Strategic Programs for Innovative Research Field 1 Supercomputational Life Science

BioSupercomputing Newsletter

2013.9 Vol.9

CONTENTS

SCLS
SUPERCOMPUTATIONAL LIFE SCIENCE

Open Up the Innovative Possibilities of Computational Life Sciences

"Multifactorial," "dynamic," and complex life phenomena to analyze in depth by the "K computer"

Program Director (PD), HPCI Program for Computational Life Sciences, RIKEN Toshio Yanagida
Deputy PD, HPCI Program for Computational Life Sciences, RIKEN Akinori Kidera
Deputy PD, HPCI Program for Computational Life Sciences, RIKEN Yukihiro Eguchi

Two years and a half have almost passed since the "Supercomputational Life Science(SCLS)" started with the aim of elucidating the complex life system, in which various hierarchies such as genomes, biomolecules, cells, and organs are involved, by using the innovative High Performance Computing Infrastructure (HPCI) of Japan with the "K computer" at the core. A discussion meeting was held between Dr. Toshio Yanagida, Program Director (PD), Dr. Akinori Kidera, Deputy PD; and Mr. Yukihiro Eguchi, Deputy PD, on the progress of the programs toward their objectives of gathering numerous researchers from different fields to explore the possibility of predicting and controlling life phenomena by means of computational life sciences, and of linking the study results to the contribution to medical treatment and drug design.

Computational sciences required to comprehend the multifactorial and complex life system

Yanagida: During these two years and a half, the direction of study has been focused more specifically on all the four themes. I believe that the themes went smoothly from the very beginning. A road map for the use of high performance computers is finally about to be developed in the field of life sciences where there had been no such advanced methods so far.

Kidera: It is quite spectacular that, with the foundation that was established in the Life Science Grand Challenge (Next-Generation Integrated Simulation of Living Matter), which ended in March 2013, the full-fledged study began using the advanced computational resource called the "K computer" in this Strategic Programs for Innovative Research.

Eguchi: There is increasing attention paid to the "K computer" and computational life sciences from the aspect of dissemination of study results as well. More than ten companies including pharmaceutical companies also started a project last year to evaluate the Massively Parallel Computation of Absolute binding Free Energy "MP-CAFEE", developed in Theme 2 from the perspective of drug design sites. The participants are engaged in this project very earnestly. We understand how strongly they wish to apply computational life sciences with the "K computer" to the business of their companies.

Kidera: SCLS is a milestone project in that it aims to produce results that would contribute to medical treatment and drug design. Computational life sciences have just begun as a research area, and are still immature in that the sense of values has not been fully established yet. Under such circumstances, SCLS is about to take the initiative in forming this field in the direction of studies based on values open to the society, viz., medical treatment and drug design.

Eguchi: Human network development is also progressing smoothly in that direction. In Theme 3 and so on, the human network began to expand through collaboration with clinicians, etc. Obtaining collaboration from clinicians with computational life sciences was almost unprecedented.

Yanagida: Molecular biologists seem to have deepened their understanding of the need for studies using computers these days. One of the reasons is that it was revealed that the world of organisms is not monofactorial, but multifactorial with fluctuating conditions dependent on the environment. The multifactorial world means that it is a complex system that cannot be elucidated further only through conventional gene analysis. I believe they came to know that it is necessary to do experiments with the assistance of a new method to address the issue of complexity, specifically computational sciences.

