post-K. A vendor and a software developer are jointly doing a co-design that allows evolution not only of the software but also of machine

architecture at the same time so as to realize a faster "GENESIS".

Sugita In the future, we are planning to incorporate useful functions that the present "GENESIS" cannot do, and also add original features to "GENESIS", in other words, greater advantage over other MD programs. Furthermore, we aim to speed-up "GENESIS" even for machines other than the K computer. Improvement in computation speed of "GENESIS" for various hardware platforms will increase the number of "GENESIS" users. We believe "GENESIS" will be widely utilized for not only basic research but also industrial applications such as drug discovery.

After the interview, the interviewees answered questions from and talked freely with young researchers and students. Visit the URL below for more information.



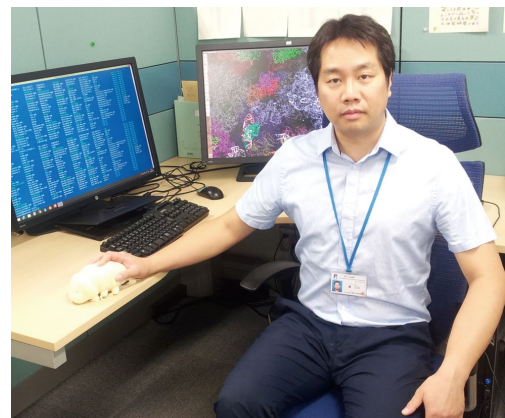
After SPECIAL TALK



Whole bacteria cytoplasm modeling and all-atom molecular dynamics simulations

RIKEN Theoretical Molecular Science Laboratory

Isseki Yu



When you hear that "Water occupies approximately 70% of the whole space in a cell", you may imagine an environment with sufficient space but actually, it is not. The space where 30% of the whole volume is occupied by molecules except those of water is extremely crowded environment, and a number of proteins or small molecules are packed in it like a crowded train. The question of how the proteins and small molecules move and interact each other in such a crowded environment is important for the fields of biology and drug design. However, it is difficult to measure the dynamics of molecules in cell with experimental ways, and therefore its details have not been clarified.

We aim to elucidate the dynamics of biomolecules in cell at the atomic level by large-scale molecular dynamics (MD) simulation using the K computer (simulation to reproduce the dynamics of molecules in detail by calculating forces that act on atoms, and repeatedly solving equations of motion for them) Specifically, we simulate cytoplasm of bacteria called *Mycoplasma genitalium*.

The constructed cytoplasm models

consist of 10-100 million atoms. This ultra large scale model was developed by researchers of RIKEN (Takaharu Mori and Ryuhei Harada) and research collaborator in U.S. (Prof. Michael Feig, Michigan State University)^[1]. Since all genes of mycoplasma have already been understood, the types and structure for most of the proteins can be estimated. Further, since omics data for other macromolecules and metabolites are available, we can also estimate types, structure, and concentrations of them in cell. In this way, we have successfully reproduced biomolecules (except cell membranes and DNAs), and their concentration in cell at the "atomic level", almost perfectly using various experimental data. We are simulating the "whole cytoplasm model" developed in this way with the MD simulation program "GENESIS" (developed by researchers of the Advanced Institute for Computational Science^[2]) on the K computer. GENESIS is the program into which various new technologies are incorporated so as to enable super large-scale parallel MD simulation. Its efficiency does not decline even when hundreds of thousands of CPU's run simultaneously^[3]. In other words, our

study was made possible by combining the computation power of the K computer with the high efficiency of GENESIS.

Presently, the simulation itself has almost been completed, and now we are analyzing how the biomolecules move in cells within a time period of around 100 nanoseconds. We have confirmed that the speed at which proteins diffuse corresponds well with experimental values. In addition, we have obtained new insights for the structure of enzymes and dynamics of metabolites in cells, that can never be seen in a dilute solution environment in vitro. We intend its application to the field of drug design in the future, and are planning to perform longer simulations so as to investigate the interaction between proteins, and binding of metabolites and enzymes in cells.

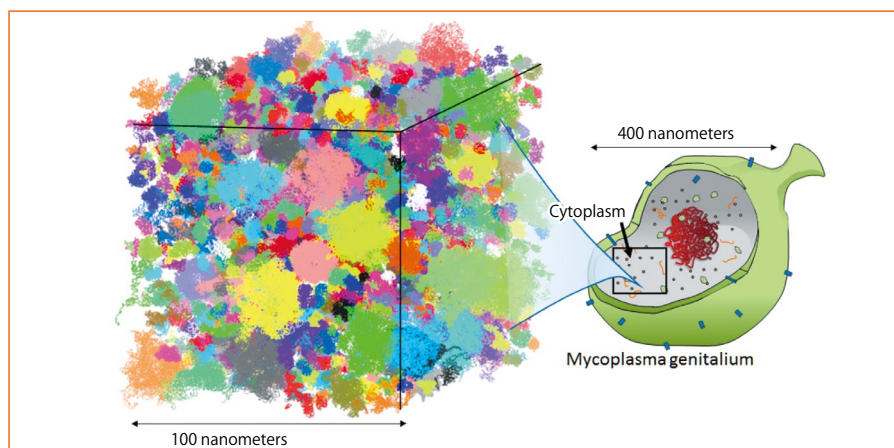


Fig. : Right: Schema of *Mycoplasma genitalium*. Left: Atomic model of mycoplasmal cytoplasm. Each macromolecule is indicated by different colors. One side is approximately 100 nanometers in length, and consists of 100 million atoms including water molecules.

