ISLiM

Proceedings of the

4th Biosupercomputing Symposium

– International Symposium for Next-Generation Integrated Simulation of Living Matter (ISLiM) –

December 3–5, 2012
Tokyo International Forum (Hall D7)

"Next-Generation Integrated Simulation of Living Matter (ISLiM)" program
commissioned by Ministry of Education, Culture, Sports, Science and Technology (MEXT)

ISLiM

Proceedings of the

4th Biosupercomputing Symposium

–International Symposium for Next-Generation Integrated Simulation of Living Matter (ISLiM)–

December 3-5, 2012

Tokyo International Forum (Hall D7)

"Next-Generation Integrated Simulation of Living Matter (ISLiM)" program
commissioned by Ministry of Education, Culture,
Sports, Science and Technology (MEXT)

Agenda

December 3 (Mon)

10:00-10:20	**Opening Talk** Koji KAYA (RIKEN)
(Session Chair: Shu TAKAGI, RIKEN/University of Tokyo)	
10:20-11:10	**ISLiM Keynote: Current status of software development in ISLiM and its future** Ryutaro HIMENO (RIKEN)
11:10-12:00	**What can Multiscale Models do for Precision Medicine?** Grace Peng (National Institute of Health)
12:00-13:00	Lunch break
(Session Chair: Ryutaro HIMENO, RIKEN)	
13:00-13:50	**Development of Multiscale Thrombosis Simulator** Shu TAKAGI (RIKEN/University of Tokyo)
13:50-14:40	**Multiscale simulations of tumor angiogenesis with therapeutic applications** Aleksander Popel (Johns Hopkins University)
14:40-15:10	Coffee break
15:10-16:00	**Multi-scale, multi-physics heart simulator "UT-Heart" for heart research** Seiryo SUGIURA (University of Tokyo)
(Session Chair: Hideo YOKOTA, RIKEN)	
16:00-16:50	**Is eukaryotic cell biology ready for supercomputing? A perspective from the Virtual Cell project** Ion I. Moraru (University of Connecticut Health Center)

December 4 (Tue)

(Session Chair: Makoto TAIJI, RIKEN)	
10:00-10:50	**K computer Keynote: Real Peta-scale computing, stairway to Exa** Motoi OKUDA (Fujitsu)
10:50-11:40	**Optimization of Life-Science applications on the K computer** Yousuke OHNO (RIKEN)
11:40-12:30	**Large Scale Biomolecular Modeling with IBM Blue Gene** Ruhong Zhou (IBM)
12:30-13:30	Lunch break
(Session Chair: Ryutaro HIMENO, RIKEN)	
13:30-13:55	**Research activity of Cell Scale Simulation Team** Hideo YOKOTA (RIKEN)
13:55-14:20	**Dissection of regulatory mechanisms for metabolic systems by quantitative imaging mass spectrometry** Makoto SUEMATSU (Keio University)

(Session Chair: Shin ISHII, Kyoto University)	
14:20-15:10	**Closing the loop: simulation of the whole sensory-motor neural network in action** Kenji DOYA (Okinawa Institute of Science and Technology 〈OIST〉)
15:10-15:40	Coffee break
15:40-16:30	**Multi-scale modelling of the motor system underlying goal-directed behaviour in a vertebrate model organism** Sten Grillner (Karolinska Institutet)
16:30-17:20	**Brain-scale neuronal network simulations on K** Markus Diesmann (Forschungszentrum Jülich)
(17:30-19:00)	**Social (optional)**

December 5 (Wed)

(Session Chair: Akinori KIDERA, RIKEN/Yokohama City University)	
10:00-11:00	**Molecular machines and nuclear processes studied by coarse-grained molecular simulations** Shoji TAKADA (Kyoto University)
11:00-12:00	**Biomolecular simulations under cellular environment** Yuji SUGITA (RIKEN)
12:00-13:10	Lunch break
(Session Chair: Mitsunori IKEGUCHI, Yokohama City University)	
13:10-14:10	**Large-scale simulations of biomolecular machines** Karissa Sanbonmatsu (Los Alamos National Laboratory)
14:10-15:10	**Use of Modeling and Simulation in Drug Discovery and Development: It is all about the Questions** Sandra R. B. Allerheiligen (Merck)
15:10-15:40	Coffee break
(Session Chair: Satoru MIYANO, University of Tokyo)	
15:40-16:05	**Supercomputing for Next-Generation Cancer Research** Satoru MIYANO (University of Tokyo)
16:05-16:30	**Large-scale protein-protein interaction network prediction by an exhaustive rigid docking system MEGADOCK** Yutaka AKIYAMA (Tokyo Institute of Technology)
(Session Chair: Yutaka AKIYAMA, Tokyo Institute of Technology)	
16:30-16:55	**Supercomputing accelerates genomic medicine** Tatsuhiko TSUNODA (RIKEN)
16:55-17:20	**Towards efficient improvement of transcriptional circuit models by Life Science Data Assimilation System (LiSDAS)** Tomoyuki HIGUCHI (Institute of Statistical Mathematics)
17:20-17:30	**Closing Talk** Ryutaro HIMENO (RIKEN)

Table of Contents

ISLiM Talks

Preface

Koji Kaya

Program Director
Computational Science Research Program, RIKEN

Profile:

1966, Doctor of Science, Tokyo University
1966, Research Staff, Theoretical Organic Chemistry Lab., RIKEN
1970, Associate Professor, Department of Chemistry, Tohoku University
1973, Member of Technical Staff, Bell Telephone Laboratories, U. S.
1981, Professor, Department of Chemistry, Keio University
1999, Director in General, Institute for Molecular Science
2004, Directors of Discovery Research Institute, and Wako institute, RIKEN
2006 - Program Director, Next Generation Computational Science Research Program

In 2006, The Next Generation Super Computer Project started at RIKEN under the leadership of President Ryoji Noyori sponsored by MEXT Japan. Since then, this project experienced and overcome many hurdles both in technical and political points. In October 2011, we reached the final goal, i.e., establishment of 10 peta-flops computation.

In parallel with the development of the super computer (hard ware), the grand challenge program for the software development for peta-flops computation in life science started in October 2006 which is named Integrated Simulation of Living Matter (abbreviated as ISLiM). ISLiM program consists of 6 teams including 4 computer simulation teams, a data analysis team and software sophistication technology team. As a result of hard works by the young reseachers, most of the developed software programs finally reached to the final stage of the development worthwhile to challenge to peta-flops computation.

In this symposium, we focus our attention on the reports and discussions of the results of our software development and possible applications of the software programs to life sciences including drug discovery, medical applications in addition to the sophisticated lectures by the invited speakers from abroad.

Finally, we wish to express our sincere gratitude to the 7 invited speakers from abroad who kindly accepted our invitation to ISLiM 4-th symposium for their great contribution to make this symposium fruitful for all the participants.

Koji Kaya
Program Director
Computational Science Research Program
RIKEN

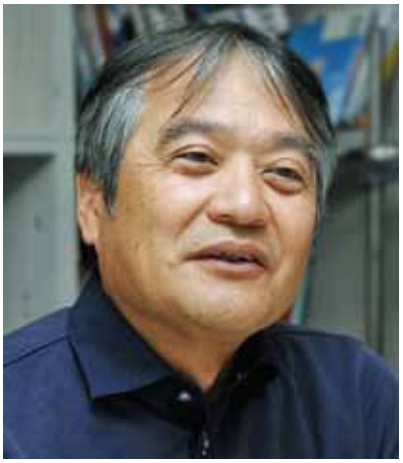
Keynotes

ISLiM Keynote:

Current status of software development in ISLiM and its future

Ryutaro Himeno

Group Director, Research and Development of Integrated Simulation of Living Matter, RIKEN



Profile:

Ryutaro HIMENO received the B.E. and the M.E. from Kyoto University, Kyoto, Japan in 1977 and 1979, respectively. He received Doctor of Engineering degree from the University of Tokyo in 1988. In 1979, he joined Nissan Motor Co., Ltd., Yokosuka, Japan, where he has been engaged in the research of applying Computational Fluid Dynamics analysis to the car aerodynamic development. From 1984 to 1986, he served as a researcher at the Institute of Space and Astronautical Science, Tokyo, Japan. In 1999, he joined RIKEN. He is now the director of Advanced Center for Computing and Communication, Deputy Program Director of Computational Science Research Program and Group Director of Research and Development of Integrated Simulation of Living Matter at RIKEN. He is a visiting Professor at the Kobe University, Hokkaido University and Tokyo Denki University. He is also well known researcher in Baseball Physics.

He was a winner of 2006 Gordon Bell Prize (Honorable Mention, Peak Performance, with Dr. Tetsu Narumi, et al) and Computer Visualization Contest in 2000 by Nikkei Science. He received both JSME Computational Mechanics Award and Computational Award by Japan Association of Computational Mechanics in 2011 as well as JSME Computational Mechanics Achievement Award in 1997 and JSME Youth Engineer Award in 1988. He was also awarded Paper Award by NICOGRAPH in 1993, Giga FLOPS Award by CRAY Research Inc. in 1990 and other awards.

Abstract

1. Introduction

Our project ISLiM started in 2006 and six years have past. The project will end in the end of March, 2013. This international symposium was planed to show what we have developed and achieved in past six years and to announce what we are going to do from now on.

2. What we plane in 2006

In 2006, we started a grand challenge project called ISLiM for K computer to demonstrate its performance. The ISLiM stands for Integrated Simulation of Living Matter to reproduce life phenomena on a supercomputer for understanding them and developing new medicine or new medical treatments. We have 6 research teams: Molecular scale team, cell scale team, organ and body scale team, data analysis fusion team, brain and neural system team and HPC team. We started a high performance software package for life science for K computer which contains 31 application software.

The target was as follows:

1) At least 2 application codes achieved effective one PetaFLOPs on K computer.
2) Innovative simulated results shall be shown using full system of K computer.
3) All codes shall be available on K computer when users want to use

3. Current status

Three codes: ZZ-EFSI, cppmd and UT-Heart achieved effective performance larger than 1 PetaFLOPs on K computer. 13 codes showed good liner scaling up to more than 10,000 nodes on K computer.
Computation using K computer is currently going on and the results will be shown from each team in this symposium. Major codes we have developed can be down-loadable from our web site and the number of available codes are increasing. We will complete the library at the end of March, 2013.
We will keep the library for users who want to use and take care of those users even after we finish our project.

References

1. http://www.csrp.riken.jp/index_e.html

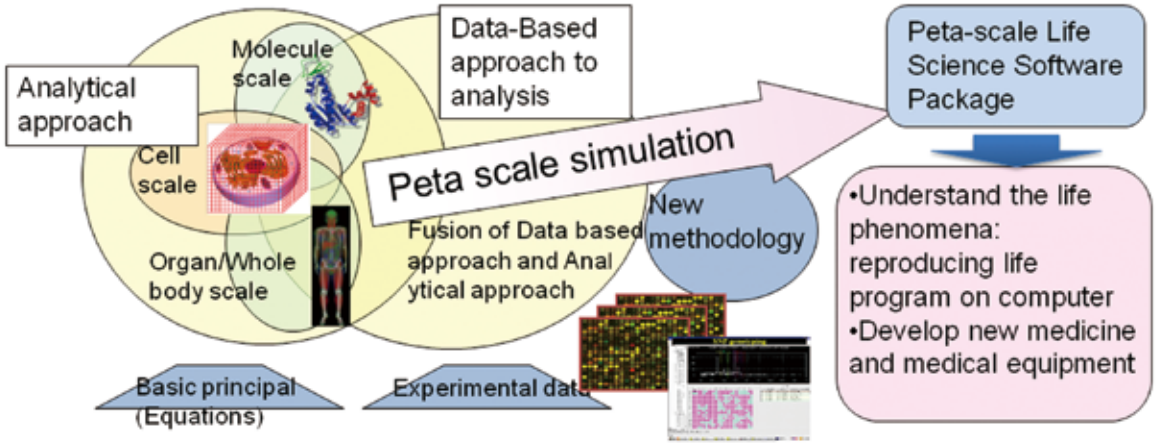


Figure1 project target and structure of research teams

Current status of software development in ISLiM and its future

Ryutaro Himeno
Group Director, Research and Development of Integrated Simulation of Living Matter, RIKEN
RIKEN
himeno@riken.jp

Motivation

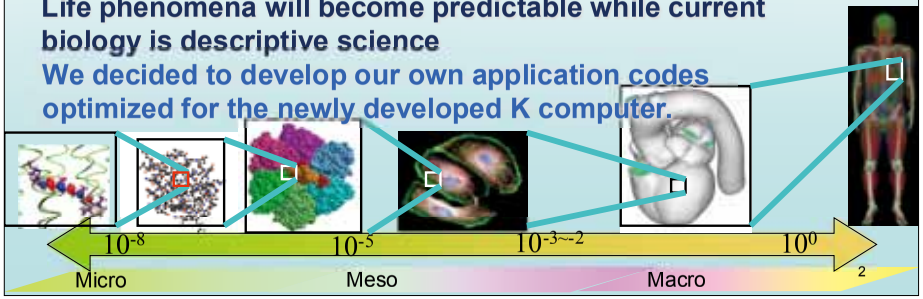
Life Phenemena are so complicated and still unknown.

But in molecular scale, they are governed by theories in Physics

Movement of many particles will be calculated on Supercomputer in some day.

Life phenomena will become predictable while current biology is descriptive science

We decided to develop our own application codes optimized for the newly developed K computer.



Next-generation Supercomputer R&D project and Life Science Grand Challenge

		FY2006	FY2007	FY2008	FY2009	FY2010	FY2011	FY2012
System		Conceptual design	Detailed design		Prototype, evaluation	Production, installation, and adjustment		Tuning and improvement
Applications	Next-Generation Integrated Nanoscience Simulation	Development, production, and evaluation					Verification	
	Next-Generation Integrated Life Simulation	Development, production, and evaluation					Verification	
Buildings	Computer building		Design	Construction				
	Research building			Design	Construction			

Life Science Grand Challenge: ISLiM

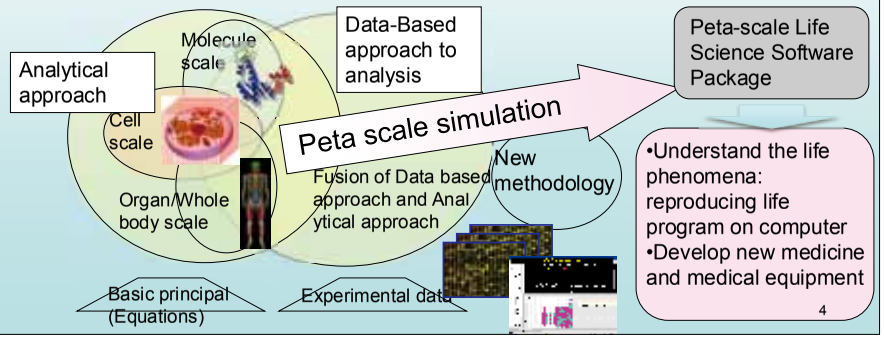
Grand Challenge in Life Science at 1st stage

Goal

Understand the life phenomena: reproducing life program on computer
Develop new medicine and medical equipment

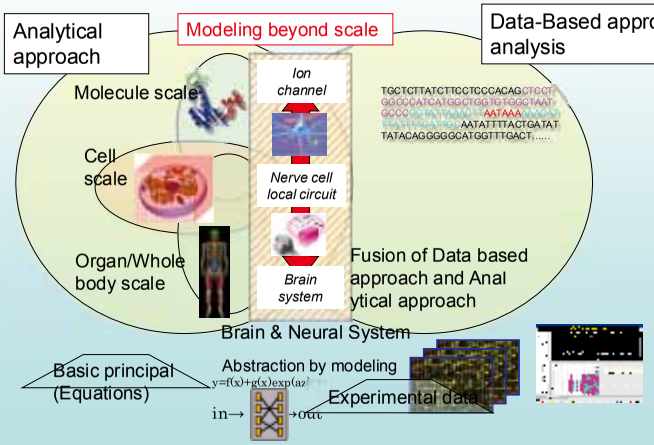
Approach

combination of analytical approach and Data-driven approach on Peta-scale computer



Grand Challenge in Life Science at 2nd stage

Added Brain & Neural System Research

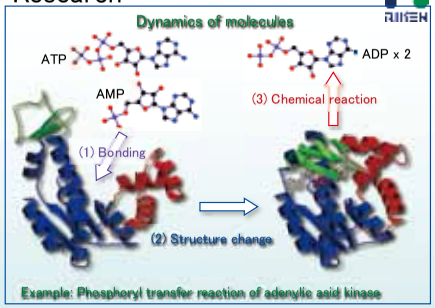


5

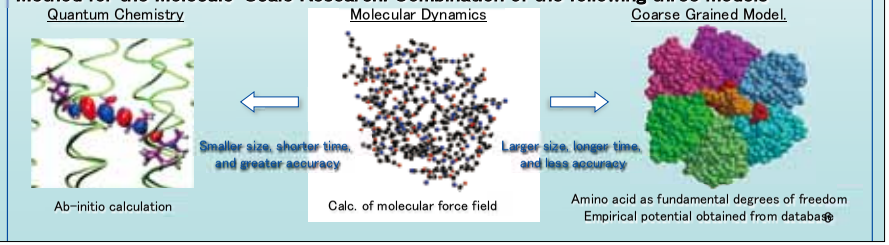
Molecule-Scale Research

Objectives

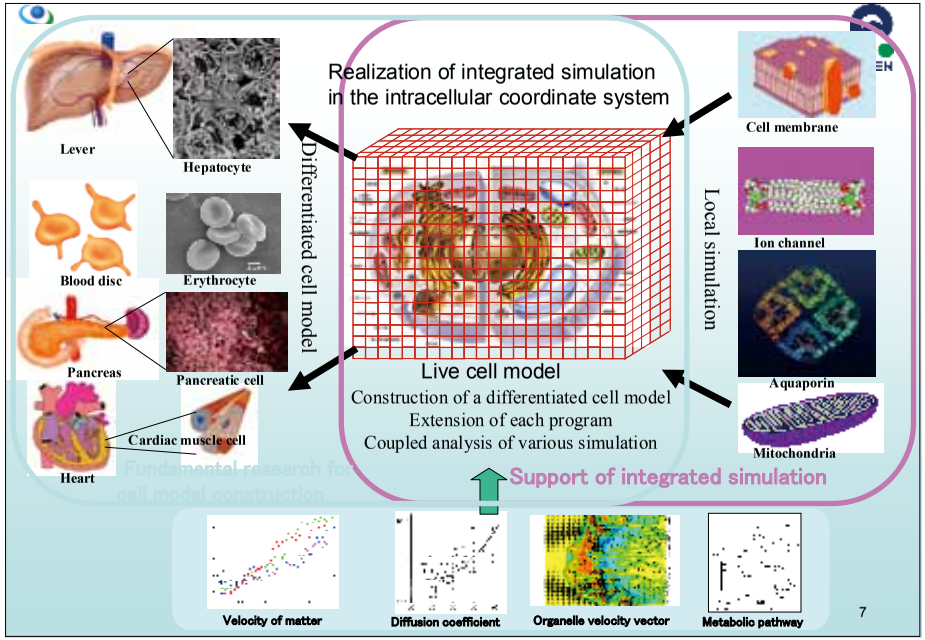
Simulating protein function in life phenomena to reveal important subjects in molecular and cell biology



Method for the Molecule-Scale Research: Combination of the following three models

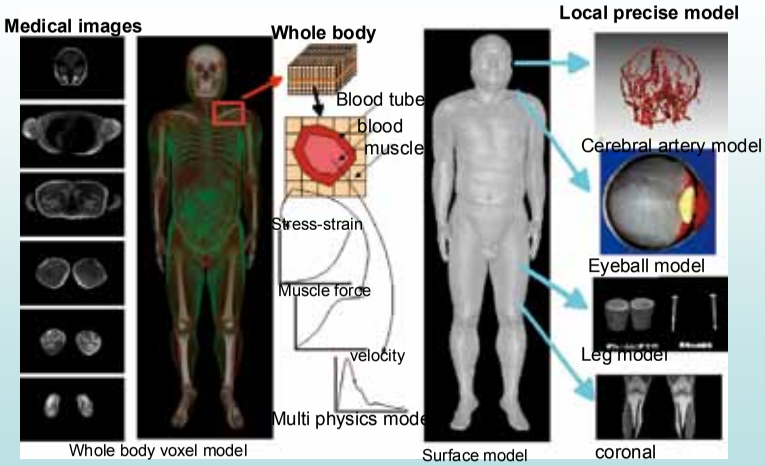


8



7

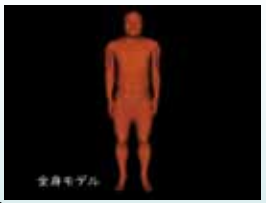
Organ/Whole body simulation



8

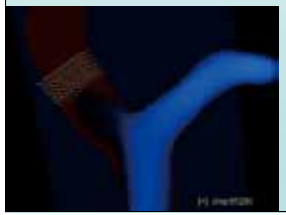
Planned simulation based on precise Human model

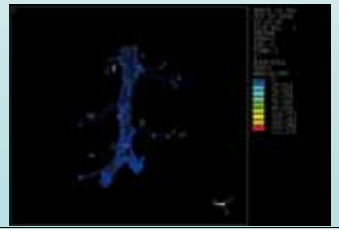
By Prof. Hisada and Prof. Sugiura



and more….

Ex) Ultrasound propagation simulation for HIFU





9

Overview of data analysis research
Fusion of large-scale heterogeneous data analysis and life-science simulation

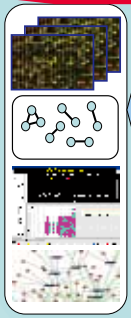
Exponential increase of experimental data
- Higher dimensions: from gene to exon
- Heterogeneity: array, structure, dynamics,…
- Invisibility: deficit of information

Peta-scale data-analysis

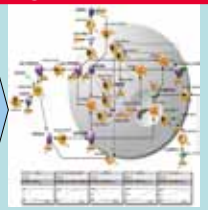
Modeling of functions

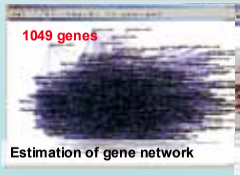
Peta scale computing

Expected outcome
- Order-made medicine and therapy based on optimized peta-scale data analysis
- Peta-scale analysis of gene network for drug target discovery and validation



Data-driven large-scale simulation technology





Estimation of gene network

In-silico analysis of drug-target gene

- Fusion of simulation models and experimental data based on data assimilation
- Automation of large-scale modeling
- Information technology for fusion of simulation and data analysis

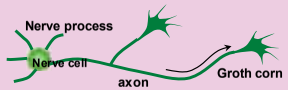
10

Brain & Neural System Team lead by Prof. Ishii (started in Oct. '08)

Challenge brain and neural system simulation combining scales: neuron, local circuit and brain. This team includes Prof. Kanzaki, Dr. Usui, Dr. Diesman and Prof. Doya.

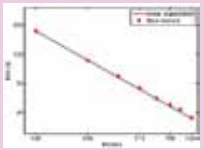
Neuron simulation

High precision neuron simulation with development & plasticity mechanism (No. of molecules: 10, 10^2 compartments)



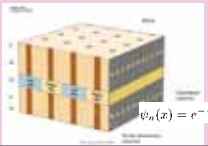
Local circuit simulation

High precision simulation of information processing & plasticity in brain cortex column (1010^{5}cells,10^{7}synapse)



Whole brain simulation

Challenge image processing of 10^5pixel picture by amphiblestrode model with10^6cells & visual cortex with columns & 10^5 cells



Grand challenge targets in long term

Establish a total model integrating input, decision and command of brain.

Investigate Neurodegenerative disorder and contribute to its medical treatment

11

HPC Team for development of common software platform and optimization on the new supercomputer

Molecular scale | Cell scale | Organ / whole body | Data-Analysis | others

Each software development

Support to derive higher performance

Common libraries + Core software

Common software platform and optimization

GUI / Workflow

Easy to use from user's point of view

Support to utilize developed codes

Software family optimized for the supercomputer

Grand Challenge Goal

Feedback after applying real phenomena

Improve by ourselves

Improve software for commercial use

Software house

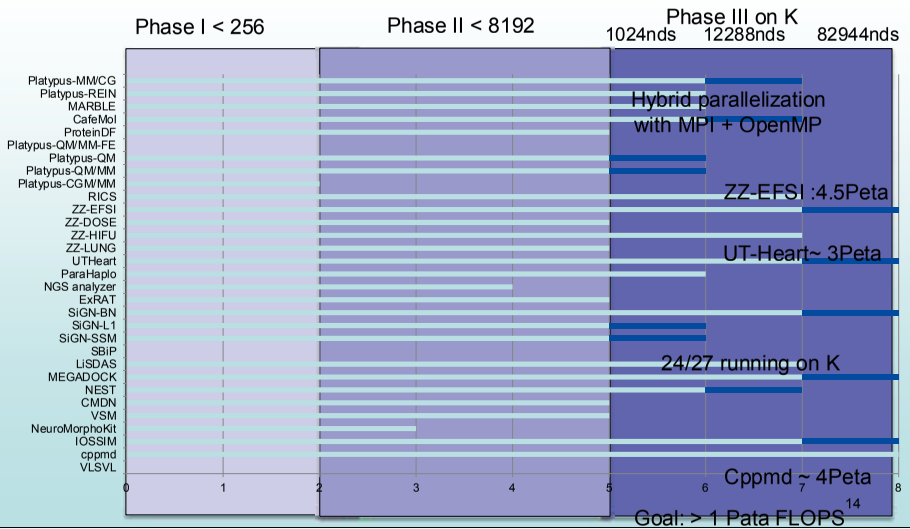
Academic and industrial usage

12

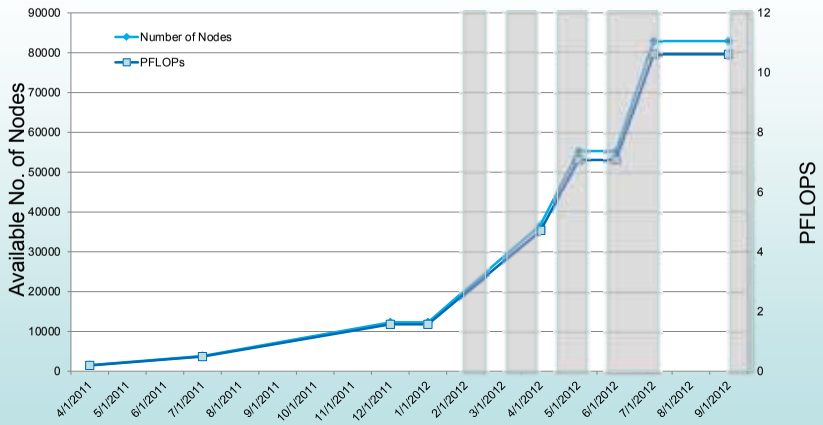
Life Science Software Package

	Applications	No. of application Software
Molecule scale	MD, Quantum Chemistry, Coarse Grain MD	9
Cell scale	Voxel-based multi-compartment transport-diffusion simulator	1
Organ/whole body scale	Heart sim., Lung sim., HIFU sim., Fluid-structure sim. for blood flow,	6
Brain & neural system	Neural sim., whole brain sim., cortical micro circuit sim.,	5
Data based analysis	Whole genome association study, data assimilation, prediction of protein-protein interaction, genome sequencer data processing, Haplotype whole genome association study	9
HPC	Parallel middleware, parallel visualization software, high-speed software core library	4

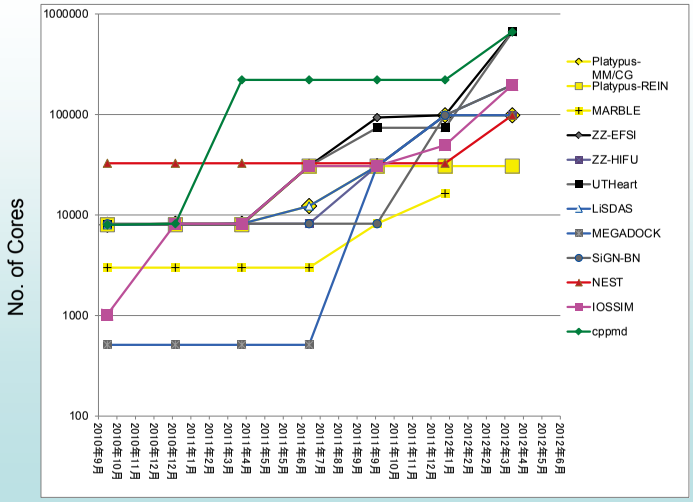
Current status of development



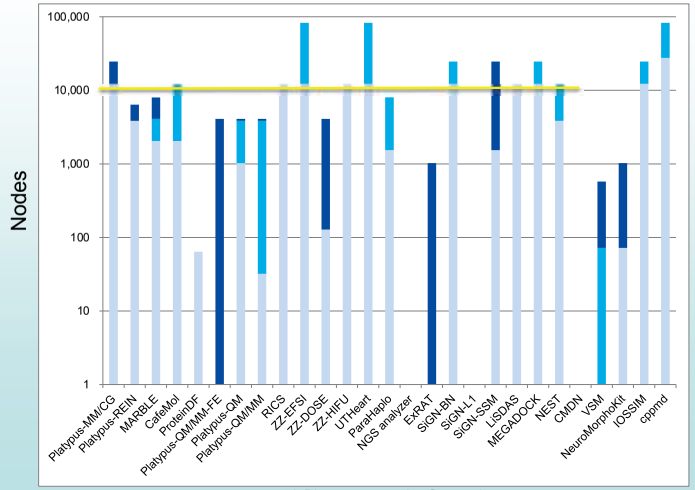
Test operation of K computer



Code development history: scalbility



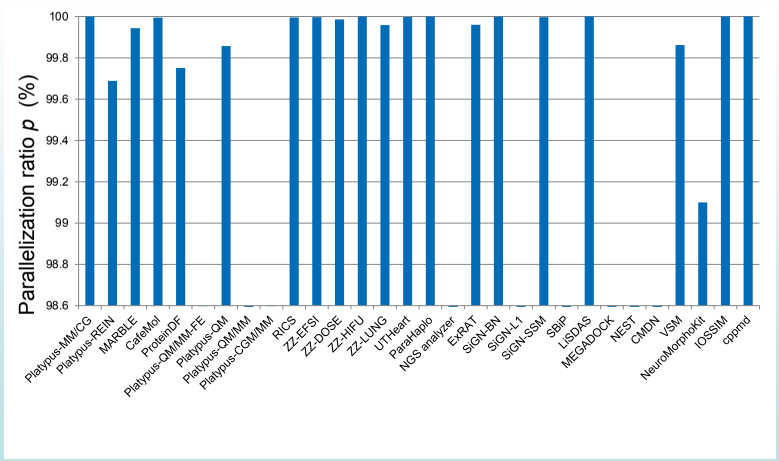
Max Node No. of Linear Speedup



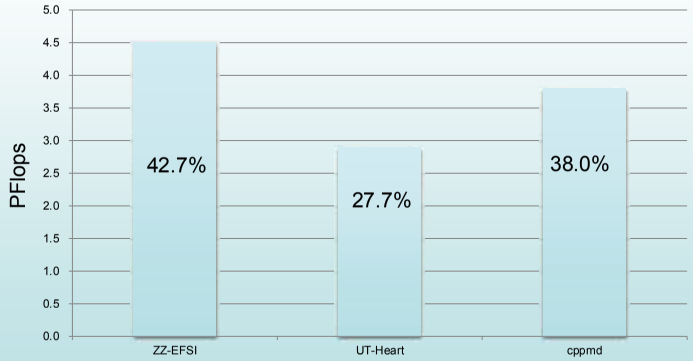
End of July, 2012

End of April, 2012

Parallel efficiency

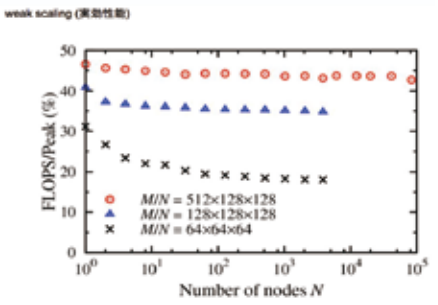


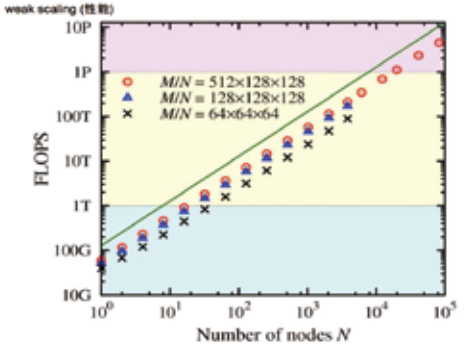
Measured performance on K computer using full system



82,944 nodes

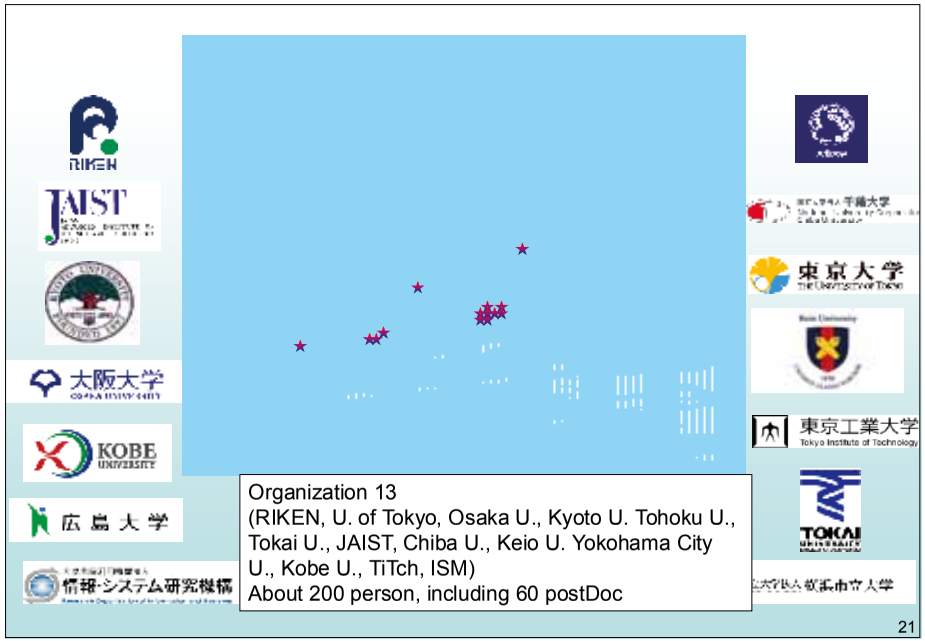
Weak Scaling on K: ZZ-EFSI





Changing solvers in ZZ-EFSI

- Using FFT to solve Poisson equation on PC
- Multi-color SOR on PC Cluster
- On K computer which has 80,000 CPUs
  - Implicit solver → explicit solver using quasi-compressible formulation
    - less communication



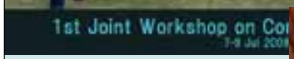
Organization 13
(RIKEN, U. of Tokyo, Osaka U., Kyoto U. Tohoku U., Tokai U., JAIST, Chiba U., Keio U. Yokohama City U., Kobe U., TiTch, ISM)
About 200 person, including 60 postDoc

21

ISLiM RIKEN

Contributors

Workshop in 2009



Workshop in 2008

22

ISLiM RIKEN

Summary

- We have developed wide variation of high performance application codes in Life Science for K computer, from molecular scale to organ/whole body scale as well as data analysis and brain/ neural system.
- 12 codes linearly scale more than 10,000 nodes of K
- ZZ-EFSI achieved 4.5 PetaFLOPs and 2 other codes achieved more than 3 PetaFLOPs.
- Currently, we are analyzing computed results right now.
- The codes are available from our web:

http://www.csrp.riken.jp/index_e.html

ISLiM RIKEN

For your references

Our homepage:http://www.csrp.riken.jp/index_e.html
Youtube:**Approach to the Life Sciences Grand Challenge-Next Generation Integrated Simulation of Living**

24

K Computer Keynote:

Real Peta-scale computing, stairway to Exa

Motoi Okuda

Executive Architect, Technical Computing Solutions Unit, Fujitsu Ltd

Profile:

Motoi Okuda Ph.D.
Executive Architect, Technical Computing Solutions Unit, Fujitsu Ltd.

Work Experience

· Planning, marketing and support of High Performance Computer
· Supervising Fj activities of national projects such as National Grid Project and Next Generation Supercomputing Projects
· Developing computational science and engineering applications, such as fluid dynamics nuclear code, crashworthiness, and molecular dynamics
· Project leader of Parallel Computing Center in Fujitsu Lab.

Education

· Ph.D., Information Science from Japan Advanced Institute of Science and Technology, Ishikawa.
· Master of Nuclear Engineering from Nagoya University.

Membership in Professional Societies includes

· Executive board member of The Japan Society for Computational Engineering and Science (JSCES)

Abstract

FUJITSU and RIKEN have been working together to develop K-computer.
After six- year project, the K-computer entered into service at the end of September and many users are now using the huge amount of computing power We are now expecting innovative outcomes of the most advanced K computer use.
My talk aims to cover the outline of the K computer project, its design concept, how to implement the system, impacts of adopted technologies on application execution, and some examples of application performance achievement. K computer's CPU, SPARC64™ VIIIfx, and Tofu interconnect are the key technologies for achieving the project targets such as performance, reliability, operability and power consumption. I will talk about these features and an example of performance on application program.
The talk also refers to our new challenges to the next step, Exa-scale computing, based on lessons we learned from K computer project and technology trends. Successful project management is crucial to promoting such a huge project. So I would like to mention it briefly for the future project planning.
Japanese government has just begun their new challenge to the future supercomputer system. I will also refer to the outline of Japanese activities.