Studies progressing smoothly in four themes

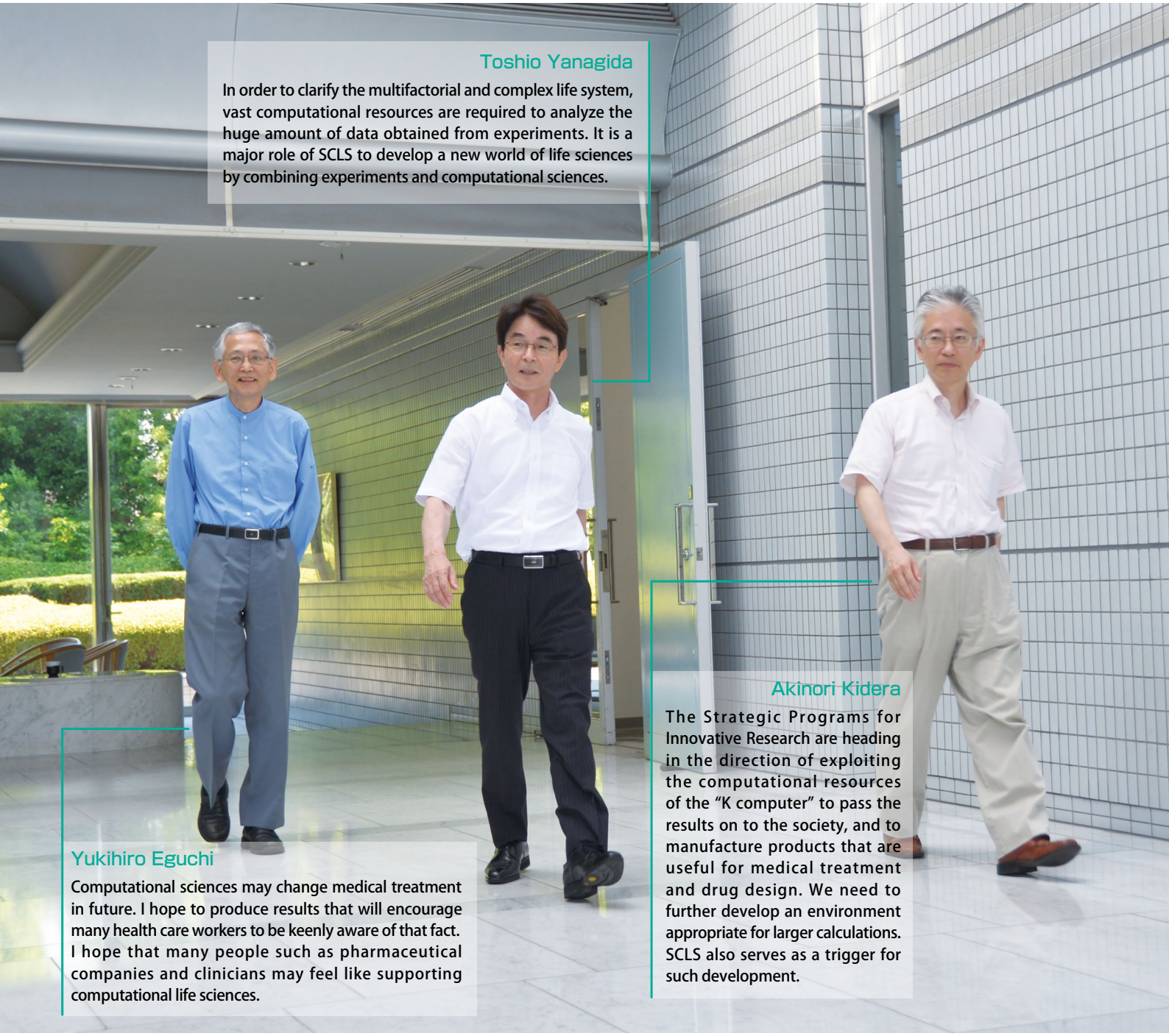
Kidera: The four themes in SCLS, which intend to elucidate life structured in a multifactorial, complex, and dynamic manner, by means of computational sciences, are largely divided into two. Themes 1 to 3 simulate objects apparently with physical substances. These approaches aim to clarify how biomolecules, cells, and organs function. On the other hand, Theme 4 aims to elucidate the regulatory network of the genome sequence system behind the life system not by means of a physical model, but by means of large-scale data analysis methodology. It remains to be seen how the two methods, which are significantly different in direction, will be linked together.

Yanagida: It is very interesting to see how the control network will function when dynamics and fluctuations are added to the life substance. It is likely that such a complex substance may actually be controlled easily in future.

Eguchi: It is not possible to connect Themes 1 to 4 in one stroke only among researchers of computational life sciences. We need to build collaboration with various researchers including biologists and clinicians. To that end, we need to produce attractive results that encourage researchers to participate in SCLS as a seemingly attractive project.