[1] Complete Atomistic Model of a Bacterial Cytoplasm for Integrating Physics, Biochemistry, and Systems Biology *Journal of Molecular Graphics and Modelling*. Michael Feig, Ryuhei Harada, Takaharu Mori, Isseki Yu, Koichi Takahashi and Yuji Sugita. *J. Mol. Graph. Model.*, 58, 1–9 (2015).

[2] GENESIS website
<http://www.riken.jp/TMS2012/cbp/en/research/software/genesis/index.html>

[3] GENESIS: a hybrid - parallel and multi - scale molecular dynamics simulator with enhanced sampling algorithms for biomolecular and cellular simulations.
Jaewoon Jung, Takaharu Mori, Chigusa Kobayashi, Yasuhiro Matsunaga, Takao Yoda, Michael Feig, Yuji Sugita.
WIREs Comput. Mol. Sci., (DOI: 10.1002/wcms.1220)



Modeling human for understanding human

Digital Human Reserach Group,
Human Informatics Research Institute,
National Institute of Advanced Industrial Science and Technology
Akihiko Murai



We create a human model on a computer, and attempt to understand human beings by estimating their state based on measured data such as motion data. Since it costs a great deal to perform computation for a very detailed human model that consists of an enormous number of elements, high-level parallel computation by, for example, a super computer is needed.

One type of information that represents the state of human beings is somatosensory information such as muscle activity. We successfully estimate the muscle activity by utilizing measurement with optical motion capture, a floor reaction force plate, an electromyograph and a human model. Regarding a human being as a rigid body link system driven by wires enables us to apply a kinematics/dynamics computation, which is developed in the field of robotics. We create a human musculoskeletal model by processing the data obtained by CT-scanning an actual skeleton sample (published as AIST adult male skeleton shape data), and positioning endpoints and passing points of muscles, tendons and ligaments based on the anatomical knowledge. Joint angle, floor reaction force, joint torque, and activity of each muscle in the whole body, are estimated by solving kinematics, dynamics and optimization

problems that minimize the quadratic sum of muscle tensions based on this model.

We develop an application known as "Magic mirror". This system realizes real-time estimation and visualization of muscle activity, and is capable of performing high-speed estimation at 60 frames per second. A subject moves his/her body in front of a display, and an image of the subject captured by a video camera and a musculoskeletal model that indicates muscle activity are superimposed on the display. Since this makes the subject feel that he/she is seeing his/her own muscles through his/her clothes, this application is called "Magic mirror". This "Magic mirror" has potential for various applications such as quantitative diagnoses of movement disorder, confirmation of rehabilitation effects, support of sports training, and so on. Conventionally, advice in these fields are given based on motion observation and empirical rules. Visualization of muscle activity allows quantitative understanding of the state of the subject, and further, the real-time visualization allows biofeedback.

We develop a neuromuscular skeleton system with an anatomical neural circuit model built on a musculoskeletal model to clarify the human motion control system. The somatosensory reflex, which

is a primitive part of human motion, is modeled, and muscle activity after a time-delay by nerve signal transfer is output when proprioceptive sensation is input. Information such as muscle length and tension, which are the input and output of this system, is computed by the above kinematics/dynamics computation, and the network between input and output is identified from experimental data. This system realizes simulation not only of a somatic reflex such as the knee jerk seen in humans, but also of reactions to tripping seen in human. It is also expected that understanding the relationship between the sensation and the kinetics allows enhancement of kinetic performance, which leads to development of sportswear and shoes that realize such enhanced performance.

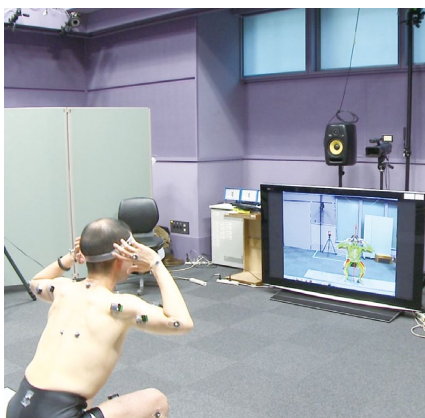


Fig.1 : Real-time muscle activity visualization system "Magic mirror". The color changes from green to red when muscle is activated.

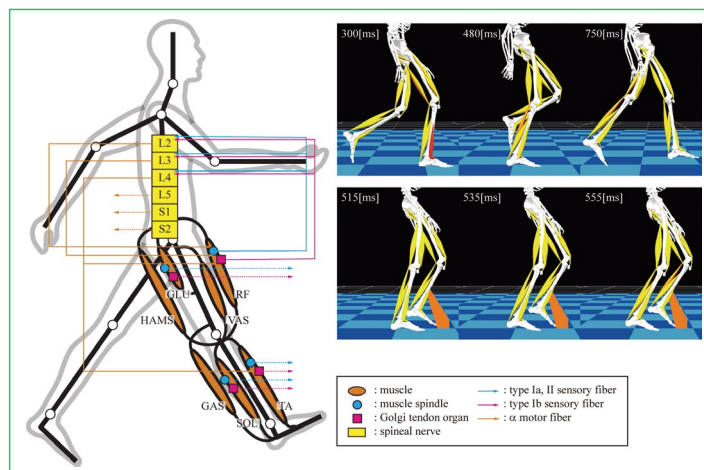


Fig.2 : Neuro-muscular skeleton system (left). Simulation of gait and tripping(right).



Neural control of human quiet standing and biped walking

Movement fluctuation and its degeneration caused by neurological disease

In the current era when humanoid robots have reached the pinnacle of their development, many readers would not be much impressed by a mathematical model that simulates physical motion in human beings and is able to stand straight or walk on two feet. However, for us, who study brain and neural control (motor control) of human biological motion, there seems to be considerable difference between the appearance and kinetic properties of the motion of humanoids and humans.