Real Peta-scale computing, stairway to Exa

Motoi OKUDA
Technical Computing Solutions Unit
Fujitsu

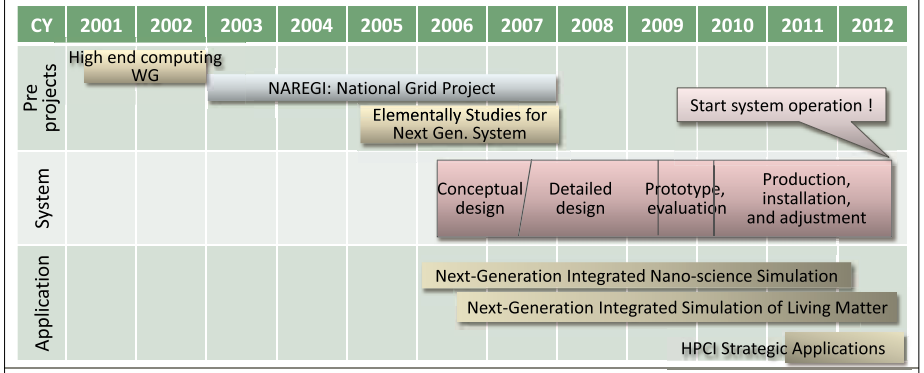
4th Biosupercomputing Symposium

Agenda

- History of the ***K computer*** project
- Design concept of the ***K computer*** and its achievements
- Aiming for Exascale computing
- Conclusion

Time-line of the *K computer* project

- In 2001, High-end computing WG was established and investigation activities started
- Following Gird project, Elementally studies started in 2005
- ***K computer*** project started in mid-2006 with two application projects
- System installation started in Oct. 2010
- Full system installation was finished in August 2011 and official operation started on 28th Sep. 2012



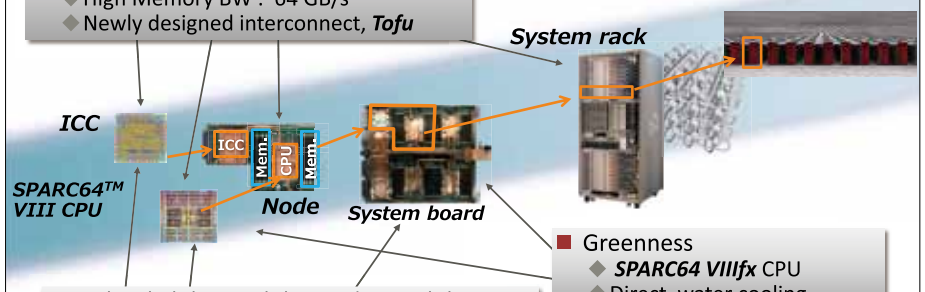
Design targets of *K computer* and its achievements

- High Performance
  - High peak performance and high performance efficiency
  - *100 times more powerful* than the fastest supercomputer in 2005
- High operability
  - Low power consumption
  - High reliability and easy to operate
- Highly parallel application performance and productivity
  - Easy to extract high performance from the highly paralleled programs without inordinate burden to programmers
  - Performance target of each strategic applications
- Time line
  - Development of the system was completed in *the end of March 2012*
- No.1 on 37th TOP500 list in June 2011 & 38th TOP500 in Nov. 2011
  - *10.51 PFlops , 12.66MW* and *93.17 %* efficiency in LINPACK BMT (Nov. 2011)
- 2011 Gordon Bell Award, Peak Performance
  - Sustained performance of *3.08 PFLOPS* (Running on 7.08PFlops system)
  - Efficiency of *43.6 %*

K computer Technology Highlights

FUJITSU

- High Performance
  - Fujitsu ***deigned SPARC64 VIIIfx*** CPU with HPC enhancement
  - High Memory BW : 64 GB/s
  - Newly designed interconnect, ***Tofu***

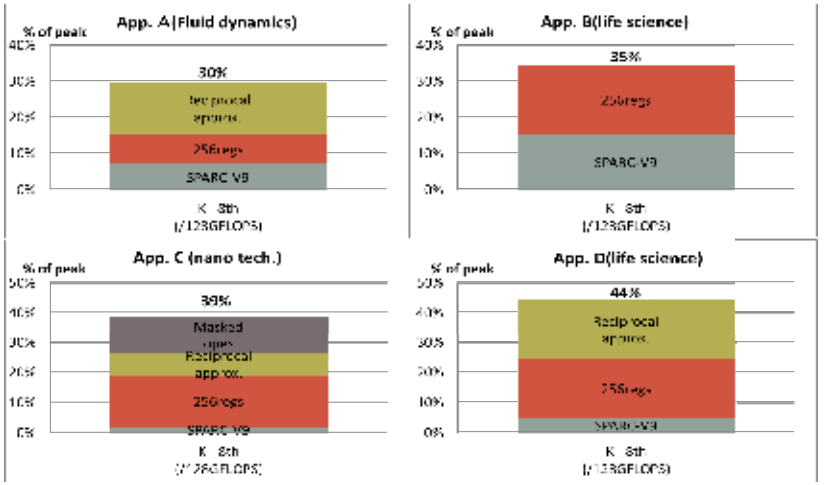


- Greenness
  - ***SPARC64 VIIIfx*** CPU
  - Direct water cooling
- High reliability, stability and operability
  - Reliable designed ***SPARC64 VIIIfx*** CPU
  - Newly designed interconnect, ***Tofu***
  - Direct water cooling

Contributions of new CPU features to efficiency improvements

FUJITSU

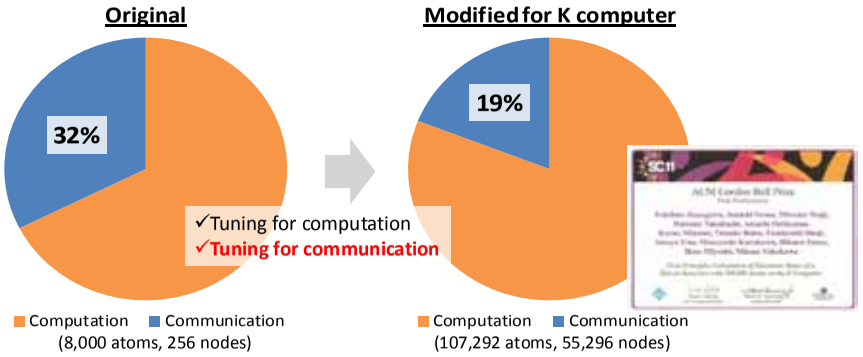
- An analysis of one CPU performance on real applications proves contribution of implemented features



Tofu Interconnect - New *allreduce* algorithm -

FUJITSU

- New *allreduce* algorithm for Tofu interconnect made a great contribution to receiving 2011 Gordon Bell Award*



- New high speed *allreduce* algorithm : Communication throughput 3.2GB/s.
- Application performance : 3.08Pflops (43.6% efficiency)

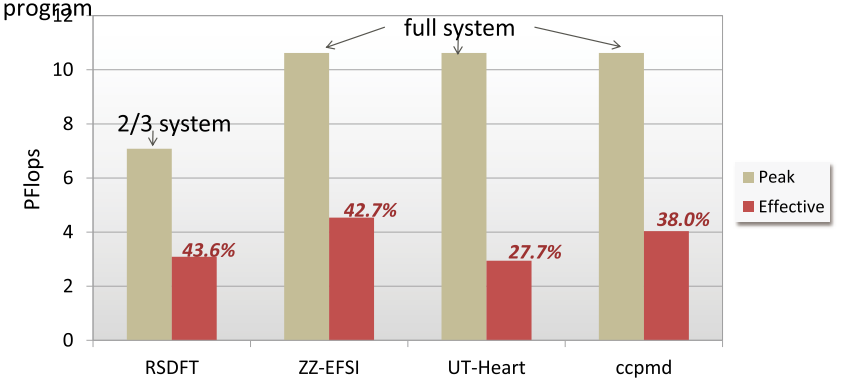
* : SC '11 Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis, Article No. 1 ,First-principles calculations of electron states of a silicon nanowire with 100,000 atoms on the K computer

Courtesy of RIKEN

Measured performance on large scale *K computer*

FUJITSU

- Several application programs have already achieved very high efficiency on large scale ***K computer*** system
- Positive outcomes of practical application are expected in early stage of HPCI program



Performance of application programs on large scale ***K computer***

RSDFT : Real Space DFT program
ZZ-EFSI : Combination of CFD and structure analysis program simulating blood flow in vessel
UT-Heart : Multi-scale and multi-physics simulating system for coronary artery circulation with capillary and metabolism in cell
ccpmd : Molecular dynamics program designed for K computer

Courtesy of ISLiM

Lessons learned from *K computer* Project

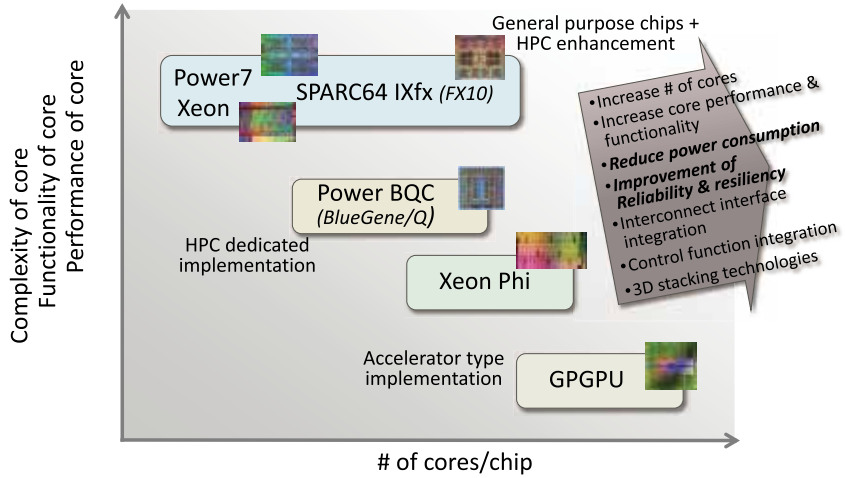
FUJITSU

- Challenges to leading edge project will bring us:
  - *Strong mind and challenging spirit*
  - End to End system development : LSI, OS, MW, system, & interconnect
  - Application software : momentum of application software asset
- Importance of the feed back from application evaluation and optimization process
  - Accumulate more expertise required for effective 1PFlops applications
  - *Transfer of the expertise to the next generation systems* development
- Project management
  - Consensus building for the *national roadmap* and securing *sustainable budget*
  - The *speed of decision-making* is really the key
  - Understanding of the nature of technology development
    - LSI development becomes very risky business
    - Combination of *assured & mature technologies* and *advanced & challenging technologies* brought the success

Approach to Exascale computing

FUJITSU

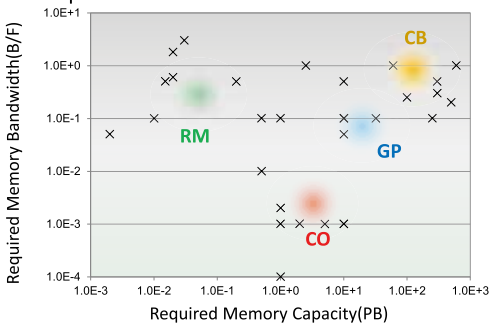
- Power consumption and reliability & resiliency may become key issue
- Various types of architectures are being used to achieve Exascale computing



Japanese Approaches to Exa-Scale Computing

FUJITSU

- In 2011, Japanese Government started the Exa-scale computing project
  - In 2011, application team reviewed the applications with architecture team and wrote up the application requirement report
  - In 2012, four two-year-Feasibility study (FS) themes were implemented



Requirements from applications
(from the Application review report)

X : Each program requirement
GP : General Purpose
CB : Capacity –Bandwidth oriented
RM : Reduced Memory
CO : Compute oriented

Outline	Leading members and vendors
Assessing the following ***architectures in terms of applications***	RIKEN
Similar architecture to K Computer's (for Advanced & Efficient Latency Core-Based Architecture)	The University of Tokyo FUJITSU
Arithmetic Accelerators (for compute oriented Applications)	University of Tsukuba HITACHI
Vector Supercomputers (for High-Bandwidth Applications.)	Tohoku University NEC

Conclusion

FUJITSU

- Success in the K computer project brought us valuable and important expertise
  - in Project management, system development and application development
  - and reminded us of importance of co-design, co-development and time line
- New challenges to Exascale computing has already started with lessons and expertise we acquired through the ***K computer*** project
- ***Demonstration of Petascale computing power*** with real & practical application programs is the key.

Invited Talks

What can Multiscale Models do for Precision Medicine?

Grace C.Y. Peng

Program Director, National Institute of Biomedical Imaging and Bioengineering (NIBIB), National Institutes of Health (NIH)

Profile:

Grace C.Y. Peng received the B.S. degree in electrical engineering from the University of Illinois at Urbana, the M.S. and Ph.D. degrees in biomedical engineering from Northwestern University. She performed postdoctoral and faculty research in the department of Neurology at the Johns Hopkins University. In 2000 she became the Clare Boothe Luce professor of biomedical engineering at the Catholic University of America. Since 2002, Dr. Peng has been a Program Director in the National Institute of Biomedical Imaging and Bioengineering (NIBIB), at the National Institutes of Health. Her program areas at the NIBIB include mathematical modeling, simulation and analysis methods, and next generation engineering systems for rehabilitation engineering, neuroengineering and surgical systems. In 2003, Dr. Peng lead the creation of the Interagency Modeling and Analysis Group (IMAG), which now consists of program officers from ten federal agencies of the U.S. government and Canada (www.imagwiki.org/mediawiki). IMAG has continuously supported funding specifically for multiscale modeling (of biological systems) since 2004. IMAG facilitates the activities of the Multiscale Modeling (MSM) Consortium of investigators (started in 2006). Dr. Peng is interested in promoting the development of intelligent tools and reusable models, and integrating these approaches in engineering systems and multiscale physiological problems.

Abstract

1. Introduction – What is Precision Medicine?

In 2011, the National Academy of Sciences (NAS) published a report entitled, "Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease."[1]This report has spurred many talks and discussions in the scientific arena in the United States to examine how we can manage the explosion of disease-relevant data, and better integrate the knowledge gained from basic biomedical research with medical histories and health outcomes of individual patients. The NAS Committee concluded it is time now to modernize the human disease taxonomy that informs healthcare decisions, by more precisely defining and classifying diseases. A *New Taxonomy* would precisely define diseases based on their intrinsic biology, in addition to traditional "signs and symptoms"; and incorporate a deeper understanding of disease mechanisms, pathogenesis, and treatments in a dynamic, iterative fashion – continuously incorporating newly emerging disease information. The New Taxonomy would require an *Information Commons* in which data on large populations of patients become broadly available for research use, and a *Disease Knowledge Network* that adds value to these data by highlighting their inter-connectedness and integrating them with evolving knowledge of fundamental biological processes. The result would be "Precision Medicine".

1.2. IMAG and the MSM Consortium

Since 2003, the Interagency Modeling and Analysis Group (IMAG), a coalition of program officers from 10 government agencies in the United States and Canada, has been promoting funding activities in the area of modeling and analysis of biomedical, biological and behavioral systems[2], with a particular emphasis on multiscale modeling. Since 2006, IMAG has facilitated the activities of the Multiscale Modeling (MSM) consortium, which is made up of investigators in the field. Each year, the IMAG MSM Consortium meets to discuss timely issues that concern the field of multiscale modeling. This year, on October 22-23, 2012, the MSM Consortium decided to focus its discussions on "Multiscale Modeling for Precision Healthcare"[3]. Four sets of MSM Consortium panelists were asked to read the NAS report on Precision Medicine[1], collect other sources of relevant information, and comment on the following questions: 1) Data-driven models, physiological models and structural models – can they be tailored to individuals for precision healthcare? 2) Can computational biology facilitate precision healthcare? 3) How can we utilize clinical data to inspire MSM research for precision healthcare? 4) How do we incorporate verification, validation and uncertainty quantification in MSM for precision healthcare? The panelists led the MSM Consortium in an interactive discussion on each of these questions, and that the views expressed below are those of the MSM Consortium, not the NIH or other government agency members of IMAG.

2. MSM and Precision Medicine

Though the NAS report on Precision Medicine[1] does not refer to computational models, the MSM Consortium concluded that multiscale models will be necessary platforms to derive knowledge from clinical and scientific data and to integrate knowledge for the purpose of informing diagnoses, supporting therapeutic decisions and predicting clinical outcomes. At the same time the improved data resources, through the New Taxonomy, Information Commons, Disease Knowledge Networks recommended in the NAS report, will all be critical in enabling the application of multiscale models to precision medicine. A recently published review article by Winslow et al.[4] defines *computational medicine* as the means to use quantitative multiscale models to understand altered structure and function in disease, and develop new methods for disease diagnosis and treatment.

The MSM Consortium also discussed the use of computational (systems) biology models, and its unique ability to link molecular scales with macro-scale information. Though computational biology models have already shown success in correlating high-dimensional, high-throughput data with disease

diagnosis, there is a need to increase the influence of mechanistic hypotheses in the modeling cycle[5]. The MSM Consortium also discussed the need for good clinical data appropriate for models, as well as a need to better engage the clinical community in the modeling discussion. Finally, the MSM Consortium conducted an extensive discussion on model validation, verification and uncertainty quantification (VV&UQ). The Consortium identified the need to develop model evaluation criteria at the beginning of the model development process, and that VV&UQ should be addressed in the context of a model's use case and level of criticality of the clinical decision being made.

In conclusion, multiscale modeling has the potential to play a critical role in implementing the goals of precision medicine. While most diseases present clinically at tissue, organ and whole body scales; they are frequently treated with molecular interventions. Multiscale models are among the most rigorous tools available to integrate a hierarchy of information at different scales and predict the dynamic state of a patient and the progression of disease.

References

1. Committee on a Framework for Development a New Taxonomy of Disease, National Research Council. **"Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease"**, The National Academies Press (2011), http://www.nap.edu/catalog.php?record_id=13284.
2. Peng, G. C.Y. **"Editorial: What Biomedical Engineers Can Do to Impact Multiscale Modeling"**, IEEE Transactions on Biomedical Engineering, Volume 58, Issue 12, Part 2, Page(s): 3440 – 3442 (2011)
3. **2012 Multiscale Modeling Consortium Meeting**, http://www.imagwiki.nibib.nih.gov/mediawiki/index.php?title=2012_MSM_CONSORTIUM_MEETING
4. Winslow, R. L., Trayanova, N., Geman, D., Miller, M. I. **"Computational Medicine: Translating Models to Clinical Care"**, Sci Transl Med, Volume 4, Issue 158 (2012)
5. An, G. **"Closing the Scientific Loop: Bridging Correlation and Causality in the Petaflop Age"**, Sci Transl Med, Volume 2 Issue 41 (2010)

Multiscale simulations of tumor angiogenesis with therapeutic applications

Aleksander S. Popel

Professor, Biomedical Engineering, Johns Hopkins University School of Medicine

Profile:

Aleksander S. Popel, Ph.D., is a Professor of Biomedical Engineering at the Johns Hopkins University School of Medicine. He holds joint appointments as a Professor of Oncology in the School of Medicine, and Professor of Mechanical Engineering and Chemical & Biomolecular Engineering in the Johns Hopkins Whiting School of Engineering. His research areas are systems biology, computational medicine & biology, angiogenesis and microcirculation. He published over 250 scientific papers in these areas. He served as a Visiting Professor at MIT and Harvard University, and a Visiting Fellow at the Isaac Newton Institute, University of Cambridge, U.K. His honors and awards include the Eugene M. Landis Award from the Microcirculatory Society, C. Forbes Dewey Distinguished Lectureship in Biological Engineering at the Massachusetts Institute of Technology, A.C. Suhren Lectureship at Tulane University, Robert M. and Mary Haythornthwaite Distinguished Lectureship at Temple University, and Kawasaki Medical Society Lectureship in Japan. He delivered a keynote address for The Virtual Physiological Human (VPH) European Union Physiome Project in 2010; he received the William H. Huggins Excellence in Teaching Award from Johns Hopkins University. He is an elected Fellow of the American Institute of Medical and Biological Engineering, American Heart Association, American Physiological Society, and American Society of Mechanical Engineers, and an Inaugural Fellow of the Biomedical Engineering Society. He has been a member of editorial boards of biological and biomedical engineering journals, such as Annals of Biomedical Engineering, American Journal of Physiology, Microcirculation, Microvascular Research, Cellular & Molecular Bioengineering, Wiley Interdisciplinary Reviews on Systems Biology & Medicine, Frontiers in Computational Physiology & Medicine, and American Journal of Cancer Research. He is Co-Chairman of the Physiome and Systems Biology Committee of the International Union of Physiological Sciences and Co-Leader, Multiscale Systems Biology Working Group for the Interagency Modeling and Analysis Group (including NIH). He has served in an advisory role to biotech companies. He regularly serves on grant review boards and advisory panels at the National Institutes of Health, National Science Foundation, and other US and international funding agencies.

Abstract

1. Introduction

Tumor growth and metastasis are complex dynamic processes that so far have largely evaded attempts to control and manage them; a grand challenge of modern biology and medicine is understanding their mechanisms at the molecular, cellular and tissue levels and learning how to control the disease [1]. Considering the enormity of the task and the complexity of the subject, systems and computational biology are absolutely necessary for achieving these goals. To make a crucial progress in the understanding of cancer and the role of new vasculature that sustains tumor growth and metastasis [2, 3], new approaches are necessary that would allow us to modularize the problems, and to identify and quantify the interactions between the modules. We need to combine reductionist approaches, down to the gene and protein levels, with integrative or systems approaches. What should make the problem feasible is its modularity, ie the ability to define and study the modules individually as a first approximation, then combine them to study interactions between the modules, up the hierarchy of biological scales and complexity.

2. Unraveling the Complexity of Cancer by Multiscale Multimodular Modeling of Tumor and Vascular Dynamics

What are the major processes involved and how do we approach the problem? Below is a list of the processes where we have made initial progress.

To be specific, here we focus on breast cancer, but the methodologies developed should apply to solid tumors and their metastatic dissemination more generally. Computationally, these are presented as modules describing different biological scales as well as integration:

1. Bioinformatic analysis of the anti-angiogenic motifs in the human proteome and computer-aided optimization of anti-angiogenic peptides [4, 5].
2. Bioinformatic analysis of protein-protein interactions (PPI) with the aim of classifying and organizing proteins important for angiogenesis and identifying novel drug targets [6-8].
3. Mechanistic models of cell signaling networks, including growth factor receptor-ligand interactions (Vascular Endothelial Growth Factor – VEGF) [9-11], intracellular signaling [12], and enzymatic processes that involve cell-extracellular matrix interactions specifically with matrix metalloproteinases, MMP [13-15].
4. Mechanistic models of transcriptional regulation; Hypoxia-Inducible Factor HIF1α [16].
5. Agent-based three-dimensional modeling of angiogenesis and tumor growth [17, 18].
6. Multiscale image-based modeling of tumor blood flow and molecular transport [19].
7. Molecular-based mechanistic models of therapeutic interventions in cancer [20, 21].
8. Computational modularity and systems integration [22, 23].

It is important that every module be validated individually, to the best extent possible, using experimental methods, from molecular biology, to in vitro cellular assays, to animal experiments. However, some of the pathways may interact (cross-talk) and even exhibit synergistic behavior, e.g. [24]; in this case the hierarchical complexity of the interactions needs to be built up gradually so as not to lose the ability to validate at every step.

Although our studies and those of other laboratories demonstrate growing technical ability to approach the systems biology of cancer, admittedly they only scratch the surface in our quest to quantitatively understand cancer. Analyzing and quantifying the modularity of multiple signaling pathways in the tumor microenvironment, e.g. cancer cells, tumor endothelial cells, and stromal cells, including immune cells will be a challenging task. Important and significant spatial heterogeneities exist within individual tumors that are the hallmark of cancer; these heterogeneities are manifested in genetic variability, physico-chemical properties, and drug response. These issues need to be confronted in future work. We predict that computational biology of cancer will remain a major research frontier; it will require novel methodologies and is likely to stretch the capabilities of modern supercomputers.

Work supported by the National Institutes of Health grant R01 CA138264, The Safeway Foundation for Breast Cancer, The Thome Foundation.

References

[1] D. Hanahan, R.A. Weinberg, Cell, 144 (2011) 646-674.

[2] P. Carmeliet, R.K. Jain, Nature, 473 (2011) 298-307.

[3] R.K. Jain, P. Carmeliet, Cell, 149 (2012) 1408-1408 e1401.

[4] E.D. Karagiannis, A.S. Popel, Proceedings of the National Academy of Sciences of the United States of America, 105 (2008) 13775-13780.

[5] C.G. Rivera, E.V. Rosca, N.B. Pandey, J.E. Koskimaki, J.S. Bader, A.S. Popel, Journal of medicinal chemistry, 54 (2011) 6492-6500.

[6] L.H. Chu, C.G. Rivera, A.S. Popel, J.S. Bader, Physiological genomics, 44 (2012) 915-924.

[7] C.G. Rivera, L.-H. Chu, J.S. Bader, A.S. Popel, Applications of Network Bioinformatics to Cancer Angiogenesis., in: A.S. Azmi (Ed.) Systems Biology in Cancer Research and Drug Discovery, Springer, 2012, pp. 231-244.

[8] C.G. Rivera, S. Mellberg, L. Claesson-Welsh, J.S. Bader, A.S. Popel, PloS one, 6 (2011) e24887.

[9] F. Mac Gabhann, A.S. Popel, Microcirculation, 15 (2008) 715-738.

[10] F.T. Wu, M.O. Stefanini, F. Mac Gabhann, A.S. Popel, Methods in enzymology, 467 (2009) 461-497.

[11] F.T. Wu, M.O. Stefanini, F. Mac Gabhann, C.D. Kontos, B.H. Annex, A.S. Popel, Journal of cellular and molecular medicine, 14 (2010) 528-552.

[12] W.H. Tan, A.S. Popel, F. Mac Gabhann, PloS one, (2013).

[13] P. Vempati, E.D. Karagiannis, A.S. Popel, The Journal of biological chemistry, 282 (2007) 37585-37596.

[14] P. Vempati, F. Mac Gabhann, A.S. Popel, PloS one, 5 (2010) e11860.

[15] E.D. Karagiannis, A.S. Popel, Journal of theoretical biology, 238 (2006) 124-145.

[16] A.A. Qutub, A.S. Popel, Molecular and cellular biology, 28 (2008) 5106-5119.

[17] A.A. Qutub, F. Mac Gabhann, E.D. Karagiannis, P. Vempati, A.S. Popel, IEEE engineering in medicine and biology magazine, 28 (2009) 14-31.

[18] A.A. Qutub, A.S. Popel, BMC systems biology, 3 (2009) 13.

[19] E. Kim, S. Stamatelos, J. Cebulla, Z.M. Bhujwalla, A.S. Popel, A.P. Pathak, Annals of biomedical engineering, 40 (2012) 2425-2441.

[20] S.D. Finley, A.S. Popel, AAPS J, 14 (2012) 500-509.

[21] M.O. Stefanini, F.T. Wu, F. Mac Gabhann, A.S. Popel, Cancer research, 70 (2010) 9886-9894.

[22] G. Liu, A.A. Qutub, P. Vempati, F. Mac Gabhann, A.S. Popel, Theoretical biology & medical modelling, 8 (2011) 6.

[23] F. Mac Gabhann, M.O. Stefanini, A.S. Popel, Simulating therapeutics using multiscale models of the VEGF receptor system in cancer., in: T.L. Jackson (Ed.) Modeling Tumor Vasculature: Molecular, Cellular, and Tissue Level Aspects and Implications., Springer, 2012, pp. 37-53.

[24] J.E. Koskimaki, E. Lee, W. Chen, C.G. Rivera, E.V. Rosca, N.B. Pandey, A.S. Popel, Angiogenesis, (2012).

Is eukaryotic cell biology ready for supercomputing? A perspective from the Virtual Cell project

Ion Moraru

Director, High Performance Computing Facility and Associate Professor, Cell Biology, University of Connecticut School of Medicine

Profile:

Dr. Moraru is the Director of the High Performance Computing Facility and an Associate Professor of Cell Biology at the University of Connecticut School of Medicine. Prof. Moraru received an MD from the Bucharest Institute for Medicine and Pharmacy in 1988 and a PhD in Cell Biology from Carol Davila University in 1992. His experimental studies in signal transduction led him to mathematical modeling and computer simulations, and he eventually dedicated himself full-time to computational biology. His current main interest is the development of methods to bridge the gap between quantitative, "bottom-up", spatially-resolved, detailed simulations of cellular functions, and large-scale, pathway- and "omics"-derived systems biology models.

Abstract

1. Introduction

Logic-derived modelling has been used to map biological networks and to study arbitrary functional interactions, and fine-grained kinetic modelling can accurately predict the detailed behaviour of well-characterized molecular systems, however, at present, neither approach comes close to unravelling the full complexity of a cell. It is possible nowadays to create a whole-cell model of a simple prokaryotic cell that accounts for all molecular components and their interactions, from electrolytes and metabolites to proteins and ribosome assemblies. A huge effort of painstaking data mining is required to assemble such models, but the mathematical representations are straightforward and the computational cost is low: an entire cellular lifecycle can be simulated with as little as 1 TFlop*hr.

In contrast, we are far from being able to simulate eukaryotic cells at a similar level of detail: the tools, data, and computational power are all still inadequate. Non-linear dynamics of intracellular molecular interactions (especially in signalling networks, but often also in metabolic or gene regulatory networks) cannot be ignored. Emergent properties are due not only to network topology, but to the detailed kinetic rate laws and quantitative parameters. Multiple phosphosites can create a combinatorial complexity of regulatory actions, and generate difficulties in mapping functional states to measured observable quantities. Compartments, scaffolds, and diffusion create spatial inhomogeneities and microdomains, which have critical functionality in most eukaryotic cells (often also in prokaryotes). And to top it all off, there has been increasing evidence that parts of the cellular machinery employ fleeting, non-stoichiometric, pleiomorphic assemblies of molecules to carry out vital processes.

2. The Virtual Cell Project

The Virtual Cell (VCell) is a unique computational environment for modeling and simulation in cell biology developed at the University of Connecticut's Center for Cell Analysis and Modeling, an NIH-designated National Research Resource. The Center integrates new microscope technologies for making quantitative in vivo live cell measurements with new physical formulations and computational tools that will produce spatially realistic quantitative models of intracellular dynamics. The latter are being made available for the use of researchers worldwide through their integration into the public, web-accessible, VCell software framework. The creation of models in VCell can range from the simple, to evaluate hypotheses or to interpret experimental data, to the complex, where multi-layered models can be used to probe the predicted behavior of large, highly non-linear systems. VCell has been designed from the beginning to work with spatially-resolved models, and it has a wide range of capabilities to specify 2D or 3D geometries for compartments. Geometries can be re-used from other VCell models (including shared models from other users). New analytic shapes can be created by defining analytic equations that describe compartment distributions in space or by manually defining compartments on a regular grid of points in space (image painting). An important VCell feature is a possibility to use geometries from external resources. Rasterized 2D or 3D (Z-stacks) images can be imported, as well as CAD specifications such as stereolithography and constructive solid geometry. Additionally, surface mesh files can be imported and sampled within a spatial domain to define compartments. Users can build complex models which specify compartmental topology and geometry, molecular characteristics, and relevant interaction parameters. VCell automatically converts the biological description into a corresponding mathematical system, deterministic (concentrations, ordinary and/or partial differential equations) or stochastic (locations, process probabilitites). VCell will then generate the code to perform and analyze simulations of the system using numerical solvers. The distributed version of VCell maintains a database of models and simulation data that can be private, shared with specific users, or public.

VCell has been continuously and rapidly growing in features and complexity over the past several years. Compartmental and spatially-resolved

models of reaction, diffusion, membrane transport, electrophysiology, advection/flow and events can be simulated both deterministically and stochastically with a large choice of solvers, including hybrid stochastic-deterministic solvers, and advanced parameter scan and estimation tools are available.

3. Bridging the Systems Biology Gap

New developments will be presented that expand VCell's capabilities towards the systems biology domain, such as automated model building from pathway databases, rule-based modeling, and network-free simulations. In parallel, we will discuss recent developments of systems biology modeling tools that expand their capabilities towards the detailed biochemistry domain, such as automated differential equations generation from Boolean networks and combined data-driven and pathway-driven methods of network inference.

We now need efficient ways for seamless interfacing between these different tools in order to combine all of the required different mathematical formalisms – a promise brought by community standards for model exchange that are currently being developed (such as SBML level 3 extensions). But will it be computationally tractable to build and simulate comprehensive quantitative whole-cell models of eukaryotic cells? A good benchmark towards that goal (perhaps a 'grand challenge'?) would be the ability to predict the phenotype produced by random mutations in the case of a simple multi-cellular organism such as C. *Elegans* (much in the same way it was recently done for the bacterium M. *Genitalium*). Based on known –omics data, we can estimate that a hypothetical simulation with currently available algorithms of the development lifecycle (from egg to 959-cell worm) would require at least 100 EFlop*hr. This is not realistic right now, but not out of reach, either; it seems reasonable to expect that the overall required computational infrastructure (algorithms, standards, software) would be available by the end of the decade. But will there be enough experimental data to parameterize, constrain and validate such models?

Modeling Nanotoxicity: Large Scale Molecular Simulations of Protein-Nanoparticle Interactions with IBM BlueGene

Ruhong Zhou

Manager, Soft Matter Theory and Simulation, Computational Biology Center
IBM Thomas J. Watson Research Center, IBM

Profile:

Ruhong Zhou is currently a Research Staff Scientist and Manager of Soft Matter Theory and Simulation Group, Computational Biology Center, IBM Thomas J. Watson Research Center, and an Adjunct Professor at Department of Chemistry, Columbia University. He received his Ph.D. with Prof. Bruce Berne in chemistry from Columbia University in 1997. He joined IBM Research in 2000, after spending two and a half years working with Prof. Richard Friesner (Columbia Univ) and Prof. William Jorgensen (Yale Univ) on polarizable force fields. He has authored and co-authored 120 journal publications and 15 patents, delivered 150+ invited talks at major conferences and universities worldwide, and chaired and co-chaired many conferences in computational biology, computational chemistry and biophysics fields. He is part of the IBM Blue Gene team who won the 2009 National Medal on Technology and Innovation. He has won the IBM Outstanding Innovation Award in 2011, IBM Outstanding Technical Achievement Award (the highest technical award within IBM) in 2008 and 2005, IBM Research Division Award in 2005, Columbia University Hammett Award in 1997 (for best graduates), and American Chemical Society DEC Award on computational chemistry in 1995. His current research interests include development of novel methods and algorithms for computational biology and bioinformatics, and large scale simulations for protein folding, ligand-receptor binding, protein-protein interaction, and protein nanoparticle interactions. He was elected to AAAS Fellow (American Association of Advancement of Science) and APS Fellow (American Physical Society) in 2011.

Abstract

1. Introduction

Nanoscale particles have become promising materials in various biomedical applications, such as cancer therapeutics, diagnosis, neuro-imaging, drug delivery, and biosensors. However, in order to stimulate and facilitate these applications, there is an urgent need for the understanding of the nanotoxicity and other risks involved with these nanomaterials to human health. Some of our recent molecular modeling works with IBM Blue Gene supercomputer will be presented in this talk. We show that a pristine carbon nanotube, one form of hydrophobic nanoparticles, can interact and disrupt the structures and functions of many important proteins such as WW domains, SH3 domains, and human blood serum proteins. In some extreme cases, such as the WW domains, the carbon nanotube can unexpectedly plug into the hydrophobic core of the protein to form stable complexes. This plugging of nanotubes disrupts and blocks the active sites of WW domains from binding to the corresponding ligands, thus leading to the loss of protein functions. In other cases, nanotubes compete with ligands for the receptor binding sites involved in the signaling and regulatory pathways. Different adsorption capacities of human serum proteins on carbon nanotubes, on the other hand, result in different cytotoxicity. The hydrophobic interactions between the carbon nanotubes and hydrophobic residues, particularly aromatic residues through the so-called π−π stacking interactions, are found to play key roles. In addition, the molecular mechanism of tumor metastases inhibition by metallofullerenol Gd@C82(OH)22 (i.e. toxic to cancer cells) has been studied with both experimental and theoretical approaches. These findings might provide a better understanding of "nanotoxicity" at the molecular level and help design better therapies with nanomedicine. In the following, two specific examples are provided to illustrate these important protein-nanoparticle interactions.

2.1 Competitive Binding of Carbon Nanotubes to Human Serum Proteins

In this example, we have used both experimental and theoretical approaches, including AFM images, fluorescence spectroscopy, CD, SDS-PAGE, and molecular dynamics simulations, to investigate the interactions of SWCNTs with human serum proteins (BSA, γ-Ig, Tf, and BFg), and find a competitive binding of these proteins with different adsorption capacity and packing modes. The π–π stacking interactions between SWCNTs and aromatic residues (Trp, Phe, Tyr) are found to play a critical role in determining their adsorption capacity. Additional cellular cytotoxicity assays reveal that the competitive bindings of blood proteins on the SWCNT surface can greatly alter their cellular interaction pathways and result in different cytotoxicity for these protein-coated SWCNTs. These findings have shed light towards the design of safe CNT nanomaterials by comprehensive preconsideration of their interactions with human serum proteins.

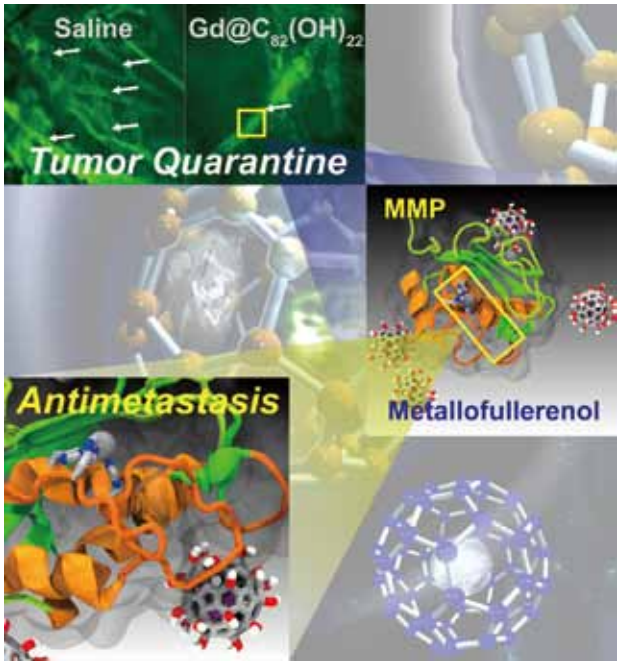


Figure 1. Molecular mechanism of pancreatic tumor metastases inhibition by metallofullerenol Gd@C82(OH)22

2.2 Pancreatic Tumor Metastases Inhibition by Gd@C82(OH)22

Pancreatic adenocarcinoma is probably the most lethal of the solid tumors and the fourth leading cause of cancer-related death in North America. Matrix

metalloproteinase (MMP) has long been targeted as a potential anti-cancer therapy due to its seminal role in angiogenesis and extracellular matrix (ECM) degradation for the tumor survival and invasion. The inhibition specificity to MMPs and the molecular level understanding of inhibition mechanism, however, remains largely unresolved. Moreover, finding better alternatives to traditional medicines with emerging nanomaterials (nanomedicine) is of great current interest.

