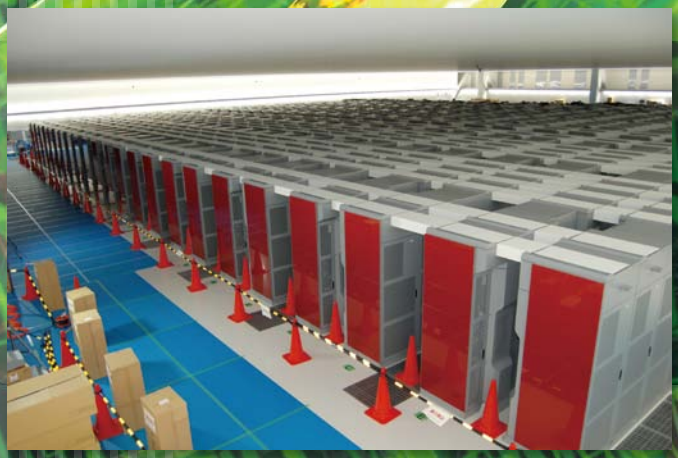
Next-Generation Integrated Simulation of Living Matter
HPCI Strategic Program Field 1 Supercomputational Life Science

BioSupercomputing Newsletter

2011.11 Vol.5

June 2011 Supercomputer "K computer" Takes First Place In World



Preparation status of "K computer" as of June 2011

1st place certificate in the TOP500 ranking

CONTENTS

The time has come for biosupercomputing to get results with the world's No. 1 supercomputer “K computer”, and take up the challenge of "prognostic biology".

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

One third of the software has started tests in the environment of “K computer”.

In 2006, the grand challenge (ISLiM:Next-Generation Integrated Simulation of Living Matter) started with the mission of understanding the super n-body multi-level problem, which is an extremely complex life phenomenon, using K computer's peta-scale computation ability with unprecedented performance, and develop software that will contribute to drug discovery and medical care. Five years has passed, and software development has now reached a new phase to proceed with super-large scale parallelization and machine tuning using “K computer”.

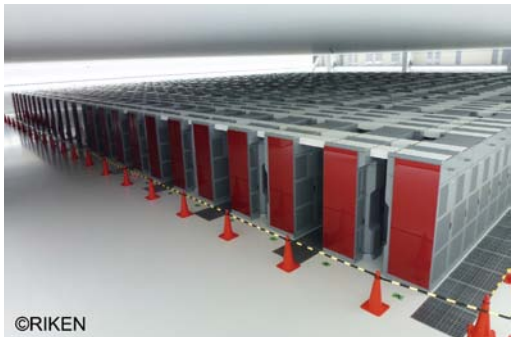
According to the 37th TOP500 announced June of this year (2011), “K computer” has measured 8.162 PFLOPS on the LINPACK benchmark (execution efficiency of 93.0%), and achieved 1st place even though it was still under construction. Currently, our 31 software projects that are to run on this highly acclaimed “K computer” are under development. 11 of these have already passed the development phase for large scale parallelization (8,192 parallel), and are being tested on “K computer”. The "MD core program for large-scale parallel computers (cppmd)" is one of them, has an execution efficiency that exceeds 30%, and reported 1.3PFLOPS as of the end of March. It is required for these software developments to make as much use of the performance of "K computer" as possible. We want to have a high line-up of software that exceeds at least 1PFLOPS when using the whole machine (10PFLOPS), and as for the mean time, one of the softwares has already achieved the target performance. Including cppmd, four softwares (cppmd, ZZ-EFSI, CafeMol, ZZ-HIFU) have exceeded an execution efficiency of 20%, and two (cppmd, ZZ-EFSI) have reached approximately 40% as of now (October 2011). Keeping in mind that we were able to start working on “K computer” from the beginning of this year, performance tuning has progressed at a rapid pace, and performance results are rather satisfactory. However, further improvements are under way in some softwares which have poor single performance values or have good parallel performance but do not have good single performance. Of course, “K computer” itself is a machine still under construction, and compilers and network libraries will be enhanced. These enhancements will probably improve the software performance, but we have no time to waste. The time allocated to us to have priority access to “K computer” is until the end of October, 2012, when its shared services will start. Until then, we must organize calculations and produce results that have a strong scientific value. Therefore, we need to make our best efforts, such as optimizing code, to withdraw the highest performance in the shortest amount of time.

We of course are planning to run on “K computer” all of the 31 softwares that are being developed and get results. But since time is limited, we will have to decide on a detailed plan for how to allocate calculation size and time. There may be 10, or maybe even less, that can be tested using the entire system. Furthermore, we need to debate on the setup, such as reducing calculation size but spending a longer time. One big challenge will be devising a way to maximize our accomplishments by utilizing calculation resources in a limited time.

Grand challenge sows seeds for opening doors to the new generation of life science

In parallel to the grand challenge, the "HPCI STRATEGIC PROGRAM Supercomputational Life Science", started this year. The main mission of the grand challenge is software development which is science-oriented, and not needs-oriented. With our desire to lead the entire life science field towards computer simulation and data analysis, we have created software packages for a wide variety of research fields, such as genetics, biomolecules, organs, and also the whole body. By producing further results and publicizing them, our goal is to lower the threshold of this genre, and thus allow it to be used by people who wish to pursue new research. Therefore, our software lineup does not necessarily consist of software that may be desired by the pharmaceutical and medical industries. On the other hand, for the HPCI Strategic Program, we will maximize the use of high computing resources such as “K computer”, and aim to produce results that will have social value. Therefore, taking the same computational software for molecular dynamics as an example, development on “K computer” is its mission to meet the grand challenge, and as for the Strategic Program, it is expected to get results that will be beneficial to drug discovery. Like "development" and "usage", its use may be entirely different even if the same topic is addressed. It is as though we have planted seeds and are striving our best for beautiful flowers to bloom. With this analogy, we can say it is the job of the Strategic Program to nurture it until it bears fruit, and then to harvest it.

We, in a sense, have been running towards a goal that has been set up without considering the needs of society. However from now on, it is also expected for the project to be applied for actual drug discovery and for development of medical equipment, and it will be required to add new functions, work along those who want use it for collaborative research, and to convene lectures. Through these activities, I believe the results of the grand challenge will be carried on with the Strategic Program. There already are high hopes in the findings by the molecule-scale research team and the data analysis integration research team, especially in the drug discovery field. Attention has been drawn to the fact that this software can already be useful. It may take some more time for development by the Organ and Body Team, but they too are also expected to contribute to drug discovery and medical care. I hope that by adding new functions and combining them based on the software, the software will grow in unpredictable ways. But to begin with, what we must first do is to enhance and perfect the software. With these efforts, there is no doubt we will achieve research results that will surprise the world, namely, the blossoming of a beautiful flower of science.



An array of the quadrillion computer "K computer" in the computer room (left). One chassis is equipped with 24 system boards (right).

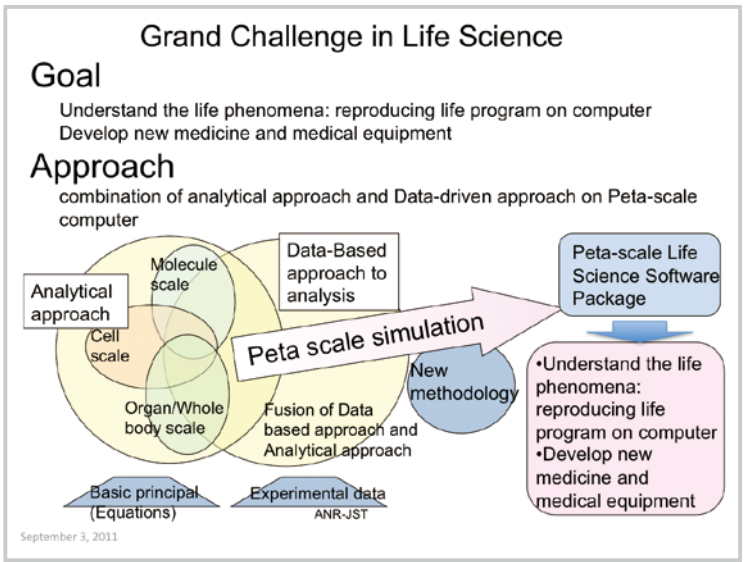
Building of the Advanced Institute for Computational Science in which "K computer" is being installed.

Being able to use "K computer" has strengthened a sense of unity between researchers

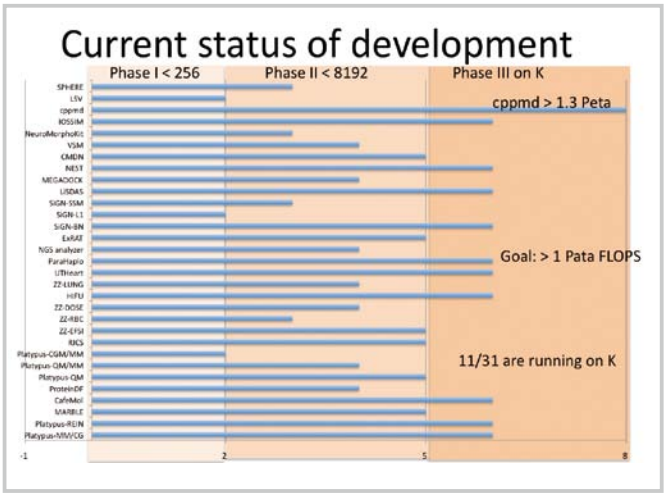
When the grand challenge was in blueprint, an overwhelming majority of people around us felt that, in order to accelerate life science research, enhancing experimental devices would be much more effective than spending money on supercomputer. The grand challenge started with voiced doubts as to whether life science software could produce decent results using a supercomputer. Already at this point, however, one third of the software being developed had started testing on "K computer", and produced results that exceed 10 thousand parallels. The number of softwares with 10% or higher execution efficiency is rising. Of course, this is only stating the performance values for the software program, and accomplishments that have scientific value will be coming henceforth. But we can say for certain that the opinions of people concerned with this project are changing. Expectations in the possibilities of supercomputing are growing, thick and fast.