When we maintain upright posture and stand still, the posture of the body moves infinitesimally. This is called postural sway during quiet standing. The upright posture is something like an inverted pendulum that is mechanically unstable, and the central nervous system of human beings stabilizes it by generating muscle forces so that the pendulum does not fall down. Given this perspective, one may think that "the larger the postural sway amplitude, the lower the stability is". However, we have proved that this intuition may not be necessarily correct. We are studying the standing posture and walking movement of patients who suffer from Parkinson's disease, which is a neurological disease accompanied by movement disorder, in collaboration with the Department of Neurology, Osaka

University Hospital and Toneyama National Hospital. The postures of some of these patients become unstable, and they tend to fall down easily. Comparison of the postural sway (foot pressure center CoP variation) of such patients and that of healthy young persons has revealed that the postural sway of healthy young persons is greater than the other, contrary to intuition (see figure). In other words, the posture of a healthy young person is flexible, and that of the above patients is rigid.

We have built a mathematical model of postural control (computer simulation model), and attempted to explain the results of the experimentally observed behaviors by the logic of mechanical dynamics and systems science. As a result, it was found that the human brain does not always continuously perform control to pull back displacement from the upright posture and there are time zones in which a human being yields its position to gravity intermittently without controlling it, which allowed us to elegantly explain the relatively large postural sway seen in healthy young persons. In fact, assuming continuous control allows us to explain the small postural sway seen in the above patients. Further, in continuous control, even a small variation in control parameters of the

brain (as a motion controller) may make the upright posture unstable, whereas we clarified that in intermittent control, the upright posture can be stabilized in a robust manner even when the control parameters vary greatly, and that intermittent control is much better in energy

efficiency. The latest study revealed that brain/neural control of biped ambulation can be explained in the same way as that of standing posture.

It is often said that "flexibility of vital functions is higher in human beings than in artificial devices such as robots". Flexibility leads to kinetic fluctuation. We believe that a basis for health science will be obtained by clarifying mechanisms of generation and loss of this movement fluctuation. Can a detailed and large-scale mathematical model of the central nervous system, muscle and skeletal system of humans built in a supercomputer reproduce flexible motion similar to that of a human being (motion fluctuation)? Can the degeneration of motion fluctuation observed in patients be reproduced by the variations in the control function that are thought to affect Parkinson's disease patients? Our challenge using the K computer continues.

Department of Mechanical Science and Bioengineering,
Graduate School of Engineering Science, Osaka University

Taishin Nomura
Chunjiang Fu

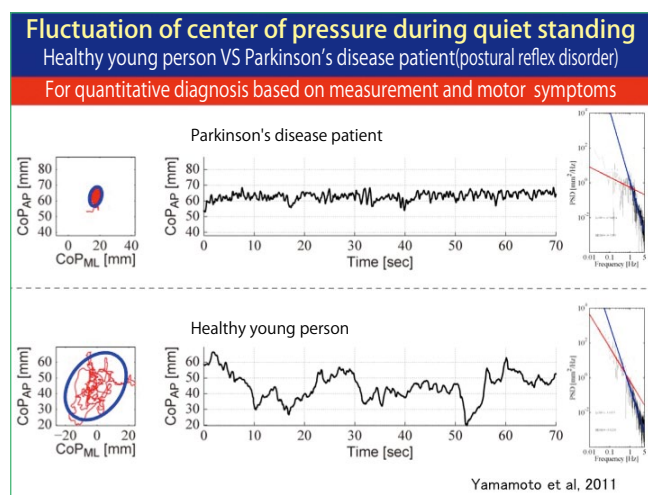
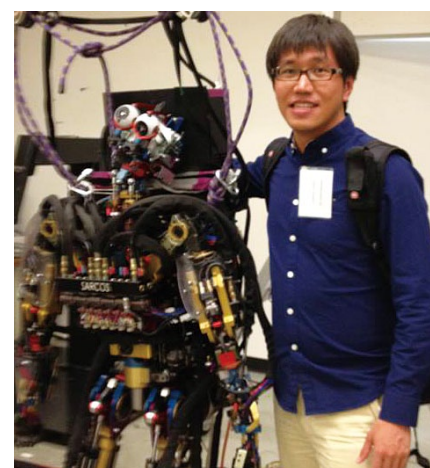
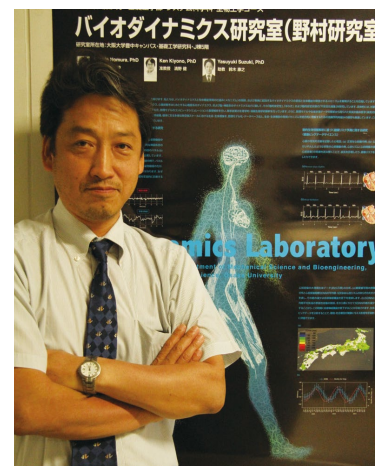


Fig. 1: Fluctuation of center of pressure during quiet standing. Upper figure: Postural sway in a Parkinson's disease patient with postural reflex disorder. Lower figure: Postural sway of a healthy young person.