In this example, we find that endohedral metallofullerenol Gd@C82(OH)22 can successfully inhibit the neoplastic activity of pancreatic cancer with experiments at animal, tissue, and cellular levels. Gd@C82(OH)22 effectively blocks tumor growth in human pancreatic cancer xenografts in nude mice model. Enzyme activity assays also show Gd@C82(OH)22 not only suppresses the expression of MMPs but also significantly reduces their activities. We then further applied large scale molecular dynamics simulations to uncover the molecular mechanism by studying the Gd@C82(OH)22-MMP-9 interactions at atomic detail. Our data demonstrated that Gd@C82(OH)22 inhibits MMP-9 mainly via an exocite interaction while the well-known zinc catalytic site only plays a minimum role. Steered by non-specific electrostatic, hydrophobic and specific hydrogen bonding interactions, Gd@C82(OH)22 exhibits specific binding modes near the ligand specificity loop S1′, thereby inhibiting the MMP-9 activity (see **Figure 1**). Both the suppression of MMPs expression and specific binding mode make Gd@C82(OH)22 a potentially more effective nanomedicine for pancreatic cancer than the traditional medicines which usually target the proteolytic sites directly, but fail in selective inhibition. Our findings based on a combination of in vivo, in vitro and in silico approaches provide new insights for de novo design of nanomedicines for fatal diseases such as pancreatic cancer.

References

1. C. Ge, J. F. Du, L. Zhao, L. Wang, Y. Liu, D. H. Li, Y. Yang, R. H. Zhou, Y. L. Zhao, Z. F. Chai, C. Y. Chen, Competitive binding of human serum proteins on single-wall carbon nanotubes reduces cytotoxicity, Proc. Natl. Acad. Sci, 108, 16968, 2011
2. G. Zuo, Q. Huang, G. Wei, R. H. Zhou, and H. Fang, Plugging Into Proteins: Poisoning Protein Function by a Hydrophobic Nanoparticle, ACS Nano, 4, 7508-7514, 2010
3. S. G. Kang, G. Zhou, P. Yang, Y. Liu, B. Sun, T. Huynh, H. Meng, L. Zhao, G. Xing, C. Y. Chen, Y. L. Zhao, R. H. Zhou, Molecular Mechanism of Pancreatic Tumor Metastases Inhibition by Metallofullerenol Gd@C82(OH)22 and its Implication for de novo Design of Nanomedicine, Proc. Natl. Acad. Sci., 109, 15431, 2012

Multi-scale modelling of the motor system underlying goal-directed behaviour in a vertebrate model organism

Sten Grillner

The Nobel Institute for Neurophysiology, Dept of Neuroscience, Karolinska Institutet

Profile:

Sten Grillner has unravelled the intrinsic function of the modular network organization underlying fundamental aspects of our motor repertoire. His initial work defined the basic organisation of the mammalian locomotor system in terms of supraspinal command systems, spinal networks coordinating the movements (CPGs), and the sensory control of the CPGs. To address the next level question - the molecular, cellular and synaptic design of these neuronal circuits - he developed a novel and simpler vertebrate model (lamprey). The different network interneurons, their synaptic interaction (transmitters, receptor subtypes), and their membrane properties (ion channel subtypes expressed) have been identified. The palette of different subtypes of ion channels expressed in different neurones is found to be of critical importance for network function. Through an interaction between detailed multi-faceted experimentation and large scale modelling with biophysically realistic numbers of Hodgkin-Huxley neurons, the operation of this entire motor control system has been uncovered. This conserved system is understood at a cellular/molecular level, and the conceptual gap between gene/molecule and behaviour has been bridged. We now understand the design of the neuronal subsystems coordinating goal-directed ambulation. In focus are now the forebrain systems that are responsible for selection of different patterns of behavior.
Sten Grillner has also helped develop the OECD initiated International Neuroinformatics Coordinating Facility (INCF) with the secretariat in Stockholm that promotes both the development of interoperable databases and multi-scale modelling. INCF has now 17 member countries extending from Japan, Korea and India to Europe and the US.

Multi-scale modelling of the motor system underlying goal-directed behaviour in a vertebrate model organism

Sten Grillner, Nobel Institute for Neurophysiology, Dept of Neuroscience, Karolinska Institutet, SE-171 77 Stockholm

The vertebrate brain controls a great variety of movements through dedicated networks like those controlling eye movements, expression of emotions, respirations and locomotion. These networks are to a large degree conserved through the vertebrate phylum. The neural mechanisms underlying the control of goal directed movement will be in focus, both biological background information and multiscale modelling from the subcellular and synaptic level to the microcircuit, systems and behavioural levels. The propulsive locomotor synergy is controlled from command regions in mesencephalon, which in turn control central pattern generating (CPG) networks in the spinal cord. My presentation, based on the lamprey CNS, will address the intrinsic function of the adaptable CPG that can generate different motor patterns. In addition, I will discuss the tectal mechanisms underlying steering movements with the retinotopic motor map, and the control from the output of the lamprey basal ganglia, which is of critical importance for deciding which motor program should be turned on at a given moment of time. The mechanism by which different motor programs are selected will be considered with special reference to the basal ganglia – experiments and modelling. Our recent findings establish that the structure and function of the basal ganglia have been conserved to a surprising degree throughout vertebrate phylogeny over a period of more than 500 million years from the ancient lamprey version to primates. This applies to the input to striatum (pallium, thalamus, dopamine, 5-HT, histamine input), the pallidal structures (GPi, substantia nigra reticulata (SNr), GPe and the subthalamic nucleus) and their output targets, the cellular properties of striatal and pallidal neurons and the effects of an MPTP induced dopamine denervation.

In order to simulate the basal ganglia – brainstem – spinal cord networks underlying different motor patterns including visuomotor control, we have simulated these networks with an approximate number of model neurons (around 15 000) corresponding to that observed biologically. The model neurons are of the detailed compartmental Hodgkin-Huxley type with detailed membrane properties including all subtypes of ion channels observed biologically and the different transmitter receptors. Each model neuron is designed to be very close to its biological counterpart including the variability observed in each population. The synaptic interaction is modelled as well as network properties. The variability in neuronal properties is important in that it provides for stability in network properties over a large range of frequencies. These networks have also been used to control a simulated body with the ability to move in a natural fashion and be able to perform orienting movements, avoid obstacles and approach certain goals. These simulations have been performed on an IBM blue gene supercomputer.

Kozlov A, Huss M. Lansner A, Hellgren Kotaleski J, Grillner S. (2009) Simple cellular and network control principles govern complex patterns of motor behavior. PNAS 24;106:20027-20032.

Grillner, S. (2006) Biological Pattern Generation: The Cellular and Computational Logic of Networks in Motion. Neuron 52; 751-766

Stephenson-Jones M, E. Samuelsson, J. Ericsson, B. Robertson and S. Grillner (2011) Evolutionary conservation of the basal ganglia as a common vertebrate mechanism for action selection. Curr. Biol., 21;1081-1091.

Stephenson-Jones M, O. Floros, B. Robertson, S Grillner (2012) Evolutionary conservation of the habenular nuclei and their circuitry controlling the dopamine and 5-hydroxytryptophan (5-HT) systems. Proc Natl Acad Sci U S A.

Large-scale simulations of biomolecular machines

Karissa Sanbonmatsu

Principal Investigator, Los Alamos National Laboratory

Profile:

Dr. Sanbonmatsu has been a principal investigator at Los Alamos National Laboratory since 2001. She received her BA in Physics from Columbia University in 1992 and PhD in Astrophysical, Planetary and Atmospheric Sciences from University of Colorado at Boulder in 1997. Dr. Sanbonmatsu's research is focused the mechanism of non-coding RNAs, including the ribosome, riboswitches and long non-coding RNAs. She has pioneered large-scale biomolecular simulations of nano-scale molecular machines such as the ribosome. She uses a combination of high performance computing and experimental biochemistry to construct an integrated picture of large-scale conformational changes of these complexes.

Abstract

1. Introduction

Large-scale biomolecular complexes play a key role in many aspects of cellular activity, including transcription, translation, splicing, scaffolding, transport and metabolism. New supercomputing technology has made it possible to study the dynamics of these complexes in atomic detail using large-scale molecular dynamics simulations. Our group has focused on a molecular machine called the ribosome for the past decade.[1] The ribosome plays a central role in all life forms and is responsible for protein synthesis. The ribosome must read the genetic information encoded on messenger RNA and implement these instructions by producing corresponding proteins. The ribosome performs the only non-trivial information processing operation in cell by executing a 'look-up' table operation, converting nucleic acid sequence into protein amino acid sequence. In this sense, the ribosome is analogous to the CPU of the cell and constitutes a nano-scale information processor. Approximately 50% of the antibiotic drugs used in US hospitals function by targeting the ribosome. Mechanistic understanding of ribosome activity is important for the design of new antibiotics that combat the growing antibiotic resistant 'superbug' problem present in today's hospitals. In addition, understanding how the ribosome decodes genetic information will help lay the foundation for the design of synthetic nano-scale information processors.

1.2. *The ribosome 'decoding problem'*

By performing large-scale molecular dynamics simulations of the ribosome, we have been able to examine the inner workings of this molecular machine. Our long-term goal has been understanding translocation, a large-scale massive rearrangement of the ribosome. As a first step, we focused on the more tractable problem of decoding, the process by which the ribosome decodes genetic information (also known as tRNA selection). Here, a key rearrangement that occurs is called 'accommodation', where transfer RNAs (tRNAs) carrying protein building blocks (amino acids) move into the ribosome. In our study of this process, we identified a new functional region of the ribosome ('the accommodation corridor') and predicted that certain parts of this corridor are important for ribosome function.[2] Our predictions have been validated in studies by several experimental groups.[3-5] In an additional separate set of studies that combined explicit solvent simulations, reduced model simulations, and single molecule experiments, a new picture of ribosome function has emerged.[6] Rather than the ribosome machine parts moving in lock-step, both simulations and single molecule experiments show the tRNAs making large-scale reversible excursions in a trial-and-error fashion. This picture is consistent with a dynamic energy landscape view of the ribosome.[7]

2. Translocation of tRNAs through the ribosome

After studying the relatively tractable problem of 'accommodation', we are now applying what we have learned to the mechanism of translocation, a more complicated rearrangement where several large conformational changes involving the entire ribosome and tRNA occur. This motion allows the ribosome to move exactly 3 nucleotides along the messenger RNA to the next amino acid codon. Using microsecond sampling in explicit solvent for the full ribosome, in combination with experimentally measured rates,[8] we estimate barrier heights for various motions important for translocation. We have also used coarse-grained methods to simulate the various sub-steps of translocation.[9,10] Our goal has been to use simulations of the ribosome to produce detailed energy landscapes of translocation.

References

1. Sanbonmatsu KY. Curr Opin Struct Biol. 2012;22(2):168-74.
2. Sanbonmatsu KY, Joseph S, Tung CS. Proc Natl Acad Sci U S A. 2005;102(44):15854-9.
3. Rakauskaite R, Dinman JD. RNA 2011,17:855-864.41.
4. Burakovsky DE, et al. FEBS Lett 2011,585:3073-3078.
5. Burakovsky DE, et al. RNA 2010, 16:1848-1853.

6. Whitford, et al. RNA 2010, RNA. 2010;16(6):1196-204.
7. Munro, et al. Trends Biochem Sci. 2009;34(8):390-400.
8. Whitford, et al. J Am Chem Soc. 2010;132(38):13170-1.
9. Ratje, et al. Nature. 2010;468(7324):713-6.
10. Whitford, et al. PNAS. 2011;108(47):18943-8.

Use of Modeling and Simulation in Drug Discovery and Development: It is all about the Questions

Sandra Allerheiligen

Vice President of Modeling and Simulation, Merck Research Laboratories

Profile:

Sandra (Sandy) Allerheiligen, PhD is currently Vice President of Modeling and Simulation at Merck Research Laboratories. Prior to joining Merck, she held a variety of positions over her almost 20 years at Eli Lilly and Company, including Global Sr. Director of Pharmacokinetics, Pharmacodynamics (PK/PD) and Trial Simulation, Sr. Director of Drug Disposition, and most recently, Distinguished Fellow and Chief Scientific Officer of Quantitative Pharmacology. Her research interests focus on study design and the application of mathematical methods to enable quantitative decisions for nonclinical and clinical development. She has applied PK/PD modeling to oncolytic and endocrine agents. Her recent work is on the integration of biomarkers, PK/PD modeling and trial simulation in non-clinical and clinical drug development, drug disease models and utilization of quantitative and systems pharmacology approaches. Dr. Allerheiligen received a doctorate in pharmaceutics (with a specialization in PK/PD) from the University of Texas in Austin in 1985 and completed postdoctoral fellowships at the University of Texas Health Center in San Antonio in 1986 and was a clinical assistant professor of Clinical Pharmacology from 1986 through 1990. Through her involvement in the American Association of Pharmaceutical Scientists (AAPS) and other organizations, she has worked to expand the use of PK/PD modeling and Quantitative Pharmacology methodologies in academia and across the industry. She is a Fellow of the American Association of Pharmaceutical Sciences (AAPS) and has served as Chair PPDM section, AAPS Member-at-Large, and AAPS Treasurer. She co-founded the Population PK/PD Focus Group at AAPS and the Midwest User's Forum for Population Approaches in Data Analyses. She has been a member of Editorial Advisory Boards for PharmSci, the AAPS Journal and the Journal of Pharmaceutical Research. She is an Adjunct Faculty member of the University of Osaka Medical School Department of Biostatistics and frequent lecturer on modeling and simulation topics.

Abstract

Impact of Quantitative and Systems Pharmacology (QSP) in Drug Discovery and Development: It is all about the Question

Sandra R. B. Allerheiligen, PhD

The pharmaceutical industry must reduce costs while delivering innovative medicines with improved benefits for patients. With increasing computational capability, scientists can leverage quantitative tools to answer critical questions that influence the discovery and development of important new medications. These new approaches and tools have led to the expansion of a specialized field, quantitative and systems pharmacology (QSP, sometimes called modeling and simulation).

Quantitative and Systems Pharmacology is an integrative science incorporating relationships between diseases, drug characteristics, and individual variability to leverage existing knowledge and guide future research. QSP helps define the questions and assumptions and highlights knowledge gaps. Through this approach it is possible to estimate drug efficacy, understand safety, plan experiments, and inform discovery/development strategies. Examples that highlight the ability to answer such questions as 1) should this molecule be developed as a drug, 2) what is the right dose to balance efficacy and safety, and 3) does this new drug increase bone strength will be presented. The ultimate goal is not a mathematical model but rather the ability to gain insight to optimize new medications for patients.

ISLiM Talks

Development of Multiscale Thrombosis Simulator

Shu Takagi

Team Leader, Organ and Body scale Team, CSRP, RIKEN
Professor, Department of Mechanical Engineering, The University of Tokyo

Profile:

Professional Position:

2010.4-current: Professor, Department of Mechanical Engineering, The University of Tokyo

2007.4- current: Team Leader, Organ and Body scale Team, CSRP, RIKEN

2002.1-2010.3: Associate Professor, Department of Mechanical Engineering, The University of Tokyo

1998.4-2002.1: Lecturer, Department of Mechanical Engineering, The University of Tokyo

1996.10-1998.3: Research Associate, Department of Mechano-Aerospace Engineering, Tokyo Institute of Technology

1995.4-1996.10: Research Associate, Department of Mechanical Engineering, The University of Tokyo

Education:

1995: Doctor of Engineering, Department of Mechanical Engineering, The University of Tokyo

1990: Bachelor of Engineering, Department of Mechanical Engineering, The University of Tokyo

Abstract

1. Introduction

Thrombosis is regarded as one of the most important diseases, which cause the myocardial and cerebral infarctions. It is very complicated disease affected from molecular scale protein-protein interaction to continuum scale in blood flow. Initially, platelets start to aggregate at the injured wall, where von Willebrand Factor (vWF) molecules are attached. The Glycoprotein, GPIb-α, on platelet membrane starts showing ligand-receptor type interaction with this vWF and platelets start aggregating around this spot. From this stage, very complicated activated process of platelets and interactions with blood, vessel walls red blood cells, fibrin etc. occur and they end up with the blockage of the vessels.

In the present study, we are developing the numerical model of the initial stage of thrombus formation. To analyze this process, we have developed the full Eulerian fluid-membrane coupling method [1], which is a further extension of the full Eulerian fluid-structure coupling method [2]. This method is coupled with the stochastic Monte Carlo simulation on GPIb-α and vWF interactions. The basic concept of this multiscale thrombosis simulator is shown in Fig.1.

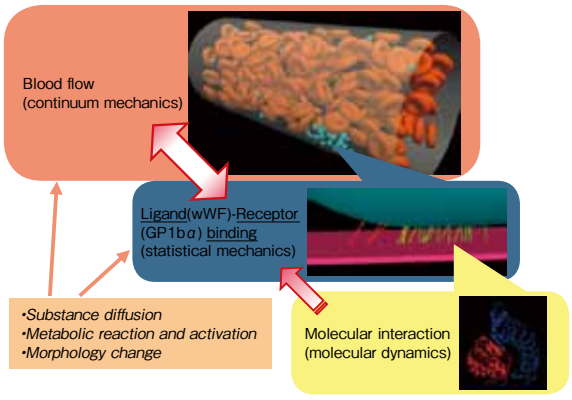
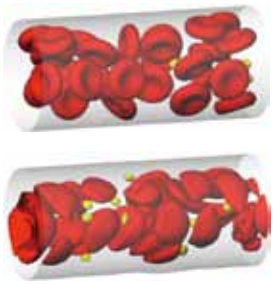


Fig.1 Multiscale modeling on the platelets adhesion

2. Simulation Results

The pressure driven flows including the multiple RBCs and platelets in a circular tube were carried out. Fig.2 illustrates the time evolution of the shape of RBCs and platelets. It is well-known that, in actual blood flows, RBCs tends to flow in the center of vessels with large deformation constructing so-called blood plasma layer, and that small platelets tends to flow in this plasma layer. It was observed that, once the platelets are coming closer to the vessel wall, they stay there and it is rarely observed that they go back to the center of the vessel.



(a) t = 0.0 [s]

(b) t = 75.0 [ms]

Fig.2 Time evolution of RBCs shape and location

Next, the effect of RBCs on the adhesion of platelets is analyzed. The simulation was conducted through the coupling of continuum mechanical simulation with Monte Carlo simulation of protein(GPIb-α) - protein (vWF) interactions. The result shown in Fig.3 illustrates the adhesion of platelets by the binding effect of proteins is achieved. This phenomenon is not observed without Red blood cells. More details will be discussed in the presentation.

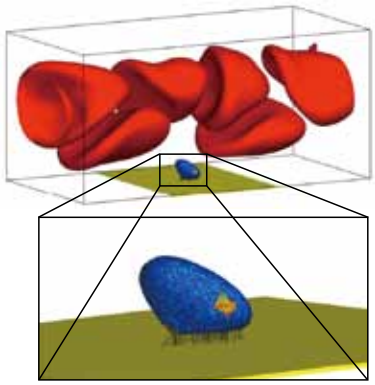


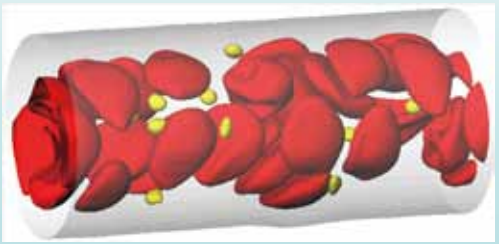
Fig.3 Platelet adhesion on Vessel Wall

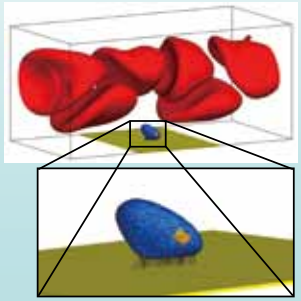
References

1. Ii, S. et al., Commun. Comput. Phys., 12, pp. 544-676. (2012).
2. Sugiyama, K. et al, J. Comput. Phys., 230, pp. 596-627 (2011)

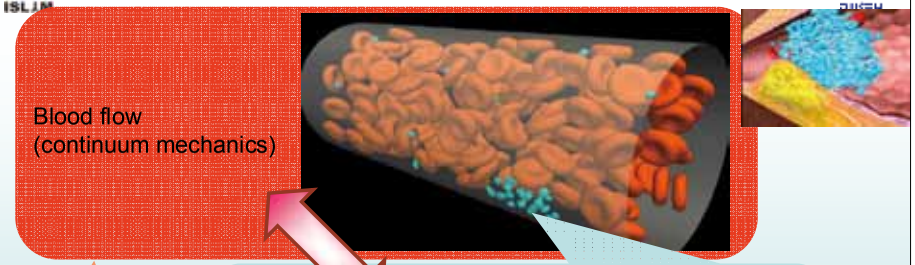
Development of Multiscale Thrombosis Simulator

Shu TAKAGI
RIKEN,
The University of Tokyo





Multiscale modeling of initial stage of thormobosis

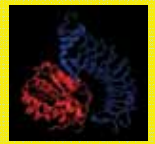


Blood flow
(continuum mechanics)

Ligand(wWF)-Receptor
(GP1bα) binding
(statistical mechanics)

•*Substance diffusion*
•*Metabolic reaction and activation*
•*Morphology change*

Molecular interaction
(molecular dynamics)



Objective

Simulate thrombosis causing a heart attack

⇩

Coronary artery diameter (d_v): 1-5 mm,
Size of Platelet (d_p): 1-3 μm

⇨ $\frac{d_v}{d_p} = \mathrm{O}\left(10^3\right)$

⇩

Require massively parallel computation
→ development of suitable numerical methods

⇩

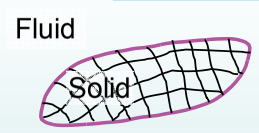
Fixed-grid, Eulerian FSI-FD method coupled with Stochastic Monte Carlo Method for recepter-ligand bindings

Previous studies for platelet adhesion: (Small Vessel Diameter)
- Tsubota et al.(2006) (Particle Method)
- Mori et al. (2008) (Stokesian Dynamics & Voigt Model),
- Mody & King (2008) (BEM & Monte Carlo),
- Fogelson & Guy (2008) (Multiscale IBM)

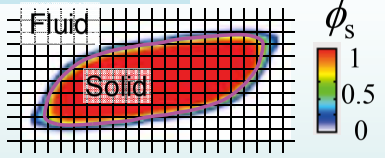
Full Eulerian approach for Fluid-Structure Interactions

Lagrangian vs. **Eulerian**

How is the two-phase distinguished ?

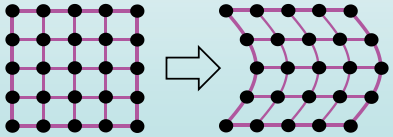


by the boundary of mesh

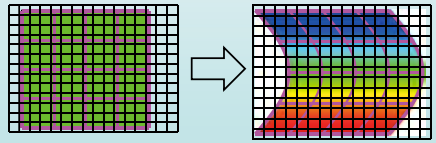


by **solid volume fraction**

How is the solid deformation described?



by the displacement of material points themselves



by **left Cauchy-Green deformation tensor**

- Sugiyama, Ii *et al.* (2011) *J. Comput . Phys.*, **230**, 596.
- Ii, Sugiyama *et al.* (2011) *Int. J. Numer. Meth. Fluids*, **65**, 150.

Basic Equations

$$\frac{\partial v_i}{\partial x_i}=0,\quad \rho_m\left(\frac{\partial v_i}{\partial t}+v_j\frac{\partial v_i}{\partial x_j}\right)=-\frac{\partial p}{\partial x_i}+\frac{\partial \sigma'_{m,ij}}{\partial x_j}+f_i,$$

Cauchy 's stress tensor $\sigma'_{m,ij}=(1-\phi_s)\sigma'_{f,ij}+\phi_s\sigma'_{s,ij}$,

$$\sigma'_{f,ij}=2\eta D_{ij},\qquad D_{ij}=\frac{1}{2}\left(\frac{\partial v_i}{\partial x_j}+\frac{\partial v_j}{\partial x_i}\right),$$

strain rate

$$\frac{\mathrm{D}\phi_s}{\mathrm{D}t}=0,$$

solid volume fraction

$$\frac{\mathrm{D}\tilde{B}_{ij}}{\mathrm{D}t}-\frac{\partial v_i}{\partial x_k}\tilde{B}_{kj}-\frac{\partial v_j}{\partial x_k}\tilde{B}_{ki}=0,$$

left Cauchy-Green deformation tensor

$$\tilde{B}_{ij}=\begin{cases}\phi_s^{\alpha}B_{ij} & \text{for } \phi_s\geq\phi_{\min}\\ 0 & \text{for } \phi_s<\phi_{\min}\end{cases}$$

$$\tilde{B}_{ij}=\phi_s^{1/2}\delta_{ij}\quad \text{at } t=0,$$

$$\mathbf{B}=\mathbf{F}\cdot\mathbf{F}^T,\qquad F_{ij}=\frac{\partial x_i}{\partial X_j},$$

Hyperelastic solid stress:

St. Venant-Kirchhoff model: $\phi_s\boldsymbol{\sigma}'_s=G(\phi_s^{1-2\alpha}\tilde{\mathbf{B}}\cdot\tilde{\mathbf{B}}-\phi_s^{1-\alpha}\tilde{\mathbf{B}})'.$

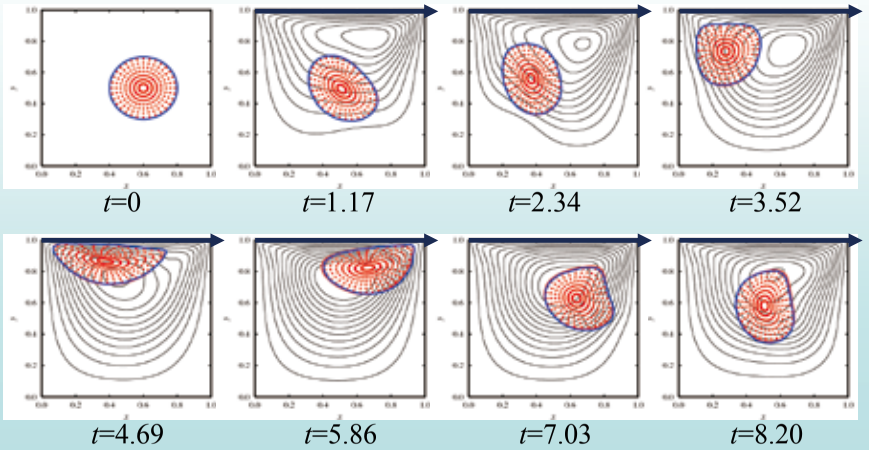
(Linear) Mooney-Rivlin model: $\phi_s\boldsymbol{\sigma}'_s=G\phi_s^{1-\alpha}\tilde{\mathbf{B}}'+2c_2\phi_s^{1-2\alpha}\left(\mathrm{tr}(\tilde{\mathbf{B}})\tilde{\mathbf{B}}-\tilde{\mathbf{B}}\cdot\tilde{\mathbf{B}}\right)'.$

Comparison with the previous study by Zhao, Freund & Moser (JCP 2008)

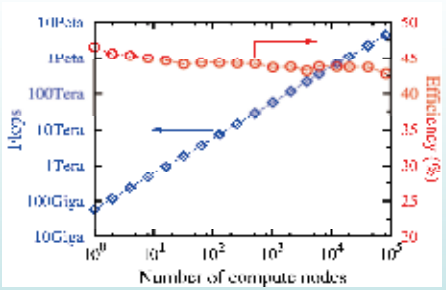
$\rho_s=\rho_f=L_s=L_f=1$, $\mu=10^{-2}$, $G=0.1$

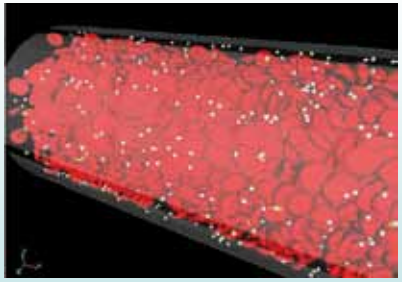
— Zhao *et al.* (Lagrangian for solid)

– Present (full Eulerian) N_xxN_y=512x512



Performance of the simulator (Program Name: ZZ-EFSI)





ZZ-EFSI achieved the actual speed of **4.5 PETA FLOPS!!!**

Software availablle at http://www.islim.org/islim-dl_e.html

Extension to Eulerian fluid-membrane coupling model*

* Ii *et al.*, *Commun. Comput. Phys.*, accepted (2011).

$$\nabla\cdot\mathbf{v}=0$$

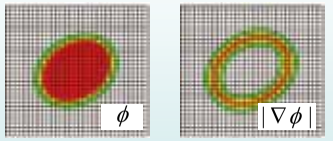
$$\rho\left(\frac{\partial\mathbf{v}}{\partial t}+\mathbf{v}\cdot\nabla\mathbf{v}\right)=-\nabla p+\nabla\cdot(\mu(\nabla\mathbf{v}+\nabla\mathbf{v}^T))+|\nabla\phi|\nabla_s\cdot(\boldsymbol{\tau}_s+\mathbf{q}_s\mathbf{n})$$

$$\frac{\partial\phi}{\partial t}+\mathbf{v}\cdot\nabla\phi=0$$

$$\frac{\partial\mathbf{B}_s}{\partial t}+\mathbf{v}\cdot\nabla\mathbf{B}_s=\mathbf{B}_s\cdot\nabla_s\mathbf{v}+\nabla_s\mathbf{v}^T\cdot\mathbf{B}_s$$

$$\frac{\partial J_s}{\partial t}+\mathbf{v}\cdot\nabla J_s=J_s\nabla_s\cdot\mathbf{v}$$

$$\frac{\partial\kappa_R}{\partial t}+\mathbf{v}\cdot\nabla\kappa_R=0$$



*Evans and Skalak, *CRC press, Boca Raton, FL.* (1980).

**Pozrikidis, *J. Fluid. Mech.*, 440 (2001) 269.

Membrane stress: $\boldsymbol{\tau}_s+\mathbf{q}_s\mathbf{n}$

➢In-plane stress (Evans-Skalak*)

$$\boldsymbol{\tau}_s=\frac{C_s}{(\Lambda_1+1)^2}\mathbf{B}_s+\left(E_s\Lambda_1-C_s\frac{\Lambda_2+1}{\Lambda_1+1}\right)\mathbf{P}$$

$$\Lambda_1=J_s-1,\quad \Lambda_2=\frac{\mathrm{tr}(\mathbf{B}_s)}{2J_s}-1$$

➢bending model (Pozrikidis **)

$$\mathbf{q}=(\nabla_s\cdot\mathbf{m})\cdot\mathbf{P}$$

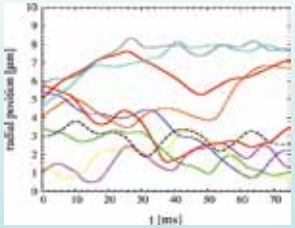
$$\mathbf{m}=E_b(\boldsymbol{\kappa}-\kappa_R\mathbf{P})$$

Simulation results for the flow wit RBCs

Gaehtgens *et al.*(1980) *Blood Cells.*, **6**, 799.

Experimental Observations
- Cell free layer near vessel wall
- RBCs with slipper shape

Simulation results



Time history of platelets position

(Platelets move along the wall, once they are coming out near the wall.)

Confidential

Ligand-Receptor bond model (Use of Transition State Theory)

- Stochastic model with energetic elasticity (Bell 1978, Dembo *et al.* 1988)

$$k_f(l) = k_{f0}\exp\left(-\sigma_{ts}\frac{(l-l_0)^2}{2k_bT}\right)$$

$$k_r(l) = k_{r0}\exp\left((\sigma_p-\sigma_{ts})\frac{(l-l_0)^2}{2k_bT}\right)$$

$$P_f = 1-\exp(-k_f\Delta t) \geq R_f \rightarrow \text{binding}$$

$$P_r = 1-\exp(-k_r\Delta t) \geq R_r \rightarrow \text{dissociation}$$

$$(R_f, R_r \in [0,1] : \text{Random numbers})$$

$$f = \sigma_p(l-l_0)$$

vWF binding site

*S.-Z. Luo, *et al.*, Blood, 109 (2007) 603.

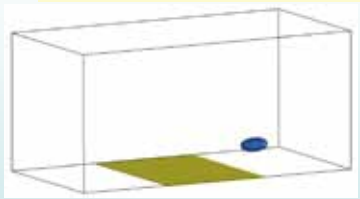
$\sigma_p = 10^{-4}$ [N/m], $\sigma_{ts} = 0.9\sigma_p$ [N/m]

$l_0 = 60$ [nm], $k_{r0} = 3$ [s^{-1}]

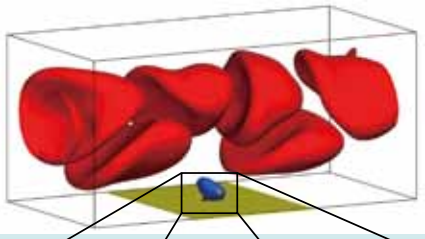
(Kim *et al.* 2011, Fox *et al.* 1988, Arya *et al.* 2005)

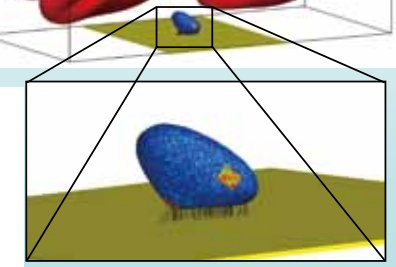
Results of the multiscale simulation about the platelet adhesion on vWF-immersed wall

without RBCs (no Adhesion simulated)



with RBCs





Red blood cells:
- Flowing in the core region of the channel
- Inducing the velocity fluctuations in the wall-normal directions

Platelet:
- Adhering on the wall due to the velocity fluctuations caused by RBCs.

Summary

- A Novel Method was developed for Fluid-Solid and Fluid-Membrane Interaction Problems, which is suitable for the massively parallel computing.
- Initial stage of thrombosis is modeled. Platelet adhesion by the binding of GP1ba and vWF molecules was simulated, through the coupling of the continuum scale simulation with the stochastic Monte Carlo simulation of molecular bindings.

On going study
- Preparation for the comparison with *in-vitro* experimental data
- Scale-up of a vessel diameter toward D=1 [mm] with activation of many chemical reaction toward heart attack simulation.

References:

- Sugiyama, el al, J. Comput. Phys., Vol.230 (2010), pp.596-627.
- Takagi et al. J. Appl. Mech. Vol.79 (2011), 010911.
- li et al., *Int. J. Numerical Method. in Fluids*, Vol.65 (2011). pp.43-66.
- li et al., Comm. in Cmuput. Phys, (2012), to appear.
- Shiozaki et al. JBSE, Vol.7 (2012), pp.275-283.
- li et al., J.Comput Phys, , Vol.231 (2012), pp.2328-2358.

Multi-scale, multi-physics heart simulator "UT-Heart" for heart research

Seiryo Sugiura, MD, PhD, FAHA

Professor, Graduate School of Frontier Sciences, University of Tokyo

Profile:

Education:

Institution	Degree	Year	Field
University of Tokyo, School of Engineering	B.S.	1975	Chemical Engineering
University of Tokyo, School of Medicine	M.D.	1982	Medicine

Medical Training and Employment:

1982-1984 Resident in Internal Medicine, University of Tokyo Hospital
1984 Staff Physician, University of Tokyo Hospital
1985-1987 Research Fellow, Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD
1987-1989 Staff Cardiologist, JR Tokyo General Hospital, Tokyo
1989-1990 Staff Cardiologist, University of Tokyo Hospital
1995-2002 Assistant Professor of Medicine, University of Tokyo
2002-present Professor, Graduate School of Frontier Sciences, University of Tokyo

Abstract

1. Introduction

Cardiology is one of the earliest fields of medical research where numerical simulation was recognized as a powerful tool for promoting our understanding on the disease processes. The fields of application cover electrophysiology, cell metabolism, cardiac mechanics and hemodynamics, but, so far, most of simulation models have only dealt with either microscopic or macroscopic aspect of the problem.

Today, we are facing the explosion of knowledge gained by the progress in molecular and cellular biology. To fully make use of these pieces of knowledge for the solution of various health problems, multi-scale modeling integrating the hierarchical biological system is required. To achieve this goal in heart research, we have developed a multi-scale, multi-physics heart simulator “UT-Heart”.

2. UT-Heart for clinical applications

Our heart simulator is based on the finite element method and its morphology was created from the multi-detector CT or MRI data. In each element of the simulator, mathematical models of cardiac excitation-contraction coupling process are implemented so that each element behaves as a virtual myocyte. The excitation starts at the pacemaker site propagates in the heart tissue to trigger the synchronous contraction and relaxation of the heart. Because the blood in the heart chamber is also modeled as the fluid elements and its interaction with the cardiac tissue is solved by the strong coupling method, we can also analyze the hemodynamic parameters. In addition, whole coronary arteries, capillaries and veins are modeled to show the flow dynamics influenced by the myocardial contraction at each level.

All these features of the model open the possibility of in silico diagnosis and treatment of the virtual heart. We have started a feasibility study by developing the patient-specific heart and torso models. With these heart models, we have successfully reproduced the actual ECG and UCG data of each patient. Prediction of therapeutic effect and optimization of the protocol for the cardiac resynchronization therapy is another target of the research. UT-Heart can also be applied to the development of novel diagnostic or therapeutic devices.