Yanagida: Such results have begun to be obtained in the respective themes.

Kidera: In Theme 4 "Large-scale analysis of life data," software technology to analyze the huge amount of genome data by the "K computer" has already been developed. On the other hand, in Theme 1 "Simulations of biomolecules under cellular environments," which intends to analyze intracellular conditions, where numerous molecules are



Toshio Yanagida

In order to clarify the multifactorial and complex life system, vast computational resources are required to analyze the huge amount of data obtained from experiments. It is a major role of SCLS to develop a new world of life sciences by combining experiments and computational sciences.

Akinori Kidera

The Strategic Programs for Innovative Research are heading in the direction of exploiting the computational resources of the "K computer" to pass the results on to the society, and to manufacture products that are useful for medical treatment and drug design. We need to further develop an environment appropriate for larger calculations. SCLS also serves as a trigger for such development.

Yukihiro Eguchi

Computational sciences may change medical treatment in future. I hope to produce results that will encourage many health care workers to be keenly aware of that fact. I hope that many people such as pharmaceutical companies and clinicians may feel like supporting computational life sciences.

jammed into a limited space at the atomic or molecular level, computation is proceeding to analyze and predict the functions of proteins within cells.

Eguchi: This is a very challenging activity, intending to simulate entire cells by making full use of the "K computer."

Yanagida: This program tries to approach cellular functions. Given the hierarchical jumps, however, it remains to be decided what should be reproduced to what extent.

Kidera: Theme 2 "Simulation applicable to drug design" succeeded in discovering compounds, which are new cancer drug candidates, after high-precision computation of approximately 300 novel compounds.

Yanagida: Theme 3 "Hierarchical integrated simulation for predictive medicine" is a study implemented to understand Parkinson's disease. It is progressing smoothly with success in integrating the simulation of neural signal generation, and the simulation of skeletal muscle contraction on induced by signal transmission.

Kidera: In the heart simulation performed with three hierarchies, viz. molecules, cardiac muscle cells, and the whole heart linked together, we simulated hypertrophic cardiomyopathy, and are about to succeed in reproducing the pathology. The results are expected to be applied to treatment at clinical sites. In addition, preparation for simulating the integration of the heart simulator and the vascular simulator is steadily progressing toward medical application to myocardial infarction.

Yanagida: Everyone wishes to be healthy. To this end, everybody has great expectations for the development of medical treatment and life sciences. I am keenly aware that they also have significant expectations of computational life sciences that have just begun. We are committed to do our best in SCLS for the remaining two years and a half, always keeping in mind that we are playing an important role in meeting such expectations.

Theme 1 Simulations of biomolecules under cellular environments

Whole cell simulation of signaling pathway

Laboratory for Biochemical Simulation, RIKEN Quantitative Biology Center
Kazunari Iwamoto

Cells have an ability to respond to the environment appropriately, so that they can survive in different environments. The response is called "signaling", and the route involved is called a signaling pathway. A signaling pathway is a network of serial reactions consisting of chemical reactions between intracellular proteins, and a cell has numerous pathways. Since it is known that breakdown of the signaling pathway makes a cell cancerous, it is an important task to reveal under what circumstances the signaling pathway breaks down.

One widely known signaling pathway is the Epidermal Growth Factor (EGF) signaling pathway. This pathway receives external stimulation from an EGF protein and transmits information into the cell to promote cell growth, division and differentiation. Recently, it has become possible to observe intracellular proteins at the monomolecular level. Observation of the EGF signaling pathway response revealed that there was a wide variety of cellular responses, even though cells with the same genes were cultivated under the same conditions and received the same stimulation (Fig. 1, left). Such a variation in response between cells (response inhomogeneity) has been suggested to have an influence on cancerous changes of the cells, drug resistance and so on. Therefore, clarification of response inhomogeneity is very crucial for understanding this influence. However, since numerous proteins participate in the signaling pathway, its understanding is becoming increasingly difficult by using only conventional biochemical or molecular biological techniques. Therefore, virtual cell simulation on a computer, called cell simulation, has come to be increasingly used. Our goal is the clarification of response inhomogeneity of the EGF signaling pathway through this technique.