Lecture at Faculty of Human Development, KOBE University “On interdisciplinary studies - the integration of biology, physics, chemistry and mathematics”

Deputy PD, RIKEN HPCI Program for Computational Life Sciences Yukihiro Eguchi

I gave a lecture titled “On interdisciplinary studies” to new undergraduate students in the Faculty of Human Development, Kobe University on 19th and 26th in June, 2015, as part of “Invitation to Human Development” whose keyword is “interdisciplinary studies”. I have been giving the opportunity to deliver such lectures over the past four years. Through the main theme --- “Rapid improvement of technologies such as

supercomputers and DNA sequencer has promoted the integration of biology, mathematics, physics and chemistry in the 21st century” --- I had talked about interdisciplinary studies until last year. In the last lecture a student asked me, “My major is liberal arts but it is said that majoring in science is helpful in the business society. What is your opinion?” Responding to this question, I decided to emphasize the importance of harmonious development of liberal arts and science, and the need of interdisciplinary studies in order to solve real-life problems in this year, the outline of the lecture being almost the same.

The concept of “Designer baby” by genome editing technology has become a big worldwide issue now. Most science researchers believe their studies are done for the good, otherwise, they could not work creatively and progressively. The idea itself is natural and correct.

But, whether further studies should now be done and whether the results can be accepted by society is another matter. Sufficient consideration should be given from the viewpoints of both liberal arts and science.

I suggest to students that “interdisciplinary studies” be thought of on a two-dimensional plane, X-axis shows academic fields including liberal arts and science, and Y-axis shows depth of exploration of a special field, giving them examples of T-shaped persons and II-shaped persons. Computational life science is a field engaged in by II-shaped persons with in-depth knowledge of computational science and life science, and it requires also knowledge of a liberal arts field such as Bioethics. In the lecture, I had unexpected feedback from a student, saying “A cuneate person would be good”. I really hope that young students will build new creative interdisciplinary research fields in the future.



The building of Faculty of Human Development, KOBE University



Third Advisory Committee Meeting

Planning and Coordination Group
RIKEN HPCI Program for Computational Life Sciences

The third advisory committee meeting was held at the RIKEN Tokyo office on July 3rd, 2015.

This is a very important committee with participation of experts from scientific and industrial fields inside and outside Japan giving advice on research content and activities in order to assist the Strategic Programs for Innovative Research Field 1, “Super Computational Life Science”. In view of advice from the first committee (in January, 2012) and the second (in September, 2013), all

the research achievements until now were presented, and then a way of taking these final achievements into the next project was mainly discussed because the current project has mostly completed its term. First, general responses to the committee's recommendations were made by the deputy PD of Research and Development, and then there were presentations delivered (final goal, progress, achievement and further future development) by each leader of four research groups. A presentation (research and development assistance, widespread research influence, network of personal contacts and manpower training) was finally made by the deputy PD in charge of Establishment of the Research System for Computational Science. After each presentation, there were many questions and answers, and proposals from members of



Photo of the meeting

the committee, which brought the committee much benefit. An important task that we must challenge from now until next March is accelerating further research for the final achievement, and successfully taking it into the next project. At the committee meeting, precious advice was given for resolution of the challenges. We would like to take this opportunity to thank all members of the committee, Dr. Mitiko Go, Prof. Peter Kohl and Dr. Masaharu Kanaoka.



Group photo

NEXT STAGE

As a "Flagship 2020" project, the Ministry of Education, Culture, Sports, Science and Technology is promoting the development of the post-K computer, which is a machine that will succeed the K computer. As a national technological infrastructure platform, nine priority issues have been set as social and scientific themes to be addressed.

In the category "Achievement of a society that provides health and longevity", organizations were selected to implement two priority issues: (1) "Innovative drug discovery infrastructure through functional control of biomolecular systems", and (2) "Integrated computational life science to support personalized and preventive medicine". Both projects launched in Feb. 2015. What epoch-making results will priority issues (1) and (2) bring to life sciences? We interviewed Dr. Yasushi Okuno and Dr. Satoru Miyano, who are supervising these issues.

Innovative drug discovery infrastructure through functional control of biomolecular systems

Priority issue 1 Supervisor
Senior Visiting Scientist **Yasushi Okuno**
RIKEN Quantitative Biology Center

The Post-K computer will realize ultra-high speed molecular simulations to achieve not only functional inhibition but also functional control of many biomolecules including factors that cause side-effects, in order to discover safe and highly effective drugs.



Aiming at full-scale use by 2020, development of the post-K computer, a next-generation supercomputer, has started. 2020 is the year of the Tokyo Olympics, and we are going for the gold medal of science and technology with the post-K as our number one weapon.