3. UT-Heart for basic science

The power of K-computer enables us to develop a further detailed model in which seamless integration of the heart activities from microscopic to macroscopic scales are realized. In this model, cycling crossbridges of sarcomere in each cardiac myocyte are modeled with its microscopic structures so that the pumping function of the heart is influenced by the force generating activity of each myosin molecule. Heart model with such detailed subcellular structure can be used as the genetically modified human heart model thus will surely contribute to the basic science.

Researchers & Collaborators

Graduate School of Frontier Sciences, Univ. of Tokyo
Takumi Washio, Junichi Okada, Hiroshi Watanabe, Toshiaki Hisada
Dept. of Cardiovascular Medicine, Univ of Tokyo
Hiroshi Yamashita, Taro Kariya, Yasushi Imai, Ryozo Nagai
Fujitsu Ltd.
Yoshimasa Kadooka, Akira Hosoi, Masahiro Watanabe, Takao Hirahara, Takashi Yamazaki, Takashi Iwamura, Machiko Nakagawa, Kohei Hatanaka, Kazunori Yoneda

Acknowledgement

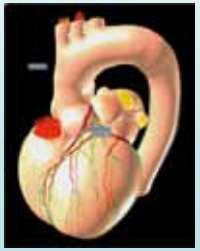
This research is supported by the Japan Society for the Promotion of Science (JSPS) through its “Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program)."

References

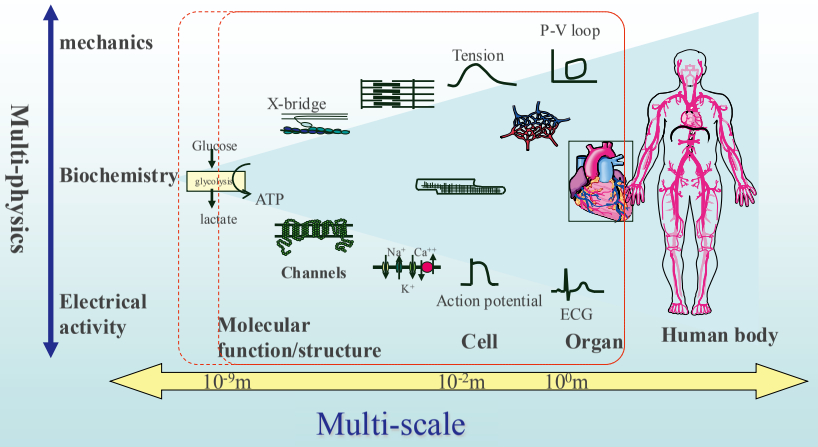
1. T Washio, J Okada, T Hisada SIAM J. SCI. COMPUT. 30:2855–2881 (2008)
2. S Sugiura, T Washio, A Hatano, J Okada, H Watanabe, T Hisada Prog. Biophys. Mol. Biol. in press (2012)

Multi-scale, multi-physics heart simulator "UT-Heart" for heart research

Seiryo Sugiura
Graduate School of Frontier Sciences
The University of Tokyo

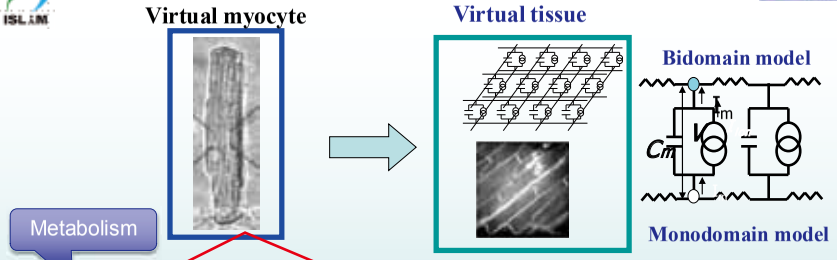


What is multi-scale, multiphysics simulator?

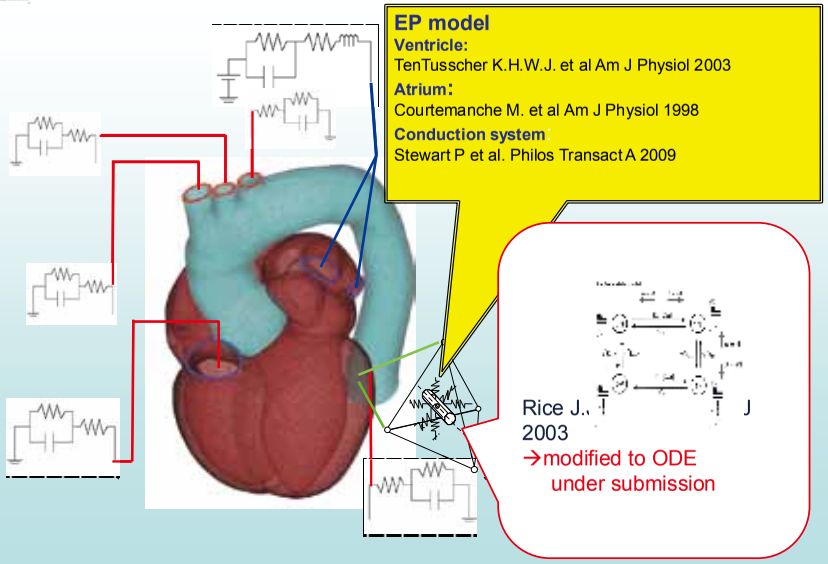


Integration of knowledge ⇒ *In silico* heart ⇒ Basic science Pt. care

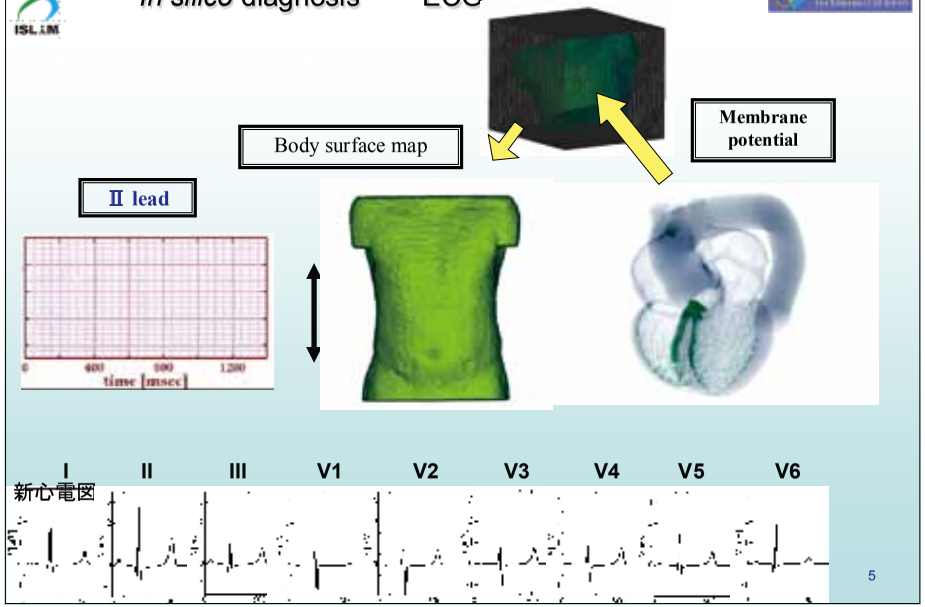
From molecules to cell and tissue



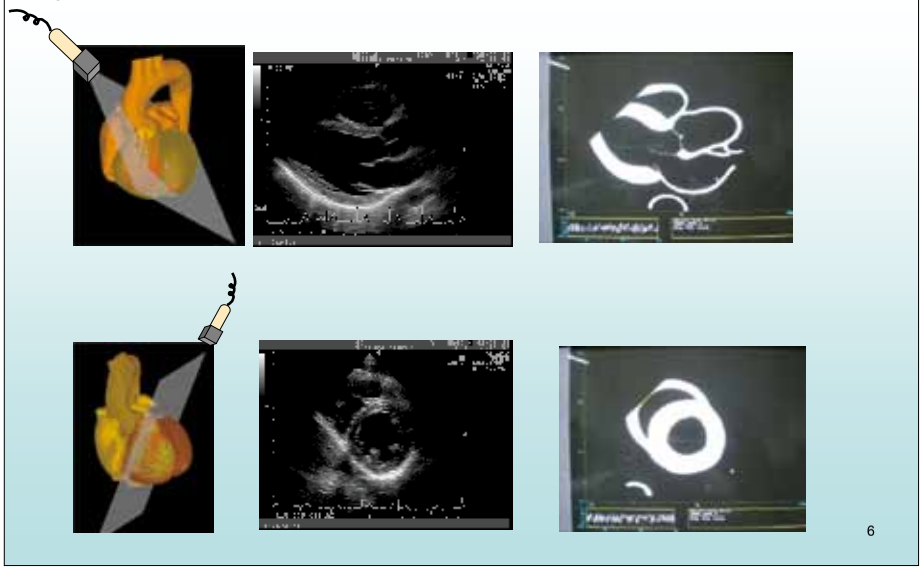
Cell models and boundary conditions



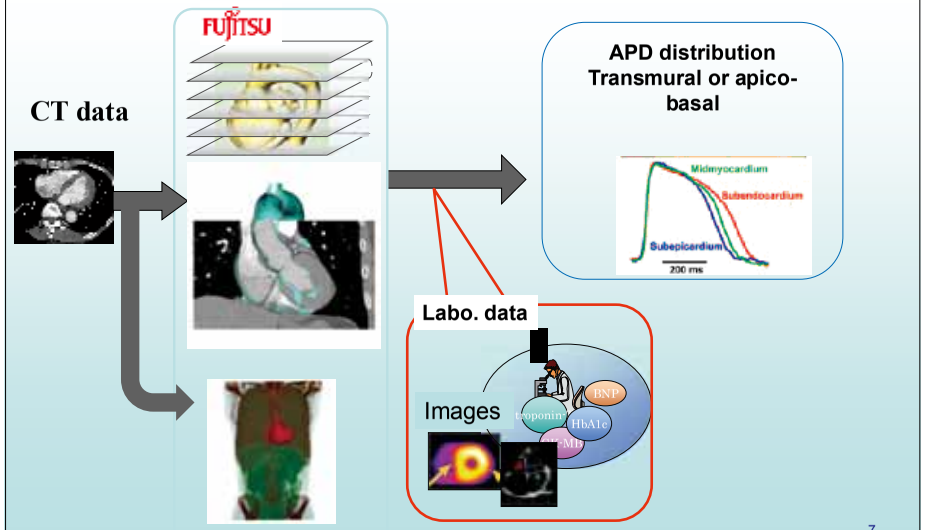
In silico diagnosis ECG



In silico diagnosis UCG

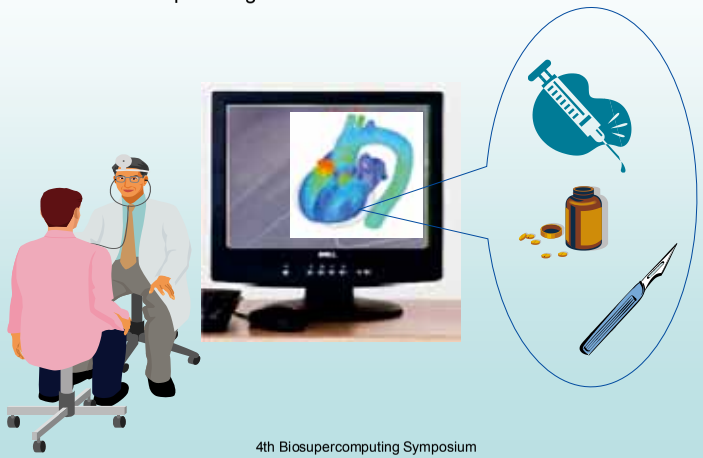


Tailor-made simulation



Purposes of patient-specific simulation

1. Prediction of outcome of the treatment
2. Optimizing the treatment



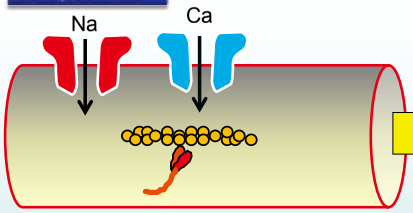
Seamless integration of the biological system

The power of "K-computer" extends the limit

Standard model

Lumped model

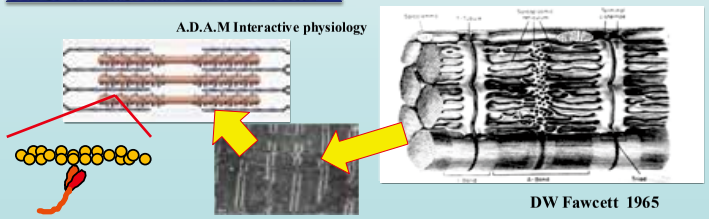
Representavie Na, Ca, ..channels
Contractile proteins
mitochodria



Excitation
Membrane potential
Contraction
Force

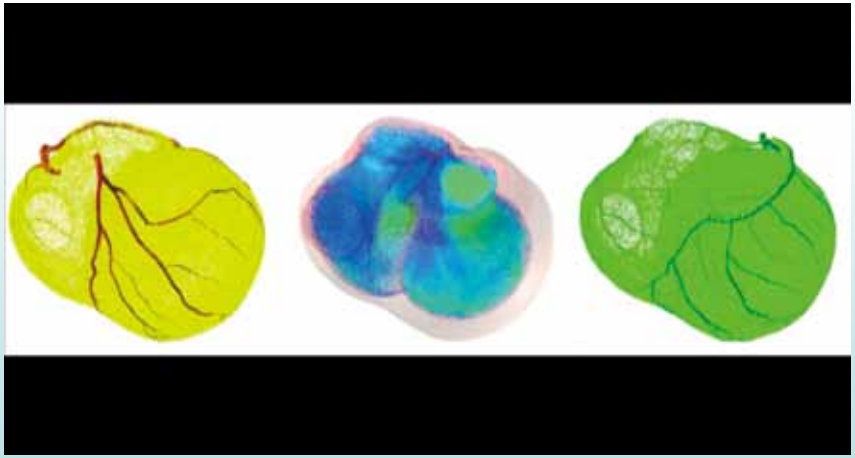
(Extended) Multi-scale model: seamless integration

Reproducing subcellular structure



DW Fawcett 1965

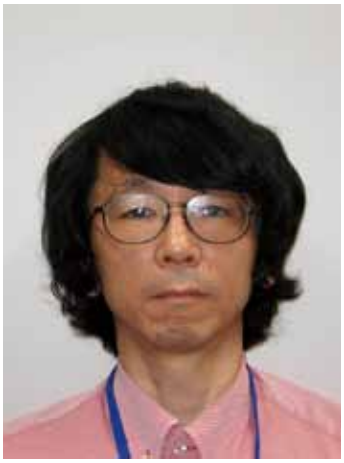
Multi-scale simulation



Optimization of Life-Science applications on the K computer

Yousuke Ohno

Senior Scientist, Computational Science Research Program, RIKEN



Profile:

March, 1996 Ph.D. in Science, Department of Astronomy, Graduate School of Science, The University of Tokyo
April, 1996 - March, 1999 Special Postdoctoral Researcher, RIKEN
April, 1999 - March, 2001 Contract Researcher, Computational Science Laboratory, RIKEN
April, 2001 - March, 2004 Research Scientist, Integrated Volume CAD System Research Program, RIKEN
April, 2004 - March, 2008 Research Scientist, Genomic Sciences Center (GSC), RIKEN
April, 2008 - current Senior Scientist, Computational Science Research Program, RIKEN
April, 2012 - current Senior Scientist, Quantitative Biology Center (QBiC), RIKEN

Abstract

1. Introduction

“Integrated Simulation of Living Matter (ISLiM) Group" is developing "Grand Challenge Applications" in the life sciences that categorized five scales/layers using the K computer and future super computers. Massive parallelization is trend of high-performance computing. In the case of the K computer, the system has 82,944 processors, 8 cores per processor, two 2-way SIMD floating-point multiply-and-add units per core. Therefore we needed to optimize applications by hybrid parallel, 80,000 MPI process, 8 threads, 2 SIMD. To achieve such high-parallelization, we must optimize application at deep layer such as structure of code and data, algorithm, etc.

High-performance Computing Team is assigned supporting optimization of applications that are developing other five teams and developing core applications/libraries for life science on the K computer.

In each category, one or more applications achieved parallelization of 24,576 processes which was maximum number of process in early access to the K computer. Three applications achieved full system of K computer at special trial. High-performance Computing Team is developing the molecular dynamics core program, one of these applications. We performed MD simulation of protein in water with 522 million atoms and achieved 4.4 PFLOPS of calculation performance using 79,872 processors.

2. Molecular Dynamics Core Program

We developed the molecular dynamics core program aimed for establishment of optimization technique, providing optimized code on the K computer.

2.1. *Optimization on the K computer*

In the MD code, the calculation of pair-wise force and potential is most significant part of calculation. The optimization of most inner loop of the force and potential (kernel loop) is hardly depend processor and compiler. On the K computer, SIMD and software pipelining are most important optimization and sensitive to structure of data and loop. We changed the data used in kernel loop from STL vector to simple C style array because the compiler can’t SIMD-optimize to calculation using STL vector. We also rewrote “if” statement as mask operation because the processor of K computer has SIMD instruction of masked operation for “if” statement. Applying such “K” specific optimization, we achieved 54% of SIMD ratio and 56% of efficiency at kernel loop. We use cell index method to processor parallelization. It is suitable for the TOFU, torus network of the K computer. Almost all communication in the cutoff method is local and well scales on the torus network. We performed the cutoff method with cutoff length of 28 Å and observed that it completely weak scaled at 6500 atom/node (Table 1). We also measured strong scaling. The scalability decreased 50% at the case of number of atom per node was about 100.

We achieved 4.4 PFLOPS calculation performance, 43% of theoretical peak on the simulation of 522 million atoms with 28 Å cutoff using 79,872 nodes.

Table 1 weak scaling: time consumption (ms/step) against number of node

Node	64	512	4096	32768	79872	82944
Time	109.1	110.5	111.2	111.7	112.4	112.1

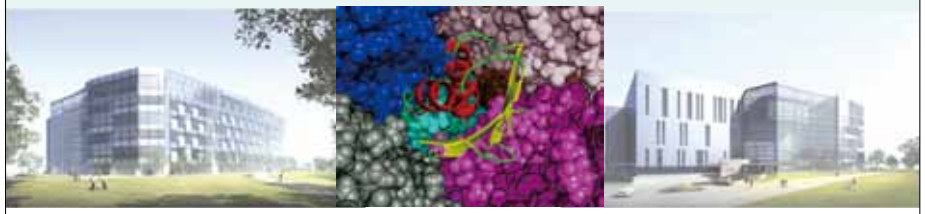
Part of the results is obtained by the K computer at the RIKEN Advanced Institute for Computational Science.

References

1. Akinori Yonezawa, Tadashi Watanabe, Mitsuo Yokokawa, Mitsuhisa Sato, and Kimihiko Hirao. Advanced Institute for Computational Science (AICS): Japanese national high-performance computing research institute and its 10-petaflops supercomputer "K". In State of the Practice Reports, SC '11, pages 13:1-13:8. ACM, 2011.
2. “Optimization of Molecular Dynamics Core Program on the K computer.” Y. Ohno, R. Yokota, H. Koyama, G. Morimoto, A. Hasegawa, G. Masumoto, T. Narumi, and M. Taiji. In Proceedings of JSST 2012 International Conference of Simulation Technology, OS8-1, Kobe, Japan, 2012.

Optimization of Life-Science applications on the K computer

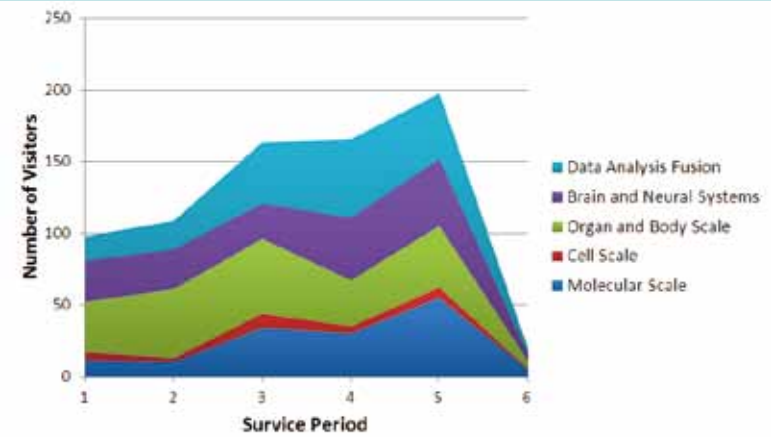
Yousuke OHNO
RIKEN



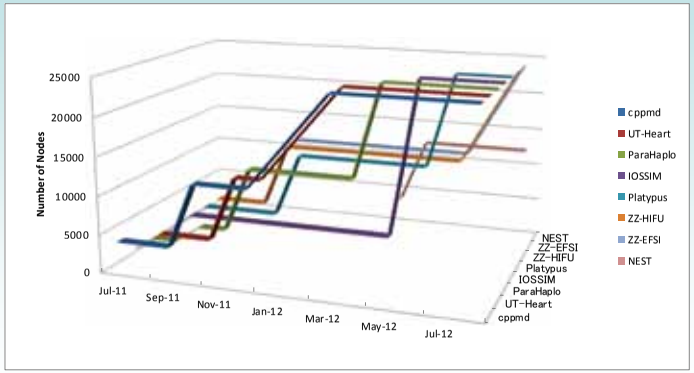
Abstract

- Support other team
- Development core programs/libraries
  - MD core program
    - Kernel tuning
    - Parallel efficiency
    - Peak performance

On-site Support : Number of Visitors



Parallelization : Number of Processors



MD Core Program

- Pair-wise Force/Potential
  - Main component of calculation
  - $O(N^2) \rightarrow O(N)$: cutoff method

$$m_i \frac{d^2 r_i}{dt^2} = -\nabla U(r_1, \cdots, r_{N_{atom}}) \quad (1)$$

$$\begin{aligned} U(\{r_i\}) &= \sum_{\text{bond}} \frac{1}{2} k_b (r - r_0)^2 + \sum_{\text{angle}} \frac{1}{2} k_a (\theta - \theta_0)^2 \\ &+ \sum_{\text{torsions}} \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)] \\ &+ \sum_{|r_{ij}|<R_c} \left\{ \epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \right. \\ &\left. + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \Psi(r_{ij}) \right\} \quad (2) \end{aligned}$$

Kernel Loop

```
#pragma omp parallel for
for (i=0;i<N_mycell;i++){ // multi thread
  for (l=0;l<nl[i];l++){// SIMD, software pipeline
    j = list[i][l];
    dx = x[i]-x[j]; dy = y[i]-y[j]; dz = z[i]-z[j];
    r2 = dx*dx + dy*dy + dz*dz;
    if(r2<cutoff2){mask=1.0}else{mask=0.0} // mask
    _r = 1.0/sqrt(r2);
    _r2 = _r * _r; _r6 = ...
    e += (LJP(_r12,_r6,) + c[i]*c[j]*CLP(_r,))*mask;
    dp = (LJF(_r14,_r8,) + c[i]*c[j]*CLF(_r3,))*mask;
    fx[i] += dx*dp; fy[i] += dy*dp; fz[i] += dz*dp;
    fx[j] -= dx*dp; fy[j] -= dy*dp; fz[j] -= dz*dp; =
  }
}
```

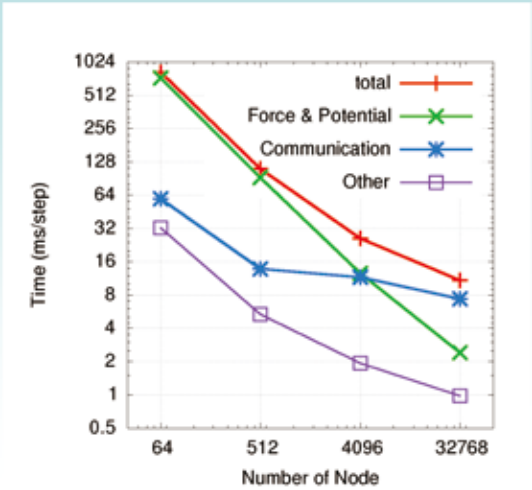
Kernel Optimization

- SIMD, Software-pipelining
  - Loop structure
    - Simple loop
    - "if" statement → mask operation
  - Data structure
    - Simple C style array
    - STL vector → C style array
  - SIMD 55 % of all instruction, 99.9% of FP instruction
  - Efficiency 56 %, 72 GFLOPS / CPU

Weak Scaling of Cutoff

Number of Node	64	512	4,096	32,768	79,872	82,944
Number of Atom	418,707	3,349,656	26,797,248	214,377,984	522,546,336	542,644,272
Total (ms/step)	109.058	110.535	111.186	111.672	112.414	112.085
Force & Potential	91.622	91.528	91.601	92.329	91.262	91.641
Communication	12.124	13.694	14.011	14.012	15.820	15.067
Other	5.312	5.313	5.574	5.332	5.332	5.376
Performance (PFLOPS)	0.003	0.025	0.201	1.599	3.871	4.031
Efficiency (%)	39.0	38.5	38.3	38.1	37.9	38.0

Strong Scaling of Cutoff



Best Performance

- 79,872 node
- 522,546,336 atom
- Coulomb and LJ with shift function
- 28 Å cutoff
- Calculation of potential : every time step
- Time consumption : 116.4 ms / step
- Performance : 4.39 PFLOPLS
- Efficiency : 42.9 %

Weak Scaling of FMM

Number of Node	64	512	4,096	9,216	32,768	73,728
Number of Atom	418,707	33,49,656	26,797,248	60,293,808	214,377,984	482,350,464
Total (ms/step)	41.1913	41.9641	42.8347	43.0937	44.0497	44.3478
Force & Potential	29.4784	29.1701	30.3507	31.4793	31.7125	31.9837
	2.12656	2.79204	3.57428	3.90615	4.38367	4.24487
Communication	7.60168	8.68757	8.36386	7.40102	8.21423	8.20732
Other	4.11113	4.10638	4.12014	4.21341	4.12298	4.15676

Conclusion

- Our team had supported optimization of application on the K computer.
- We developed MD core program for massive parallel computer.
- Our MD core program has good scalability.
- Our MD core program had performed 4.39 PFLOPS, 42.% efficiency on the K computer.

Part of the results is obtained by the K computer at the RIKEN Advanced Institute for Computational Science.

Research activity of Cell Scale Simulation Team

Hideo Yokota

Team Leader, Cell-scale Research and Development Team, RIKEN
Team Head, Bio-research Infrastructure Construction Team, RIKEN

Profile:

2012-present:Deputy Team Leader, Measurement information Laboratory, RCIIC, RIKEN
2011-present:Team Head, Bio-research Infrastructure Construction Team, ASI, RIKEN
2007-present: Team Leader, Cell-scale Research and Development Team, ISLiM RIKEN
2006:Team leader: Bio-research Infrastructure Construction Team, VCAD, RIKEN
2003:Team leader: VCAT development team, RIKEN
1999:Contract researcher: Computational biomechanics unit, RIKEN
1999: Doctor of Engineering degree from the University of Tokyo
1993:Researcher, Kanagawa Academy of Science and Technology
1993: Graduated from Nihon University of Agriculture and veterinary medicine(M.S.(agri.))

Abstract

1. Introduction

Cells are the smallest units of life. The ultimate goal of our cell-scale simulation team is to recreate intracellular and intercellular phenomena in computer simulations. However, most intracellular phenomena have yet to be elucidated, and remain a major area of research in cell biology. It is therefore currently impossible to develop a simulator that can recreate all intracellular phenomena. Hence, our cell-scale simulation team's objective is to simulate intracellular and extracellular phenomena that are reasonably well understood and for which mathematical models have been established. In addition, using the computational power of a next-generation 10-petaflop "Kei" supercomputer, our team aims to simulate intracellular environments that heretofore could not be calculated due to high computational costs. Future advances in cell biology, mathematical model construction, and simulation technologies should markedly expand the scope of cell simulation. In the hope of contributing to the development of cell simulation, a cell simulation platform that can analyze coupled phenomena by linking simulators for multiple phenomena (RICS) was developed. The cell phenomenon simulation developed by our team aims to recreate actual intracellular phenomena, and not merely run computer calculations. Experiments have been performed concurrently to gather the necessary parameters for the simulation and to verify the simulation. This multi-pronged approach will allow the simulation to discover unknown phenomena, rather than just recreate phenomena.

2. Integrated cell simulation platform

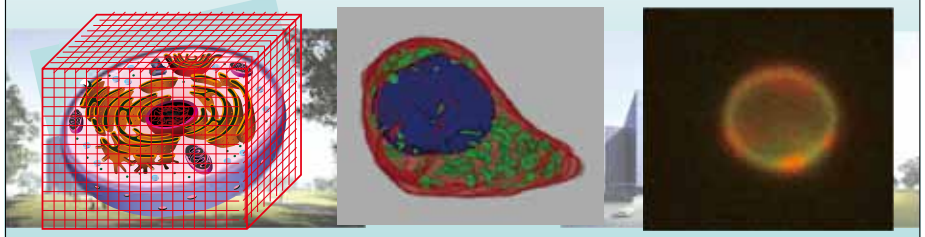
The basic design and methodologies of a common cell simulation base that can couple intracellular fields and multiple simulations were investigated. Basic fields inside the solver were 3-dimensional scalar fields consisting of the number of various molecules and the volume ratio of the medium. As a general rule, different simulators working on the common base calculate temporal changes in the number of molecules and the volume of medium inside each voxel. The medium represents organelles, and the system was designed to deal with organelle movements in the future. The system was also designed to withstand large-scale parallel processing by appropriately compartmentalizing calculation spaces. In addition to designing the cell simulation platform, metabolism simulation was also carried out. In metabolism simulation, "E-cell3" was used because of its track record in intracellular metabolism simulation. Furthermore, since membrane functions are important cellular functions, the following cell membrane functions were added: "substance penetration", "penetration on/off switch (ion-channel)", "active transport (pump)", and "cell membrane receptor function (internal enzymes activate when substances bind outside)". Moreover, a function to analyze the membrane potentials of neurons and other cells is being developed. These multiple phenomena were examined by weak coupled analyses based on sequential computations per time step.

3. Conclusions

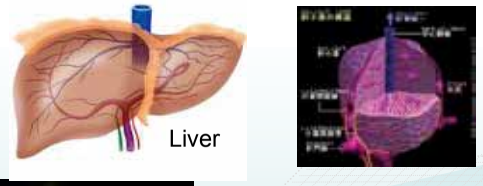
We have developed a system to enable spatiotemporal simulation of the cell. The resulting RICS system was designed not to only to simulate specific cell phenomena, but to be a universal spatiotemporal simulator of the cell. In addition to enabling simulation of biochemical reactions, diffusion, transport, membrane function, membrane potential, and advection within the cell, the RICS platform can couple simulations of multiple phenomena. Furthermore, with the ability to create geometric models that replicate the shapes of actual cells, it becomes possible to create simulations that account for the physical geometry of individual cells and to verify simulation results against actual observations. Because this approach is computationally expensive, it would have been extremely difficult to do using conventional computers. However, spatiotemporal simulations of cells are now possible by utilizing the large-scale parallel processing power of the next-generation "Kei" supercomputer.

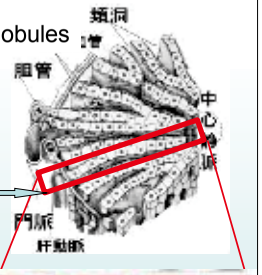
Research activity of Cell Scale Simulation Team

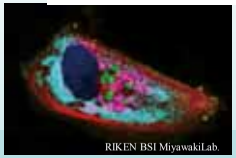
Hideo YOKOTA
RIKEN



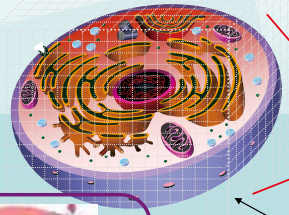
Need :spatiotemporal simulation of the cell



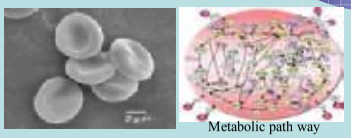




Microscopic Image
Almost cell: Un-homogeneous



Portal vein
Central vein



Red blood cell: homogeneous

RICS

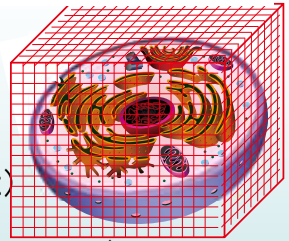
RIKEN Integrated Cell Simulation Platform

nm

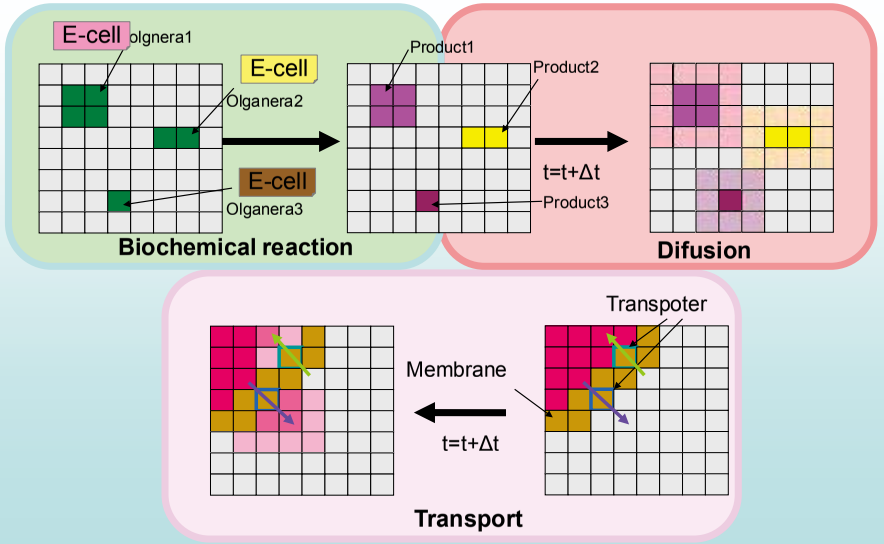
- Biochemical reaction（Metabolism）
  - E-cell
- Diffusion（Passive transport）
  - concentration gradient
- Membrane transport（Active transport）
  - Membrane transport
- Membrane Functions（Selective permeability）
  - ion channels, transporters, receptor, et al.
- Blood flow （Computational Fluid dinamics）
- Membrane potential（Hodgkin-Huxley）

mm

- Simulation of cells cluster
- Integrated simulation above phenomenon.

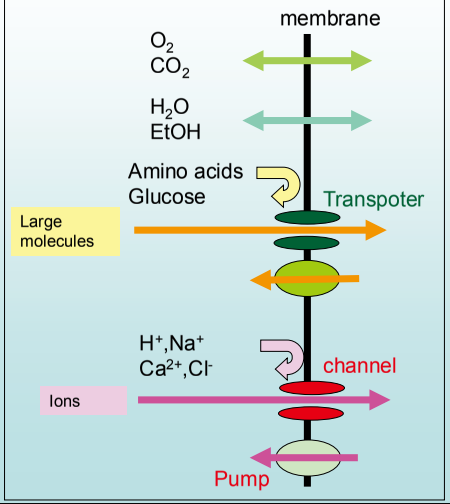


Coupled simulaton in RICS



Functions of membrane

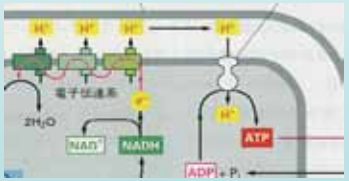
Selective permeability



Region of biochemical reaction

Receptors

Integrated enzymes



Flux of membrane transporters

$$\frac{\partial \phi}{\partial t} = \nabla \bullet (C \nabla \phi) - \nabla \bullet f_m + [BioChemical\ \mathrm{Re}action]$$

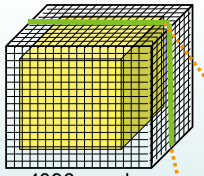
$$f_m = \sum_{Func} SW^{Func} f_m^{Func}$$

Func = {penetrate, channels, pumps, transporters, gap junctions}

SW = regulatory function of each transporter

Transport of the substance through the membrane is the sum of the flow velocity for each transport.