Our cell simulation technique is different from conventional ones, and describes random movement, collision, and reaction (called fluctuation or noise) per intracellular protein molecule. Such a simulation was difficult due to the limitation of a computer's computing power. However, it is now becoming possible by utilizing the computing power of the K computer. Our laboratory is developing cell simulation software, "pSpatiocyte," and conducting research through the use of this software.

This is the latest result of our simulation. In the EGF signaling pathway, information is transmitted through the route, "EGF → Raf → ERK." The ERK protein eventually accumulates in the cell nucleus (Fig. 1, Center). Fig. 1 right shows the simulation result of visualized ERK protein. These two cells were under the same conditions, but their ERK response was found to be different. The number of protein molecules was different among EGF, Raf and ERK (EGF>Raf<ERK). Raf molecules are the smallest in number. In general, noise associated with signal transduction is more amplified as the number of molecules becomes smaller. Therefore, it is believed that noises amplified at Raf are transmitted downstream to ERK without modification, and produced intercellular variation. Presently, we cannot assure that real response inhomogeneity occurs through the same mechanism, but we are aiming at the whole cell simulation of the EGF signaling pathway while demonstrating those findings.

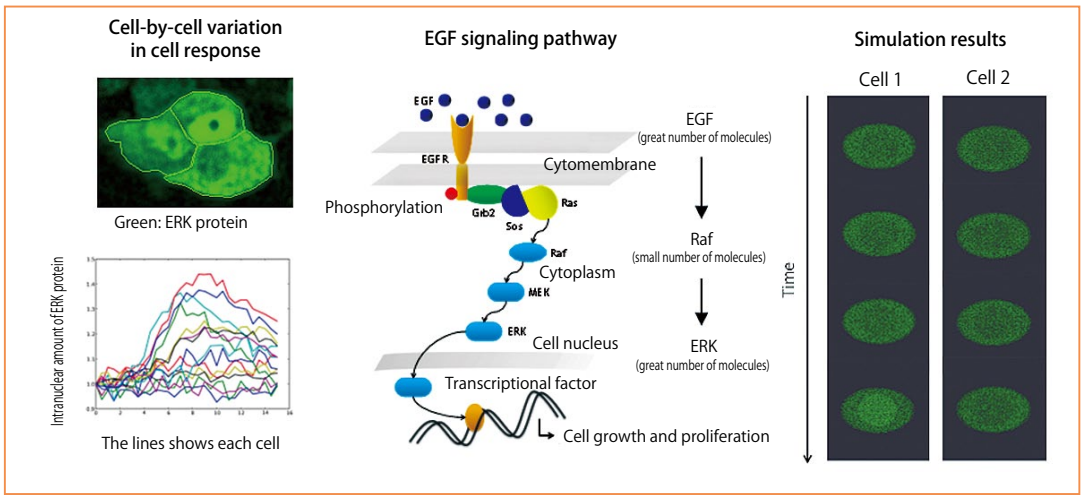


Fig.1: Cell-by-cell variation in cell response (Left), Pattern diagram of EGF signaling pathway (Center), An example of simulation results (Right)

See the URL on the right for detailed information on this report. http://www.kobe.riken.jp/stpr1-life/en/newsletter/Vol.9/zoomin01.html

Theme 2 Simulation Applicable to Drug Design

Molecular recognition mechanism observed through simulation and drug design

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

Our study is focused on "Drug Development by use of the supercomputer." We search for molecules acting strongly on disease-causing proteins through large-scale computation using the supercomputer, and apply the analysis results to drug development. For this purpose, we also develop new analysis techniques and theories. Simulation Applicable to Drug Design ranges from the basic/academic aspect to applicative/technical aspects. (Fig. 1)

From the perspective of the academic aspect, drug design may be summarized as "gaining knowledge about the molecular recognition mechanism." Many pharmaceutical products serve as drugs by suppressing the function of disease-causing proteins. What would happen if a drug affects not only the target protein, but also various other proteins? It would trigger a life-threatening adverse reaction. Therefore, drug must recognize and attack only the target protein (Fig. 2).