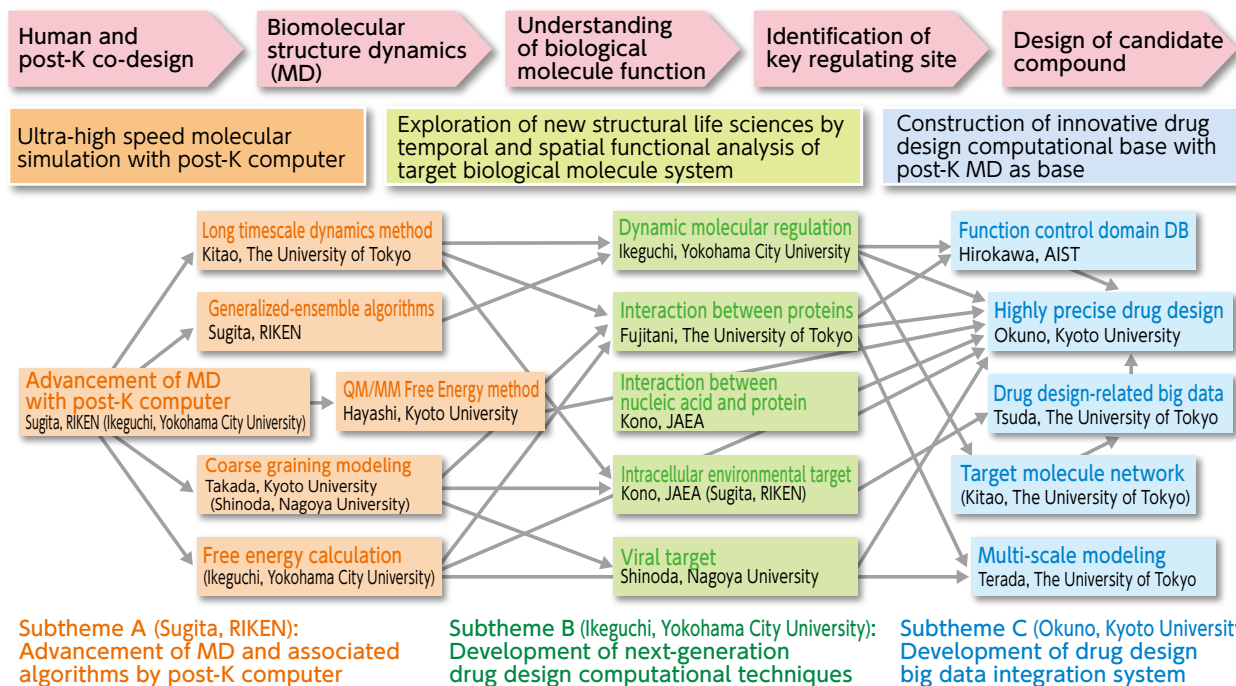
The post-K development focuses on social as well as scientific issues, and we are developing a computational technique (application) that would bring innovation to nine priorities such as drug development, medicine, weather, manufacturing, space and so on. I am in charge of supervising "Innovative drug discovery infrastructure through functional control of biomolecular systems", which is one of these nine priorities. At this opportunity, let me introduce the study goal and an outline of our drug development with the post-K.

Over the past dozen years or so, the number of approvals of new medicines remains at the same level (around 20 items / yr) in the drug industry while research and development expenses continue rising, which has become a serious bottleneck. Therefore, "to accelerate new medicine

development while holding down development cost" has become a most important issue for drug development and the medical field. Our final goal is to contribute to overcoming obstacles that the drug companies encounter, by using the overwhelming power of the post-K. In the post-K, we achieve a computational speed (molecular dynamics calculation) to simulate biomolecular behaviors that is dozens of times faster than that of the K computer. It allows us to aim at reproducing *in vivo* molecule behaviors over longer time periods (millisecond level), and promoting drug development simulation for more molecules *in vivo*. Based on such development work, we are exploring a new drug design method to which dynamic control of disease-causing proteins and multiple drug development-related proteins are added. Specifically, we are developing new structural life sciences and a next-generation drug development computational technique (Subtheme B) through the development of a new molecular simulation technique that allows operation of the post-K at its maximum computational power (Subtheme A).

Moreover, we aim to establish a highly precise and ultra-high speed innovative drug development computational base while developing an integrated system for which these core computation techniques are connected along the drug development computation flow (Subtheme C). Use of the drug development computational base with the post-K as a key in actual pharmaceutical scenarios leads to promotion of efficiency in drug development processes. For example, the experimental processes of past drug development are replaced by computer simulations, and this dramatically reduces the development cost of medical supplies and thus medical cost. We are also exploring an innovative approach for new drug development, such as drugs for dynamic functional control of target proteins, epigenetic drug development that targets the macromolecular assemblage of protein-nucleic acids, and systemic drugs based on the simulation of a super large-scale biological system. It allows creation

Organization for direct application of post-K to drug discovery



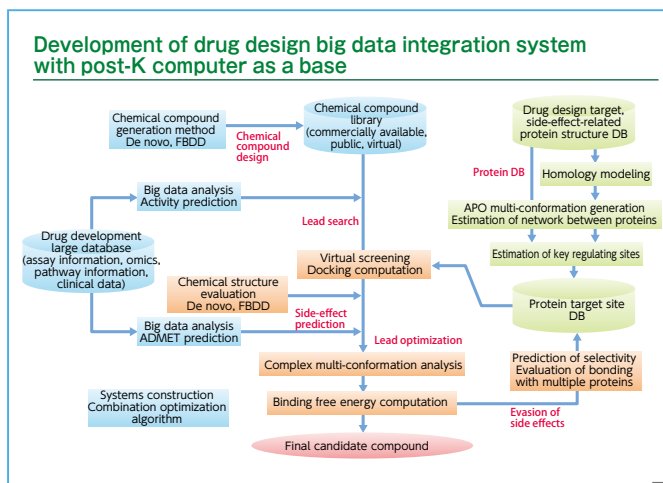
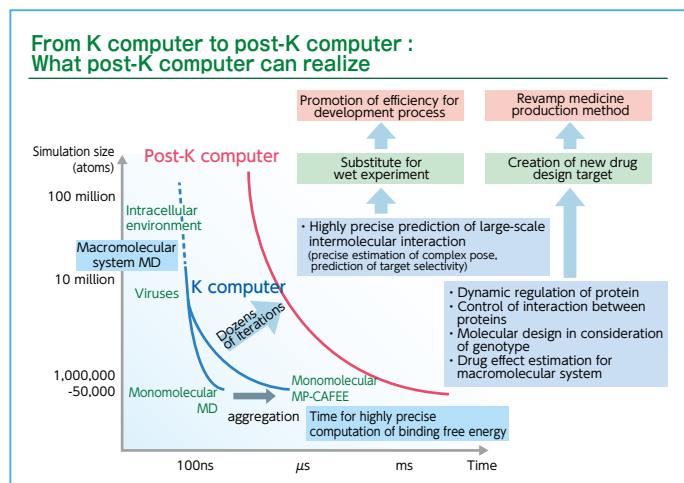
Research Implementation System for priority issue 1

of new medicines that so far were difficult to develop, and speeds-up development of pharmaceuticals. As a result, they are expected to contribute to the needs of patients for new medicines.