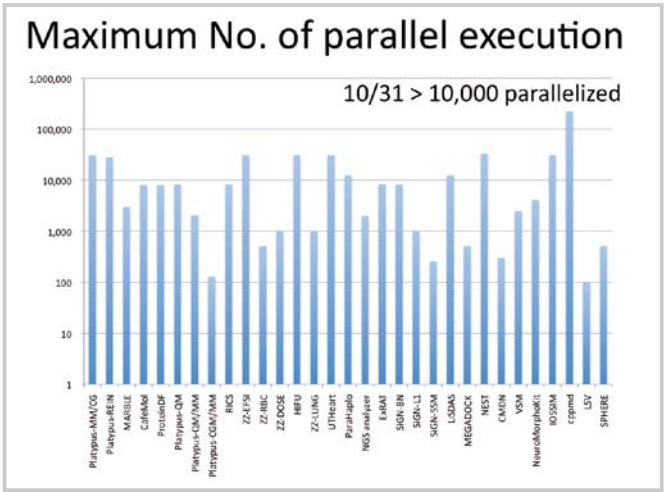
At the same time, the project members have also changed. Since the members came from a wide range of research fields, many only had interests in their own field, and in the beginning when we started where there were no common topics to discuss. Facing the same table, we were able to identify each other's faces and were able to have some understanding among one another, but a dramatic change occurred when "K computer" became available to us. "What can be done in order to achieve high performance with "K computer"?" This question is shared by every researcher involved in the project. The sense that they are together in driving the same project forward with "K computer" at its heart has been born through shared experiences and a common language, and the bond between each member has strengthened. One example can be seen in the mailing list recently founded for young researchers working on "K computer". Passionate messages can be seen there, such as, ""K computer" is now the world best's in hardware, next it's for us to achieve something with it", and "it's up to us to try our best". Once again, it was made clear to me that the project is moving forward with everyone having a high sense of purpose and togetherness.

Our mission is to produce results at the same time "K computer" goes into action. The time left for us is only one more year. The final year should be the most important to make a beautiful flower blossom, and go on to the next phase, "usage". In getting ready for the prospective year, the biggest concern is next year's budget. To secure our budget is also a major challenge.



Aiming at development of software for contributing to the understanding of life phenomena and medical care by utilizing the full performance of the "K computer".





Of 31 kinds of software under development, 11 achieved the scheduled massive parallelization and are being tested with "K computer" (above). There are 10 kinds of software which have achieved 10 thousands parallelization or more (below) (as of October 2011).

What should we do to promote industrial use of sophisticated computer resources and development applications?

Chief Coordinator of Foundation for Computational Science
Masahiro Fukuda (left photo)
Chief Researcher of Urban Innovation Institute and
Executive Board Member and Bureau Chief of BioGrid Center Kansai
Ryuichi Shimizu (right photo)

Why computer simulation is not widespread in private companies?

You two are members of the Preparation working group on promotion of industrial use in the HPCI consortium. First, would you tell us what kind of work you are doing now? Incidentally, Mr. Fukuda was engaged in the development and operation of the Numerical Window Tunnel (NWT), the supercomputer that once achieved the world speed record at the National Aerospace Laboratory of Japan, and then has consistently worked in computational science and contributed to its development.

FUKUDA (Honorifics omitted) **:** Now I belong to the Foundation for Computational Science which was established to motivate Japanese industry by disseminating computational science and simulation technology, and help various private companies with research and development utilizing supercomputers. Our foundation also has a supercomputer for industry, the "FOCUS Supercomputer," which promotes advanced technology in the industry. I hope you make extensive use of it.

SHIMIZU : I'd like to speak about the BioGrid Center Kansai I am involved in. Originally, Osaka University promoted the "BioGrid Project." It aimed at running development applications for bio and medical care in a grid computing environment. In order to get the technology and research products widely utilized in, for example, the industrial arena, we established the BioGrid Center Kansai (NPO). In 2004, we started a testbed and appealed for its use, but initially pharmaceutical companies took no action. We reserved Osaka University's supercomputer, installed applications in it and were ready for running them, but pharmaceutical companies would not use them, saying "we cannot provide our data" or "we are not allowed to connect our network to outside." So, we decided to get results by ourselves. We cooperated with a venture company from Osaka University which began operation and started a Development Project for Creation of an Intellectual Cluster. The year after, the National Institute for Biomedical Innovation asked us if we could do in silico drug discovery with them. Then, the "Drug Discovery Value Chain Project" (We named it "Pharmaceutical Innovation Value Chain") which creates pipeline compounds based on computer simulation was started. It was a time when the trend toward in silico drug discovery was about to take off. That might be also helpful.

FUKUDA : Were applications available at the BioGrid Center Kansai developed by professors at Osaka University?

SHIMIZU : We use various applications. But we basically use a docking simulation (myPresto) developed by Professor Haruki Nakamura, from the Institute for Protein Research, Osaka University, and Mr. Yoshifumi Fukunishi, from the National Institute of Advanced Industrial Science and Technology.

FUKUDA : After starting the project, were detailed calculations done by people from the pharmaceutical companies?

SHIMIZU : We outsourced them to a company. Pharmaceutical companies did not participate in the first place. We planned the development without pharmaceutical companies.

FUKUDA : I see. This is why I asked. I often hear that although researchers develop applications, private companies cannot master them easily. They need supporters teaching them step-by-step how to use such applications. Therefore, those applications do not get used by private companies.

SHIMIZU : There are certain levels of acceptance in the pharmaceutical industry. About 10 of Japan's largest companies have personnel who have solid knowledge of computational science and enough ability to use supercomputer by themselves. As they want to know the details, they ask for explanations on applications. After that, they only need some basic training. Among those companies, about three companies are actively using university supercomputers for computing with us.

FUKUDA : Several companies have already gotten that far, right?

SHIMIZU : Although I don't know what level of computational resources they have, I think major companies have an in-house compound library and a system good enough to do reactive tests by themselves to find the compound. However, there are many companies which ask for teaching from the basics up. Some companies say they want to study to work on drug discovery using supercomputers like major companies in the future, and participate in our research project as observers, of course after signing confidentiality agreements.

FUKUDA : There are some companies which are determined to take a proactive approach to supercomputers among middle-class pharmaceutical companies, right?

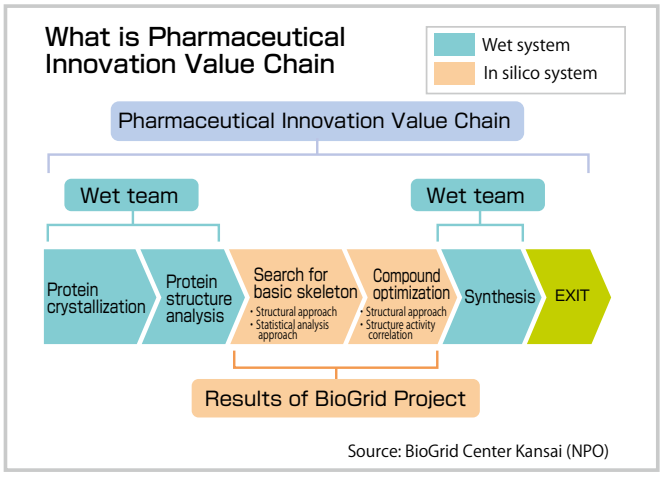
SHIMIZU : Yes, but I doubt such a proactive approach will spread throughout the pharmaceutical industry. I get the impression that it has not yet become the mainstream, and that many companies do a bit of simulation for drug discovery when their experimental trials go nowhere.

FUKUDA : You mean that it is not yet the mainstream that the in silico team first does a simulation, which is finished up by experiment?

SHIMIZU : There seems to be a problem of time. They say they can see results faster by doing experiments than waiting for computational results for half a month. Due to such time-lag problems and other things, the synthesis team and in silico team do not collaborate well in their systems, I think. Since some have been doing what they can do with their conventional computational resources, they don' t think they can do large-scale computation and are discouraged from the very start. From the beginning, they cannot imagine doing a simulation aggressively with an unusual resource like "K computer".

Pharmaceutical companies are changing their understanding of simulation.

Is it still the same, the situation you mentioned before that they cannot take data outside the company, or they are banned from



connecting to the network?

SHIMIZU : I think their attitude is changing little by little. There may be circumstances that they cannot afford to concentrate their personnel only on computation. It seems that there is a growing movement to allow outsourcing of routine computation as far as an interorganizational confidentiality agreement is signed.

FUKUDA: I think reliable organizations must be provided that take care of computation for important industries. For example, RIKEN takes care of computation for the pharmaceutical industry. Such a system should be prepared as national policy. By doing this, you can make clear who is responsible. Not to mention cost-effectiveness, you can gather all sorts of know-how. The FOCUS supercomputer of our foundation was introduced exactly with the concept of being used by the industrial sector. Although its level is 25TFLOPS meaning not so powerful, I am pretty sure there is enough demand after visiting many companies to talk to them. Many companies we have visited are in the field of CAE (Computer Aided Engineering). We have not yet visited so many pharmaceutical companies, but I think there is enough demand from what I hear from you.

SHIMIZU : The pharmaceutical industry is now in a transition stage. They had been defensive, but they are starting to think they should use outside resources. On the other hand, No.1 in the world, "K computer" is starting to gear up for industrial use. I feel the time is right. Only we have to present successful examples, or when a tangible achievement is made, some pharmaceutical companies may get interested and begin using supercomputers.

FUKUDA : In this regard, the result of the RIKEN grand challenge has become very crucial.