Simulation of membrane transport with bioreaction



Cytosol
Substance A: 0 mM
Substance B: 0 mM
Substance C: 0 mM

Center
Enzyme E1 : 1 mM

Cell membrane
Enzyme E2 : 1 mM

extracellular area
Substance A : 100 mM

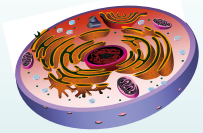
A —Ez1→ B —Ez2→ C

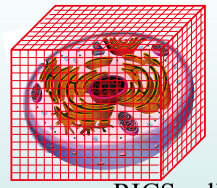
Ez1: Vmax = 1.0, Km = 0.5

Ez2: Vmax = 2.0, Km = 1.2

Section of a cell in voxels

Virtual cell shape (CAD,CG)

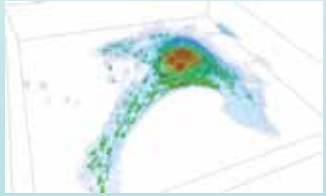




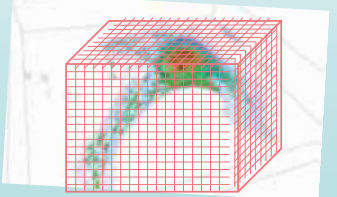
RICS cell model

3D rebuilding Cell data from microscopic images

RICS Preprocessor



Sectional images



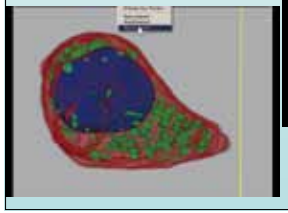
RICS cell model

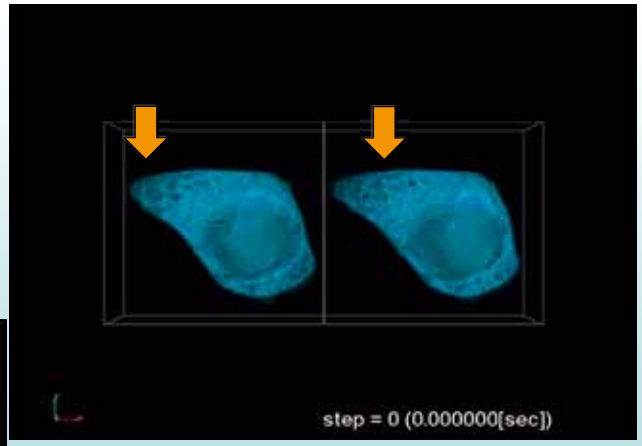
©RIKEN : Live Cell Modeling Project

Ca^{+} kinetics simulation（HepG2）

Ca^{2+}reaction model

reaction：24
material：10



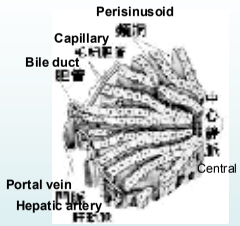


diffusion+reaction
（apparent diffusion rate 28μm/s）

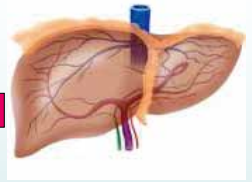
Only diffusion
（diffusion rate 32μm/s）

9

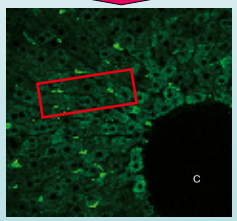
Liver cell simulation



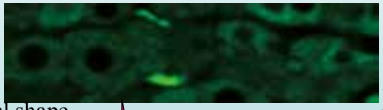
Hepatic lobules



Liver



Microscopic image



Real shape

Hepatic cords

10

Summary

- RICS (Riken Integrated Cell Simulation Platform)
  :Spatiotemporal simulation of the cell
- Couple simulations of multiple phenomena
  :biochemical reactions, diffusion, transport, membrane function, membrane potential, and advection
- Real cell shape simulation
  : verify simulation results against actual observations
- 京(kei) available
  : large-scale parallel more than 24,000nodes

Open source release : December 2012

11

Acknowledgement

Keio Univ.
Ayako Yachie
Makoto Suematsu

Tokai Univ.
Yohei Nagasawa
Shinya Goto

Osaka Univ.
Shingo Murakami
Hiroshi Hibino
Yoshihisa Kurachi

Kobe Univ.
Tadao Shibasaki
Susumu Seino

RIKEN
Yasuhiro Sunaga
Shin Yoshizawa
Shigeho Noda
Takashi Koyama
Yuki Tsujimura
Koichi Takahashi
Taiji Adachi

Dissection of regulatory mechanisms for metabolic systems by quantitative imaging mass spectrometry

Makoto Suematsu

Professor, Department of Biochemistry, School of Medicine, Keio University, JST ERATO Suematsu Gas Biology Project

Profile:

2009-present: Leader, Japan Science and Technology Agency, ERATO Suematsu Gas Biology Project

2007-present: Dean, School of Medicine, Keio University

2007: Leader, Global Center of Exellence for Life Sciences, Human Metabolomic Systems Biology from MEXT

2003: Leader, National Leading Project for Biosimulation by Ministry of Edication, Sciences and Technology

2001: Professor and Chair, Department of Biochemistry and Integrative Medical Biology, Keio University School of Medicine

1991: Bioengineer Step IV, Institute for Biomedical Engineering, University of California San Diego (Supervised by Professor Benjamin W Zweifach and Professor Geert W Schmid-Schoenbein)

1983: Graduated from Keio University School of Medicine (MD)

Abstract

Collection of the data in reality is absolutely necessary to empower the capability of large-scale biosimulation of metabolic systems, while it has been difficult to collect quantitative data of many of metabolites in tissues in vivo. We have applied advanced metabolome technology including imaging mass spectrometry to this effect. Application of metabolome analyses based on CE-MS allowed us to predict and demonstrate roles of hemoglobin allostery in O2-sensing mechanism in human erythrocytes, and now extending the technology to analyses of cancer metabolism in vivo. Metastatic progression of cancer does not only upregulate their glucose metabolism but might utilize metabolic properties of host organs that benefit cancer metabolism, although such a hypothesis remains elusive. Newly developed microscopic imaging mass spectrometry combined with MS^2 analyses allowed us to collect micrographs of many different metabolites in a single frozen section, and combination with CE-MS data collected from the serial section provided semiquantitative information of individual signals. The current method revealed that human colon cancer xenografts metastasized in livers of super-immunodeficient NOD/scid/γnull (NOG) mice deprives L-alanine to support their metabolic demands for synthesizing glutathione and nucleotides. In this model, hepatic metastasis triggered regenerative responses of the host liver concurrently with hypoglycemia and accumulation of glutathione and nucleotides in the tumor-bearing liver. MS^2 analyses under loading $^{13}C_3$-L-alanine provided evidence for earlier filling of glutathione with $^{13}C_2$-γ-glutamylcysteine structure in metastases than surrounding liver parenchyma. The $^{13}C_3$-L-alanine loading also caused an increase in $^{13}C_2$-UDP in metastatic foci and in the host liver. MS^2 analyses to assess the pathways for ^{13}C incorporation revealed that L-alanine not only undergoes gluconeogenesis in the host to synthesize ribose but serves as a substrate to supply glutamate and pyrimidine carbons for nucleotide synthesis occurring in the metastases. Our results suggest that human colon cancer metastases utilize gluconeogenic substrates of the host not only through pentose phosphate pathway but through glutaminolysis, supporting their metabolic demand of glutathione and nucleotides to cause hypoglycemia. In order to explore novel molecular targets that play a crucial role for survival of cancer cells, combination of human cancer xenograft models with large-scale computer simulation of metabolic systems might serve as a powerful stratagem in future.

References

1. Soga, T., et al. Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. **J. Biol. Chem.** 281(24) 16768-16776, 2006.
2. Kinoshita, A., et al. Roles of hemoglobin allostery in hypoxia-induced metabolic alterations in erythrocytes: simulation and its verification by metabolome analysis **J. Biol. Chem.** 282(14), 10731-10741, 2007.
3. Hattori K, et al. Paradoxical ATP elevation in ischemic penumbra revealed by quantitative imaging mass spectrometry. **Antioxid Redox Signal** (News & Views), 13(8), 1157-1167, 2010.
4. Ishimoto T, et al. CD44 variant regulates redox status in cancer cells by stabilizing the xCT subunit of system xc- and thereby promotes tumor growth. **Cancer Cell** 2011, 19(3), 387-400.
5. Morikawa T, et al. Hypoxic regulation of the cerebral microcirculation is mediated by a carbon monoxide-sensitive hydrogen sulfide pathway. **Proc Natl Acad Sci USA** 2012, 109(4), 1293-1298.

Dissection of regulatory mechanisms for metabolic systems by quantitative imaging mass spectrometry

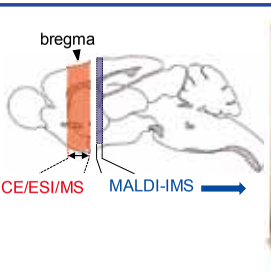
~ Application to ischemic diseases and cancer models ~

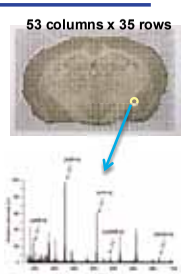
Makoto Suematsu, MD, PhD
Professor & Chair
Department of Biochemistry, School of Medicine, Keio University
Japan Science Technology Agency, ERATO Suematsu Gas Biology
gasbiology@z6.keio.jp

KEIO 150
Design the Future

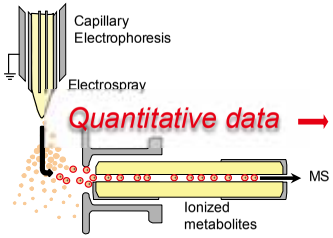
ISLiM Symposium 2012

Semiquantitative imaging mass spectrometry: IMS Combination with CE/ESI/MS





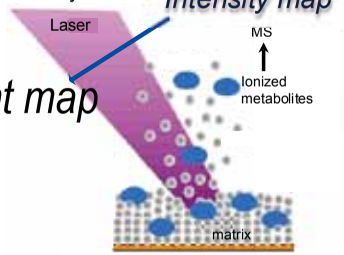
Complete Ionization by ESI



Quantitative data → Content map

Soft Ionization by MALDI

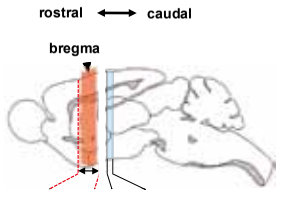
Intensity map



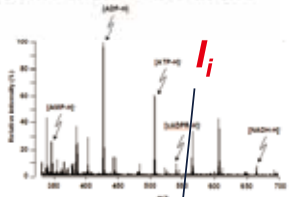
http://biopedia.org/images/1/1e/Maldi-TOF-MS.jpg

Hattori *et al.*, Antioxid Redox Signal, 2010

Semi-quantification: apparent contents of a specific metabolite at the i^{th} spot of tissue (C_i) was estimated:



$$C_i = \frac{I_i}{\bar{I}} C'$$

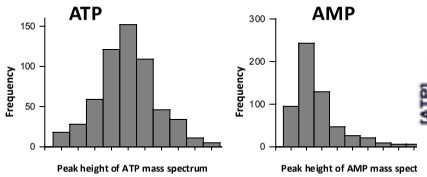


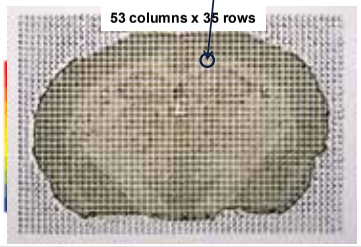
C': the metabolite content of tissue from a contralateral hemisphere determined by the CE/ESI/MS

I_i: the maximum intensity among mass spectra in a specified range at the i^{th} spot

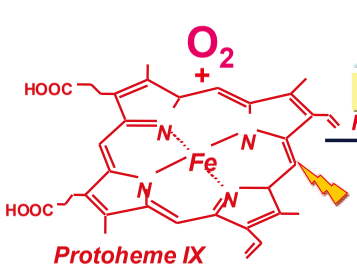
$\bar{I}$: the median of maximum intensities of a metabolite from all the spots in the contralateral hemisphere

$\bar{I}$ = median





Liver and Brain: Gigantic resources of CO with distinct gas-responsive receptors



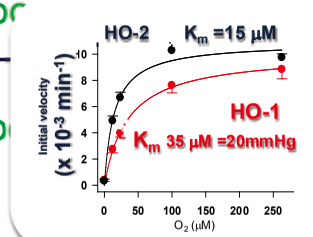
Protoheme IX

HO-1
HO-2
Heme oxygenase

Cytokines, ROS
cAMP, heavy metals, etc

Fe (II)

CO



Liver: soluble guanylate cyclase (sGC)
Microvascular pericytes in sinusoids (Ito cells)

↑ cGMP

Vasodilation

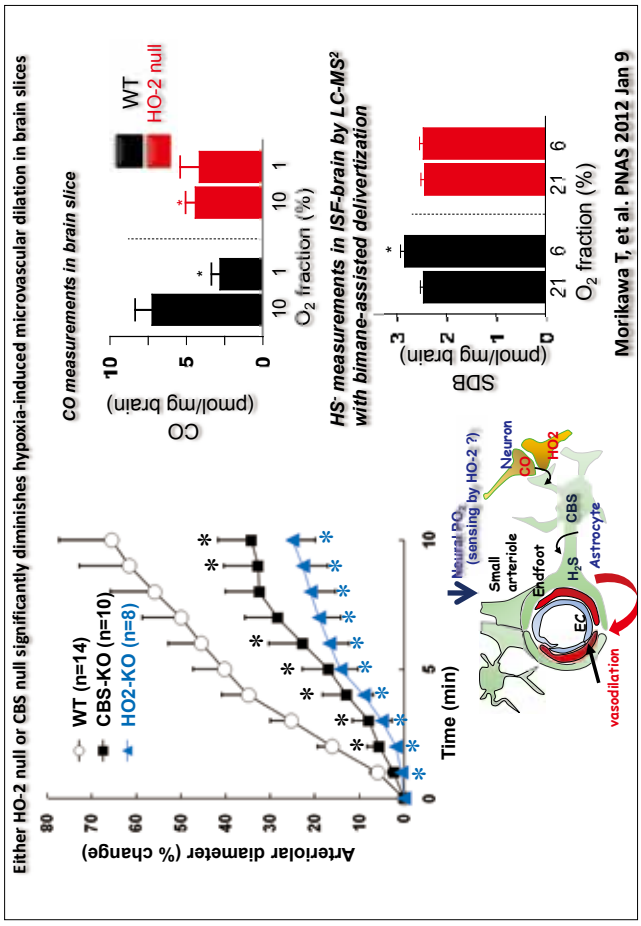
Suematsu M, et al. BBRC 1994, J Clin Invest 1995
Goda N, et al. J Clin Invest 1998
Kyokane, et al. Gastroenterology 2001

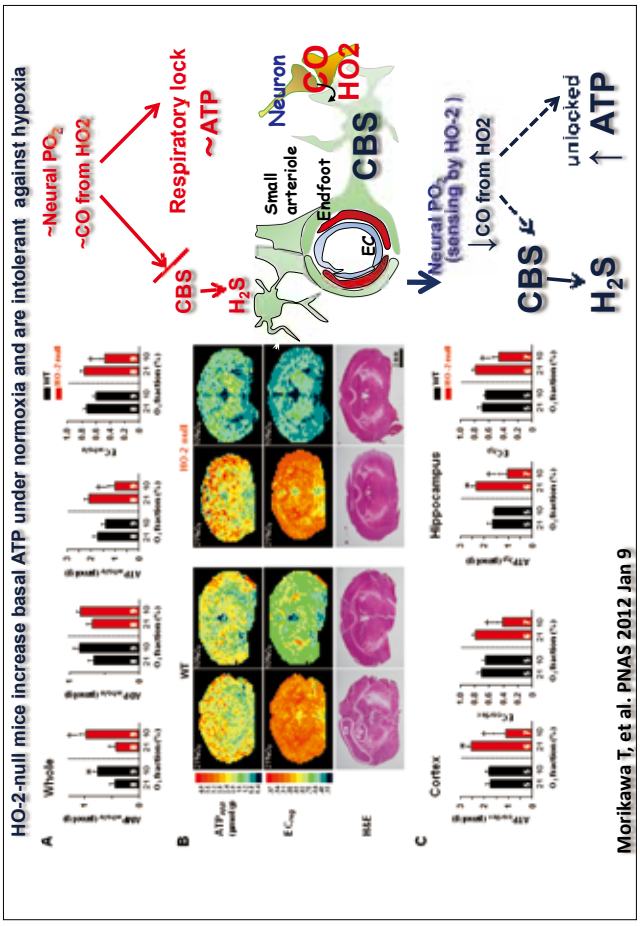
Brain: cystathionine β-synthase (CBS)
Astrocytes in neurovascular units

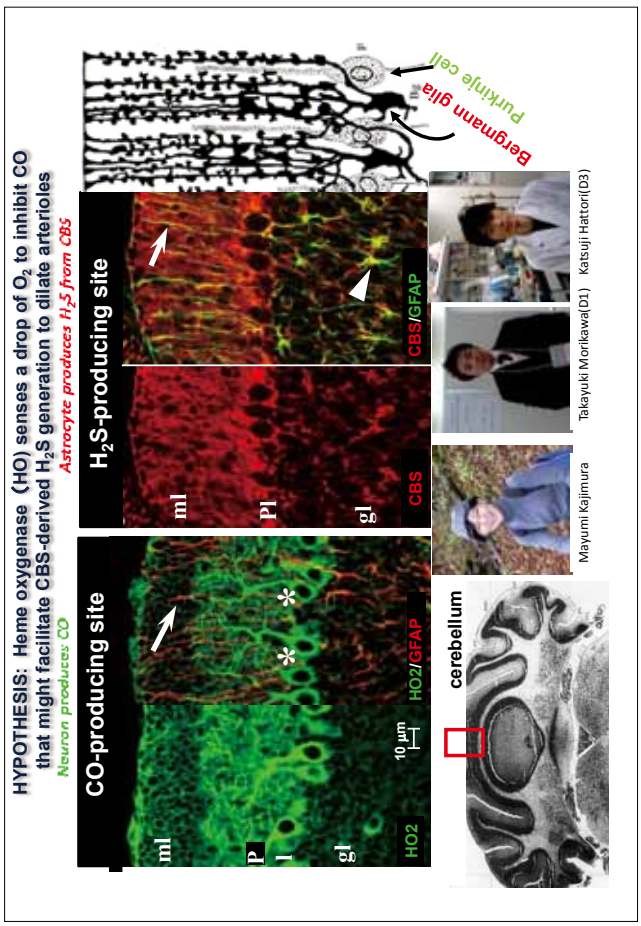
↓ H_2S

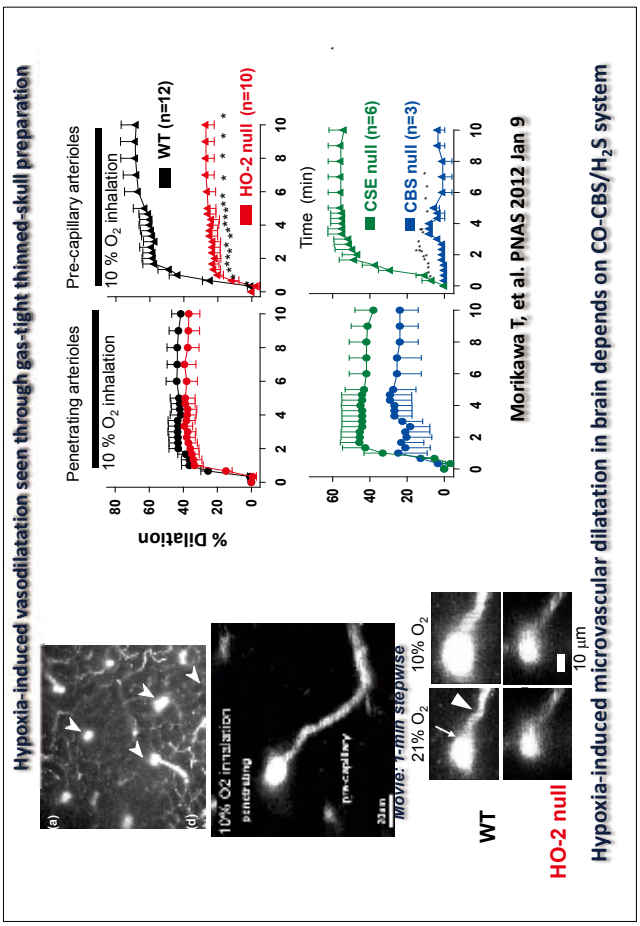
Vasoconstriction

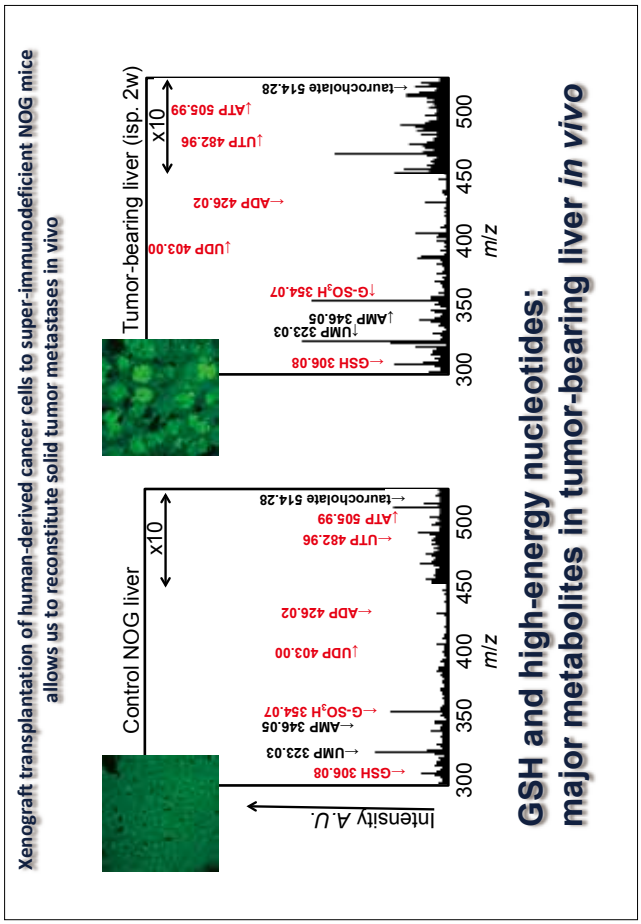
Ishikawa M, et al. Circ Res 2005
Shintani T, et al. Hepatology 2009
Kajimura M, et al. ARS 2010
Morikawa T, et al. PNAS 2012

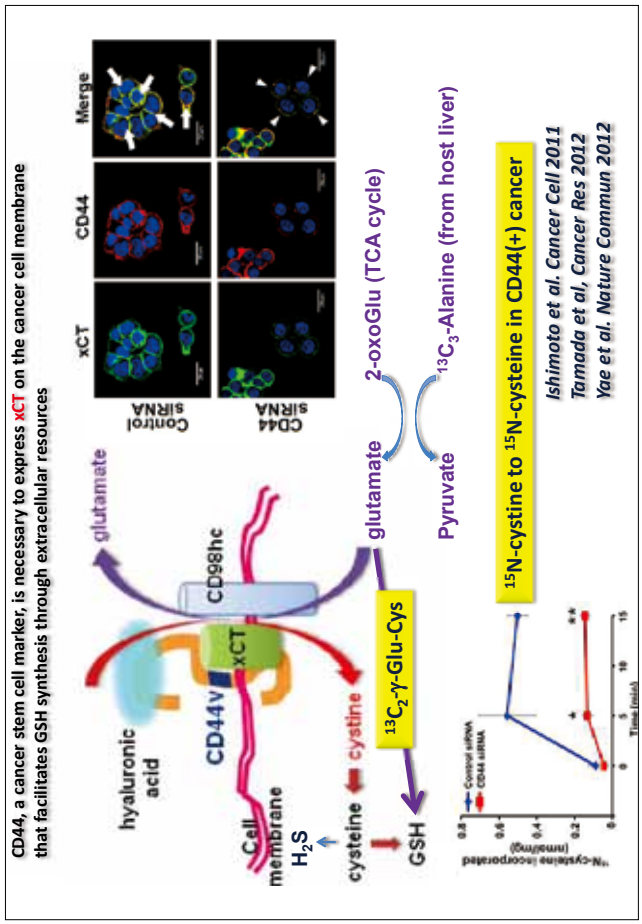

Either HO-2 null or CBS null significantly diminishes hypoxia-induced microvascular dilation in brain slices
Arteriolar diameter (% change)
WT (n=14)
CBS-KO (n=10)
HO2-KO (n=8)
Time (min)
Neural PO2 (sensing by HO-2 ?)
Small arteriole
Endfoot
EC
H2S
CBS
Astrocyte
Neuron
CO
HO2
vasodilation
CO measurements in brain slice
CO (pmol/mg brain)
WT
HO-2 null
O2 fraction (%)
HS- measurements in ISF-brain by LC-MS2 with bimane-assisted delivertization
SDB (pmol/mg brain)
O2 fraction (%)
Morikawa T, et al. PNAS 2012 Jan 9

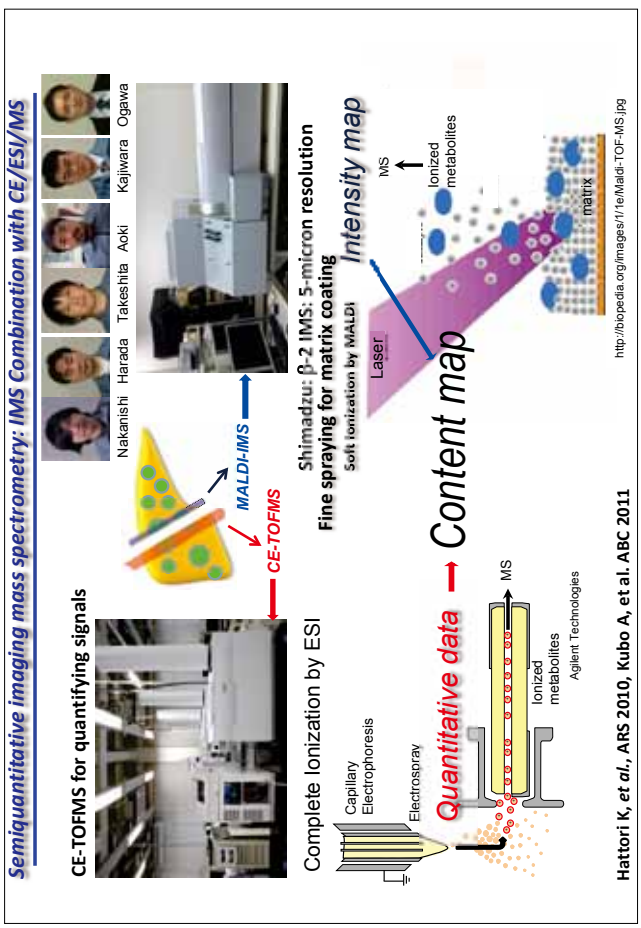

HO-2-null mice increase basal ATP under normoxia and are intolerant against hypoxia
~Neural PO2
~CO from HO2
Respiratory lock
~ATP
CBS
H2S
Small arteriole
Endfoot
Neuron
CO
HO2
EC
CBS
Neural PO2 (sensing by HO-2)
CO from HO2
CBS
H2S
unlocked
↑ ATP
A
Whole
B
WT
HO-2 null
Hippocampus
C
Cortex
Morikawa T, et al. PNAS 2012 Jan 9

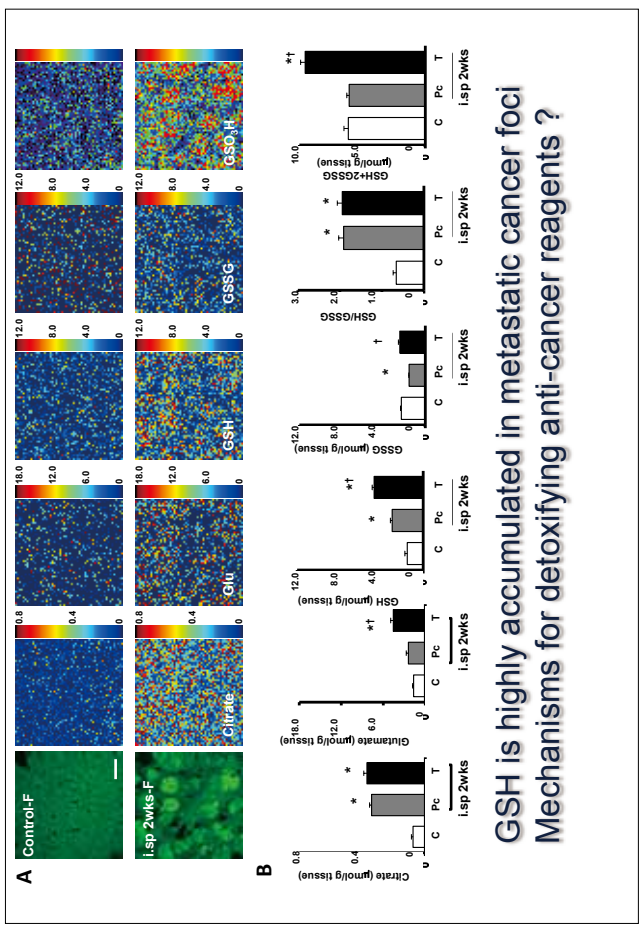

HYPOTHESIS: Heme oxygenase (HO) senses a drop of O2 to inhibit CO that might facilitate CBS-derived H2S generation to dilate arterioles
Neuron produces CO
Astrocyte produces H2S from CBS
CO-producing site
H2S-producing site
ml
P
I
gl
HO2
HO2/GFAP
CBS
CBS/GFAP
10 μm
Bergmann glia
Purkinje cell
cerebellum
Mayumi Kajimura
Takayuki Morikawa(D1)
Katsuji Hattori(D3)


Hypoxia-induced vasodilatation seen through gas-tight thinned-skull preparation
Penetrating arterioles
Pre-capillary arterioles
10 % O2 inhalation
% Dilation
WT (n=12)
HO-2 null (n=10)
Time (min)
CSE null (n=6)
CBS null (n=3)
Morikawa T, et al. PNAS 2012 Jan 9
10% O2 inhalation penetrating
pre-capillary
Movies 1 min stepwise
21% O2
10% O2
WT
HO-2 null
10 μm
Hypoxia-induced microvascular dilatation in brain depends on CO-CBS/H2S system


Xenograft transplantation of human-derived cancer cells to super-immunodeficient NOG mice allows us to reconstitute solid tumor metastases in vivo
Tumor-bearing liver (isp. 2w)
Control NOG liver
x10
Intensity A.U.
m/z
GSH and high-energy nucleotides: major metabolites in tumor-bearing liver in vivo


CD44, a cancer stem cell marker, is necessary to express xCT on the cancer cell membrane that facilitates GSH synthesis through extracellular resources
CD44 siRNA
Control siRNA
xCT
CD44
Merge
glutamate
hyaluronic acid
CD44v
CD98hc
xCT
Cell membrane
cystine
cysteine
GSH
2-oxoGlu (TCA cycle)
Pyruvate
$^{13}C_3$-Alanine (from host liver)
$^{13}C_2$-γ-Glu-Cys
^{15}N-cystine to ^{15}N-cysteine in CD44(+) cancer
Ishimoto et al. Cancer Cell 2011
Tamada et al, Cancer Res 2012
Yae et al. Nature Commun 2012


Semiquantitative imaging mass spectrometry: IMS Combination with CE/ESI/MS
CE-TOFMS for quantifying signals
Nakanishi Harada Takeshita Aoki Kajiwara Ogawa
CE-TOFMS
MALDI-IMS
Complete Ionization by ESI
Capillary Electrophoresis
Electrospray
Quantitative data
MS
Ionized metabolites
Agilent Technologies
Shimadzu: β-2 IMS: 5-micron resolution
Fine spraying for matrix coating
Soft Ionization by MALDI
Laser
Intensity map
Content map
matrix
http://biopedia.org/images/1/1e/Maldi-TOF-MS.jpg
Hattori K, et al., ARS 2010, Kubo A, et al. ABC 2011


A
Control-F
i.sp 2wks-F
Citrate
Glu
GSH
GSSG
CSO_3H^-
B
Citrate (μmol/g tissue)
Glutamate (μmol/g tissue)
GSH (μmol/g tissue)
GSSG (μmol/g tissue)
GSH/GSSG
GSH+2GSSG (μmol/g tissue)
C Pc T
i.sp 2wks
GSH is highly accumulated in metastatic cancer foci
Mechanisms for detoxifying anti-cancer reagents ?

Closing the loop: simulation of the whole sensory-motor neural network in action

Kenji Doya

Professor, Okinawa Institute of Science and Technology

Profile:

KENJI DOYA took BS in 1984, MS in 1986, and Ph.D. in 1991 at U. Tokyo. He became a research associate at U. Tokyo in 1986, U. C. San Diego in 1991, and Salk Institute in 1993. He joined ATR in 1994 and became the head of Computational Neurobiology Department, ATR Computational Neuroscience Laboratories in 2003. In 2004, he was appointed as the principal investigator of Neural Computation Unit, Okinawa Institute of Science and Technology (OIST) and started Okinawa Computational Neuroscience Course (OCNC) as the chief organizer. As OIST re-established itself as a graduate university in 2011, he became a professor and the vice provost for research. He serves as the co-editor in chief of Neural Networks from 2008. He is interested in understanding the functions of basal ganglia and neuromodulators based on the theory of reinforcement learning. Contact: doya@oist.jp, 1919-1 Tancha, Onna, Okinawa 904-0495, Japan.

Abstract

1. Introduction

Realistic simulation of the brain is a grand challenge toward understanding how the brain works and also in establishing how to simulate complex dynamical systems. Detailed computer simulations of the brain have so far been limited to the cellular or local circuit levels, such as the Blue Brain Project and Cortical Columns in Silico. The aim of our project is to reconstruct the whole brain network including sensory inputs and motor outputs and to test how the brain functions under dynamic interactions with the environments.

The use of super computers has mostly been limited to batch operations in which huge jobs are put into a queue and the results are given after indefinite hours or days waiting. An important rationale for a locally concentrated super computer like K, as opposed to world-wide distributed computing grids, is its use for time-critical missions. Simulation of the brain of a humanoid robot is a scientifically important and computationally demanding challenge. Despite big progresses in movement control and appearances, current humanoid robots are far from human-like because of their limited behavioral responses, in other words, because of their poor brains. Human-like behaviors require huge computations, from sensory processing and motor control to language and emotion. And natural human interaction requires timely responses. Real-time simulation of the brain of a humanoid robot is an ultimate test of what we know (and do not know) about the brain function and also an ultimate test in high-throughput computing.

As the first step in implementing a humanoid brain, we focus on the oculomotor control loop, which is one of the best-studied model system in neuroscience.

2. Oculomotor Network Model

The mammalian oculomotor system has a hierarchical organization with the retina – superior colliculus – brain stem – eye muscles as the most basic pathway, over which the cerebellum, the basal ganglia, and the cerebral cortex impose more sophisticated control. Here we focus on the basic pathway with the emphasis on the neural circuit of the intermediate layer of the superior colliculus, which is regarded as the center for the control of saccadic eye-movements.

The superior colliculus model [1] consists of the visual input layer, burst neuron layer (output), buildup neuron layer, and deep inhibitory layer. The neurons were implemented by integrate-and-fire models and the excitatory and inhibitory connections were set up according to the anatomical literature and to reproduce the activity features of different types of neurons.

The model was implemented using a spiking neural network simulation tool NEST [2]. The model of one side of the superior colliculus consisted of about 100,000 neurons and the computing time was about four times the simulated biological time on 800 nodes of the Intel cluster at RICC. The result replicated the known velocity and duration features of saccadic eye movements, known as ‘main sequence.’ The model suggested the role of spreading activities of buildup neurons in deciding when to stop the eyes [1].

3. Robot Eye Movement Experiment

We connected the left and right copies of the superior colliculus model and the brain stem network model with a multi-simulation coordinator MUSIC [3]. The integrated model ran at RICC was connected through the internet to the eye control system of a humanoid robot CB-i at ATR [4]. The CB-i could interact with a human partner by tracking by its eyes a marker moved by the partner. Although the K computer was not available for such an interactive use, our result demonstrates the feasibility and usefulness of high-throughput computing for real-time simulation of the brain for live interaction with humans.