Working with the present the K computer, I received a very strong impression that greatly changed my view of computational sciences. The supercomputer, on which the nation put high priority, not only brings industrial benefits and lets us reach the top of world supercomputer rankings, but also

gives a big dream to young people who are the scientists of tomorrow. Indeed, it is often said that supercomputers have the performance of future general purpose computers that will be available 15 years from now. This means that the post-K, which is to be completed in 2020, brings us the future computational environment of 2035. Therefore, we have great expectations of the post-K and a huge responsibility. We, who are involved in the field of drug development, are vigorously

advancing research and development so as to construct of a world-class drug development computational base. We hope you will not simply look at the post-k as a machine available in 2035 evaluated by 2015 viewpoints, but continuously provide us with your prospective and constructive opinions. Your continued support is greatly appreciated for successful development of the post K, which will support continuous development of industry and science in Japan.



See the URL on the right for detailed information on this report.

<http://www.scls.riken.jp/eng/newsletter/Vol.13/nextstage01.html>



Integrated computational life science to support personalized and preventive medicine

Priority issue 2 Supervisor
 Professor **Satoru Miyano**
 The Institute of Medical Science, The University of Tokyo



Creation of the new paradigm called integrated computational life science supports personalized and preventive health care, and the super computer supports a healthy, long-life society.

Disease appears as a phenomenon of disorders in related organs. As a background of disease, we need to consider our genome, which is often called “the blueprint of life”, and various data of molecules such as epigenome, RNA, protein, etc. These are collectively-referred to as “omics” and drive cells for creating biological functions. Further, each cell in our individual has a specific state of cell called “cellular context” that changes with the environmental factors and aging. Cellular contexts vary from cells to cells, and therefore, the functions of organs also depend on their cellular contexts. Moreover, for the genome, which was previously considered not to change through our life, it has been recently reported that genomic mutations accumulate along with aging steadily in our hematopoietic stem cells. Cancer develops in the long organism-spatiotemporal process while being affected by various environmental factors. It is no exaggeration to say that this process begins since a person is born. Not only the variations in population but also everybody has a variety in its life. It may be said that this is the biological reality of the aging society that genomic science has unraveled.

On the other hand, the genome sequencing technologies, which can also produce epigenomic data and RNA sequence data, have already produced genome sequence data in petabyte order. Moreover, high precision clinical data such as imaging or physiological data have also made a large-scale data mountain. In the Grand Challenge Program and the following Strategic Programs for Innovative Research Field 1, the world’s top level simulation technologies had been developed that include the heart simulator UT-Heart that successfully used the full power of the K computer. We had also development of technologies for large-scale analysis of life data that allowed us to analyze all genes and non-coding RNAs of size more than dozens of terabytes. These technologies led us to important discoveries in cancer biology and development of various prediction methods (e.g. for drug resistance, cancer survival) that could not be achieved without the K computer.

Of course, it was a very epoch-making scientific contribution that the multi-scale simulator UT-Heart achieved a full simulation of 1.5 heartbeats from sarcomere

to blood ejection. It should be emphasized that this simulation took 17 hours by continuously employing almost all nodes of the K computer just for 1.5 heartbeats. In cancer research, international collaborations are building a catalog of cancer mutations with frequency more than 5% for major 50 cancer types/subtypes. The world-wide total computing power spent for this catalog building was similar to the power of the K computer. However, in order to understand “personal cancer”, we need to carry out the comprehensive analysis of mutations with frequency less than 1% by whole genome sequencing. This analysis requires, theoretically, 5,000 days by the K computer, an unrealistic analysis. As described above, while human diversity has been identified from various viewpoints, a strong need is to develop technologies to fill the gap between the molecular mechanisms of individual pathologies and the large-scale imaging and physiological data, and further, health and genome information. The above heart

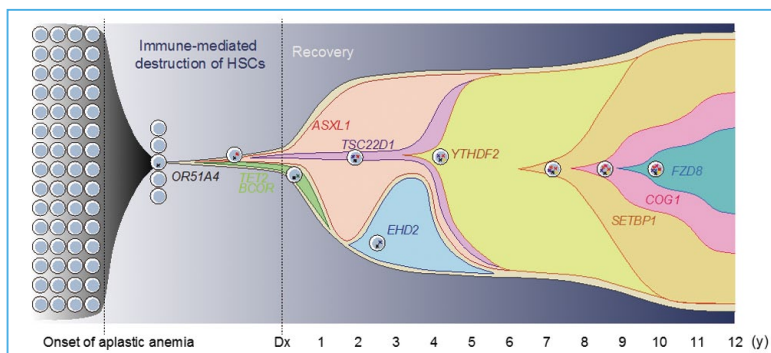


Fig. 1 : Evolution of clone during 12 years when hypoplastic anemia worsened to myelo dysplasia syndrome