SHIMIZU : No other country has put more effort into the development of biosimulation than Japan. I can say Japan is the world leader in this field. Since the wisdom of the Japanese people has come this far, it would be a waste not to use it. What a shame. We should all publicize it.

Sprouts do not appear without sowing seeds, although you water the ground.

— We cannot produce new drugs easily today, but what is the expectation for computer simulation in the pharmaceutical industry as a method to find candidate compounds?

SHIMIZU: I don't know the full details, but I think a willingness to do computer simulation is being fostered. However, they don't have the necessary know-how and might still be hesitating. Also, top executives may mistrust us. For example, they may think, "Can drugs be developed in such a manner?" I think it is our challenge to change their attitude.

FUKUDA : There was a story when I still worked in the National Aerospace Laboratory of Japan. When we introduced a supercomputer, we told major companies in the industry that simulation must be used more in the development of aircraft. However, the middle managers at that time expressed little enthusiasm. What I did next was talk with young frontline engineers who were eager to absorb new technologies. Then, I had them use applications we had developed as much as they liked. Next, I went to see the director of the company and said, "Japanese technology won't go anywhere unless we start now". However, he still didn't lift a finger. Then, I said, "Your young and active engineers are very enthusiastic". Then, he got interested, saying, "I would listen to your idea if we have such young engineers," and I was able to proceed further. That was a real experience. It is crucial to find out how to input data in the company. It is not always good to rush things and negotiate only with senior management. Sprouts do not appear without sowing seeds, although you water the ground as much as you can.

SHIMIZU : In fact, similar things are going on in pharmaceutical companies. While we are undertaking joint research with them by using applications developed at the university, they will bring in their assignments, and are going to feed them back into the computation.

FUKUDA : After getting a result, someone in a higher position only has to give the top management a kick in the butt (laugh).

— Do you have something you would like researchers developing applications to know as regards industrial use?

FUKUDA : Applications are easy to understand when a sample calculation is ready. For example, this input results in this computation and so on. In addition, it would be good to have additional functions that can respond to various requests from companies, for example a program that can be applied to a slightly different calculation. By meeting such requests willingly, a relationship of trust would be built up fostering a willingness to work together in a team environment. Of course, it does not mean that you do everything the company suggests from A to Z.

SHIMIZU : I want them first to build tidy applications and then to prepare basic manuals, because the user interface and so on are not the area of researchers. If possible, the application should focus on versatility.

FUKUDA : It is not good that an application can be used for one group of compounds but not another.

SHIMIZU : Anyway, we aim at understanding and widespread use of supercomputer simulation, and would like to spread the word mostly among pharmaceutical companies that you can do these applications or that we have these great applications.

— We may still have big challenges to overcome, but I feel I can look forward to the future. Thank you for coming today.

Analysis of molecular mechanism of enzymatic reactions by QM/MM Free Energy Method

Graduate School of Science, Kyoto University
Shigehiko Hayashi
(Molecular Scale WG)

Our group is developing the QM/MM free energy method for analyzing the molecular mechanism of enzyme reactions. Many molecular functions of biomolecules are coupled to chemical reactions with enzyme catalytic activity. Therefore, understanding the mechanism of enzymatic chemical reactions is important in control and design of biological molecular functions. The mainstream method for this purpose is currently the hybrid QM/MM approach in which quantum mechanics (QM) method is combined with the molecular mechanics (MM) method. In this method, very high computational efficiency is attained by describing the local active site of the enzyme catalyst by the QM method, and the other vast protein environment by the computationally cheap MM method. So far our group has elucidated molecular mechanisms of various enzyme reactions such as photoreceptor and molecular motor proteins by using the QM/MM approach.

In spite of the usefulness of the QM/MM approach, the calculation of enzymatic chemical reactions coupled to molecular function still remained difficult. In many cases, molecular function appears with accompanying by large conformational changes in biomolecules. Such conformational change is usually analyzed by molecular dynamics (MD) simulation based on the MM method. However, calculation of enzymatic chemical reactions coupled to conformational change requires a description based on the QM/MM method. In the QM/MM approach, however, QM calculation is computationally much more expensive compared to MM calculation, even though its application is limited to local active sites. Therefore, it is impossible to obtain enough MD sampling time to adequately take into account the slow relaxation of the biomolecular system.

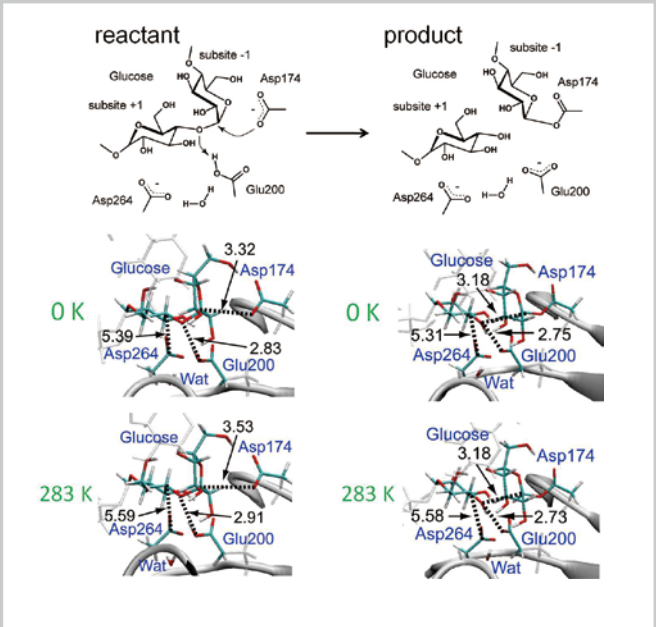


Fig 1 : Enzyme reaction scheme of α-amylase (above), and structural change from reactant to product (middle and below). Many parts of optimized structure on the free energy surface are different from the structure on the potential energy surface (0K), but it is also observed that the distance between atoms which have strong electronic interaction in the product is maintained in a thermal fluctuation.

In order to solve this problem, we have developed a novel QM/MM free energy method (QM/MM-RWFE-SCF method). In the QM/MM free energy method, the optimum free energy structure of the active site molecule treated by the QM method is determined on a free energy surface defined by the structural distribution of the MM region sampled by MD simulation. We developed a highly accurate and very efficient method by combining a mean field approximation with a statistical reweighting method, and then considering appropriately long-range coulomb interaction between QM and MM by the Ewald method. Especially in the scheme of this method, the calculations of the QM/MM method part are completely separate from the MD simulation part for MM structure sampling. Therefore, existing sophisticated MD programs can be applied to the MD simulation, which achieves great flexibility in performing calculations.

As a test of the method, it was applied to hydrolysis of the glycoside bond of α-amylase shown in the reaction scheme in Fig 1. The protein system in water represented by a periodic boundary condition consists of 68,000 or more atoms, and QM molecules of the active site are described with abundant base functions of over 600 bases. By using this method, we determined the optimum energy structures of the reactant and the product of the reaction. In the result, as shown in Fig 2, their optimum structures were determined by the accompanying 90- and 21-ns large structural relaxation of protein loops surrounding active sites, respectively. Such a QM/MM free energy structural optimization calculation close to sub-microseconds has no precedent. More specifically, the structural optimization calculation performed by this method can follow the relaxation process of a protein structure ten-thousand times longer compared to direct QM/MM MD calculation, and describe relaxation 100 to 1000 times longer compared to other similar QM/MM free energy calculations. In addition, this calculation suggests the possibility that adjacent loops assume different structures (Fig 2). Such a major change of protein structure correlating with the enzyme reaction is a finding which cannot be obtained through conventional methods. It is believed to offer a new avenue for elucidating, controlling and designing new molecular mechanisms of enzyme reactions.

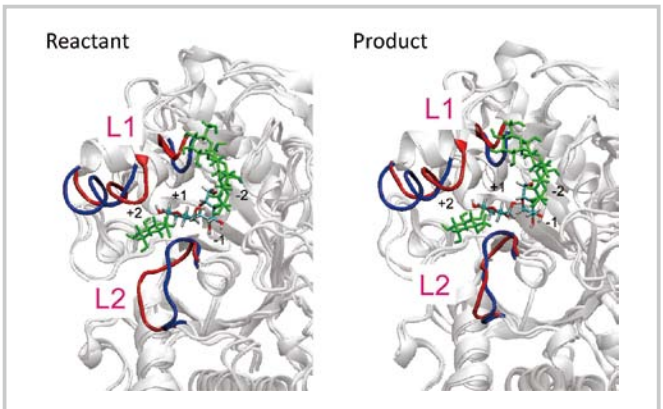
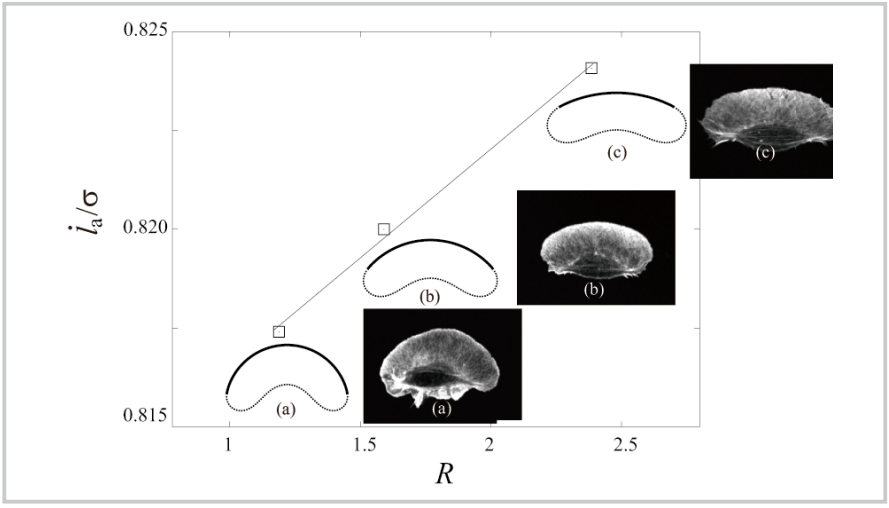


Fig 2 : Change of protein structure between reactant and product. The structure of the loops adjacent to active sites (L1 and L2 in red) obtained by free energy structure optimization are largely different from the structure before structural optimization (in blue). The structure of the L2 loop differs substantially between reactant and product.