"Absolute binding free energy" is one of the physical quantities that indicate how strongly a molecule recognizes its target protein. A molecule with high absolute binding free energy can strongly recognize and attack its target protein, so it is a promising drug candidate. We believe that accurate prediction of absolute binding free energy will be an effective tool for drug development. We are making an attempt to apply the idea to a real cancer-causing protein by the computation of absolute binding free energy, proposed by Fujitani et al.[1] In this computational method, even an assessment of absolute binding free energy between a molecule and its target protein requires a massive computational resource. We have to make calculations for many molecules to find the one best suited for a drug, but it has been very difficult by use of conventional computers. Nowadays, it has become practicable to some extent due to the advent of the K computer.

Thanks to super large-scale calculation by use of the K computer, we are gradually accumulating new knowledge, though we sometimes face unexpected problems. We believe we are able to understand the molecular recognition mechanism "more deeply" and "more accurately" by gaining such knowledge. We consider that we will be able to step out of conventional empirical drug development and develop drugs more simply through a theoretical approach when we come to understand the molecular recognition mechanism.

One of the merits of molecular simulation is that we can view what we cannot see directly, for example proteins in the water, as if we are observing it through a high-precision microscope. What makes it more interesting is that we can create physically-impossible virtual states. By making the most of the virtual states, we can contrive ways to calculate real physical quantities at higher speed. In reality, we use this idea in our computation of absolute binding free energy. The development of a computational simulation method that is possible only by using computers is also the great attraction of this field.

Reference

[1] H. Fujitani, Y. Tanida, and A. Matsuura: *Phys. Rev. E*, **79** 021914 (2009), H. Fujitani, K. Shinoda, T. Yamashita, and T. Kodama: *J. Phys. Conf. Ser.* **454** 012018(2013)

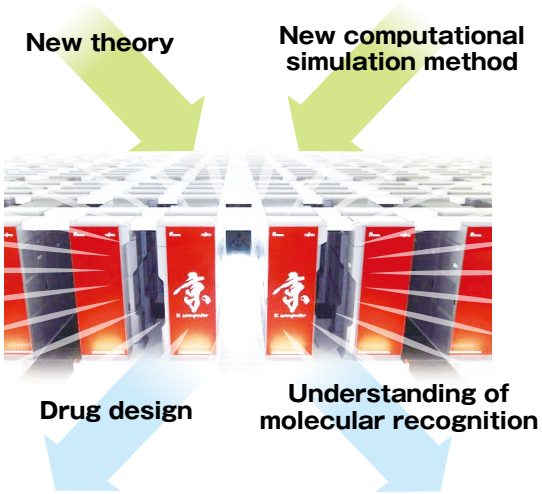


Fig. 1: Image of Simulation Applicable to Drug Design

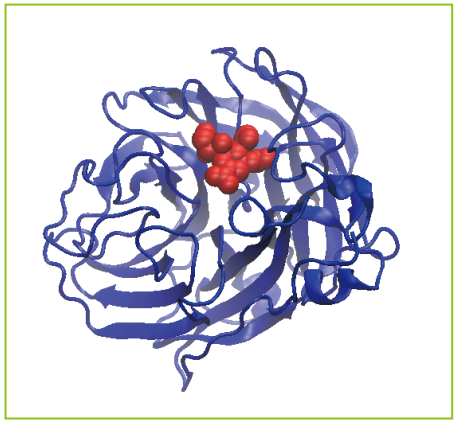


Fig. 2: Relenza (shown in red) bound to the influenza protein (shown in blue)

See the URL on the right for detailed information on this report. http://www.kobe.riken.jp/stpr1-life/en/newsletter/Vol.9/zoomin02.html

Theme 4 Large-scale analysis of life data

Massively parallel sequence analysis pipeline for next-generation sequencers

Akiyama Laboratory, Department of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology
Masanori Kakuta

DNA, responsible for genetic information, consists of chain combinations of monomers called nucleotide. Identification of the nucleotide sequence enables the estimation of the RNAs synthesized based on the DNAs, and the proteins synthesized based on the RNAs. Since DNA, RNA and proteins interact and participate in various vital functions and chemical reactions, nucleotide sequence determination gives a clue to the clarification of complex interactions between biomolecules. In addition, similar nucleotide sequences are sometimes observed both within and across species. Their comparison provides knowledge about the functions and evolution of similar parts of the sequences.