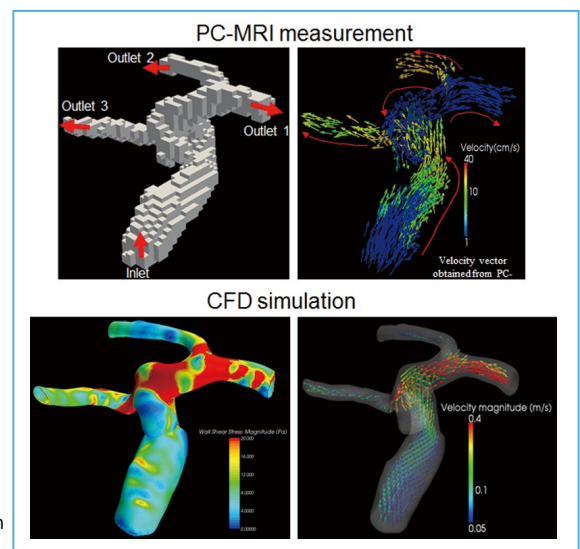
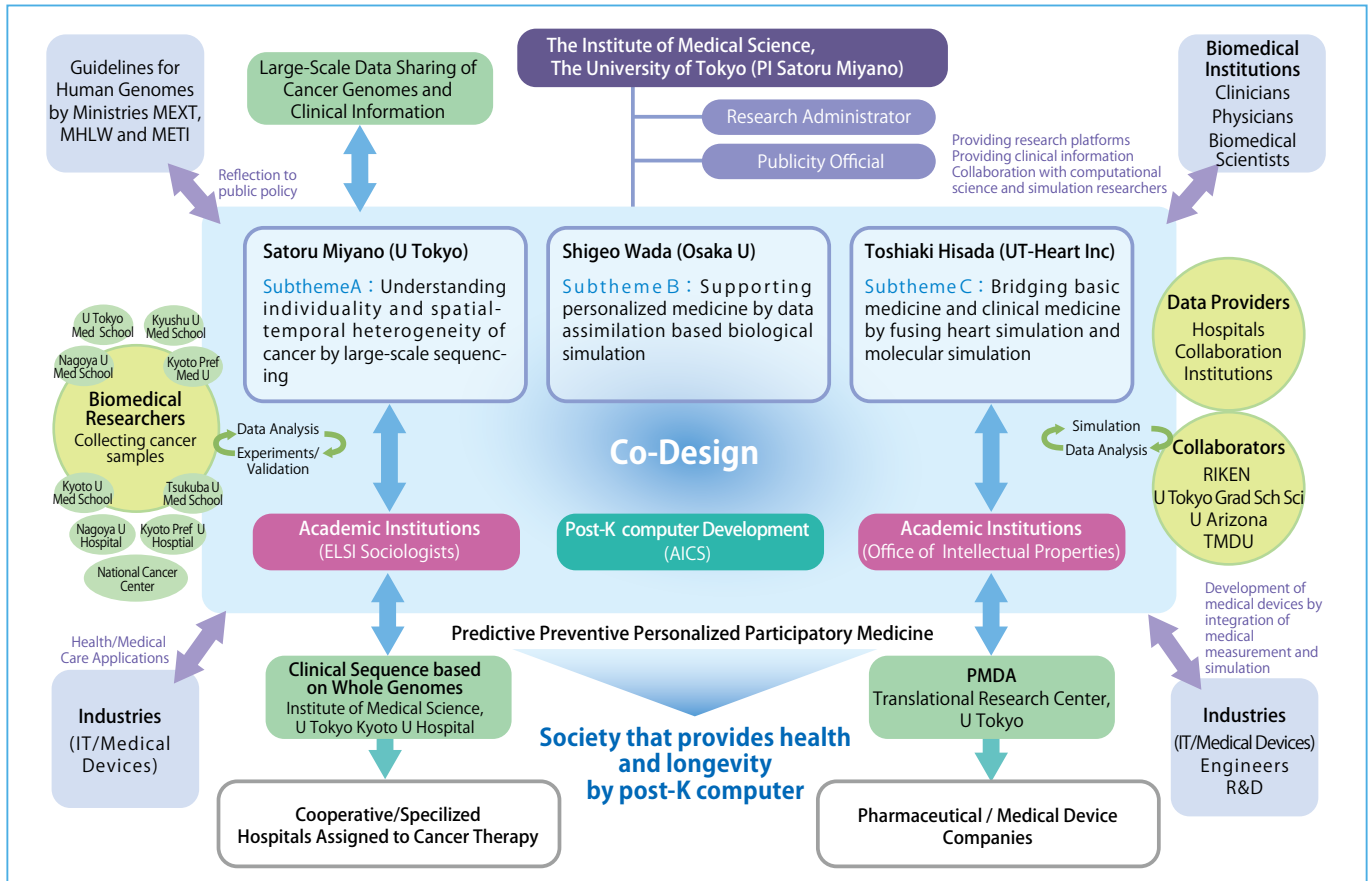


Fig. 2 : Data assimilation biological simulation



Research Implementation System for Priority Issue 2

simulation and cancer genome research are some of the extremely important scientific/ biomedical issues that Japan and humans must challenge. However, even for these cases alone, we need a new computing environment exceeding the K computer, and simultaneously we will face with a big mission that we develop technologies to fully exploit it for overcoming the difficulties. It has been gradually recognized that the complexity of integrative understanding of diseases is far beyond human abilities. This is one of the reasons why development of the post-K computer is inevitable.

In priority issue 2, our goal is to estab-

lish the integrated computational life science that constitutes the basis for personalized/preventive medicine. This requires a methodology for comprehensive understanding of pathological states and exploration of their effective treatments through a view from genome to the whole body, both environment- and organism-spatiotemporally. We consider that this methodology can be realized by "information technology", "application of physics principles", and "utilization of big data" enhanced with the post-K scale computation.