Computational Mechanobiology of Actin Cytoskeleton

Institute for Frontier Medical Science, Kyoto University
Yasuhiro Inoue
(Cell Scale WG)

The actin cytoskeleton forms higher structures such as a network structure and bundle structure through interaction of various auxiliary proteins using actin filaments as the basic skeleton, in which actin monomers are coupled with one another, and constitutes an important mechanical and chemical base which supports cell shape and motion. For example, actin beneath the cell surface forms structures for motility such as filopodia and lamellipodia. They are essential for formation of neuronal growth cone and cell motion. In actomyosin filaments consisting of actin and myosin, the motor motion of myosin generates a force, which is crucial in cell deformation in morphogenesis, cytoplasmic division by a contractile ring, and turnover of cell adhesion. Those structures emerge from molecular systems centered on actin, and can change their structures and characteristics dynamically by going the round of mechanically and chemically metastable states. We are studying the relation of actin cytoskeleton dynamics to cell functions by using a computational mechanobiology approach. Our recent studies revealed that cell shape in cell motion is coupled with actin polymerization via thermal fluctuation [1], and the relation between the force generated in the actomyosin network and the myosin density is bilinear in response to network structure change [2]. These studies give important insights in the relation between cell-scale expression of function and molecular-scale dynamics. Since actin mechanobiology can be simulated by using computers, it is not a totally unrealizable dream to recreate a whole cell on the computer. Of course, we have a lot of challenges to overcome. As for these challenges, we have to work with researchers inside and outside the Next-Generation Integrated Simulation of Living Matter Group. A Computational Science Algorithm Study Group is regularly held to create such an opportunity. There, researchers share knowledge of mathematical models and computational approaches beyond the scope of their fields, and pursue collaboration with researchers in related boundary areas.



[References]
[1] Inoue, Y and Adachi, T. Biomech. Model. Mechanobiol. 2011 10:495-503
[2] Inoue, Y. Tsuda, S., Nakagawa K., Hojo, M., Adachi T. J Theor Biol. 2011 21;281(1):65-73

Fig 1 : Computational simulation of migrant cell: Cell shape obtained by simulation (solid line) coincides with actually-observed cell shape (fluorescence image). In addition, it was found that the actin polymerization speed at the tip of the cell (longitudinal axis) is in proportion to the curvature radius at the tip of the cell (horizontal axis).

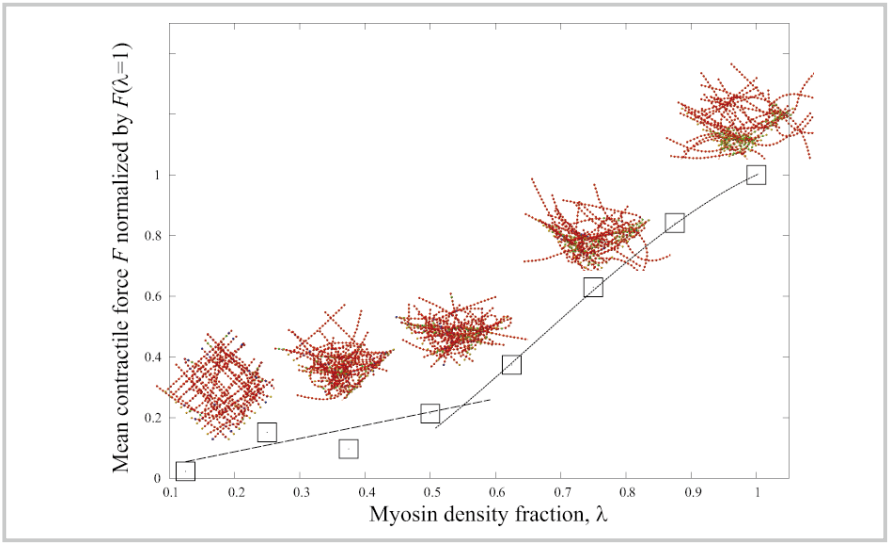


Fig 2 : Autonomous structural change of actomyosin network: Due to motor motion of myosin in the actomyosin network, the structure of actomyosin network changes autonomously. Simulation showed that the force generated in the network according to restructuring is bilinear with respect to myosin density.

Development of Blood Flow Analysis Method for Simulation of Thrombus Formation

Department of Mechanical Engineering, The University of Tokyo
Satoshi Ii
(Organ and Body Scale WG)

When the inner wall of a blood vessel is damaged, a thrombus is formed at the site of damage. Due to the thrombus, bleeding from a slight injury usually stops after a certain amount of time. This is the major role of a thrombus. The thrombus formed is lysed after bleeding stops. For some reason, it sometimes does not disappear and becomes enlarged. Then, it may block the blood vessel, or the enlarged thrombus may slough off and occlude a more distal vessel to block blood flow. Such a thrombus causes brain infarct and cardiac infarct. It is also a cause of the development of economy-class syndrome or flight-related DVT. For these reasons, anti-thrombosis agents are used in clinical practice. However, since they have the potential to affect normal thrombus involved with hemostasis, they should be given cautiously. However it is known to be very difficult to assess the effect of a medicine in the bloodstream of the human body in advance. Against this background, our group is aiming to reproduce thrombus formation through simulation. For example, we are thinking of its use for understanding drug efficacy in the bloodstream in developing new drugs.

In the thrombus formation process, adhesion of platelets to the damaged vessel wall, agglutination of platelets and coagulation incorporating erythrocytes, involving various factors such as the bloodstream and cell metabolism, occur in a step-by-step manner. In this research, we focus on the early process, platelet primary aggregation in which platelets adhere to the damaged cell wall. In this process, platelets adhere to the damaged cell wall through binding of vWF (von Willerbrand factor) attached to collagen beneath vascular endothelial cells and GPIbα (glycoprotein Ibα) found on the platelet surface. In order to analyze such phenomena, we are developing an approach to deal with bloodstream phenomena including erythrocytes and platelets, and GPIbα-vWF binding phenomena between the platelet surface and damaged cell wall in a unified manner (Fig.1).

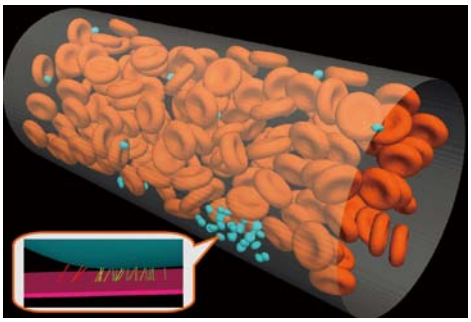


Fig 1 : View showing a frame format of analysis toward platelet primary aggregation

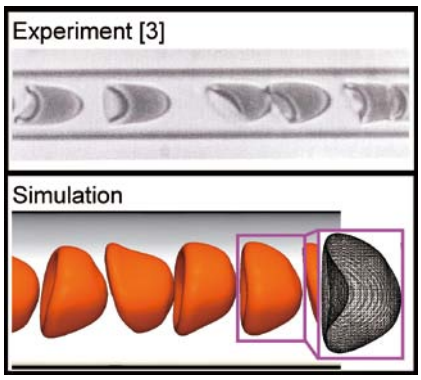


Fig. 2 : Deformation behavior of erythrocytes in the flow field [3], and result of simulation

Since erythrocytes have a very high bulk ratio of around 45% in the bloodstream, when the vessel diameter is micrometer-size, they affect the flow properties of the blood stream and become strongly involved in platelet aggregation. Since erythrocytes are very flexible and deformable, they cause an interacting phenomenon with the flow field. Since these phenomena must be analyzed in a system containing numerous erythrocytes, our group proposes a new analytical approach which excels in calculation efficiency and has a good compatibility with blood flow analysis [1,2]. We performed an analysis by use of an experimentally-obtained physical property of erythrocytes. As shown in Fig 2, we obtained a valid result as compared to the deformation behavior of real erythrocytes in the flow field.

In the analysis of platelet primary aggregation, binding of GPIbα on the platelet surface and vWF in the damaged cell wall has to be considered. They are molecular-scale phenomena, but their interactive analysis in combination with blood flow is currently impossible. Therefore, we introduced a technique in which we can deal with such binding phenomena statistically, and combined it with blood flow analysis. Fig 3 shows an example of the analysis. It reproduces how binding of GPIbα on the platelet and vWF at the bottom of the wall (simulating a damaged vessel wall) takes place, and how platelets adhere to the cell wall.

Hereafter, we will deal with aggregation of platelets by modeling the metabolic response of platelets, and reproduce the thrombus formation process in a step-by-step manner.

[1] K. Sugiyama, S. Ii, S. Takeuchi, S. Takagi and Y. Matsumoto, A full Eulerian finite difference approach for solving fluid-structure coupling problems, J. Comput. Phys., 230 (2011) 596-627.
[2] S. Ii, X. Gong, K. Sugiyama, J. Wu, H. Huang and S. Takagi, A full Eulerian fluid-membrane coupling method with a smoothed volume-of-fluid approach, Commun. Comput. Phys. (2011) accepted.
[3] P. Gaehtgens, C. Dührssen and K.H. Albrecht, Motion, deformation, and interaction of blood cells and plasma during flow through narrow capillary tubes. Blood Cells, 6 (1980) 799-817.