Thanks to the development of the DNA sequencer, an instrument that automatically reads nucleotide sequences, the speed of sequencing and the cost per base are improving rapidly. A commercially-available sequencer has the capacity to read about 100 billion base pairs per day. The human genome consists of about 3 billion base pairs, so its speed is obviously very high. Since it is relatively easy to obtain an enormous amount of DNA information, a high-speed analysis system for such information is now required. Our team is therefore developing a high-speed analysis algorithm and information-processing system which enables the analysis of an enormous amount of DNA information through parallel computing using multiple computation nodes.

Recently, metagenome, genomes in microbial communities, in the environment (surface and inside of the human body, soil, water sphere, etc.) has been actively studied following the progress of the DNA sequencing techniques. By comparing the composition of environmental microorganisms (phylogenetic analysis) or gene composition (functional analysis) between different environments, researchers are trying to obtain knowledge about the relationship between microorganisms and human health, the interaction between the environment and microbial communities, or the interaction among microorganisms.

In order to estimate the microorganisms present in the environment and their genes, we search sequence databases for sequences similar to the one we read. It was the time-consuming process.

We developed GHOSTX, which is more than 100-times faster than conventional techniques, through improvement of the search algorithm, and then developed GHOST-MP, which performs parallel analysis based on GHOSTX. GHOSTX constructs a data structure called a suffix array both in query sequences and database sequences. By using it as an index, we can search for similar sequences without scanning all sequences. GHOST-MP reduces the search time by overlapping communication and computation. The many computation nodes of the K computer and its high-speed inter-node network allow to process large-scale metagenomic data in a short time. By using GHOST-MP as a search technique, we are developing an analysis pipeline, which performs phylogenetic analysis and functional analysis based on the sequences obtained from the microbial communities, and analyzing real data utilizing the analysis pipeline.

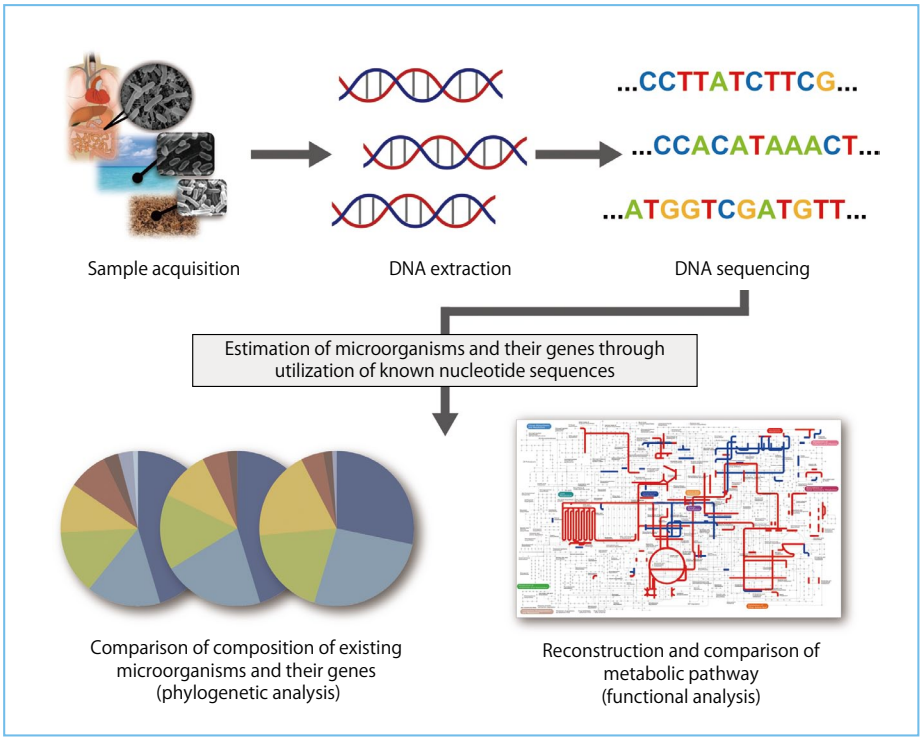


Fig. 1: Pattern diagram of metagenome analysis

See the URL on the right for detailed information on this report. http://www.kobe.riken.jp/stpr1-life/en/newsletter/Vol.9/zoomin03.html

SCLS Gotcha!