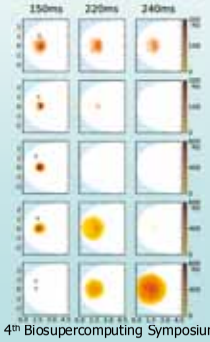
References

1. Moren J, Shibata T, Doya K (2010). Toward a spiking-neuron model of the oculomotor system. International Conference on Simulation of Adaptive Behavior, 2010, Paris. (Springer Lecture Notes in Computer Science 6226, pp. 104-113.)
2. http://www.nest-initiative.org
3. http://software.incf.org/software/music.
4. http://www.cns.atr.jp/bri/en/robot/cb-i/

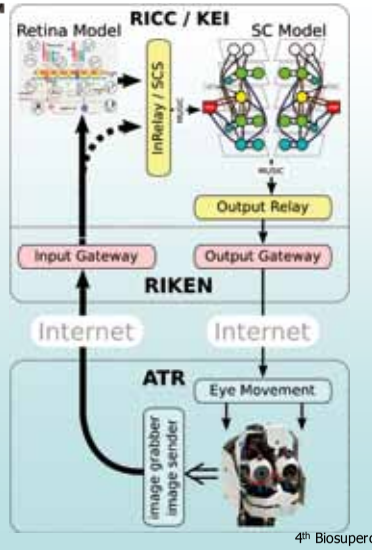
Closing the Loop: Simulation of the Whole Sensory-Motor Neural Network in Action

Kenji Doya

Okinawa Institute of Science and Technology



Goal of the Project



- **Simulate the whole sensorimotor loop**
  - oculomotor network
- **Connect a super computer with a robot**
  - live human interaction

Outline

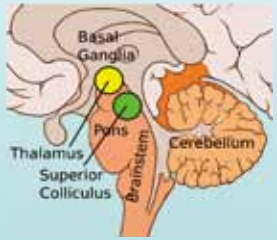
- **Retina model**
- **Superior colliculus model**
- **Off- line simulation**
- **On- line simulation**

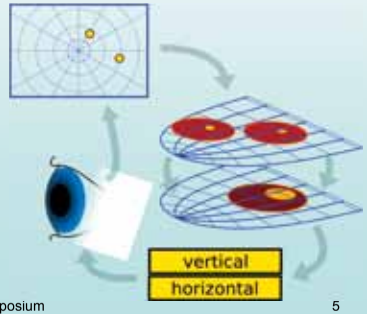
Retina Model

- **by Usui Group at RIKEN BSI**
  - eye optics
  - photo receptors
  - interneurons
  - ganglion cells

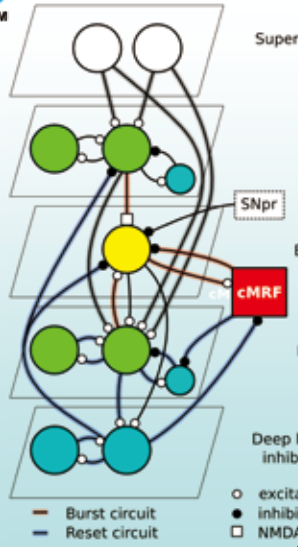
Superior Colliculus (SC)

- **The center of saccadic eye movement**
  - superficial layer : sensory input processing
  - intermediate layer: motor pattern generation
    - retinotopic coordinate
    - spreading activation





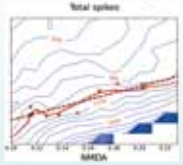
Anatomy



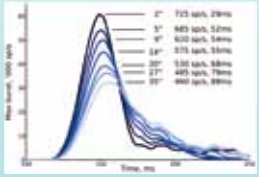
- **Visual input**
  - Superficial SC
  - Quasi Visual layer
- **Burst neurons: output**
  - inhibitory inputs
    - substantia nigra
    - reticular formation
    - SC deep layer
- **Buildup neurons**
  - spreading activity

Physiology

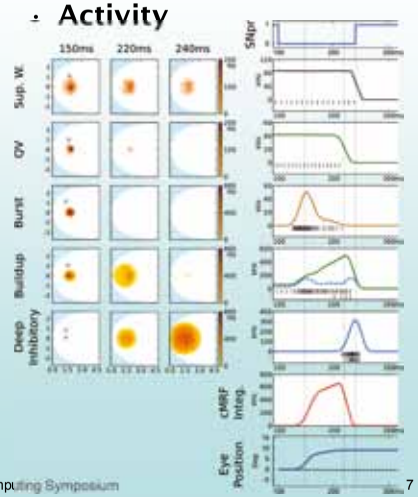
- constant total spikes



- near target: higher rate
- far target: longer duration

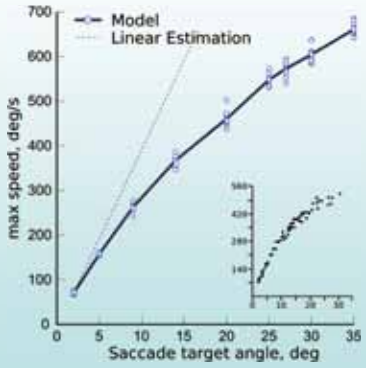


- **Activity**

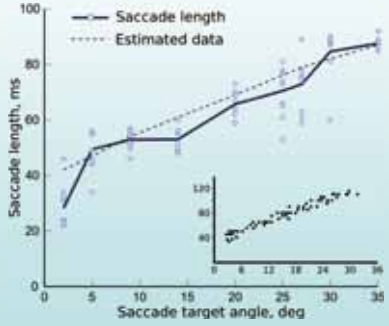


Saccadic *Main Sequence*

- **Peak velocity**

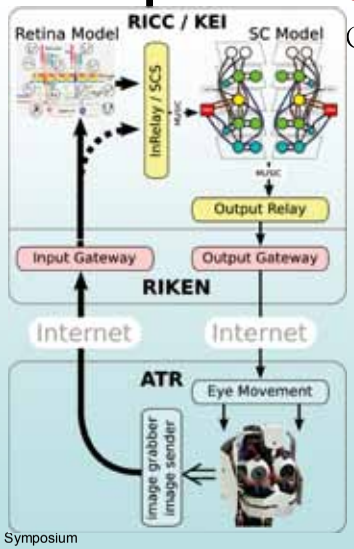


- **Duration**



Closing the Loop

- **MUSIC**
  - multi simulation coordinator
- **Real- time simulation**
  - at RICC
- **Connection over internet**
  - TCP socket
- **Robot eyes**
  - CBi at ATR



In Action

- **Outside view**
- **Robot's eye view**

Conclusion

- **We built the first spiking neuron model of the intermediate SC based on experimental data.**
- **The model reproduced:**
  - burst neuron spike profiles
  - buildup neuron spreading activity
  - saccade main sequence
- **The model predicts:**
  - spreading activity to stop the eye accurately
- **Real- time simulation of a robot's brain**
  - enables interactive testing of the model

Brain-scale neuronal network simulations on K

Markus Diesmann

Professor, Director of Institute of Neuroscience and Medicine (INM-6) Computational and Systems Neuroscience, Juelich Research Centre (Germany)

Profile:

Markus Diesmann received his Diploma and PhD in Physics from the University of Bochum (Germany) in 1994 and 2002. The PhD work was conducted at the Weizmann Institute of Science, Rehovot (Israel) and continued at the University of Freiburg (Germany). In 1999 he joined the Max-Planck-Institute for Dynamics and Self-Organization, Goettingen (Germany) for a staff position. From 2004 to 2006 Markus Diesmann was appointed Junior professor for Computational Neurophysics at the University of Freiburg. In September 2006 he moved to the RIKEN Brain Science Institute, Wako (Japan) for the position of a Unit and later Team Leader. Since March 2011 Markus Diesmann is Professor for Computational Neuroscience at the Medical Faculty of RWTH Aachen University (Germany) and serves as the director of the Institute of Neuroscience and Medicine (INM-6) Computational and Systems Neuroscience, Juelich Research Centre (Germany). His interests include the correlation structure of cortical networks and large-scale simulations.

www.csn.fz-juelich.de

www.nest-initiative.org

Abstract

1. Introduction

The human brain comprises about 10^{11} neurons that are sparsely and specifically connected by about 10,000 outgoing synapses each, mediating electrical pulses (spikes). In computational neuroscience, the bottom-up approach starts from a mathematical description of the single neurons and their interactions in order to investigate the emergent network dynamics [1]. The NEST simulator [2] is tailored to this resolution. Individual neurons with a single or few compartments are represented as small systems of differential equations, which interact by δ-impulses to form coupled networks of natural size and complexity. For the neuron models frequently used in the field, the time evolution of the (essentially linear [3]) dynamical equations can often be integrated exactly, requiring only a few floating point operations per neuron to implement precise update schemes [4]. The mammalian cortex, which carries out higher brain functions, is of particular interest. A network of about 10^5 neurons, called the local cortical microcircuit, is required to represent the layer- and cell-type specific connectivity with all local synapses and can be interpreted as an elementary unit of the cortex. Fundamental dynamic properties, like the layer-dependent firing rate are explained at this scale [5]. The top-down approach starts from an abstract description of a particular brain function and investigates how this function is implemented at the level of neurons and synapses (e.g. [6]). The functional circuits of the cortex typically involve several brain areas [7]. The well-studied visual cortex of primates comprises on the order of 100 million neurons, hierarchically organized in areas that can exceed 10 million neurons and 100 billion synapses. Previously, simulations were constrained to models of the local cortical microcircuit or non-spiking macroscopic models. With the technological advances of recent years, the representation of the full multi-scale connectivity of the brain has come within reach. The memory demands of such simulations [8] are only met by distributed simulation software and supercomputers, such as the K computer in Kobe.

2. NEST on peta-scale supercomputers

NEST is subject to incremental and iterative development driven by the needs of the neuroscience community. These efforts are coordinated by the NEST Initiative (www.nest-initative.org), a non-profit organization. The storage of synaptic connections of brain-scale networks dominates the memory consumption of spiking network simulations, because synapses outnumber neurons by a factor of 10^4. Only a parallel computer offers the required memory. In NEST the synapses are stored on the machine where the target cell is located. This scheme keeps the communication load between processors low, as only the identifiers of neurons that emitted a spike need to be communicated to the other machines, employing collective MPI communication.

The kick-off meeting of the Brain and Neural Systems Team (BNT) took place on November 10th, 2008, at the RIKEN Marunouchi office in Tokyo. First trials of the simulation kernel developed for small clusters (2nd generation kernel, 2g) executed on the BG/L at RIKEN Wako scaled to 1024 processors for a network of 10^5 neurons. At the 3rd BNT meeting October 6th, 2009, we presented scaling data for 10^6 neurons and up to 32,768 core of the JUGENE BG/P computer at Juelich. These data exposed memory consumption as the limiting factor. To investigate the issue systematically, we developed a mathematical model which describes the memory usage per process as a function of the total number of processors and network size [8]. Guided by the model, we identified the data structures dominating the memory usage of NEST. On supercomputers, the 10^4 or more compute nodes outnumber the synapses per neuron, such that on average each node stores only one synaptic target per source neuron. Moreover, only a small fraction of the total number of neurons is local to a given node. By February 2010 we had defined a roadmap for the K computer where along 3 milestones network size increases from 10^6 (1) to 10^8 (2) and above (3). On September 28th, 2010, just 2 days before the 5th BNT meeting, the simulation language interpreter compiled on a prototype machine. On November 12th 2010 the complete NEST compiled

for the first time as reported at the 6th meeting on April 18th of the following year. On May 6th, 2011, NEST compiled on the K test system and finally on September 1st on K as reported at the 7th BNT meeting on the 29th of the same month. In the 3g kernel, we employ data structures which account for the sparseness of synapses and enable the efficient storage of information about non-local neurons [8]. With these improvements to the fundamental data structures the second milestone came into sight at the 8th meeting on March 13th, 2012, and was reached in May [9] utilizing just above 12,288 compute nodes which is less than 14 per cent of K. The simulation of realistic neuronal networks requires the representation of different forms of synaptic plasticity [10,11]. The outgoing synapses of a given neuron are stored in a set of homogeneous data containers. Distinguishing different synapse types therefore requires an intermediate vector-like data structure. However, this distinction is inefficient for computers with 10^4 nodes, where typically for a source neuron only a single synapse is stored. In the second step of our redesign we therefore built an adaptive framework which chooses the optimal container from a predefined set depending on number and type of the synapses. The set of possible containers is created by recursive template meta-programming [12]. In the third step, we improved the memory layout of the synapse objects. When a connection is established, it is checked whether the involved neurons and synapse types match. The improved handshake mechanism does not require a virtual function table pointer in the synapse. In addition the alignment of synaptic parameters was improved. The last two steps of the redesign are combined in the 4g kernel presented at the 9th BNT meeting on September 25th. Fig. 1 shows the network size which just fits on a given number of compute nodes and the runtime for the corresponding network simulation. The 4g kernel enables simulations of 10^9 neurons on K. The increase of runtime with increasing number of nodes is due to the collective communication scheme. Good scaling of network setup and simulation phase are achieved with a hybrid code which combines fine grained parallelism employing OpenMP threads and distributed parallelism based on MPI. The 3g and

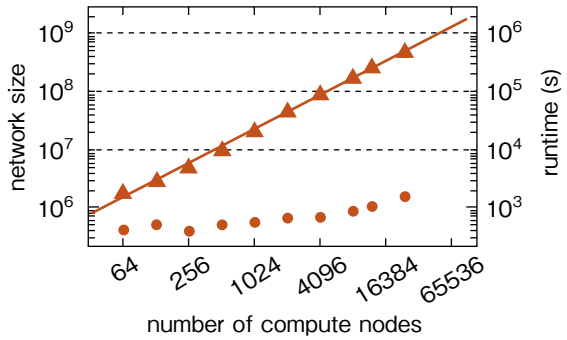


Fig 1. Maximal network size (triangles) and simulation time (dots) as a function of the number of employed nodes for the 4g kernel.

4g kernels do not compromise on generality of NEST; functionality and user interface remained the same [13,14].

Acknowledgements

Partly supported by early access to K at RIKEN AICS, Next-Generation Supercomputer Project of MEXT, Helmholtz Association: HASB and portfolio theme SMHB, JARA and EU Grant 269921 (BrainScaleS).

References

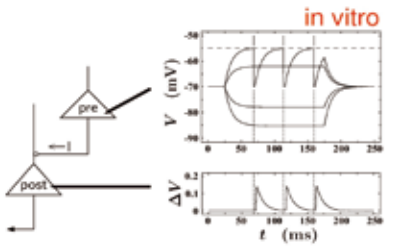
1 Lansner and Diesmann (2012) Chap. 10 In Computational Systems Biology, Le Novére ed., Springer
2 Gewaltig and Diesmann (2007) Scholarpedia, 2(4):1430
3 Morrison et al. (2008) Biol. Cybern. 98:459-78
4 Hanuschkin et al. (2010) Front. Neuroinform. 4:113
5 Potjans et al. (2012) Cerebral Cortex, in press
6 Potjans et al. (2011) PLoS Comput Biol 7(5): e1001133
7 Wagatsuma et al. (2011) Front. Comput. Neurosci. 5:31
8 Kunkel et al. (2012) Front. Neuroinform. 5:35
9 Helias et al. (2012) Front. Neuroinform. 6:26
10 Potjans et al. (2010) Front. Comput. Neurosci. 4:141
11 Pfeil et al. (2012) Front. Neuromorph Eng 6:90
12 Abrahams and Gurtovoy (2004) C++ Template Metaprogramming, Addison-Wesley Longman
13 Eppler et al. (2009) Front. Neuroinform. 2:12
14 Gewaltig et al. (2012) Chap. 18 In Computational Systems Biology, Le Novére ed., Springer

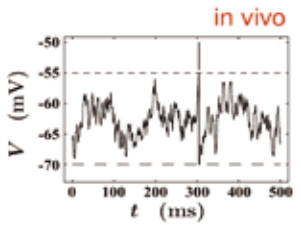
Brain-scale neuronal network simulations on K

Markus DIESMANN
Jülich Research Centre

Brain and Neural Systems Team

Fundamental interactions





- current injection into pre-synaptic neuron causes excursions of membrane potential
- supra-threshold value causes spike transmitted to post-synaptic neuron
- post-synaptic neuron responds with small excursion of membrane potential
- inhibitory neurons (20%) cause negative excursion

- each neuron receives input from up to 10,000 other neurons
- causing large fluctuations of membrane potential
- emission rate of 1 to 10 spikes per second

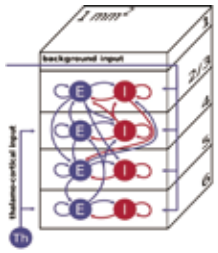
Local cortical microcircuit

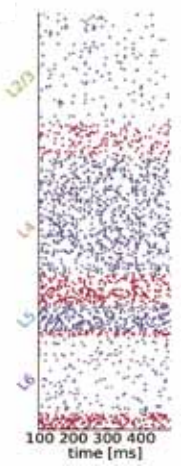
- 10^5 neurons and 10^9 synapses
- excitatory (80%) and inhibitory population (20%) of neurons in each layer

taking into account layer and neuron-type specific connectivity is sufficient to reproduce experimentally observed:

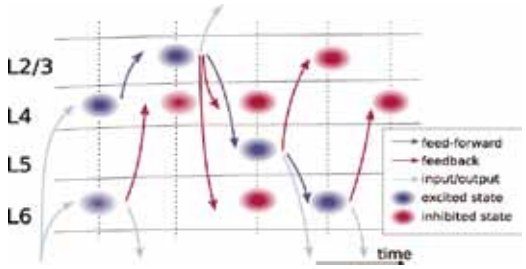
- asynchronous-irregular spiking of neurons
- correct distribution of spike rates across layers

Potjans TC & Diesmann M (2012) The cell-type specific connectivity of the local cortical network explains prominent features of neuronal activity. *Cerebral Cortex*, in press



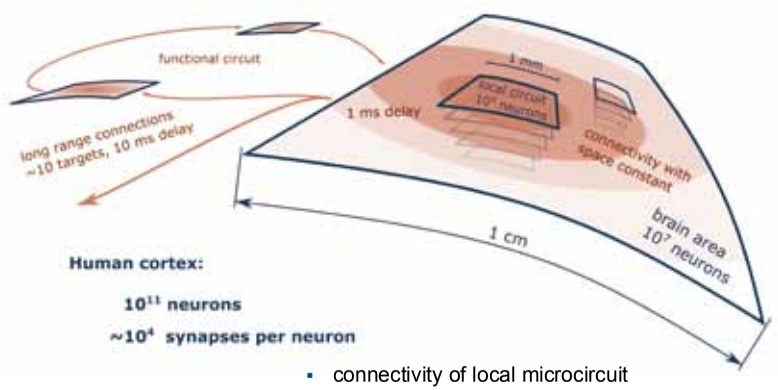


Cortical flow of activity



- Potjans TC & Diesmann M (2012) The cell-type specific connectivity of the local cortical network explains prominent features of neuronal activity. *Cerebral Cortex*, in press
- **building block for further studies:** Wagatsuma N, Potjans TC, Diesmann M and Fukai T (2011) Layer-dependent attentional processing by top-down signals in a visual cortical microcircuit. *Front Comput Neurosci* 5:31

Architecture of the human cortex



At least three levels of organization:

- connectivity of local microcircuit
- within-area connectivity with space constant
- long-range connections between areas

Need for brain-scale networks

- only 50% of incoming synapses of a neuron come from local microcircuit
  - underconstrained model limits predictive power
  - only brain-scale model represents majority of inputs to a neuron
  - self-sustained activity
- brain functions are distributed over many areas
  - brain-scale models required to close functional circuits
- requires work on software technology

Collaboration

Major goals:

systematically publish new simulation technology

produce public releases under GPL

- collaboration of several labs (since 2001)
- teaching in international courses

Scale-up to networks of 10^9 neurons

- scale-up on K guided by 3 milestones
  1. port NEST software to K (2nd generation simulation kernel)
  2. scale of 10^8 neurons (3g simulation kernel)
  3. towards full brain (4g simulation kernel)
- scale 10^8 relevant because:
  - larger than largest area (V1)
  - enable visual cortex model respecting relative sizes
  - larger networks: long delays, sparse macroscopic connectivity
- memory overhead increases with cores
- memory not simulation time limits network size
- use full memory resources: maximum-filling scaling
- analysis based on mathematical model of memory consumption:
  - Kunkel S, Potjans TC, Eppler JM, Plesser HE, Morrison A and Diesmann M (2012) Meeting the memory challenges of brain-scale network simulation. *Front Neuroinform* 5:35

Memory layout of 3g and 4g kernel

3g memory layout

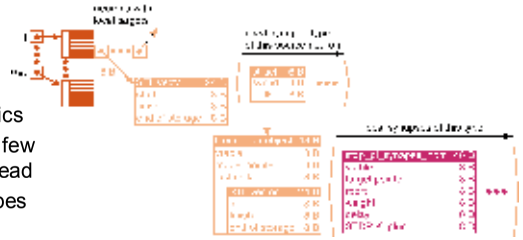
- accounts for sparseness in neuronal and connection data structures

published as:

Helias M, Kunkel S, Masumoto G, Igarashi J, Eppler JM, Ishii S, Fukai T, Morrison A, Diesmann M (2012) **Supercomputers ready for use as discovery machines for neuroscience** *Front Neuroinform* 6:26

4g memory layout

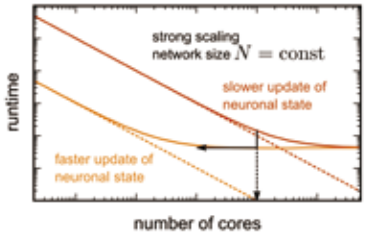
- data structures account for heterogeneity of synaptic dynamics
- for > 10,000 cores, neurons with few local targets cause severe overhead
- novel adaptive data structure copes with short target lists
- not compromising on generality



How to measure scalability

Strong scaling

- fixed size of neuronal network
- ideally: simulation time decreases proportional to number of cores
- saturation due to communication
- **faster element update leads to earlier saturation** (comm. dominates runtime already on fewer cores)
- however, same network can be simulated faster on fewer cores



Faster integration of neuronal dynamics can be accomplished by:

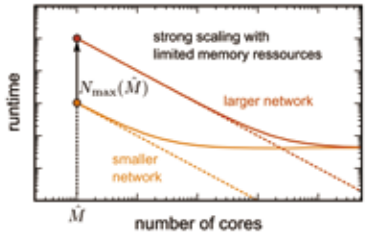
- improved solver
- compromising on accuracy of the solution

How to measure scalability

Strong scaling

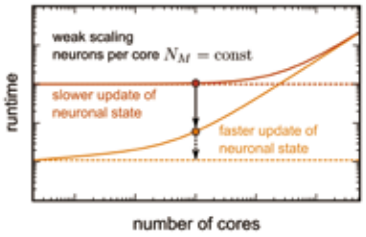
regime of limited memory resources

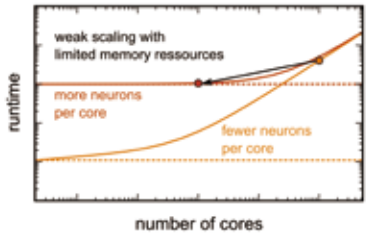
- network just fits on $\hat{M}$ cores
- improving memory consumption
  - larger network on $\hat{M}$ cores
  - higher work-load per core
  - saturation shifted to larger number of cores
  - better scaling



How to measure scalability

Weak scaling



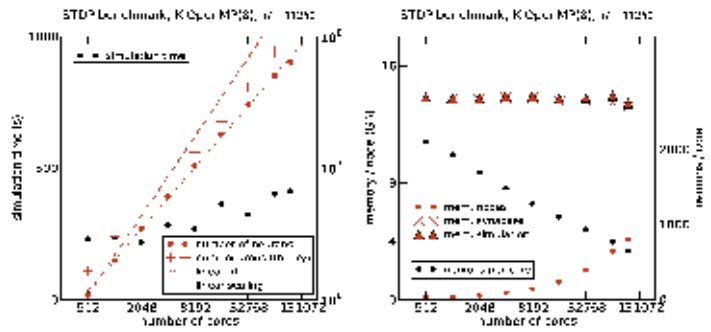


- fixed number of neurons per core
- ideally: constant runtime
- increase due to communication
- **again, faster element update leads to worse scaling**

in case of limited memory resources:

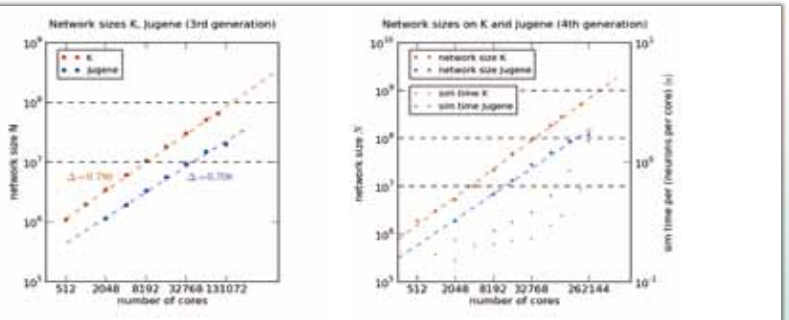
- again, reduced memory consumption improves scaling
- **extreme case: same network can be simulated faster on fewer cores (dots indicate equally sized networks)**

Maximum filling scaling of 3g on K



- neuroscientist interested in maximum use of resources
- hold memory close to maximum available on node (16 GB)
- neurons per core drop
- simulation time increases due to increased communication

Performance of 3g and 4g kernel on Jugene and K



- end of dashed lines indicate machine size
- Jugene largest number of MPI jobs tested 65,536 = 262,144 cores
- K largest number of MPI jobs tested 24,576 = 196,608 cores
- wall clock time 1 biological second of the largest 4g simulation on K: 1520s = 25min

Summary

- 10^8 milestone achievable with 3g kernel (milestone 2)
- 10^9 barrier in sight with 4g kernel (milestone 3)
- no compromise on generality
- short run times enable use of K as discovery machine

- next steps:
  - prepare manuscript on 4g kernel
  - make 4g kernel available to partners
  - concepts for exa-scale machines

Acknowledgements

Moritz Helias

Susanne Kunkel

Jochen M Eppler

Hans Ekkehard Plesser

Marc-Oliver Gewaltig

Abigail Morrison

Shin Ishii

Tomoki Fukai

Jun Igarashi

Makoto Taiji

Gen Masumoto

Yousuke Ohno

Molecular machines and nuclear processes studied by coarse-grained molecular simulations

Shoji Takada

Associate Professor, Department of Biophysics, Graduate School of Science, Kyoto University

Profile:

1990	Graduate School of Science, Kyoto University (Master degree)
1991-1998	Institute of Molecular Science, (Technical officer)
1998-2007	Faculty of Science, Kobe University (Lecturer, Assoc. Prof)
2007-	Graduate School of Science, Kyoto University (Assoc. Prof.)

Abstract

1. Introduction

Living matter is intrinsically hierarchic. In molecular scales, atomic structures made revolution in the field, but a gap between atomic resolution information and cellular biological interest is still large. Experimentally, large fluctuation involved in huge biomolecular systems makes the structural analysis harder. Thus, molecular dynamics simulations can augment experiments hopefully filling a gap between high resolution structural information and cellular biology. Yet, atomic simulation cannot easily deal with long time dynamics. To this end, for long-time simulations of huge systems, coarse-graining is useful.

Given these situation, our purpose is, on the basis of atomic resolution structural data and simulation technology, to

1) systematically construct CG models,
2) develop a software CafeMol that implements the CG models, and,
3) apply them to various biological phenomena at molecular scale important in cellular biology.

1.2. Development of CG models and a software CafeMol

In 2007, we started coding CafeMol, a generic software for CG simulations of biomolecular systems. Dr. Kenzaki primarily coded the major part of the software, and others contributed to write some portions. CafeMol is written in Fortran 90 and is parallelized with openMP and MPI. On K computer, we tested it showing a good scaling up to 98000 cores and 33% efficiency in a single node.

CafeMol uses CG models that has a one- bead-per- amino acid resolution for proteins, and three-bead- per- nucleotide resolution for nucleic acids. Lipid model is still premature. We developed some multiscale protocols to derive CG parameters from the all-atom force field and experimental structural data.

2. Applications to various biological phenomena

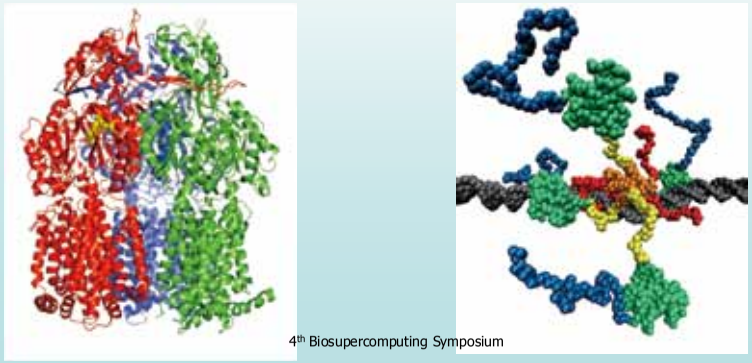
In parallel to the development of CG models and the software CafeMol, we have been applying them to various biological phenomena in molecular scale important in cellular biology. In particular, applications include a few molecular machines, such as multidrug transporter AcrB and molecular motor kinesin, and simulations of nuclear processes, such as nucleosome dynamics and transcription factor p53 search dynamics.

References

1. H. Kenzaki, N. Koga, N. Hori, R. Kanada, W. Li, K. Okazaki, X. Yao, & S. Takada, CafeMol: A coarse-grained biomolecular simulator for simulating proteins at work, ***J Chem Theo Comp***, **7**: 1979-1989, 2011.
2. W. Li, P. G. Wolynes, & S. Takada, Frustration, specific sequence dependence, and nonlinearity in large-amplitude fluctuations of allosteric proteins, ***Proc Nat Acad Sci USA***, **108**: 3504-3509, 2011.
3. W. Li, T. Terakawa, W. Wang, & S. Takada, Energy landscape and multiroute folding of topologically complex proteins adenylate kinase and 2ouf-knot, ***Proc Nat Acad Sci USA***, 2012 in press.
4. N. Hori & S. Takada, Coarse-Grained Structure-Based Model for RNA-Protein Complexes Developed by Fluctuation Matching, ***J Chem Theo Comp***, **8**: 3383-3394, 2012.
5. X. Yao, H. Kenzaki, S. Murakami, & S. Takada, Drug export and allosteric coupling in a multidrug transporter revealed by molecular simulations, ***Nature Comm***, **1**:117 (8pages), 2010.
6. T. Terakawa & S. Takada, Multiscale ensemble modeling of intrinsically disordered proteins: p53 N-terminal domain, ***Biophys J***, **101**: 1450, 2011.
7. T. Terakawa, H. Kenzaki, & S. Takada, p53 searches on DNA by rotation-uncoupled sliding at C-terminal tails and restricted hopping of core domains , ***J. Am Chem. Soc.***, **134**: 14555, 2012.

Molecular machines and nuclear processes studied by coarse-grained molecular simulations

Shoji Takada
Dept Biophysics, Grad School of Science, Kyoto University



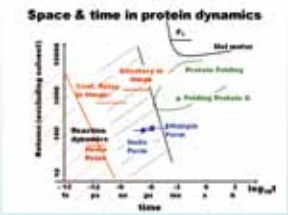
Background & Purpose

- Background

  Living matter is hierarchic

  In molecular scales, atomic structures made revolution in the field, but a gap between atomic resolution and cellular scale is still large. Large fluctuation makes the structural analysis harder.

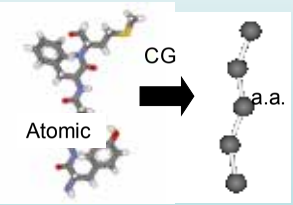
  For long-time simulations of huge systems, coarse-graining is useful.



- Purpose

  On the basis of atomic resolution structural data and simulation technology, we

  1) systematically construct CG models,
  2) develop a software CafeMol that implements the CG models, and,
  3) apply them to various biological phenomena at molecular scale important in cellular biology.



CafeMol

- Summary
  - By CG models, it simulate long-time dynamics of huge biomolecular systems.

- Approach
- Method
  - Clasical molecular dynamics with CG model
- Simulations
  - Time propagation based on Langevin dynamics and others
- Parallel programming
  - Neighbor list, replica exchange
- Computer language and library
  - Fortran90, MPI, OpenMP
- Code release
  - Freely downloadable from http://www.cafemol.org
- Target size
  - Simulations that correspond to milliseconds and longer.

CafeMol : Current status

- Current status of CafeMol
  - Completed

    Implementation of CG models for protein, NA, lipid

    Parallel computation up to 98000 cores on K

    Single node efficiency up to 33% on K
  - To do

    Accurate derivation of and more accuracy in CG models →

    Multiscale algorithm for CG parameter derivation from atomic level
  - Applications

    Multidrug transporter, kinesin,

    nucleosome, transcription factor (p53 etc)
  - Experimental verification

    Multidrug transporter, collaboration with S. Murakami (Tokyo Inst Tech)

    Kinesin, collaboration with M. Tomishige (Univ Tokyo)

Atomic interaction based CG model

Deriving CG para from AA via a multiscale method

AICG1; Li, Wolynes, & Takada PNAS 2011
AICG2;Li, Terakawa, Wang, & Takada PNAS 2012

AICG2 model

$$V = \sum_I k_b (r^I - r_0^I)^2 + \sum_I V_a^I(\theta_I) + \sum_I V_{dih}^I(\varphi_I) + \sum_{J+2\le I\le J+3} \varepsilon_{loc}^{IJ} \exp\left(-\frac{(r^{IJ}-r_0^{IJ})^2}{2w^2}\right) + \sum_{I>J+3}^{native} \varepsilon_{nloc}^{IJ}[5(r_0^{IJ}/r^{IJ})^{12} - 6(r_0^{IJ}/r^{IJ})^{10}] + \sum_{I>J+3}^{non-native} \varepsilon(C/r^{IJ})^{12}$$

Local contact

Flexible local potential
(Statistical potential)

Energy decomposition

pairwise contact energies

Fluctuation matching

rmsf bet. AA and CG

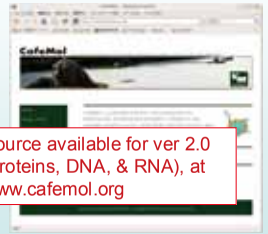
Gohlke et al., 2003, Chu and Voth, 2006 Li et al., 2010,

CafeMol, software for CGMD

Kenzaki
J. Comp. Theo. Chem. 2011

Simulating "protein at work"

Also for nucleic acids, membrane, and their complex



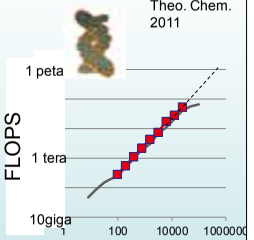
Source available for ver 2.0 (proteins, DNA, & RNA), at www.cafemol.org

Fortran 90, MPI & openMP

Running on K-computer

1-node efficiency 33%

Parallel performance REMD @RICC parallel 99.995%

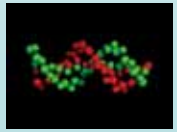


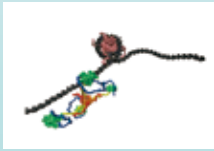
Protein: Structure-based, switching

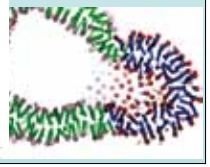
DNA: Structure-based+ P-Chem

Protein-DNA complex

Membrane : P-Chem

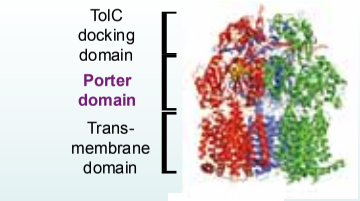




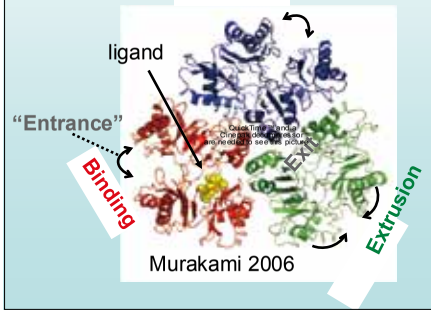


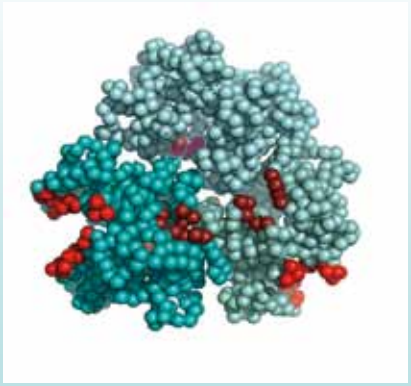
Application 1) Multidrug transporter

Yao



Access

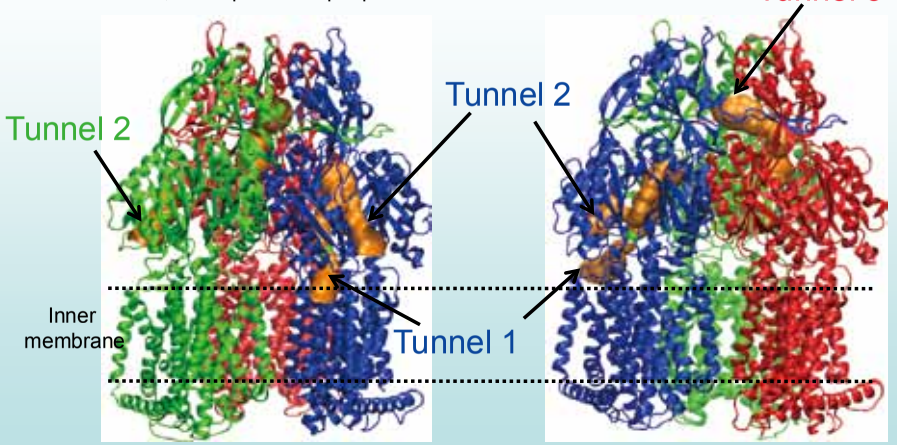




Yao et al, Nature Comm 2010

Drug uptake pathways

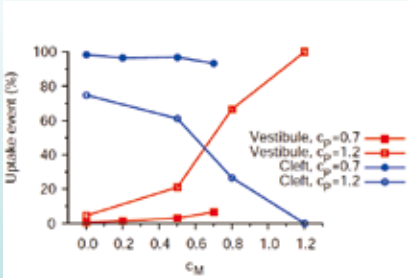
Tunnel 1 , Long path from membrane surface
Tunnel 2 , Short path from periplasm

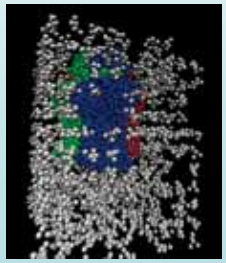


Dynamic simulation of drug uptake and export

- 273 ligand molecules in a 160x160x300 (Å) box
- 300K Langevin
- 100 trajectories
- Switching
  ABE → BEA → EAB

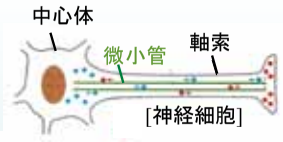
We found
drug-dependent uptake pathways

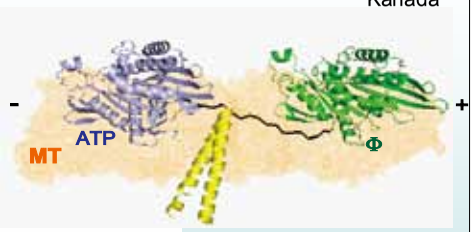


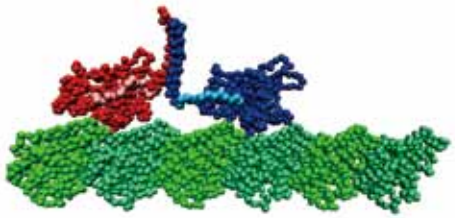


Application 2) Kinesin

Kanada



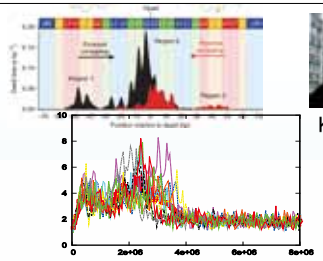




CG MD simulations together with high-resolution modeling of K-MT complex

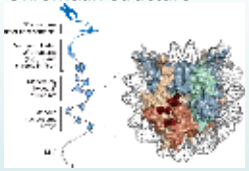
In collaboration with experimentalists (Tomishige, Kikkawa)

Application 3) Nucleosome dynamics

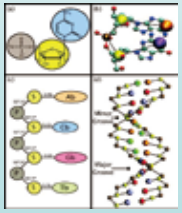


Kenzaki

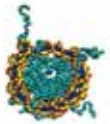
Towards modeling of Chromatin structure



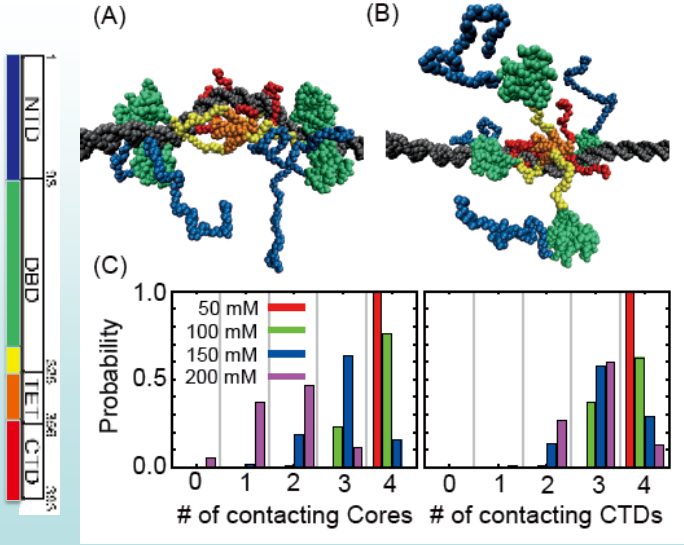
Assembly & disassembly of nucleosome



CG model of DNA
de Pablo et al



Application 4) full-length p53



Terakawa

BPJ 2011
JACS 2012

Biomolecular simulations under cellular environment

Yuji Sugita

Chief Scientist, RIKEN

Profile:

Education and degree:

Ph. D from Kyoto University (March, 1998)

Professional record:

Postdoctoral fellow, RIKEN, April – August, 1998

Research associate, Institute for Molecular Science (IMS), September, 1998 – March, 2002

Lecturer, Institute of Molecular and Cellular Biosciences, the University of Tokyo, April, 2002 – March, 2007

Associate Chief Scientist, RIKEN, April, 2007 – March, 2012

Chief Scientist, RIKEN, April, 2012 - Present

Team Leader (concurrent post), RIKEN Advanced Institute for Computational Science, October, 2010 - Present

Team Leader (concurrent post), RIKEN Quantitative Biology Center, April, 2011 - Present

Abstract

1. Introduction

At the cytoplasm in cells, 70% of the volume fraction is occupied by water molecules and in the rest of the volume, a large number of proteins, RNA, or other metabolites exist. This environment is often called as 'macromolecular crowding' in cell and has been paid a lot of attention by many theoretical and experimental researchers. One of the key questions in this issue is whether protein is stabilized or destabilized in such crowded environment. Most of the previous studies suggest that protein is more stabilized in crowded environments rather than that in dilute solution. However, recent experimental studies suggest that protein stability significantly dependent on the crowding environments. The hydrogen-exchange experiments in NMR spectroscopy shows that protein stability significantly depends on the types of crowding agents: In another experiment, in-cell NMR spectroscopy, Inomata et al. found that ubiquitin is destabilized inside of the cell [Inomata et al. *Nature* (2009)]. These experiments suggest that protein-protein interaction in crowded environment is essentially important and is not fully examined based on the previous theoretical studies in which only volume exclusion effect is considered as the major crowding effect. We therefore performed atomistic molecular dynamics (MD) simulations of proteins under crowded conditions in explicit water.