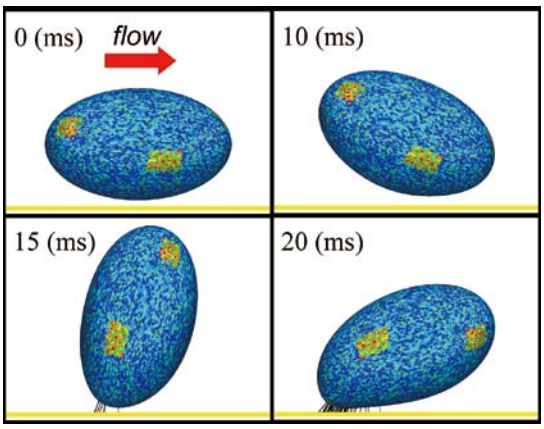


Fig 3 : How platelets adhere in fluid analysis taking GPIbα-vWF binding into account

Development of Data Assimilation Technology for Simulation of Living Things

The Institute of Statistical Mathematics
Tomoyuki Higuchi
(Data Analysis Fusion WG)

Through the East Japan Great Earthquake (the big earthquake that hit Eastern Japan this year), we researchers recognized that it is difficult to understand and provide protection against or control complicated systems, and that we continuously need to bring together our wisdom to solve problems. When we try to understand and control complicated systems such as the earth as well as living things, it is effective to evaluate and correct the progress of research by using our ability to predict phenomena on the assumption that information about an object is always incomplete. This approach has been demonstrated in statistics, and it has always contributed greatly to human prosperity on earth.

Prediction ability is an integrated index of effectiveness, which consists of two major functions. One of them is the descriptive power of a forward computing model, and the other is the cognitive power that catches the present state of the object (current state). To put it simply, forward computing represents a repeat assignment operation. For example, it is like a computing method where, when a value is put into the right side of an equation, the resulting value comes out of the left side, then the resulting value is also put into the right side, and the resulting value of the next step comes out of the left side. Many simulation computations explicitly solving time development adopt this method, and long-term prediction is achieved by repeating this forward computing. On the other hand, the latter is related directly to innovation of measurement methods. New instruments can provide larger amounts of more precise information than ever before by innovations with epoch-making measurement methods, and they are a great attraction to researchers in any field, especially life sciences, where new instruments have driven development.

However, leaning toward research and development of measurement devices is not a good strategy from the viewpoint of improving prediction ability. This is not only because directly measuring the whole object has limitations in theory, but it is extremely effective to strengthen the descriptive power of the forward computing model for increasing prediction ability. In the research fields that have a long history of simulation, such as earth and space sciences and solid-state physics, governing equations that are the basis of forward computing have usually been established, and it is important for success to implement calculations approximately based on the governing equations on a super computer. Improvement of this approximate calculation is equivalent to improvement of the forward computing model. One of the major goals of BioSupercomputing is to dramatically improve the descriptive power of this forward computation model by taking full advantage of the scale of computing hardware. Unfortunately, it is not an exaggeration to say that there is no principle corresponding to the governing equation in life sciences, so forward computing models themselves must be based on a wide variety of ideas and become less general.

Then, is the systematic improvement of forward computing model difficult in life sciences? As mentioned above, the ability to predict phenomena can be improved by enhancing both the measurement method and the forward computing model. Therefore, it looks more natural to arrange the forward computing model so that prediction ability may improve than to modify the model in accordance with its own evaluation criteria. This means that learning functions fed back from measurement data are added to the forward computing model. In fact, in the weather and oceanography fields that are state-of-the-art areas of simulation research, it is usual that weather forecasting services improve prediction performance by integrating large quantities of space data collected hourly from all over the world and the largest world scale simulated calculation result on a super computer using Bayesian statistics, and then improving simulation models in real time. Also, it has been pointed out that the simulator SPEEDI, which estimates the effects of atomic radiation and became a hot topic recently, was not able to demonstrate its power sufficiently partly because it had no function to reflect real observation data in the simulator in real time.

The integration of observation data and the result of the model is referred to as data assimilation, and it has recently been attracting attention in the field of simulation science. If the idea of data assimilation is applied to the simulation of living things, it at least will lead to a steady improvement of prediction ability, and consequently, help understanding and control of the complex system. With this fervent desire, we have worked on research and development of data assimilation technology for simulation of living things as members of the Data Analysis Fusion Team every day. Current living thing simulation models are like ready-made clothes when compared to clothes. Even if there is a variation, there might be a difference in size like S, M, and L at most. On the other hand, each human body system is different. We are looking forward to the day when a custom-made or even semicustom-made living matter simulator, which is suitable for a patient, can automatically be built from medical information about patients who suffer from side-effects of medicine or a treatment method.

Analysis and prediction based on a simulation personalized for an individual or an environment are important in the future. Data assimilation is the integral technology for realizing this objective.

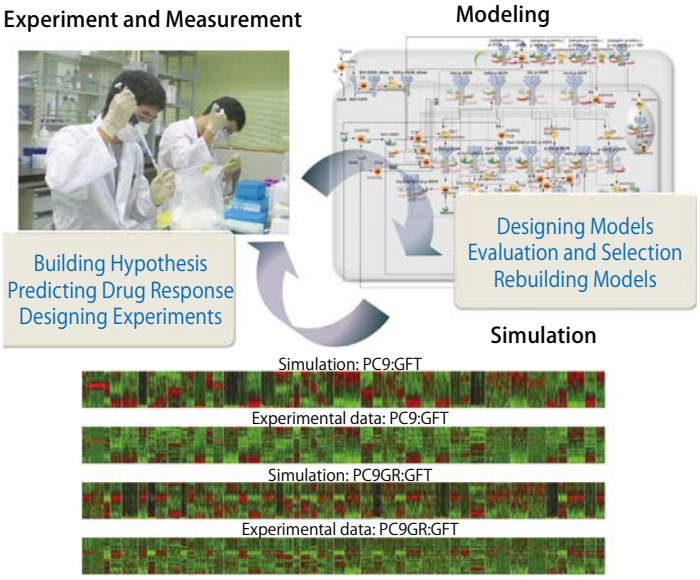


Figure 1 : Conceptual diagram that shows how to use the application under development, LiSDAS (Life Science Data Assimilation Systems). LiSDAS took its name from the well-known Automated Meteorological Data Acquisition System, AMeDAS. Data acquired from an experiment or a measurement site (upper left part) and an existing set of models (upper right part) are combined to perform the data assimilation. Data assimilation allows for evaluation and rebuilding of models at the same time. The calculation result and the measured data in the assimilated simulator are shown in the lower part. The result of the data assimilation is utilized to build a new hypothesis, and to design the following experiment. LiSDAS is such a calculation platform which achieves a series of intelligence cycles.

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


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

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

The HPCI Strategic Program, which aims to produce the world's best research products in a wide variety of fields by utilizing Japanese High-Performance Computing Infrastructures (HPCI) centered on "K computer" and returning their fruits to society, got into full swing this year. What breakthrough will Field 1, "Supercomputational Life Science" bring to life science? We had Director Toshio Yanagida, Deputy Program Director Akinori Kidera, and Deputy Program Director Yukihiro Eguchi talk about that (This article appears in this and following numbers).

Our goal is to recreate life phenomena on the computer.

YANAGIDA (Honorifics omitted) **:** In a nutshell, our dream is to reproduce life phenomena representing highly complex, non-linear and dynamic systems on the computer. It is our final target. For example, in order to reproduce various cell functions, we have to have enough data to input to the computer. Of course, we cannot understand the system only with the data, and need a concept to process and analyze the data. After formulating it properly, we set up a model based on it and reproduce cell functions with the computer. Of course, we want to reproduce not only cells but also tissues and organs, and that's not easy. Therefore, we must first have a firm understanding of partial functions such as protein functions and reproduce them on the computer. It would be wonderful if that could lead to drug discovery. Of course, we are interested in the way how complex movements of proteins lead to their function, but if we could manipulate protein functions, we would be able to expect to translate it into drug discovery.

KIDERA : I have been studying protein simulation. My final goal in the HPCI Strategic Program is the same as what Dr. Yanagida said. I want to replicate protein behavior plainly by simulating proteins in the computer. However, compared to other fields such as the materials field in which computational science is advanced, life science is still largely undeveloped. If I were to speak out without fear of being wrong, I would say that simulation of life phenomena is nearly impossible. Then, what is possible to us? I think our challenge started from here. It is difficult to replicate cell behavior. Then, how about the more limited case of a single protein molecule? Like this, we have studied what we can find by limiting the theme until we can handle and do the simulation properly. We made the theme smaller until we could handle it. As a result, what happened? It became a matter of physics, not biology. Everything starts from physics. However, many biologists ignored it completely. Then, what is different between life phenomena and physical phenomena? For example, nothing happens by putting a single protein into the computer and simulating it. A partner is needed to make something happen. To make the partner work, another partner is required. And then, that partner needs yet another one. Such a world of endless chain reactions takes place in an astonishingly complex and heterogeneous cell environment. Formerly, we were not able to observe it due to the limitation of computer power. With the increased computer power, the range that we can reproduce becomes larger spatially and temporally, and we don't have to minify the limitation. Now at last, biologists are starting to show interest. In the next step, we are going to observe something which is larger and more real with "K computer". Our research level at last is starting to be recognized by biologists doing experiments. In addition, people in the fields of medical care and drug discovery, who thought they had nothing to do with computing, are starting to say, "Simulation may be useful" or "We may be able to do something interesting with K computer". What was once thought to be impossible may become feasible or probable. Well, that's an overstatement (laugh).

YANAGIDA : If you say, "Let's improve this by five times," you will have trouble dealing with the reality, but if you say, "By one hundred times," you will feel relieved and get on with the research (laugh).