Fiscal 2013 Period 1 Peer-reviewed Results for Research Themes on the Use of SCLS Supercomputer System

Supercomputational Life Science(SCLS) aims to broadly form user communities for computational life sciences through research support, human resource development, and so on, as well as to promote creation of landmark results in life science areas including medical care and drug development through the use of HPCI. Towards achievement of this objective, as one way to make use of HPCI easy for external life science researchers and engineers, SCLS seeks research theme on the use of SCLS supercomputer systems highly compatible with the "K computer" on a peer-reviewed basis. Upcoming applications will be posted in September. For details, see the Information on Acceptance of Applications from the Public to Use SCLS supercomputer system (http://www.kobe.riken.jp/stpr1-life/outreach/sclsuse/application.html).

SCLS Theme No.	Name	Representative's affiliation	Theme title
201301101	Toru Hyakutake	Faculty of Engineering, Yokohama National University	Development of numerical technique on estimation of oxygen transport effect by tailor-made artificial red blood cell
201301102	Kenji Takizawa	Waseda Institute for Advanced Study	Development of Computer Modeling Techniques for Patient-Specific Cardiovascular FSI
201301103	Kenji Mizuguchi	National Institute of Biomedical Innovation	Construction of a method for screening drug candidates based on protein-interaction prediction
201301104	Ryuichi Shimizu	BioGrid Center Kansai	Utilization of parallel computational software,and Study on possibility for the software to facilitate drug discovery
201301105	Kazuya Yasuo	Shionogi & Co., Ltd.	Improvement of binding affinity prediction with Quantum mechanics
201301106	Huynh Quang Huy Viet	Graduate School of Environmental and Life Science, Okayama University	Parallelization and evaluation of an incompressible Navier Stokes equation solver for blood flow simulation
201301107	Akira Suyama	Graduate School of Arts and Sciences, The University of Tokyo	Stability of nucleic acid single-strand base-stacking
201301108	Takuya Hayashi	RIKEN Center for Life Science Technologies	Exploring functional brain network using non-invasive brain images and rapid and parallelized computation

SCLS Gotcha!

Co-host Symposium with Okayama University

In Supercomputational Life Science(SCLS), in order to encourage life science communities to understand computational life sciences in HPCI environments with "K computer" at the core and to form human networks, SCLS largely divides Japan into six blocks, and holds symposia in cooperation with hub universities and related academic societies in the blocks.

On Friday, June 1, SCLS held a symposium entitled 'Promotion of Fusion and Exchanges among Different Fields Engaged in Life Sciences: "K Computer" and Life Sciences' in cooperation with Okayama University.

Professor Matsuda reports his analytical result for bimolecular networks that control the browning of white fat cells.

At this second symposium held this year, lectures were given by 12 researchers, who are engaged in boundary study fields between computational sciences and life sciences, with the objective of setting up a forum, where researchers and engineers in different fields can exchange their views, and create new study fields and joint research around Computational Life Sciences, and discussing the methods and details of HPCI use in fields where life sciences and computational sciences are combined.

From "Theme 1: Simulations of biomolecules under cellular environments," Kazunari Kaizu, postdoctoral researcher (RIKEN), reported on the molecular scale simulation in a signaling pathway in consideration of the cellular environment. Then, from "Theme 4: Large-scale analysis of life data," Professor Hideo Matsuda (Osaka University) reported on his analytical result of bimolecular networks obtained through the use of the "K computer."

Dr. Kazunari Kaizu, postdoctoral researcher of RIKEN, introduces a parallel computation method for molecular scale simulation.

*He belongs to the same research team as Kazunari Iwamoto, postdoctoral researcher of RIKEN whose article appeared in "Zoom in" on page 4 of this issue.

We had more than 100 participants not only from universities such as students and teaching staff, but also from a wide range of affiliations from public research institutions to enterprises.