To achieve the goal, we are tackling three subthemes. In Subtheme A, "Understanding individuality and spatial-temporal heterogeneity of cancer by large-scale sequencing" (Supervisor: Satoru Miyano) (Fig.1), we will analyze large-scale data of an unprecedented scale in the field of life sciences. In Subtheme B, "Supporting personalized medicine by data assimilation based

biological simulation" (Supervisor: Shigeo Wada (Graduate School of Engineering Science, Osaka University) (Fig. 2), we are developing a technique to assimilate individual data for a high-level, organism hierarchy integrative simulation. Parallel to the approach based on the large-scale data, in Subtheme C, "Bridging basic medicine and clinical medicine by fusing heart simulation and molecular simulation" (Supervisor: Toshiaki Hisada (UT-Heart Inc) (Fig. 3), we are integrating studies on the molecular cellular level and those on individual organ levels to develop a simulation model for quantitative capture by associating micro with macromechanics.

In the present era when we will witness the advent of the super-aged society, we believe that the solutions by the priority issue 2 will constitute an essential basis to support the healthy and longevity society. Further, for various diseases that occur along with aging, we believe that contributions of the new paradigm, "Integrated Computational Life Science", to the health of citizens should be of great social significance.

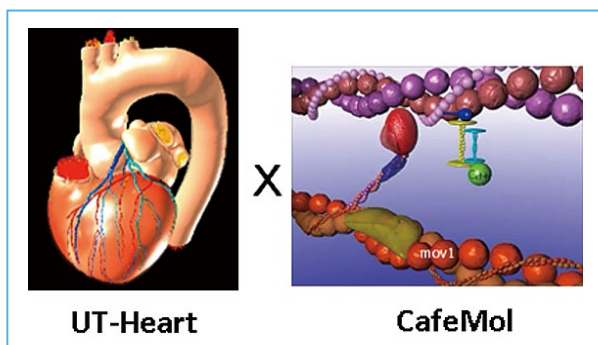


Fig. 3 : Fusion of heart simulation and molecular simulation

See the URL on the right for detailed information on this report.

<http://www.scls.riken.jp/eng/newsletter/Vol.13/nextstage02.html>



RP WS Science and Society pioneered by K computer
Supercomputational Life Science 2015

- **Date** : October 20 (Tue) – 21 (Wed) in 2015
- **Location** : Takeda Hall, The University of Tokyo (Hongo, Bunkyo-ku, Tokyo)
- **Participation fee** : free (Reception fee will be charged)



Read this QR code for more information about programs, application, etc.

The 1st day 10:00 a.m. - 5:50 p.m. International Workshop on Current Topics (6:00 p.m. - Reception)

The 2nd day 9:30 a.m. - 6:25 p.m. The reporting session by Strategic Programs for Innovative Research Field 1 "Supercomputational Life Science"

SM Remote lecture with Kobe University Education Center on Computational Science and Engineering
Foundations of Computational Life Sciences II

- **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)

For more details, please visit our website <http://www.scls.riken.jp/e-scls/2015-2016.html>

News & Events

SP **The 53rd Annual Meeting of the Biophysical Society of Japan**

- **Date** : September 13 (Sun) – 15 (Tue)
- **Location** : Kanazawa University Kakuma Campus (Kanazawa-shi, Ishikawa)

Symposium September 13 (Sun) a.m.

Biosupercomputing opened by next-generation supercomputer post-K

Organizer : Mitsunori Ikeguchi (Yokohama City University)

EX **RIKEN Open House Day at RIKEN AICS (Kobe)**

- **Date** : October 24 (Sat)
- **Location** : RIKEN advanced institute for Computational Science (Kobe-shi, Hyogo)

RP **The Second Project Report Meeting of the HPCI System Including K computer in the fiscal year 2015**

- **Date** : October 26 (Mon)
- **Location** : National Museum of Emerging Science and Innovation (Koto-ku, Tokyo)

SM **CBI Annual Meeting 2015**

- **Date** : October 27 (Tue) – 29 (Thu)
- **Location** : Tower Hall Funabori (Edogawa-ku, Tokyo)
- **Sponsored Session** : October 28 (Wed) 4:00 p.m. - 5:30 p.m.
Chairman : Yukihiro Eguchi (RIKEN HPCI Program for Computational Life Sciences)

SM **Informatics in Biology, Medicine and Pharmacology 2015**
The 2015 Annual Conference of Japanese Society for Bioinformatics

- **Date** : October 29 (Thu) – 31 (Sat)
- **Location** : Uji Campus Kyoto University (Kyoto-shi, Kyoto)
- **Sponsored Session** : October 30 (Fri) 1:30 p.m. - 3:30 p.m.

Integration between bioinformatics and physical chemistry in life science

Wide view point on bioinformatics

- **Tutorial Session** : October 30 (Fri) 3:45 p.m. - 5:15 p.m.
Course on Parallel sequence similarity search program "GHOST-MP"

EX **Science Agora 2015**

- **Date** : November 14 (Sat) – 15 (Sun)
 – the period holding the exhibition
- **Location** : National Museum of Emerging Science and Innovation (Koto-ku, Tokyo)

WS **BMB 2015 Biochemistry and Molecular Biology**

- **Date** : December 1 (Tue) – 4 (Fri)
- **Location** : Kobe Portopia Hotel (kobe-shi, Hyogo)
- **Workshop** : December 4 (Fri) 2:00 p.m. - 4:30 p.m.

New Paradigm Creation in Personalized and Preventive Medicine

- the Gift from Big medical data and Supercomputer -

Organizers : Satoru Miyano (Human Genome Center, The Institute of Medical Science, The University of Tokyo)

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

SP **The 9th K Computer Symposium (tentative title)**

- **Date** : January 29 (Fri) in 2016
- **Location** : Yomiuri Otemachi Hall (Chiyoda-ku, Tokyo)

SP **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)



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.13 2015.9

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