KIDERA : Since the level of the computer we used so far was 10TFLOPS and that of "K computer" is one thousand times higher, 10PFLOPS, more and more researchers should show an interest (laugh).

YANAGIDA : Aside from joking, it's actually very important that researchers think something is "interesting." If they think simulation is interesting and useful, they come to think along those lines and their research may take a new direction.

We will show that life science is becoming an information-driven science.

EGUCHI : Dr. Yanagida said it's a dream, but I hope the day would come, when many researchers who once performed many experiments with trial and error say, "As I can now find the result of this experiment through computer simulation, I will leave it to simulation." For example, the safety of automobiles is tested by collision experiments. It was once actually done with many automobiles to change collision angles and speeds, but today most experiments are done through simulation. This means that computer simulation has become that much more reliable. I hope something similar would happen in life science.

YANAGIDA : Right, in the field of manufacturing, half of the experiments are done by simulation. It is common to use simulation in experiments. In this regard, life science may be behind in computer science. However, life science will also follow the same pattern. It is impossible to replace everything, but we have to omit experiments and use simulation with determined parameters as much as possible. Then, the part we cannot simulate should be solved by experiments to increase efficiency, and especially in drug discovery. There are too many combinations for trial and error.

KIDERA : There is another important thing in addition to what Mr. Eguchi mentioned. I think it's to show that life science can be shown to be a science which utilizes information. This means improving on the limited way of conventional information processing with "K computer" change the image of information itself, and make what was once thought impossible, possible. I think it's one of the roles of the HPCI Strategic Program we are addressing to clearly point the way in this direction, and greatly expand the possibilities. It's important to show how to do data analysis or prototypes using massive computer resources like "K computer", and widely demonstrate its possibilities and results to the outside. Also, I'd like to speak a little on what Dr. Yanagida said: life science lags behind manufacturing. Actually we are doing the same level of computation. However, it's not enough when we come to life science. So it seems we can only handle small things. For example, there is whole body-scale simulation. The human body basically looks similar to an automobile, but it's different from automobiles which are a combination of steel sheets and an iron frame - even an organ is a gathering of a lot of very, very many kinds of cells. In addition, heat and substances go in and out of the organ for chemical reactions. We must change our computations depending on what we observe. Anyway, it's a very difficult problem.

YANAGIDA : Its crucial difference from the other fields is that life science does not have fundamental equations. We can only search for parameters and find a certain predisposition or state of things. It's also difficult that what is going on has multiple factors. Since the targets of life science are very complex and difficult, we still have a long way to go. In addition, a high computing power is required due to many combinations. However, if we clear those obstacles and our research level reaches that of other fields, computer simulation would be able to do a lot of things, i.e., contribute to drug discovery.

Can you do your studies without enjoying them?

EGUCHI : It's tough enough to analyze only proteins. Cells have all sorts of other data. Dr. Yanagida, who is conversant with the difficulty of complex, varied and non-linear life phenomena, is head of the project and is always issuing communications. I think this is very important. By doing this, I believe that new ideas will come out of new data to lead to further research. In order to enliven this future Field 1, we will need to rebuild the path so that R&D teams bring up issues one after another, or newcomers can participate. This is a 5-year project, but I'd like to create a framework in which more new trends appear after the 6th year and there are more young researchers who want to participate in this field. Although the research is hard, I want to get it across to the life science community and young researchers' communities that all members are actively getting results by making the best use of new computer power, and expanding the horizons of research.

YANAGIDA : Not only are researchers who are engaged in computer science backing this field, but many biologists who use computers today. So, promoting this trend leads to further improvement of science itself. However, it's no use only to say that computers are important. We need something playing the role of a driving force or becoming a base. I think that symbol is "K computer" which demonstrated a world-beating computing ability. Then, an actual project using "K computer" starts rolling, impressive research is being done there, and Japan has world-class researchers. I believe that is what is important.

EGUCHI : Now, "K computer" has become No. 1 in the world. Because of that, I'd like professors to get good research results.

YANAGIDA : Since this research resource became No. 1, it, of course, also hints that Japanese science research itself might become the world's No.1. To achieve this, it's crucial to expand the horizons of interdisciplinary studies, for example a successful marriage of Science and Engineering.

EGUCHI: It's great to do research with world-beater "K computer". I would like to see an environment in which researchers from various fields participate, compete with and complement each other.

KIDERA : It's true that along with an increase in computer power, we can not only do research in our field, but also do more advanced research. We can use "K computer". This offers a great advantage for our research. For this reason, it may be the challenge of the future to build a system in which users do their research with a sense of thrill and excitement, not pressure, although it's my position to urge them to produce results (laugh).

EGUCHI : Please create an atmosphere in which everyone gets excited and wants to team up (laugh).

KIDERA : I really felt "K computer" was great when I put it into use. After experiencing the speed, I really felt it is a terrific computer.

YANAGIDA : The triple-digit difference from conventional computers makes you feel not only a quantitative difference but also a qualitative change.

KIDERA : Everyone gets the feeling they can get great results with it. That' s why they also feel pressured. I don't want to say, "Hurry up and get results," as far as I can avoid it, but··· (laugh)

YANAGIDA : As I recall, at the review board the other day someone said about a research presentation, "Research must be carried out with a sense of joy. Are you happy when you do your research?"

KIDERA : Of course we have responsibility, but after all you cannot do great things without enjoying them.

(Continued in the next issue)

Simulation Applicable to Drug Design

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

Hideaki Fujitani

(Field 1- Program 2)

The examination of interaction between a protein and a compound that specifically binds to the protein (ligand) is not only a major challenge in the basic study of life phenomena, but also extremely important from the aspect of application. Accurate evaluation of the free energy of binding between proteins and drug molecules is a longstanding challenge in the field of drug discovery. Although various methods including Docking-Simulation have been tried, there has been no method that predicts the free energy of binding to a protein with an accuracy effective enough for designing a drug molecule. For finding the difference in free energy between two different states of thermal equilibrium, one in which the protein and its ligand bind together in water (cell fluid) and the other in which they separately dissolve in water, a method for accurate incorporation of the entropy change due to conformational change of the protein and molecular motion is required. For such a calculation, the molecular dynamics method is used and its accuracy depends strongly on the force-field parameter used.

Torsion interaction relative to the Ramachandran angle of the protein main chain is the most important force-field parameter in defining the cubic structure of proteins, and its energy profiles in a glycine peptide are compared in Fig 1[1]. The black solid line shows the profile defined through a molecular orbital calculation which is currently the most accurate (LCCSD). It is compared with conventional molecular force fields such as AMBER (ff99, ff03) and OPLS-AA. Ff99sb is a force field in which only the main chain torsion interaction of ff99 released in 2006 was modified, and defined so that it conforms to energy by high-accuracy molecular orbital calculations at some angles. For this reason, the energy values are close to those shown by the black line when φ and ψ are 0 degrees. However when φ is 80 to 180 degrees, an energy barrier higher by 1kcal/mol is observed compared to the energy shown by the black line. By use of the electric charge of AMBER and the van der Waals parameter, we developed a method to assign force-field parameters in a unified manner to proteins, DNA, RNA or arbitrary compounds, and named it the FUJI force-field[2].

With a relational expression between the free energy difference ΔG and non-equilibrium work W[3], we performed molecular dynamics calculations independently at multiple intermediate states from the state in which there is full interaction between the ligand and the other molecule to the state in which there is no interaction between them[4], and calculated the free energy from the Work W required for shifting to the next stage (Fig2). With the MP-CAFEE method and the FUJI force field, it is possible to perform an accurate computation of protein-ligand binding free energy.

[1] H. Fujitani, A. Matsuura, S. Sakai, H. Sato, and Y. Tanida : *J. Chem. Theory Comput.*, **5**, 1155 (2009).
[2] H. Fujitani, Y. Tanida, and Azuma Matsuura : *Phys. Rev. E,* **79**, 021914 (2009).
[3] C. Jarzynski : *Phys. Rev. Lett.*, **78**, 2690-2693 (1997).
[4] H. Fujitani, Y. Tanida, M. Ito, G. Jayachandran, C. D. Snow, M. R. Shirts, E. J. Sorin, and V. S. Pande : *J. Chem. Phys.*, **123**, 084108 (2005).