2. Protein stability changes due to different protein crowders

In the first applications of crowding simulations, we examined the effect of cellular crowding by performing MD simulations of chymotrypsin inhibitor 2 (CI2) in the presence of either lysozyme or bovine serum albumin (BSA) crowder molecules. This is the same protein and crowder proteins in which Pielak et al recently studied by the hydrogen-exchange experiment in NMR spectroscopy (Miklos, A. C.; Sarkar, M.; Wang, Y.; Pielak, G. J. J. Am. Chem. Soc. **2011**, 133, 7116). The simulations confirm a destabilization and significantly slowed diffusion of CI2 in the presence of lysozyme and indicate that this observation is a result of extensive, non-specific protein-protein interactions between CI2 and lysozyme. CI2 interacts much less with BSA crowders corresponding to a weak effect of crowding. Energetic analysis suggests an overall favorable crowding free energy in the presence of lysozyme while weaker interactions with BSA appear to be unfavorable.

3. Hydration and protein-protein interactions in crowded conditions

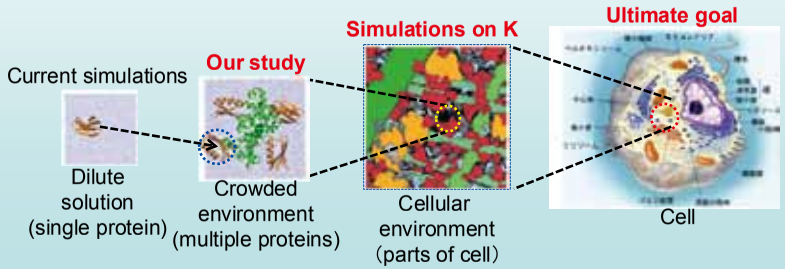
In the next applications of crowding simulations, we performed explicit solvent MD simulations of a series of protein G and protein G/villin systems at different protein concentrations. In this study, we focused on hydration structure and protein stability. Hydration structure was analyzed in terms of radial distribution functions, three-dimensional hydration sites, and preservation of tetrahedral coordination. Analysis of hydration dynamics focused on self-diffusion rates and dielectric constants as a function of crowding. The results show significant changes in both structure and dynamics of water under highly crowded conditions. The structure of water is altered mostly beyond the first solvation shell. Diffusion rates and dielectric constants are significantly reduced following linear trends as a function of crowding reflecting highly constrained water in crowded environments. The reduced dynamics of diffusion is expected to be strongly related to hydrodynamic properties of crowded cellular environments while the reduced dielectric constant under crowded conditions has implications for the stability of biomolecules in crowded environments. The results from this study suggest a prescription for modeling solvation in simulations of cellular environments.

4. Summary and perspectives

The results provided MD simulations with explicit solvent seem to be complementary with experimental data and show new insight on the crowding effect on proteins in cell. Using K computer, we expect to simulate larger and more realistic systems of the cytoplasm to examine crowding and confinement effect on biological molecules.

Biomolecular simulations under cellular environments

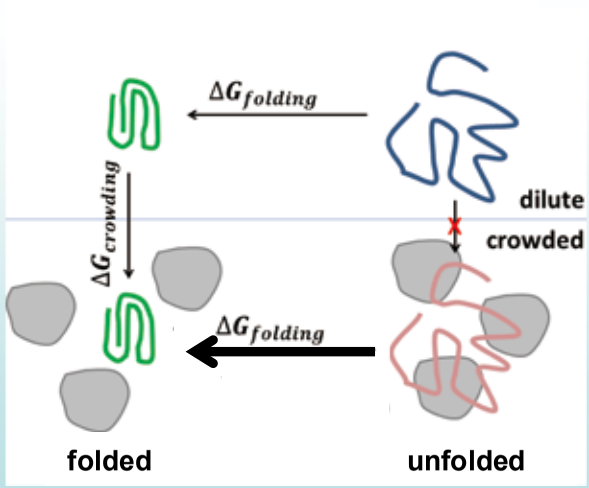
Yuji Sugita
RIKEN



Topics

- **Introduction**
  - How macromolecular crowding affects protein stability and function?
- **Protein stability changes due to different protein crowders**
  - NMR amide proton exchange experiments
  - MD simulations of crowded environments
- **Hydration and protein-protein interactions in crowded conditions**
  - Hydration properties under crowded environments
  - Protein stability changes due to non-specific protein-protein interactions
  - NMR experiments on protein-protein interactions in crowded conditions
- **Towards cellular-scale simulations**

Classical view of crowding

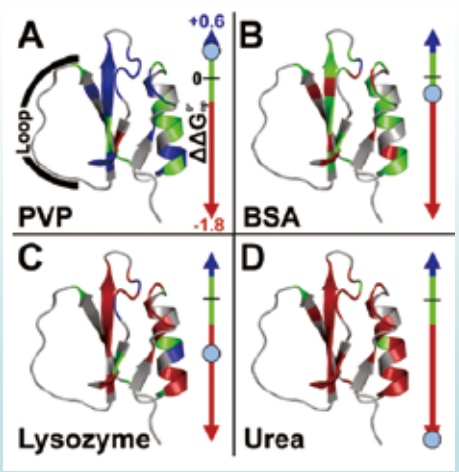


Protein Stability in Crowded Environments

NMR amide proton exchange experiments

More stable

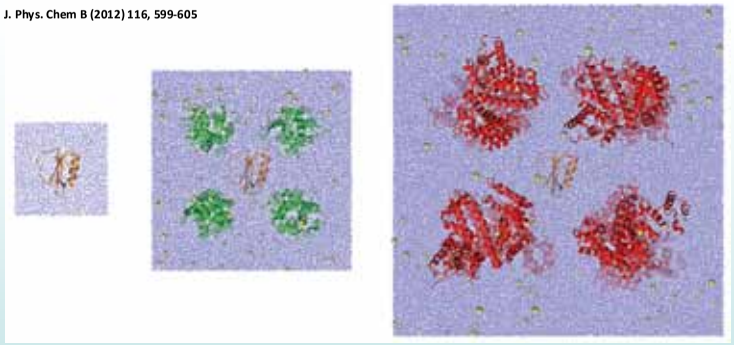
Less stable



A. Miklos, M. Sarkar, Y. Wang, G. Pielak: JACS (2011) 133, 7116-7120

MD simulations of crowded environments

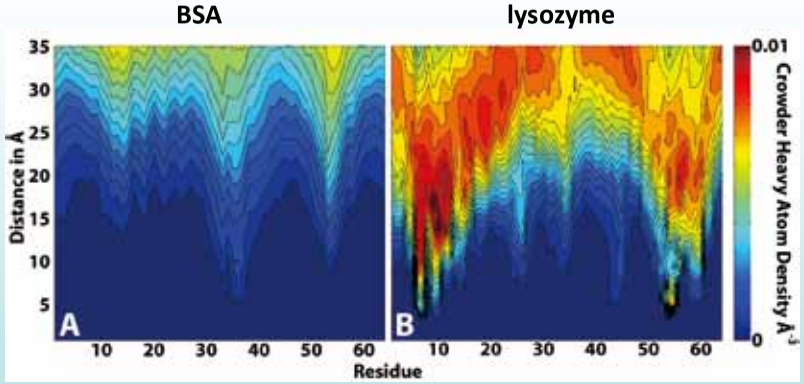
Feig, Sugita: J. Phys. Chem B (2012) 116, 599-605



CI2	CI2 + 8 lysozymes	CI2 + 8 BSAs
18K atoms	184K atoms	835K atoms
6K H_2O	56K H_2O/64 Cl^-	253K H_2O/136 Na^+
infinite dilution	108 g lysozyme/L	104 g BSA/L solution
	7% vol fraction	7% vol. fraction
~160 ns MD	~250 ns MD	~120 ns MD

5

CI2 interaction with crowded proteins

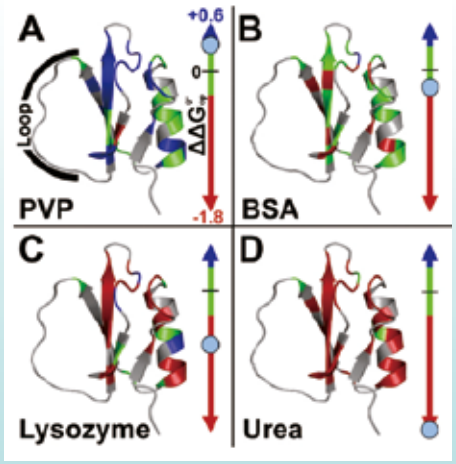


Feig, Sugita: J. Phys. Chem B (2012) 116, 599-605

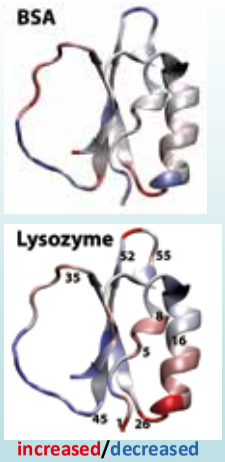
4th Biosupercomputing Symposium 6

Residue fluctuations vs stability

RMSF differences between crowded and dilute solvent



A. Miklos, M. Sarkar, Y. Wang, G. Pielak: JACS (2011) 133, 7116-7120

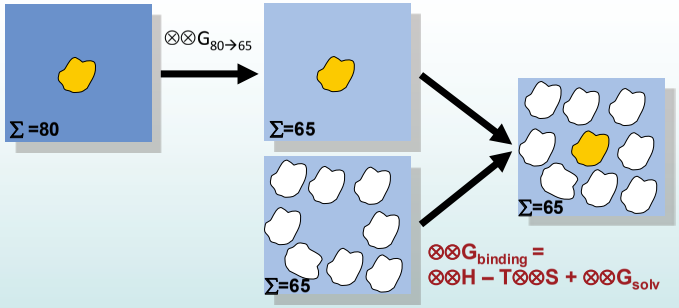


increased/decreased fluctuations in MD

Feig, Sugita: J. Phys. Chem B (2012) 116, 599-605

7

Free Energy of Crowding for CI2 in BSA/Lysozyme



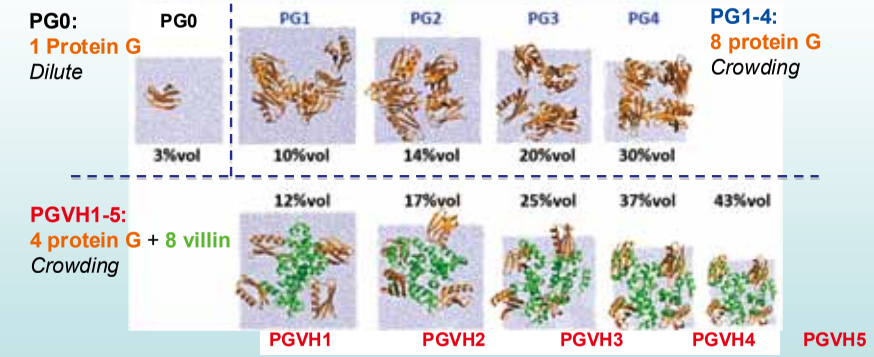
	⊗⊗H	⊗⊗G_{solv}	-T⊗⊗S	⊗⊗$G_{80/65}$	⊗⊗$G_{crowding}$
BSA	-21.3	22.9	0.3	4.2	**6.1**
Lysozyme	-194.0	186.2	0.5	4.3	**-3.0**

[kcal/mol]

Feig, Sugita: J. Phys. Chem B (2012) 116, 599-605

8

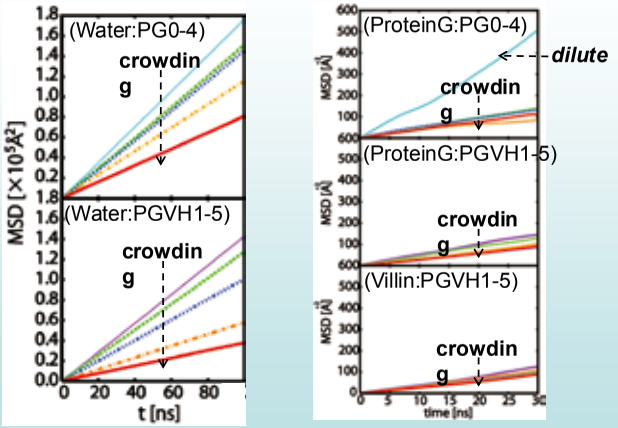
Crowding simulations of two different protein systems



Harada, Sugita, Feig, *J. Am. Chem. Soc.* 2012, 134, 4842-4849 NPT(1bar, 298K, 300ns)

Dynamics of water and proteins

$$MSD(t) = \left\langle | r(t'+t) - r(t') |^2 \right\rangle_{t'}$$

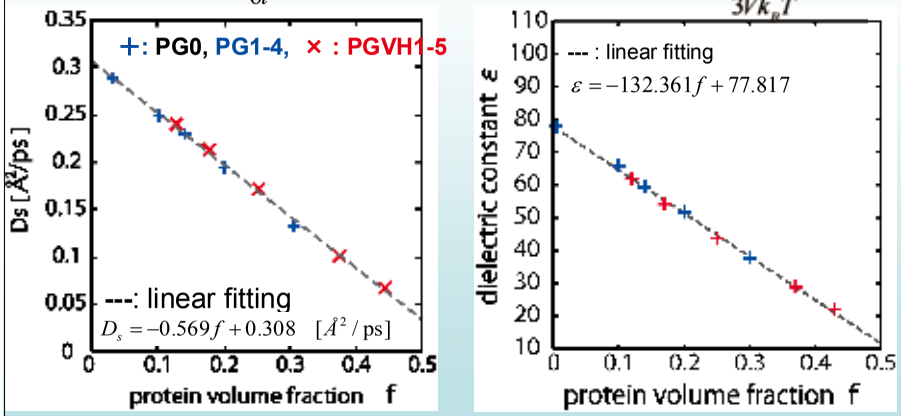


★ In crowded environment, diffusion of water and protein *significantly slow down*

Universal behaviors of diffusion coefficient and dielectric constant

$$D_s = \lim_{t\to\infty} \frac{1}{6t} MSD(t)$$

$$\varepsilon = 1 + 4\pi \frac{\left\langle \vec{M}^2_{water} \right\rangle - \left\langle \vec{M}_{water} \right\rangle^2}{3Vk_BT}$$



Summary and perspectives

- Crowding strongly altered *hydration structure* with protein volume fraction over 30%vol, reducing the number of *bulk water keeping that of the first solvation shell*.
- Decreases in *diffusion coefficient* and *dielectric constant* were observed in crowded systems
  - Reduction of diffusion coefficient
    → *Modulations of rates of biochemical reactions*
  - Reduction of dielectric constant
    → *Thermodynamic properties*
    (directly protein-protein interactions ↑)

Acknowledgement to collaborators

This work has been carried out in collaboration with Prof. Michael Feig (Michigan State University) and Dr. Ryuhei Harada (RIKEN AICS).

Supercomputing for Next-Generation Cancer Research

Satoru Miyano

Professor of Human Genome Center, The Institute of Medical Science, The University of Tokyo

Profile:

Satoru Miyano, Ph.D., is a Professor of Human Genome Center, The Institute of Medical Science, The University of Tokyo. He received the B.S., M.S. and Ph.D. all in Mathematics from Kyushu University, Japan, in 1977, 1979 and 1984, respectively. He joined Human Genome Center in 1996. His research mission is to create computational strategy for systems biology and medicine towards translational bioinformatics.

Abstract

Computational systems biology is gradually seeping in cancer research. We present a challenge for uncovering systems in cancer by supercomputer. SiGN-L1 (NetworkProfiler) (based on L1-regularizaon) is a method that will exhibit how gene networks vary from patient to patient according to a modulator, which is any score representing characteristics of cells. We defined an EMT (epithelial-mesenchymal transition) modulator and analyzed gene expression profiles of 762 cancer cell lines. The computation took 3 weeks on 1024 CPU cores. Network analysis unraveled global changes of networks with 13,508 genes of different EMT levels. By focusing on E-cadherin, 24 genes were predicted as its regulator, of which 12 have been reported in the literature. A novel EMT regulator KLF5 was also discovered in this study. We also analyzed Erlotinib resistant networks using 160 NSCLCs with GI50 as a modulator. Hubness analysis exhibited that NKX2-1/TTF-1 is the key gene for Erlotinib resistance in NSCLCs. Our microRNA/mRNA gene network analysis with Bayesian network method called SiGN also revealed subnetworks with hub genes (including NKX2-1/TTF-1) that may switch cancer survival. The supercomputer was also applied for modeling dynamics in cancer cells from time-course gene expression profiles and revealed dynamic network changes against anti-cancer drugs and network differences between drug-sensitive and drug-resistant cancer cells. For dynamic system modeling, we devised a state space model (SSM) with dimension reduction method for reverse-engineering gene networks from time-course data, with which we can view their dynamic changes over time by simulation. We succeeded in computing a gene network with prediction ability focused on 1500 genes from data of about 20 time-points. We applied this SSM model to human normal lung cell treated with (case)/without (control) Gefitinib, and we identified genes under differential regulations between case and control. This signature of genes was used to predict prognosis for lung cancer patients and showed a good performance for survival prediction.

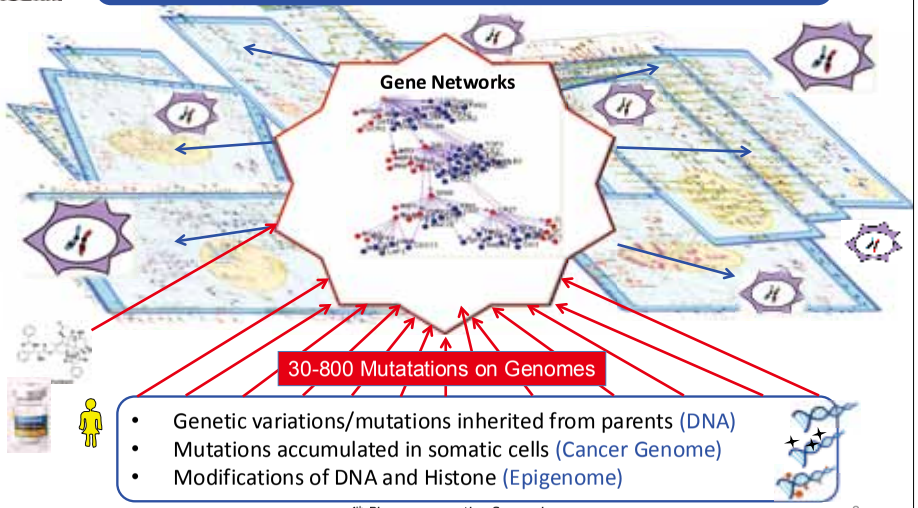
References

1. Shimamura T, Imoto S, Shimada Y, Hosono Y, Niida A, Nagasaki M, Yamaguchi R, Takahashi T, Miyano S. A novel network profiling analysis reveals system changes in epithelial-mesenchymal transition. PLoS ONE. 6(6): e20804, 2011.
2. Tamada Y, Imoto S, Araki H, Nagasaki M, Print C, Charnock-Jones DS, Miyano S. Estimating genome-wide gene networks using nonparametric Bayesian network models on massively parallel computers. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 8(3): 683 - 697, 2011.
3. Tamada Y, Yamaguchi R, Imoto S, Hirose O, Yoshida R, Nagasaki M, Miyano S. SiGN-SSM: open source parallel software for estimating gene networks with state space models. Bioinformatics. 27: 1172-1173, 2011.
4. Yamauchi M, Yamaguchi R, Nakata A, Kohno T, Nagasaki M, Shimamura T, Imoto S, Saito A, Ueno K, Hatanaka Y, Yoshida R, Higuchi T, Nomura M, Beer DG, Yokota J, Miyano S, Gotoh N. Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage I lung adenocarcinoma. PLoS ONE. 7(9): e43923, 2012.
5. Yamaguchi R, Imoto S, Yamauchi M, Nagasaki M, Yoshida R, Shimamura, T, Hatanaka Y, Ueno K, Higuchi T, Gotoh N, Miyano S. Predicting differences in gene regulatory systems by state space models. Genome Informatics. 21:101-113, 2008.

Supercomputing for Next-Generation Cancer Research

Satoru MIYANO
Human Genome Center
The Institute of Medical Science, The University of Tokyo

These variations affect malignancy of cancer, therapeutic response/effect, and adverse reaction by anticancer drugs



- Genetic variations/mutations inherited from parents (DNA)
- Mutations accumulated in somatic cells (Cancer Genome)
- Modifications of DNA and Histone (Epigenome)

SiNG-L1

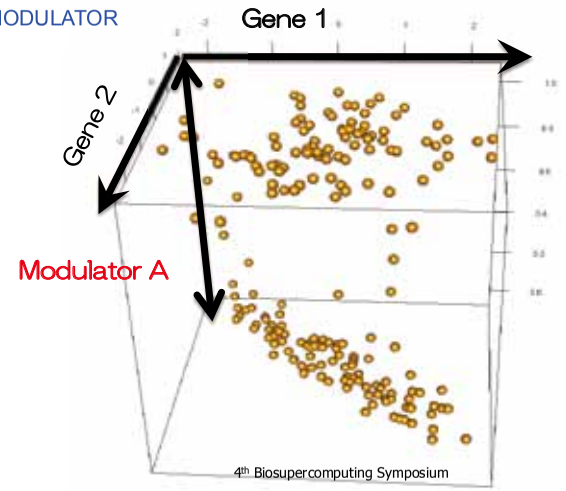
NetworkProfiler

Modulator-dependent Graphical Model Using Structural EQ

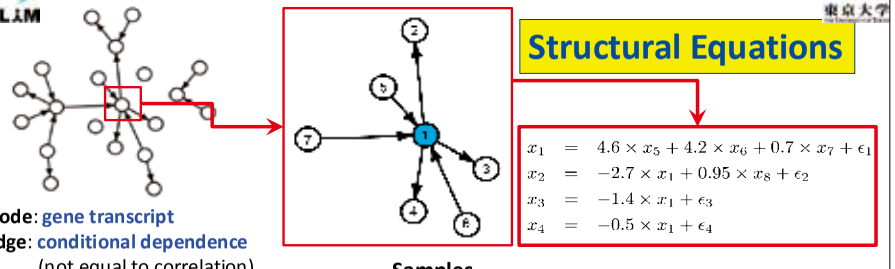
Q: How is "My" gene-gene causal network structured in My Cancer?

Change can be found if we define some MODULATOR and look data along it

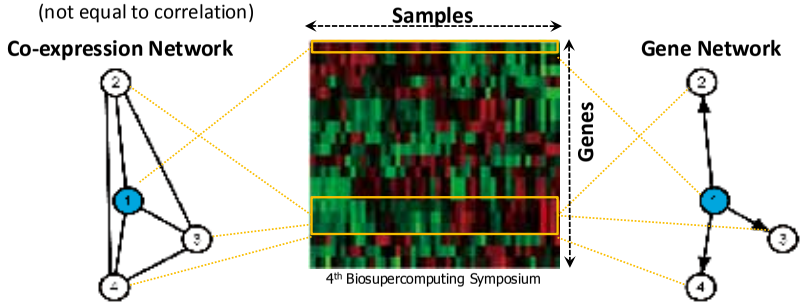
Role of MODULATOR



Modeling with Structural Equation

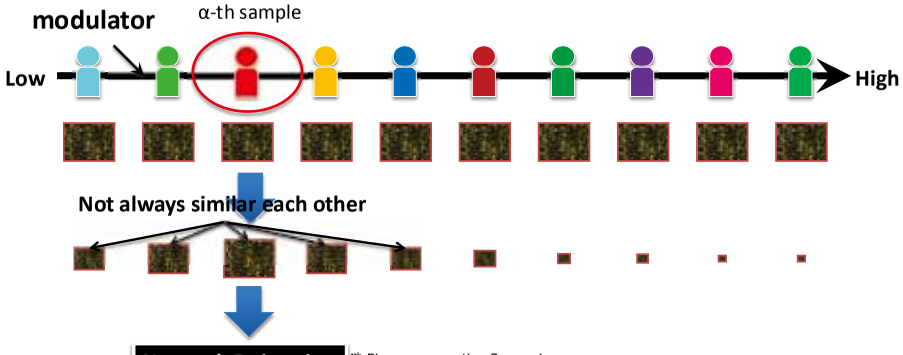


Node: gene transcript
Edge: conditional dependence
(not equal to correlation)



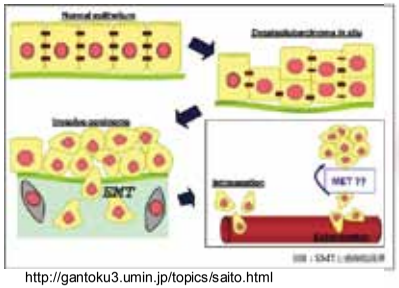
Estimate coefficients for α-th sample $\beta_{jk\alpha}$

Difficulty: Data for α-th sample is only **one**!!
Idea for solution: Data integration by **sample weighting**



Epithelial-Mesenchymal Transition (EMT)

- Key developmental remodeling program, where cells alternate between Epithelial-like and Mesenchymal-like phenotypes
- Relate to tumor grade, metastasis, drug resistance



http://gantoku3.umin.jp/topics/saito.html

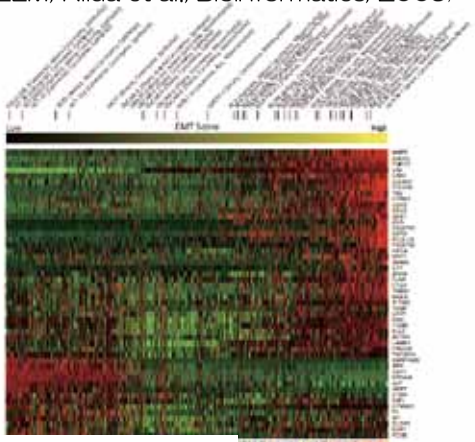
?
Some gene inducing EMT have been known. However, its mechanism is still not well understood.

Modulator for EMT

We selected coherent 50 genes from 122 EMT signature genes to define the modulator for EMT (EEM, Niida et al., Bioinformatics, 2009)

Signature-based hidden modulator extraction algorithm

1. Selected 122 genes labeled "EMT UP", "EMT DN", "JECHLINGER EMT UP", and "JECHLINGER EMT DN"from Molecular Signatures Database v2.5 ([6]; http://www.broadinstitute.org/gsea/msigdb/index.jsp).
2. Then, narrowed the set to 50 functionally coherent genes with $p < 10^{-5}$ by using the extraction of expression module (EEM) .
3. Computed the first principal component of these 50 genes as hidden values of the EMT-related modulator



Run Supercomputer for 3 Months!

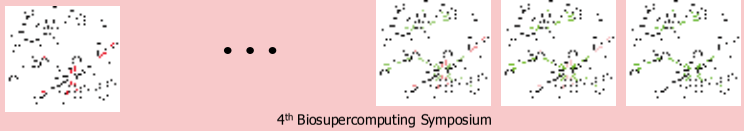
Input

- Transcriptome data of 762 cancer cells focused on 13,508 genes
  13,006 mRNAs + 581 miRNAs
- EMT Modulator

Output

762 gene networks with 13,508 genes exhibited how networks structurally change from low to high EMT modulator score

Low ⟵ ⟶ High



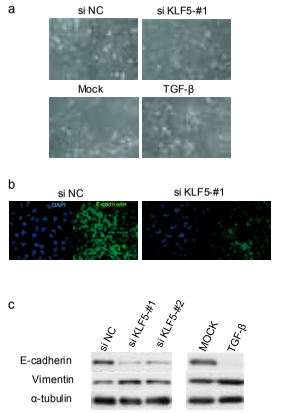
4th Biosupercomputing Symposium

EMT Regulators Predicted by Supercomputer
A New Regulator KLF5 of EMT

- Krueppel-like factor 5 (KLF5) from a list of the remaining candidate regulators and conducted *in vitro* validation experiments. As a result, we found that knockdown of KLF5 by siRNA significantly decreased E-cadherin expression and induced morphological changes characteristic of EMT (Validated by Takahashi Lab, Nagoya University)
- These 12 highlighted of genes are known to have regulatory effect among EMT in the literature.

regulator	type	regulatory effect change	Evidence
ZEB1	A		Nat Cell Biol 10(5): 593-601, 2008
GRHL2	A		Cancer Res 72(9):2440-53, 2012
ZEB1 (MPC)	I		
LSR	I		
miR-200c	A		Nat Cell Biol 10(5): 593-601, 2008
KLF4	A		J Biol Chem 285(22):16854-63, 2010
OVOL2	A		
miR-200b	A		Nat Cell Biol 10(5): 593-601, 2008
FOXA2	A		Cancer Res 70(5):2115-25, 2010
TCF4 (E2-2)	I		J Cell Sci 122(Pt 7): 1014-24, 2009
ELF3	A		
ZNF217	A		
MYC	A		
KLF5	A		PLoS ONE 6(6): e20804, 2011
miR-192	A		FEBS J 274(3): 1932-7, 2007
FOXA1	A		Cancer Res 70(5):2115-25, 2010
ZNF165	A		
NKX2-1	A		Cancer Res 69(7):2783-91, 2009
WNT5B	A		
TFF3	A		
ZEB2 (SIP1)	I		Oncogene 24(14):2375-85, 2005
TRIM29	I		
SNAI1	I		Cancer Res 62(6): 1613-8, 2002

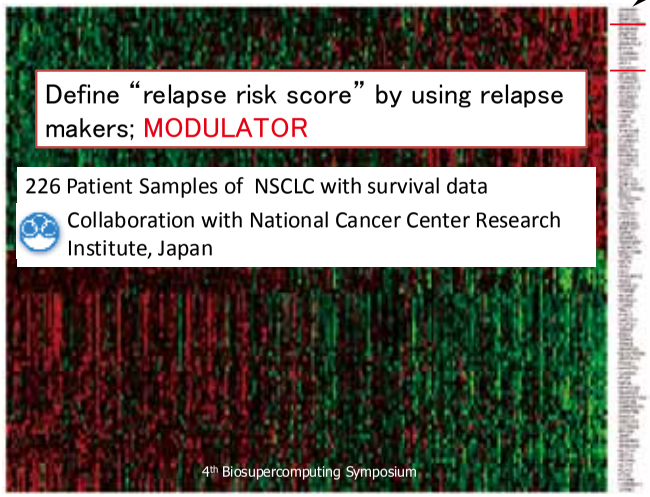
Shimamura T, Imoto S, Shimada Y, Hosono Y, Niida A, Nagasaki M, Yamaguchi R, Takahashi T, Miyano S. A novel network profiling analysis reveals system changes in epithelial-mesenchymal transition. PLoS ONE. 6(6): e20804, 2011.



4th Biosupercomputing Symposium

Relapse Risk

low ⟵ ⟶ high



Define "relapse risk score" by using relapse makers; MODULATOR

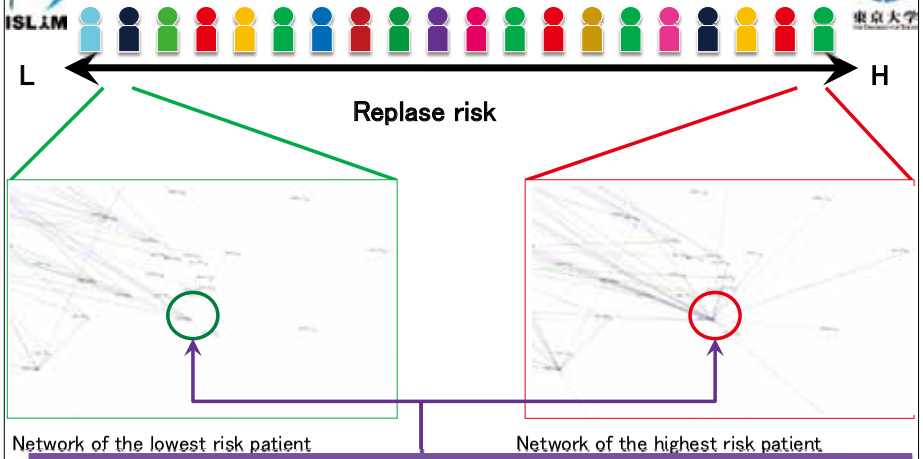
226 Patient Samples of NSCLC with survival data

Collaboration with National Cancer Center Research Institute, Japan

4th Biosupercomputing Symposium

11

Compute each patient network



CTGF (Connective Tissue Growth Factor)

CTGF has been shown to be associated with tumor development and progression. CTGF may regulate cancer cell migration, invasion, angiogenesis, and anoikis.