We are committed to continue to hold similar symposia to set up a forum for reporting studies using the "K computer" and HPCI environments in computational life sciences, as well as a forum for information exchange and transmission.

News & Events

SP Symposium RP Reporting session EX Exhibition SM Seminar WS Workshop

SP **September 19 (Thu) 13:00**

Development of a new field of life science and the "K computer"

We discuss development of a new field of life science involving computing science with you. ● Participation fee : Free

Location Centennial Hall Kyushu University School of Medicine (Fukuoka-shi, Fukuoka)

http://www.agr.kyushu-u.ac.jp/sympk/

RP **October 2 (Wed)—3 (Thu)**

Mid-term reporting session on research themes using HPCI systems with the "K computer" at the core in 2013

Location TIME 24 building (Koto-ku, Tokyo)

For more details, visit our website.
https://www.hpci-office.jp/pages/858

EX **October 9 (Wed)—11 (Fri)**

Bio Japan 2013 —World Business Forum—

Location Pacifico Yokohama (Yokohama-shi, Kanagawa)

Presentation at the RIKEN booth

EX **October 19 (Sat)**

RIKEN Open House Day at RIKEN AICS (Kobe)

Location RIKEN Advanced Institute for Computational Science (Kobe-shi, Hyogo)

Five fields come together to run a panel exhibition and interactive booth. (6F Science Square)

Open House Day in 2012

SM **October 28 (Mon) 12:30—13:20**

The 51st Annual Meeting of the Biophysical Society of Japan, Luncheon Seminar

Future seen from the aspect of Life Science and Computational Science

- Large-scale computing may be surprisingly important?
- Significance of large-scale computing in computational life sciences
- Invitation to the "K computer" and HPCI

Location Kyoto International Conference Center Room B-1 (Kyoto-shi, Kyoto)

SP **October 29 (Tue) 13:00—16:00**

The 51st Annual Meeting of the Biophysical Society of Japan

Structure, dynamics, and function of nucleosomes and chromatin in nuclear crowded environment

● Organizer : Yuji Sugita (RIKEN), Koichi Takahashi (RIKEN)

Location Kyoto International Conference Center (Kyoto-shi, Kyoto)

SP **October 31 (Thu) 10:05—11:20**

The 2nd Joint Conference on Informatics in Biology, Medicine and Pharmacology, Organized Session

Supercomputing with the "K computer" and drug design

HPCI Educational Program of Supercomputational Life Science(SCLS)

● Organizers: Masao Tanaka (Osaka University), Tsuneaki Sakata (Osaka University)

Location Tower Hall Funabori, 401 Meeting Room (Edogawa-ku, Tokyo)

EX **November 9 (Sat)—10 (Sun)**

Science Agora

RIKEN research centers in the Kobe area come together to make a joint presentation with SCLS.
Booth name: Because it is a cell

Location National Museum of Emerging Science and Innovation (Koto-ku, Tokyo)

EX **November 17 (Sun)—22 (Fri)**

SC13

A lecture on SCLS is scheduled at the "K computer" booth.

Location Colorado Convention Center (Denver, Colorado, USA)

WS **December 4 (Wed) 13:15—15:45**

The 36th Annual Meeting of The Molecular Biology Society of Japan, Biology of Diseases Workshop

Leveraging Supercomputers for Hierarchical and Systems Understanding of Life towards Strategic Intervention against Diseases

● Organizers: Satoru Miyano (The University of Tokyo), Seiji Ogawa (The University of Tokyo)

Location Kobe International Conference Center, 4F 401+402 (Kobe-shi, Hyogo)

SP **December 17 (Tue)**

(Tentative title) Strategic Programs for Innovative Research Field 1 x Field 2 Symposium

Computing Biomolecular Complex System What is brought about by interaction

● Participation fee : Free

Location Nagoya University (Nagoya-shi, Aichi)

We will notify the details on the website of SCLS as soon as they are fixed.

SCLS SUPERCOMPUTATIONAL LIFE SCIENCE

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.9 2013.9

RIKEN 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

Issued: September 2013

RIKEN 2013-071