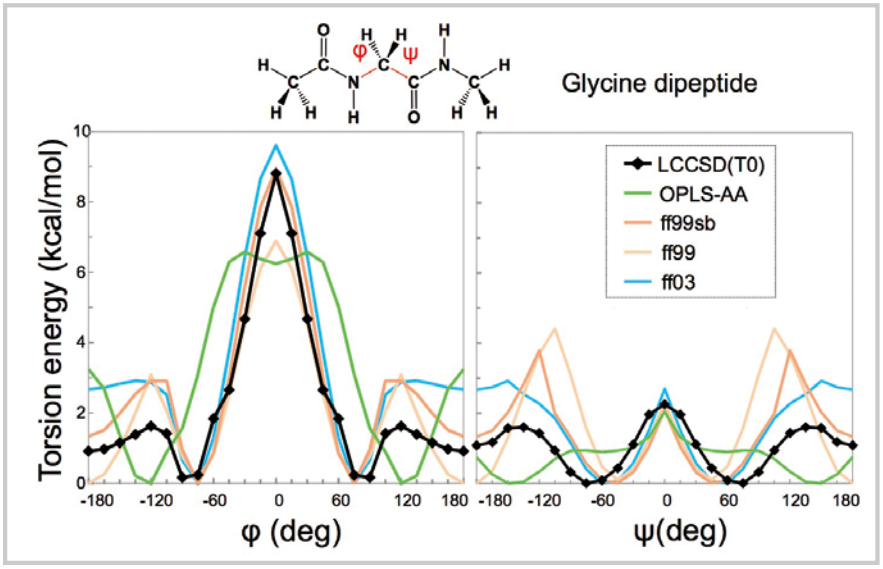


Fig 1 : Comparison of torsion interaction of protein main chain

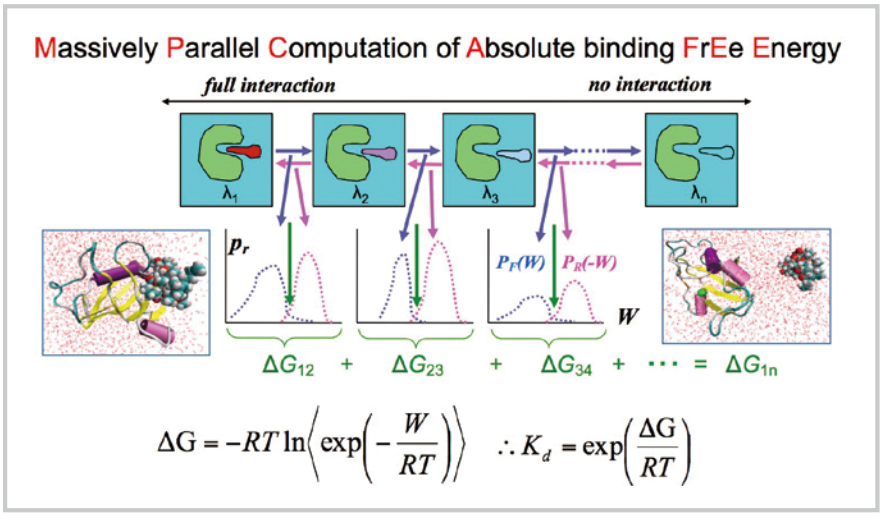


Fig 2 : Massively parallel computation of absolute binding free energy, MP-CAFEE

An Ultra-fast Analysis System for Next-Generation DNA Sequencer Data

Graduate School of Information Science and Engineering,
Tokyo Institute of Technology
(From the above) Yutaka Akiyama, Takashi Ishida
Masanori Kakuta, Shuji Suzuki
(Field 1- Program 4)

The Next-Generation DNA Sequencer allows DNA sequencing of about 1 trillion bases with a single run by a dramatic technological innovation. Although the "number of bases sequenced per dollar" doubled with a marvelous extension rate in about 19 months with what we call a first-generation sequencer, it doubled in only about 5 months after the appearance of Next-Generation Sequencers which are coming into wide use everywhere. The largest bottleneck now is the calculation of semantic interpretation for read sequences. It is common knowledge that supercomputers are required for integrated advanced calculation such as gene network prediction. But now, large-scale calculations are also required for elementary processes for DNA fragments from sequencers.

A high-speed algorithm which is suitable to explore an exact match with a known sequence or highly similar character string can be used for the study of subtle variations of human genes, and large-scale personal computer clusters have managed to address it. However, in the study of a comprehensive analysis of enteric microbe's genome fragments of humans or the study of significant sequence mutations, fragment mapping procedure (finding correspondence with a portion of known genome sequence) is not possible without detection of sequence similarity among distant or largely mutated sequences. For these studies, incomparably large-scale calculation is required, and personal computer clusters are not effective.

Human health is maintained by the properly balanced work of many symbiotic microbes in the body. In addition, the balance is considered to have several locally stable points. Comprehensive sequencing of genome fragments of microorganism gastro-intestinal which exist in the body's internal environment including oral, digestive, skin and urogenital system has been implemented for about 500 people in the Human Microbiome Project (HMP) in the USA and for about 300 people in the MetaHIT Project in Europe. However, when analyzing these valuable data, the rate of non-analyzable sequence fragments to be excluded will be significantly increased only with insensitive methods hypothesizing highly similar character strings.

We firstly established the super parallel execution environment of the existing BLASTX software with the cooperation of Prof. Ken Kurokawa (Tokyo Institute of Technology), who is an expert in genomic analysis. A scalability up to 16,000 cores was confirmed on TSUBAME 2.0 of the Tokyo Institute of Technology (Fig. 1). Binary-tree transfer was used for applying the bandwidth of networks at the initial delivery of the database to each calculation node. After translating DNA fragment sequences into amino acid sequences in six possible frames, we flexibly compared them with existing amino acid sequences. Event though the target DNA sequence fragments mutated at the DNA sequence level, they are often consistent with residues of existing sequences at the amino acid sequence level, or are changed into residues with similar physicochemical properties. Therefore, our flexible method which compares using a distance matrix between characters improves sensitivity for sequence search to explore distant genes with the same functions.

We then created a software, GHOSTX, as a substitute for BLASTX, with drastic speed-up on the algorithm [1, 2]. GHOSTX can find similar short strings between database and query sequences immediately by multito-multi comparison by describing a partitioned database stored in memory using Suffix Array, translating a query sequence to Suffix Array, and comparing them e ffectively (Fig. 2). While BLASTX compares only character strings with a fixed-length, character strings for comparison can be extended with variable-length until they satisfy a similarity threshold score in GHOSTX, which improves the sensitivity per calculation cost. GHOSTX achieved about 20 times processing speed even when an equivalent sensitivity an equivalent sensitivity with BLASTX was required. When the sensitivity is decreased to a degree of no difficulty, GHOSTX attained over 100 times the speed. Its sensitivity is much better and it is 2 to 3 times faster than the BLAT algorithm which had high speed but was not able to be used because of poor sensitivity. We established a super parallel analysis system named "GHOST-MP" by implementing the GHOSTX algorithm on the ten quadrillion speed computer "K computer", and integrating thread parallelization using OpenMP and data parallelism between nodes using MPI.

For the principal parts of homologous search, good scalability up to 1,536 nodes (12,288 cores) was measured with "K computer". Although the execution time for each sequence fragment varies widely, a scalability of up to several tens of thousands of nodes is expected because of the load balancing mechanism. In GHOST-MP, I/O operation is not completely intensive because its calculations are proportional to the product (multiplication) of length of the database and length of query sequence sets. Even though, careful I/O design is required. In the case of "K computer", I/O transfer to and from I/O nodes takes place in the z axis direction. We aim to achieve highly- efficient implementation up to several tens of thousands of nodes by adjusting data input using representative nodes and a broadcast mechanism, and by improving dispersion and reduction of I/O load.

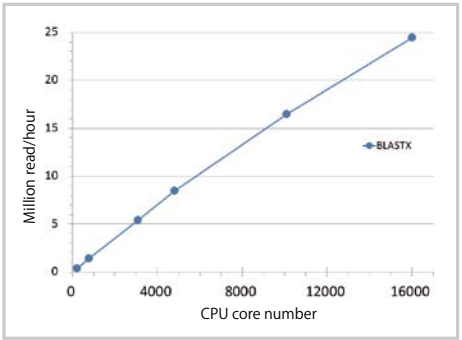


Fig 1 : Parallel performance in BLASTX-based system

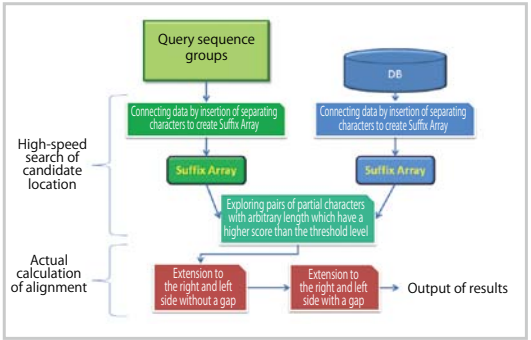


Fig 2 : Processing flowchart in GHOSTX algorithm

References
[1] Suzuki, Ishida, Akiyama (2010) IPSJ Technical report, 2010-BIO-23(21):1-6
[2] Suzuki, Ishida, Akiyama (2011) IPSJ Technical report, 2011-BIO-25(32):1-8

BioSupercomputing Summer School 2011

Computational Science Research Program, RIKEN Yasuhiro Ishimine (Organ and Body Scale WG)
Research and Development Center for Data Assimilation, Institute of Statistical Mathematics Masaya Saito (Data Analysis Fusion WG)
Niigata University of International and Information Studies Eisuke Chikayama (Cell Scale WG)
School of Medicine, Tokai University Yohei Nanazawa (Cell Scale/Organ and Body Scale WG)
Computational Science Research Program, RIKEN Takashi Handa (Brain and Neural Systems WG)
Computational Science Research Program, RIKEN Gen Masumoto (High-performance Computing Team)
Computational Science Research Program, RIKEN Kei Moritsugu (Molecular Scale WG)
(In Japanese alphabetical order)

On September 26 and 27, 2011, the "BioSupercomputing Summer School 2011" was held at the Awaji Yumebutai International Conference Center in Awaji, Hyogo Prefecture. This is the third time as a young researchers' workshop planned and organized by research associates in the Next-Generation Integrated Simulation of Living Matter (ISLiM) project, following last year's summer school and winter school. This school was renamed in connection with the support from the "BioSuperComputing Research Community" (BSCRC).

The ten-petaflops supercomputing, "K computer", now being installed in Kobe, Hyogo Prefecture, was ranked number one in the world at TOP500 in this June. ISLiM is developing softwares for numerical simulation and massive data analysis for life science so that the K computer can achieve its full performance. This summer school was held to demonstrate what K computer is able to do and how it is able to contribute to life sciences, and then consisted of two sessions, "Computer and computational methodology" and "Application for life science". A total of 56 research associates in the project and graduate students from the associated research fields attended to have a close relationship with each other.

First, as keynote lectures, Prof.Koji Kaya, Director of the Next-Generation Computational Science Research Program, delivered a lecture about the history of the ISLiM project and its future prospects. Then, Prof. Ryutaro Himeno, Vice Director of the Next-Generation Computational Science Research Program, delivered a lecture on the current performance and status of the application softwares now being developed at ISLiM, and the future schedule for the usage of the K computer.

In the following "Application" session, Prof. Masahiro Kami, the Institute of Medical Science, University of Tokyo, gave a lecture as an invited speaker and the three members from the ISLiM teams talked on living matter softwares under development. Prof. Kami gave a talk from the viewpoint of a researcher as well as a clinician on genomic analysis, tailor-made medical care, the doctor shortage problem, and his activity in Fukushima Prefecture hit by the Great East Japan Earthquake. It is a known fact for computational scientists that they must have data to analyze, but his point was impressive that it is important to collect data actually by way of extending the community of his own. Dr. Masaya Saito from the Data Analysis Fusion Team spoke about the analysis of the mammalian circadian cycle transcriptional network using LiSDAS, and introduced data assimilation library. Dr.Yasuhiro Sunaga from the Cell Scale Team introduced the Cell simulation integrated platform (RICS), and its applied research. Dr.Tomoki Kazawa from the Brain and Neural Systems Team told us about a bombycid olfactory and motor nerve systems simulation and real time imaging of nerve circuits.

In the subsequent banquet and poster session, there were 35 poster presentations participated, and debates on their present works went on until late at night. The poster award competition was held by the choices of five participating graduate students. Proposed by Prof. Himeno, debates were also made about the forthcoming "exascale computing".

In the "Computer" session on the second day, first Dr. Ikuo Miyoshi and Dr. Yoshinori Sugizaki, both from the PA project of the Next-Generation TC Development Section at FUJITSU, delivered keynote lectures as invited speakers. Then, the three members from the ISLiM teams talked about particle and fluid systems simulation and their high-performance computing techniques. Dr. Miyoshi described the hardware configuration and software development status of the K computer and Dr. Sugizaki explained tuning techniques with the help of simple and useful examples. In the subsequent question-and-answer session, we could obtain detailed understandings from the knowledge of the K computer developers. Dr. Yusuke Matsunaga from the Molecular Scale Team described Molecular Dynamics (MD) simulation, especially how to deal with long-range forces, and introduced the platypus-MM/CG under development. Dr. Yosuke Ohno from the High-performance Computing Team explained MD speed-up techniques particularly for short-range forces by giving useful examples with the K computer. Dr. Hitoshi Ii from the Organ and Body Scale Team spoke about the simulation of platelet thrombus formation by the Euler approach, and Dr. Hiroshi Koyama from the High-performance Computing Team delivered a lecture about the numerical errors in the solution of differential equations and why they appear.

Through the three ISLiM young researchers' workshops from last year, the initial goal to encourage exchanges of the ISLiM members between different teams seems to has been achieved. We expect that in the near future this kind of workshop will become an opportunity for creating a community of young researchers in the fields of simulation and massive data analysis in life sciences, together with the participants from the Advanced Institute for Computational Science and the HPCI program.

“Next-Generation Integrated Simulation of Living Matter (ISLiM)”, a web page dedicated to new applications, has opened.

Computational Science Research Program
Integrated Simulation of Living Matter Group

Next-Generation Integrated Simulation of Living Matter started in 2006, and evolved application programs that can attain peta-flop class execution efficiency by using "K computer", a peta-scale computer, with the core mission of "developing software that can contribute to medical care with an analytical approach to basic equations, an approach to extract experimental data that looms in on unknown theorems using massive amounts of experimental data, and an approach to integrate multiple levels to fully utilize the performance of a supercomputer having an unprecedented peta-scale performance. This will make it possible to integrate and organically link research and experimental data of varying scale, and to understand the highly varied phenomena of the biological body." Throughout its research, the data and findings collected during development have been released to academic conferences and journals, and a symposium was convened to freely exchange views.

Experimental use of "K computer", a peta-scale computer, started from fiscal year 2011, and some programs under development have achieved an execution efficiency of peta-flop class. From now on, we would like many people to become familiar with these new application programs, and to use them. Please take time to look at our web page that showcases individual programs (see diagram).

This page introduces all application programs that are currently under development. As an effort to give back the project's accomplishments to the society, we are planning to gradually release their source code. It is also planned to increase the content of the download page so that the public can use actual programs.

A status report of the research and development of these application programs will be released at the “Debrief Meeting 2011 on Next-Generation Integrated Simulation of Living Matter (ISLiM)” (December 21 (Wed) – 22 (Thurs) 2011). Please feel free to attend if you are interested.

Application Program Introduction Page

Japanese page: http://www.csrp.riken.jp/application_j.html
English page : http://www.csrp.riken.jp/application_e.html

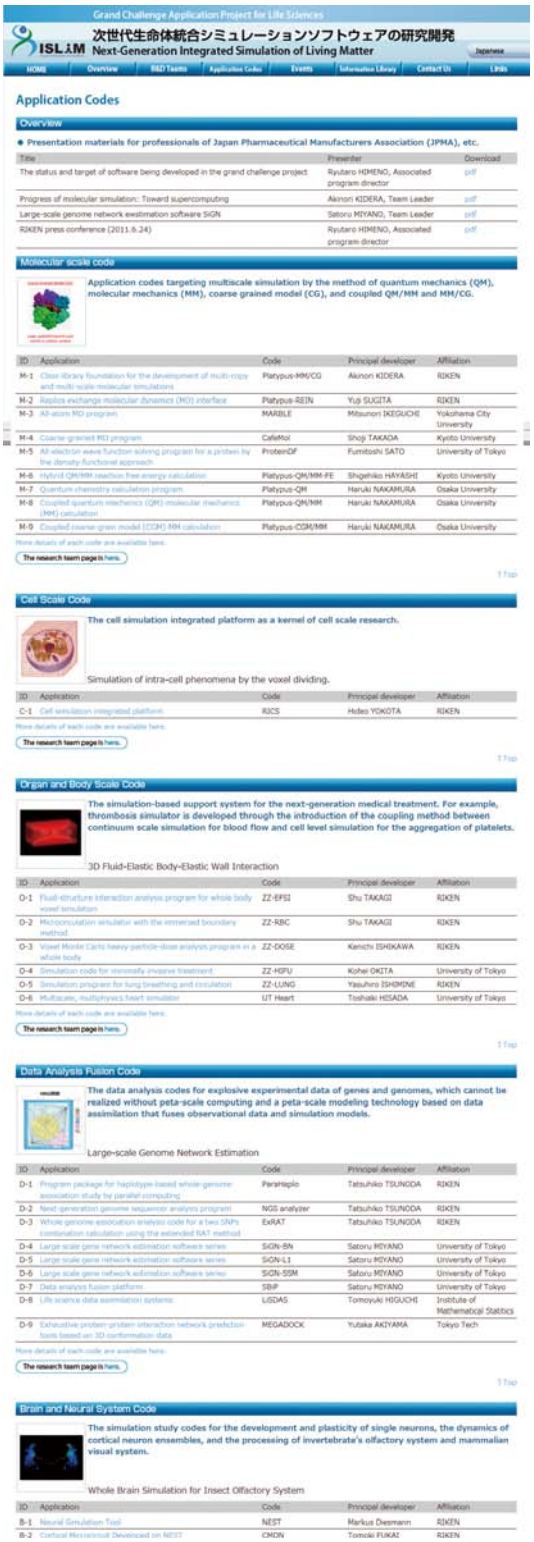
“Debrief Meeting 2011 on Research and Development of Next-Generation Integrated Simulation of Living Matter (ISLiM)”, Notice page

http://www.csrp.riken.jp/2011/islim-houkokukai2011_j.html

An introductory video for “Next-Generation Integrated Simulation of Living Matter” is available on the following page.

http://www.riken.jp/engn/r-navi/video/research.html

Video name: Approach to the Life Science Grand Challenge – Next-Generation Integrated Simulation of Living Matter –



Event Information

"Debrief Meeting 2011 on Next-Generation Integrated Simulation of Living Matter (ISLiM)"

- Date : December 21 (Wed), 22 (Thurs) 2011
- Location : The University of Tokyo, Takeda Hall (Hongo, Bunkyo-ku, Tokyo)

For debrief program and details, please refer to the web page.
http://www.csrp.riken.jp/2011/islim-houkokukai2011_j.html

4th BioSupercomputing Symposium

- Date : March 5 (Mon), 6 (Tue) 2012
- Location : Kobe city, Hyougo prefecture (TBA)

Details will be published on the web page (http://www.csrp.riken.jp) once fixed.

K computer ranked the fastest computer in the world.

"K computer" has been acclaimed No.1 in the "37th TOP500", the world's supercomputer computation performance ranking, which was released in June 2011. Its performance has marked 8.162 peta flops (performance of 8,162 tera floating point calculations per second) on the LINPACK benchmark during its preparation phase (peak performance of 8.774 peta flops). An execution performance of 93% is an outstanding record for a system of this scale. In addition, it has been proven to have top class energy efficiency due to its ranking No.6 in Green500 for its calculation performance per energy.

(Cover photo)

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

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

HPCI Strategic Program Field 1

Next-Generation Integrated Simulation of Living Matter

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

Supercomputational Life Science

The HPCI Strategic Program is a program of the Ministry of Education, Culture, Sports, Science and Technology aiming to produce ground-breaking results in computer science technology by maximizing the benefits of the HPCI (High Performance Computing Infrastructure) centered on the K computer, and encouraging developments in five research fields that need to be strategically addressed.
"Supercomputational Life Science" as a representative organization of RIKEN, has conducted research with the mission of understanding and predicting life phenomena based on large-scale simulation and advanced data analysis, and to apply these results to design and implement medicine and medical care with its research.

BioSupercomputing Newsletter Vol.5 2011.11

Issued: November 2011

RIKEN

Research Cluster for Innovation
Computational Science Research Program
2-1 Hirosawa, Wako, Saitama 351-0198, Japan
Tel : +81-48-462-1488 Fax : +81-48-462-1220
http://www.csrp.riken.jp

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