4th Biosupercomputing Symposium

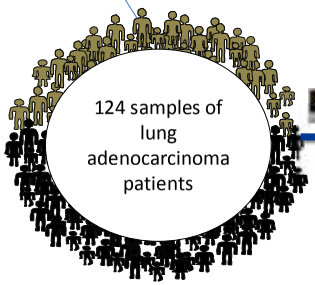
SiGN-BN

Bayesian Network and Nonparametric Regression applied to a large set of lung cancer patient gene expression profiles

Finding genes that "switch" good/poor prognosis in lung cancer

Collaboration with Takashi Takahashi (Nagoya U)

Systems Biology approach unraveled key hub genes that switch good/poor prognosis in lung adenocarcinoma

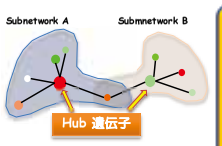


Nonlinear Bayesian Network

Supercomputer

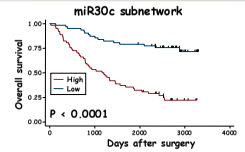
SiGN-BN

Gene Network of Lung Adenocarcinoma



Functional Analysis of Hub Genes

Survival Data



Searching for subnetworks with smaller p-values in Kaplan-Meyer diagram survival prediction and hub genes

14 subnetworks and their hub genes which are significantly corelated with survival /death

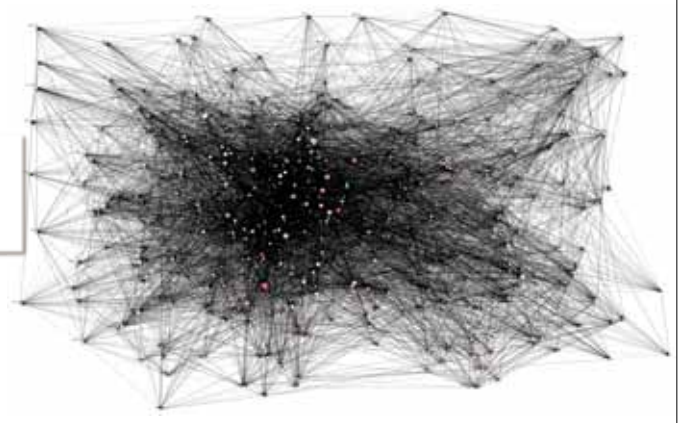
Cancer Biology

Computatioanl Systems Biology

Gene Network of 400 mRNA+32 microRNA Computed from 124 Patient Expression Profiles

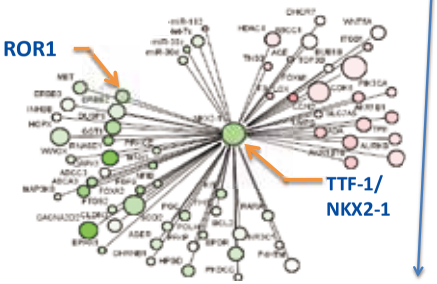
SiGN
- **Nonlinear Bayesian Network Estimation Software**

http://sign.hgc.jp/signbn/

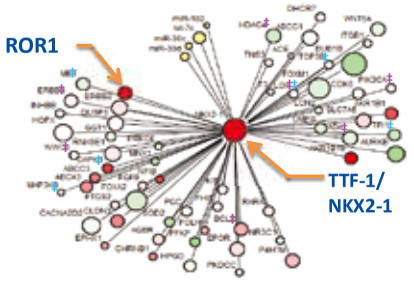


NKX2-1 and ROR1 appear clearly on the network and shows the consistency and new insights

Inferred subnetwork of TTF-1/NKX2-1 with good prognosis



After introducing TTF1/NKX2-1



Expression ratio
- High in poor prognosis
- Low in poor prognosis

NKX2-1/TITF1/TTF-1-Induced ROR1 Is Required to Sustain EGFR Survival Signaling in Lung Adenocarcinoma. ***Cancer Cell*** **2012** by Takahashi Lab (Nagoya U)

Expression ratio (TTF-1-intorudced / Vector control)
- **High in TTF-1/NKX2-1-introduced**
- **Low in TTF-1/NKX2-1-introduced**
- ○: N.A. (miRNA) ‡ : Positive binding in E19.5 mouse lung ChIP-chip assay data (Tagne et al., PLoS ONE, 2012)

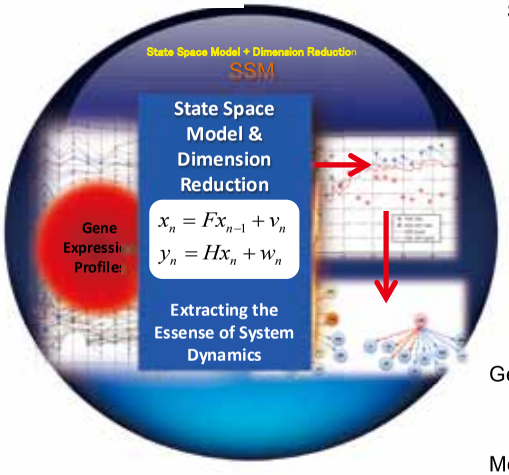
SiGN-SSM

State Space Model for Inferring Transcriptional Module Networks from Time-Course Gene Expression Data

Discovery of hub genes that "switch" survival/death for lung cancer

4th Biosupercomputing Symposium

Modeling Cancer Systems with State Space Model



State Space Model:

$x_n = Fx_{n-1} + v_n \in R^k$ System Model

$y_n = Hx_n + w_n \in R^p$ Observation Model

High-Dimensional Short Time Series Data:

$$Y = \{y_1, \cdots, y_{N_{\text{obs}}}\} \qquad N_{\text{obs}} << p \approx 10^3$$

System Estimation with Dimension Reduction:

$$\dim(x_n) = k < \dim(y_n) = p$$

$$L(\theta) = \int p(Y, X \mid \theta) dX$$

$$H^T R^{-1} H = \Lambda = \text{diag}(\lambda_1, \cdots, \lambda_k)$$

Gene Expression Prediction:

$$y_{n|n-1} = H \int x_n p(x_n \mid y_1, \cdots, y_{n-1}, \theta) dx_n$$

Module-Based Gene Network Estimation:

$$R^{-1/2}(y_n - w_n) = \Psi R^{-1/2}(y_{n-1} - w_{n-1}) + R^{-1/2} H v_n$$

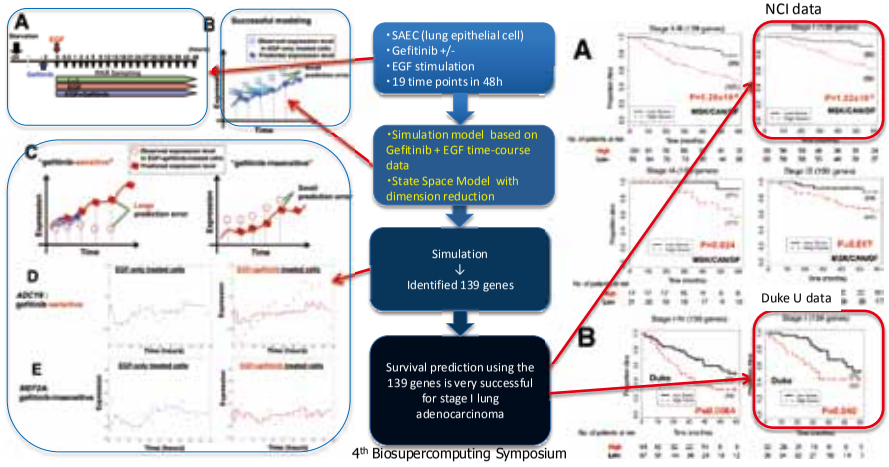
4th Biosupercomputing Symposium

18

State Space Model (SSM) Approach Captured Gefitinib Response Systems

Yamauchi M*, Yamaguchi R*, Nakata A*, Kohno T*, Nagasaki M, Shimamura T, Imoto S, Saito A, Ueno K, Hatanaka Y, Yoshida R, Higuchi T, Nomura M, Beer DG, Yokota J, Miyano S, Gotoh N. (2012) Epidermal Growth Factor Receptor Tyrosine Kinase Defines Critical Prognostic Genes of Stage I Lung Adenocarcinoma. PLoS ONE 7(9): e43923, 2012.

The power of survival prediction for the NCI validation data set, the Duke data set and NCC-Tokyo data set using the 139 genes.



4th Biosupercomputing Symposium

Large-scale protein-protein interaction network prediction by an exhaustive rigid docking system MEGADOCK

Yutaka Akiyama

Professor of Department of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology

Profile:

Yutaka Akiyama is a professor of the Department of Computer Science, Graduate School of Information Science and Engineering, and also the Director of the Education Academy of Computational Life Sciences (ACLS), Tokyo Institute of Technology. He is currently serving as the President of Initiative for Parallel Bioinformatics (IPAB), a board member of BioSuperComputing Research Community (BSCRC), and a board member of Japanese Society for Bioinformatics (JSBi).

He received B.E.(1984), M.E.(1986), and Dr. Eng.(1990) in Electrical Engineering from Keio University. His doctoral thesis was about theoretical studies and VLSI implementation of a novel neural network model called “Gaussian Machines”. On that study, he received Best Paper Award for Young Researcher of IPSJ National Convention (1988). After his enthusiastic activity to initiate world’s earliest bioinformatics WWW services (1995) on GenomeNet Japan when he was an associate professor at Kyoto University, he started his own computational biology research group (1996) as a part of the Real World Computing national project with using massively parallel computers and pioneering large-scale PC clusters. In 2001, he became the founding Director of Computational Biology Research Center (CBRC), the biggest government-supported bioinformatics research center in Japan under Ministry of Economy, Trade and Industry.

His research interest covers large-scale processing and acceleration techniques for computational biology and bioinformatics, including ultra-fast metagenome sequence analysis, protein-protein interaction prediction, virtual screening of drug compounds, machine learning for pharmacokinetics prediction, and software system for mass spectrometry analysis.

Abstract

1. Introduction

Protein-protein interaction (PPI) plays a core role in cell functions. Massively parallel supercomputing systems have been actively developed recently, that enables us to solve large-scale biological problems such as PPI network prediction based on tertiary structures.

To challenge interactome level large-scale analysis by fully utilizing protein tertiary structures, we have proposed a large-scale PPI prediction system "MEGADOCK" based on exhaustive protein docking and post-docking analysis [1][2]. We input protein structure data to the system and get predictions of possible interacting pairs among them.

We have already applied our system to 44×44 (subset of protein docking benchmark 2.0) and 89×89 (structures of bacterial chemotaxis proteins) scale analyses. In real biology problem, such as searching drug induced pathway of EGFR signaling, about 200 proteins should be examined. In our preliminary survey on the EGFR pathway and related proteins data, we found about 2000 structures corresponding to these proteins. Therefore, the PPI network prediction system needs to handle about 2000×2000 combinations of protein structures.

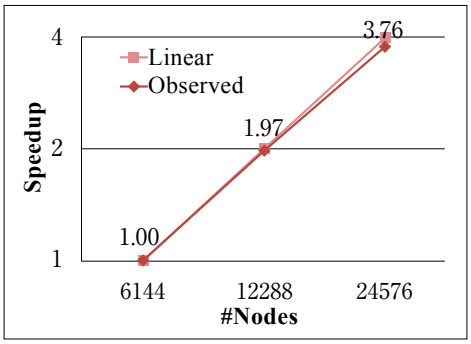
2. Implementation

To solve such large-scale problems, highly efficient computing system is necessary. We implemented MEGADOCK by using hybrid parallelization where each docking processes are calculated in parallel within one node and number of docking jobs are distributed among nodes. We also designed MEGADOCK using a simple score model to reduce calculations required for protein docking.

3. Results and Conclusion

MEGADOCK showed almost linear scaling up to 24,576 nodes on K computer (RIKEN Advanced Institute of Computer Science, Japan).

As an application to real biology pathway, PPI prediction using MEGADOCK was performed on the reconstruction of a canonical signal transduction pathway of bacterial chemotaxis (13 proteins, 89 structure data including multiple structures for each protein species), human apoptosis pathway (57 proteins, 158 structures) and human EGFR signal transduction pathway (49 proteins, 497 structures). The F-measure value was 0.44 when applied to chemotaxis pathway, 0.28 when applied to apoptosis pathway and 0.36 when applied to human EGFR pathway.



Parallel scaling of MEGADOCK on K computer.

The proposed approach to computational PPI detection is a promising methodology for mediating between structural studies and systems biology by utilizing cumulative protein structure data for pathway analysis.

References

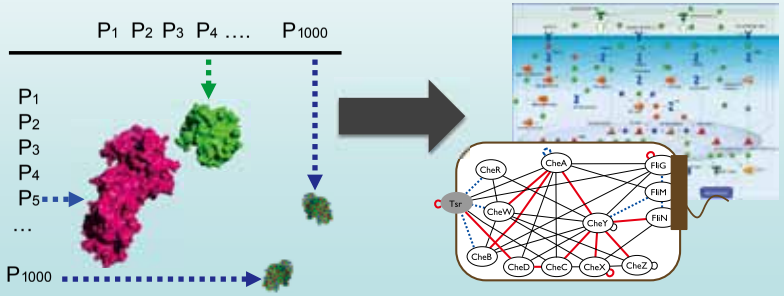
1. Matsuzaki,Y.,Matsuzaki, Y., Sato T, Akiyama Y. In silico screening of protein-protein interactions with all-to-all rigid docking and clustering: an application to pathway analysis. *J. Bioinform. Comput. Biol.*, **7**:991-1012, 2009.
2. Ohue M, Matsuzaki Y, Akiyama Y. Docking-calculation-based Method for Predicting Protein-RNA Interactions. *Genome Inform.*, **25**:25-39, 2011.

Acknowledgements

This work was supported in part by a Grant-in-Aid for Research and Development of The Next-Generation Integrated Life Simulation Software from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT). Part of the result was obtained by early access to the K computer at the RIKEN Advanced Institute for Computational Science.

Large-scale protein-protein interaction network prediction by an exhaustive rigid docking system MEGADOCK

Yutaka AKIYAMA
Tokyo Institute of Technology



PPI network prediction by exhaustive docking

Exhaustive docking & post-docking analysis

Binding partners

PPI network prediction

Objective:

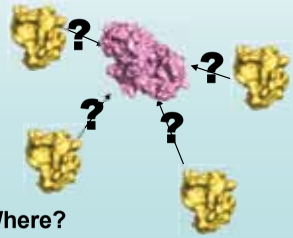
- Search novel PPIs in lung cancer pathway proteins and cancer drug related proteins
- The first large-scale PPI prediction challenge (2000 x 2000 = 4 million scale)
- To contribute finding target proteins of cancer drugs

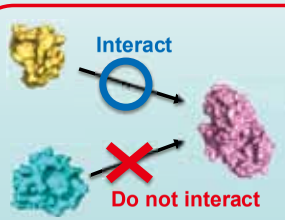
EGFR signaling network

2

PPI prediction based on tertiary structures

- Questions asked by protein docking
  - 1:1 docking: How given proteins interact?
  - Our challenge by exhaustive docking: Which protein pairs interact?
    - Developing an ultra-fast docking tool **MEGADOCK** feasible for exhaustive docking of million-order pairs





1:1 docking

Our problem

3

A compact score function of MEGADOCK

Compress three terms into one complex number

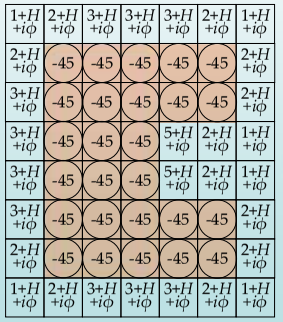
- Shape complementarity
- Hydrophobic interaction
- Electrostatic interaction

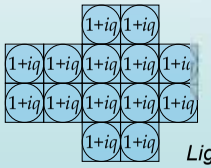
$$S(\alpha,\beta,\gamma) = \Re\left[\sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N} \boldsymbol{R}(l,m,n)\boldsymbol{L}(l+\alpha,m+\beta,n+\gamma)\right]$$

$$\boldsymbol{R}(l,m,n) = G_R(l,m,n) + w_h H(l,m,n) + i\phi(l,m,n)$$

$$\boldsymbol{L}(l,m,n) = G_L(l,m,n) + i w_e q(l,m,n)$$

$$S(\alpha,\beta,\gamma) = \mathrm{IFT}\left[\mathrm{DFT}[\boldsymbol{R}(l,m,n)]^{*}\mathrm{DFT}\left[\boldsymbol{L}(l,m,n)\right]\right]$$

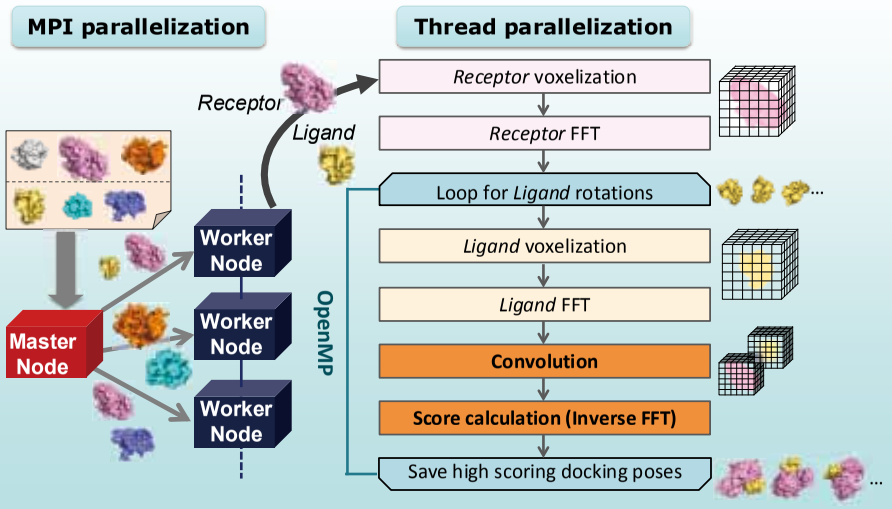




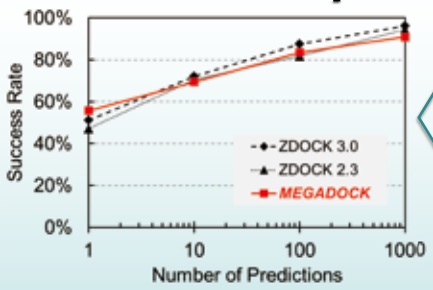
Convolution can be calculated fast by FFT (Katchalski-Katzir model)

Ohue M, Matsuzaki Y, Ishida T, Akiyama Y. *Lecture Note in Bioinformatics 7632*, 178-187, Springer, 2012.

Implementation with hybrid parallelization

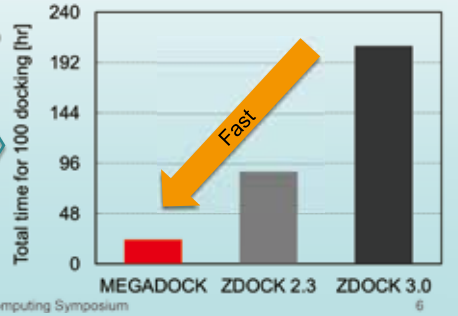


Performance of protein docking by MEGADOCK



Docking accuracy was comparable to conventional tools (Benchmark 4.0, 176 bound complexes)

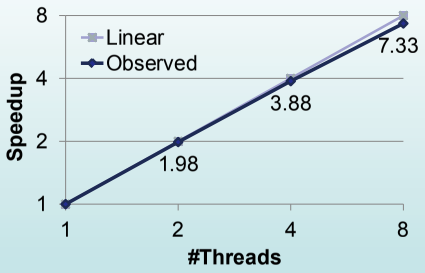
8.8 times speedup compared to ZDOCK 3.0 (single node, single thread)



Ohue M., *et al.*, *Lecture Note in Bioinformatics 7632*, 178-187, Springer, 2012.

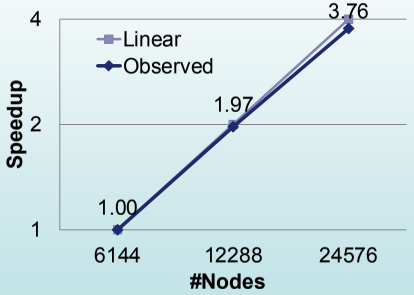
Scalability on K computer

Thread parallelization



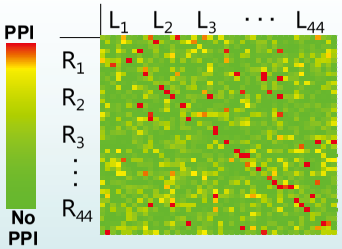
One docking of FFT size: 140 proteins

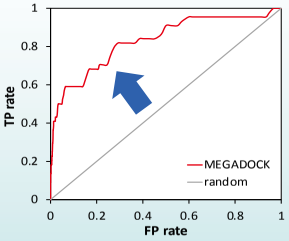
MPI parallelization

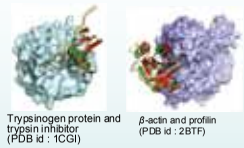


An exhausting docking of 220 proteins
Size of FFT: 140

PPI prediction using general benchmark data







Binding partner prediction from 44x44=1936 dockings and post-processing based on highest docking score (Diagonal: interacting pairs)

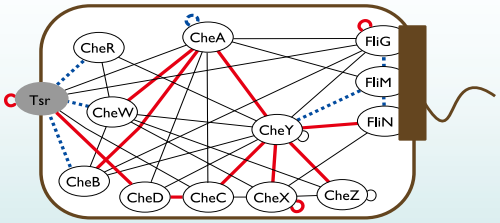
F-measure : 0.42

$$F\text{-measure} = \frac{2 \cdot \mathrm{TP}}{(\mathrm{TP}+\mathrm{FP})+(\mathrm{TP}+\mathrm{FN})}$$

PPI prediction by MEGADOCK achieved better than random performance on general benchmark dataset (monomer pair from protein-protein docking benchmark 2.0, Mintseris *et al*, *Proteins*, 2005.)

Matsuzaki *et al.*, *J Bioinform Comput Biol*, 7:991-1012 (2009).
Ohue *et al.*, submitted.

Application to bacterial chemotaxis pathway

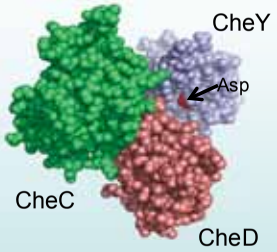


: True Positive
: False Negative
: False Positive

13 proteins (101 structures)

F-measure : 0.44

Acceptable performance was shown on a real biology pathway reconstruction problem.

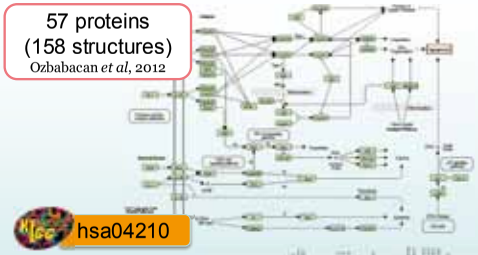


A suggestion of CheY-CheC interaction by using "False-positive" pair CheY-CheD as a mediator

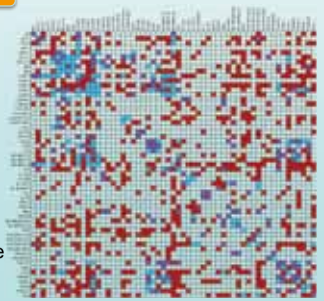
Matsuzaki *et al.*, *J Bioinform Comput Biol*, 7:991-1012 (2009).

PPI prediction results (Apoptosis)

57 proteins (158 structures) Ozbabacan *et al*, 2012



hsa04210



True Positive
False Positive
False Negative

F-measure : 0.28

PPI prediction by docking without any other knowledge showed comparable results to template-based search of interaction partners (F-measure 0.30, Ozbabacan *et al.*, *J Struct Biol*, 2012).

Prediction	Interacting	No Interaction
Positive	88	364
Negative	96	1105

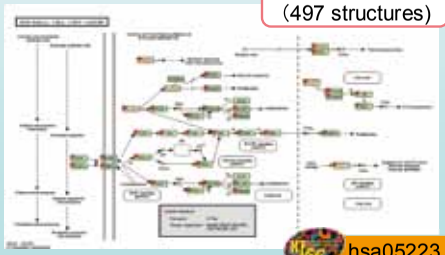
Ohue, et al., *Tech. Rep. IPSJ SIG*, 2012-BIO-32(14), 1-8, 2012.

Application to non-small cell lung cancer pathway

- Completed large-scale exhaustive docking
  - 497 structures, all-to-all docking ＝247,009 structure pairs
- Achieved high PPI prediction performance
  - Precision 0.29
  - Recall 0.47
  - F-measure 0.36

PPI prediction of about 250 thousand structure pairs showed comparable performance to the application to bacterial chemotaxis (10 thousand pairs).

44 proteins (497 structures)



hsa05223

Prediction	Interacting	No Interaction
Positive	53	131
Negative	59	747

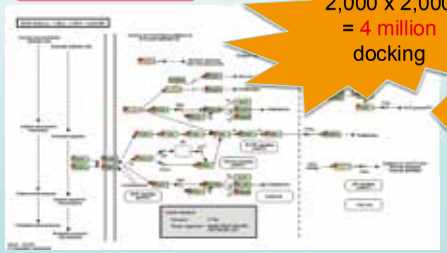
Counts are based on protein species

Interaction of cancer pathway and proteins related to Gefitinib

EGFR pathway related to non-small cell lung cancer

Proteins related to Gefitinib estimated by Miyano lab, Tokyo Univ. from microarray analysis

44 proteins (497 structures)



2,000 x 2,000 = 4 million docking

294 proteins (1424 structures)

Searching novel cancer related PPIs

Supercomputing accelerates genomic medicine

Tatsuhiko Tsunoda

Director, Research Group for Medical Informatics
RIKEN Center for Genomic Medicine

Profile:

Tatsuhiko Tsunoda, Ph.D. (Medicine) & Ph.D. (Eng.), Group Director
1985-1989: B.S., Department of Physics, Faculty of Science, The University of Tokyo
1989-1991: M.S., Department of Physics (Elementary Particle Physics), The University of Tokyo
1992-1995: Ph.D., Department of Engineering, The University of Tokyo
Degrees: Ph.D. (Medicine) & Ph.D. (Engineering)
1995-1997: Assistant Professor, Graduate School of Engineering, Kyoto University
1997-1998: Research Associate, Human Genome Center, The Institute of Medical Science, The University of Tokyo
1998-2000: Assistant Professor, Human Genome Center, The Institute of Medical Science, The University of Tokyo
2000-present: Laboratory Head, Laboratory for Medical Informatics, RIKEN Center for Genomic Medicine
(2000-2008 RIKEN SNP Research Center)
2011-present: Director, Research Group for Medical Informatics, RIKEN Center for Genomic Medicine

Abstract

Whole-genome approaches accompanied with supercomputing have opened new frontiers in medical research. Genome-wide association studies (GWAS) exhaustively explore disease-related single nucleotide polymorphisms (SNPs) and genes in the human genome. They also provide information for deciding which drugs and dosages are adequate for individuals – personalized medicine. In addition, high throughput next-generation sequencing (NGS) enables us to apply whole-genome sequencing to individual genomes with high quality. Supercomputing accelerates the processing of these huge datasets, and leads to novel findings in genomic medicine by allowing for computationally intensive sophisticated mathematical modeling and analysis. These technologies have revolutionized medical research as well as health care.

1. Introduction

In 2002, our center reported the world's first GWAS results [1]. In 2004, using gene-based data, I constructed the world's first linkage disequilibrium (LD) map, and found exotic patterns of natural selection on genes [2]. Thereafter, we participated in the International HapMap project to construct a more extended LD map and select tagging SNPs, which has been used for chips/arrays [3]. This resulted in a large increase in the number of GWAS, further accelerated by the BioBank Japan project, and the discovery of many genes related to common diseases, cancers, and drug responses.

2. Approaches to missing heritability problem

We are now facing the missing heritability problem: current GWAS results are insufficient to explain the expected heritability of common diseases. One of approaches to this problem is to increase the power of GWAS. To this end, we are now enlarging sample sizes using disease cohorts and performing meta-analysis through collaborations around the world.
Another approach is to examine multiple markers simultaneously; typical GWAS looks at only single common SNPs as markers. We have developed an efficient algorithm, ParaHaplo, for handling haplotypes, combinations of alleles at adjacent loci on the chromosome which are transmitted together, as markers [4-7]. Because the number of tests significantly drops when using haplotype blocks rather than single SNPs as units, we can expect an increase of power. To calculate the empirical p-values for significance, we apply permutation procedures and Markov Chain Monte Carlo methods, which require supercomputing power.
Another method for processing multiple markers simultaneously is based on gene-gene/SNP-SNP interaction detection. We have developed an efficient algorithm, ExRAT, which processes all SNP combinations, a very large number of comparisons, and calculates the empirical p-values using the importance sampling technique.
Beyond the aforementioned techniques for handling multiple SNPs on current GWAS platforms, we are exploring methods using other types of markers to approach the missing heritability problem. We are using genotype imputation to explore hidden SNPs with lower allele frequency, as well as to combine data from chips with different marker sets.
One promising approach for developing higher quality reference haplotypes and for exploring unknown variation is analyzing lower frequency variations, e.g. single nucleotide variations (SNVs) and copy number variations (CNVs), through NGS. To develop our analytical pipeline for NGS data, we sequenced a single genome at high coverage, resulting in the first reported Japanese individual's whole-genome sequence [8]. That work allowed us to establish methodologies for detecting multiple types of variations: SNVs, structural variations including CNVs, and novel sequences. Based on our methodology, we have constructed a pipeline NGSanalyzer for analyzing personal genomes [8], cancer genomes [9,10], as well as whole-exome analysis for common/monogenic diseases.

3. Supercomputing for genomic medicine

For analyzing huge GWAS and NGS datasets, we had to implement massively-parallel pipelines for ParaHaplo, ExRAT, and NGSanalyzer on supercomputer systems. For each algorithm, we have developed an efficient pipeline that can be run on the K computer, which allows for the processing of huge datasets of whole genomes for many cases/controls within practical time: ~ days. In summary, statistical genetics and supercomputing enables us to analyze many individuals' whole genome sequences with high accuracy, speed, coverage, and preciseness for applying specific therapies with each individual, allowing the prevention of disease in an individual by prediction – personalized medicine.

References

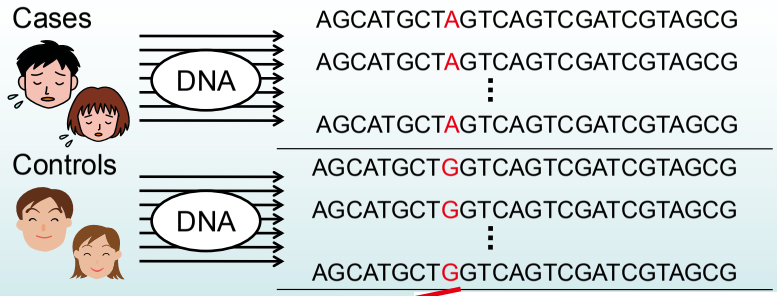
1. K.Ozaki, Y.Ohnishi, A.Iida, A.Sekine, R.Yamada, T.Tsunoda, H.Sato, H.Sato, M.Hori, Y.Nakamura, and T.Tanaka. Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nature Genetics*, **32**, 650-654 (2002).
2. T.Tsunoda, G.M.Lathrop, A.Sekine, R.Yamada, A.Takahashi, Y.Ohnishi, T.Tanaka, and Y.Nakamura. Variation of gene-based SNPs and linkage disequilibrium patterns in the human genome. *Human Molecular Genetics*, **13**, 1623-1632 (2004).
3. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**, 851-862 (2007).
4. K.Misawa, S.Fujii, T.Yamazaki, A.Takahashi, J.Takasaki, M.Yanagisawa, Y.Ohnishi, Y.Nakamura, and N.Kamatani. New correction algorithms for multiple comparisons in case-control multilocus association studies based on haplotypes and diplotype configurations. *Journal of Human Genetics*, **53**, 789-801 (2008).
5. K.Misawa and N.Kamatani. ParaHaplo: A program package for haplotype-based whole-genome association study using parallel computing. *Source Code for Biology and Medicine*, **4**, 7 (2009).
6. K.Misawa and N.Kamatani. ParaHaplo 2.0: a program package for haplotype-estimation and haplotype-based whole-genome association study using parallel computing. *Source Code for Biology and Medicine*, **5**, 5 (2010).
7. K.Misawa and N.Kamatani. ParaHaplo 3.0: A program package for imputation and a haplotype-based whole-genome association study using hybrid parallel computing. *Source Code for Biology and Medicine*, **6**, 10 (2011).
8. A.Fujimoto, H.Nakagawa, N.Hosono, K.Nakano, T.Abe, K.A.Boroevich, M.Nagasaki, R.Yamaguchi, T.Shibuya, M.Kubo, S.Miyano, Y.Nakamura, and T.Tsunoda. Whole genome sequencing and comprehensive variant analysis of a Japanese individual using massively parallel sequencing. *Nature Genetics*, **42**, 931-936 (2010).
9. The International Cancer Genome Consortium. International network of cancer genome projects. *Nature*, **464**, 993-998 (2010).
10. A.Fujimoto, Y.Totoki, T.Abe, K.A.Boroevich, F.Hosoda, H.H.Nguyen, M.Aoki, N.Hoshono, M.Kubo, F.Miya, Y.Arai, H.Takahashi, T.Shirakihara, M.Nagasaki, T.Shibuya, K.Nakano, K.Watanabe-Makino, H.Tanaka, H.Nakamura, J.Kusuda, H.Ojima, K.Shimada, T.Okusaka, M.Ueno, Y.Shigekawa, Y.Kawakami, K.Arihiro, H.Ohdan, K.Gotoh, O.Ishikawa, S.Ariizumi, M.Yamamoto, T.Yamada, K.Chayama, T.Kosuge, H.Yamaue, N.Kamatani, S.Miyano, H.Nakagama, Y.Nakamura, T.Tsunoda, T.Shibata, and H.Nakagawa. Whole Genome Sequencing of Liver Cancers Identifies Etiological Influences on Mutation Patterns and Recurrent Mutations in Chromatin Regulators. *Nature Genetics*, **44**, 760-764 (2012).

Supercomputing accelerates genomic medicine

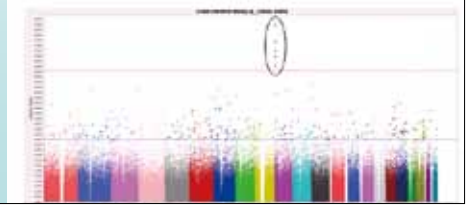
Tatsuhiko Tsunoda
Center for Genomic Medicine, RIKEN

4th Biosupercomputing Symposium

Genome-wide Association Study (GWAS)



	A	G
Cases	6173	4209
Controls	4542	5538



Test the whole genome:
Genome-wide association study

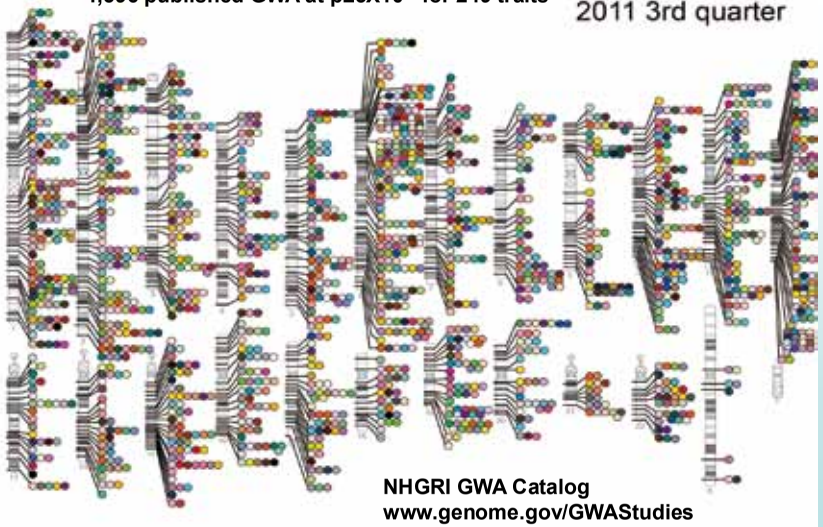
Significant milestones in GWAS

	The world	Japan (RIKEN CGM)
2000		The millennium project started
2002		The first GWAS, *Nature Genetics*.
2003		Biobank Japan Project
	The International HapMap Project	
2007	The paper on Phase II published 550,000 tagging SNPs Commercial platforms GWAS rush	Many GWAS papers published Started genotyping Biobank samples with the platforms

550,000 tagging SNPs
Chip/array platforms

Extended the power!

Published Genome-Wide Associations through 09/2011
1,596 published GWA at p≤5X10^{-8} for 249 traits



Missing heritability problem

The results from GWAS are insufficient to explain the heritability of diseases

	#loci	Heritability explained
Age-related macular degeneration	5	50 %
Crohn's disease	32	20 %
Systemic lupus erythematosus	6	15 %
Height	180 *	10 % *
Type 2 diabetes	18	6 %
HDL cholesterol	7	5.2%
Early onset myocardial infarction	9	2.8%
Fasting glucose	4	1.5%

Manolio TA *et al. Nature, 461, 747-753* (2009)
* Lango Allen H et al. Nature, 467, 832-838 (2010)

Haplotype-based analysis (ParaHaplo)

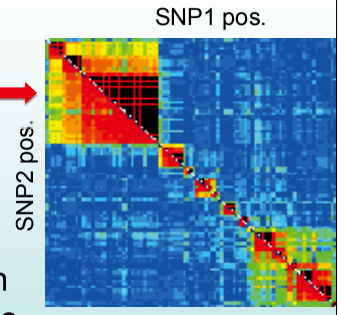
Standard methods require genome-wide significance α = 0.05 / #SNPs ~ 10^{-7} (Bonferroni correction)

↓

However, this is too strict because SNPs are in linkage disequilibrium

↓

Compare haplotype frequencies within each block b/w case and control groups



For calculating empirical p-values, perform massively parallel permutation testing on the K computer: ParaHaplo

Calculation of exact probability under the assumption of a multinomial distribution

$$P[f(X_{11}, X_{12}, \ldots, X_{1,L-1}, X_{21}, X_{22}, \ldots, X_{2,L-1}) = 1]$$
$$= \sum_{x_{11}=0}^{2n_1} \sum_{x_{12}=0}^{2n_1-x_{11}} \sum_{x_{13}=0}^{2n_1-x_{11}-x_{12}} \cdots \sum_{x_{1,L-1}=0}^{2n_1-x_{11}-x_{12}-\cdots-x_{1,L-2}}$$
$$\times \sum_{x_{21}=0}^{2n_2} \sum_{x_{22}=0}^{2n_2-x_{21}} \sum_{x_{23}=0}^{2n_2-x_{21}-x_{22}} \cdots \sum_{x_{2,L-1}=0}^{2n_2-x_{21}-x_{22}-\cdots-x_{2,L-2}}$$
$$\times f(x_{11}, x_{12}, \ldots, x_{1,L-1}, x_{21}, x_{22}, \ldots, x_{2,L-1}) \frac{(2n_1)!(2n_2)!}{\prod_{i=1}^{L}\prod_{j=1}^{2} x_{ji}!} \prod_{i=1}^{L} h_i^{\sum_{j=1}^{2} x_{ji}},$$

where $x_{1L} = 2n_1 - \sum_{i=1}^{L-1} x_{1i}$ and $x_{2L} = 2n_2 - \sum_{i=1}^{L-1} x_{2i}$.

Let us define the function f by $f(\ldots) = 1$ if the test is significant at any locus, otherwise $f(\ldots) = 0$.

Calculating the exact probability is very computational intensive.

Misawa K. *et al., J Hum Genet*, **53**, 789-801 (2008). 7

Markov Chain Monte Carlo (MCMC) method

An algorithm for sampling haplotypes under multinomial distribution.

1. As an initial state, arbitrary integer values of x_{ji} are given so that their sum is #samples.
2. j=0 or j=1 is selected in equal probability.
3. An integer value u is selected in equal probability from the integers from 1 to L.
4. If $x_{jv} > 0$, an integer value v other than u ($1 \leq v \leq L$) is selected.
5. New candidates $x^*_{jv} = x_{jv} + 1$ and $x^*_{ju} = x_{ju} - 1$ are calculated.
6. Then, the following value is calculated, $c = h_v x_{ju} / h_u(x_{jv}+1)$. If $c \geq 1$ then $\{x_{ji}\}$ is updated by substituting x^*_{jv} for x_{jv} and x^*_{ju} for x_{ju}, the step is advanced, a significance test is performed as described in (8) and the process then returns to (2).
7. If $c < 1$, then {xji} is updated by substituting x^*_{jv} for x_{jv} and x^*_{ju} for x_{ju}, with probability c, else the state is kept invariant (probability 1 - c), the step is advanced, a significance test is performed as described in (8) and the process then returns to (2).
8. A test of independence between the phenotype and alleles at each of the l loci is performed using a chi-square test

This algorithm works much faster than the exact method, but takes long time

Misawa K. *et al., J Hum Genet*, **53**, 789-801 (2008). 8

ParaHaplo analysis flow-chart

Genotype Data download from WTCCC

↓

Quality Check

↓

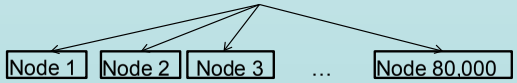
Detection of LD blocks (Gabriel et al. 2002)

↓

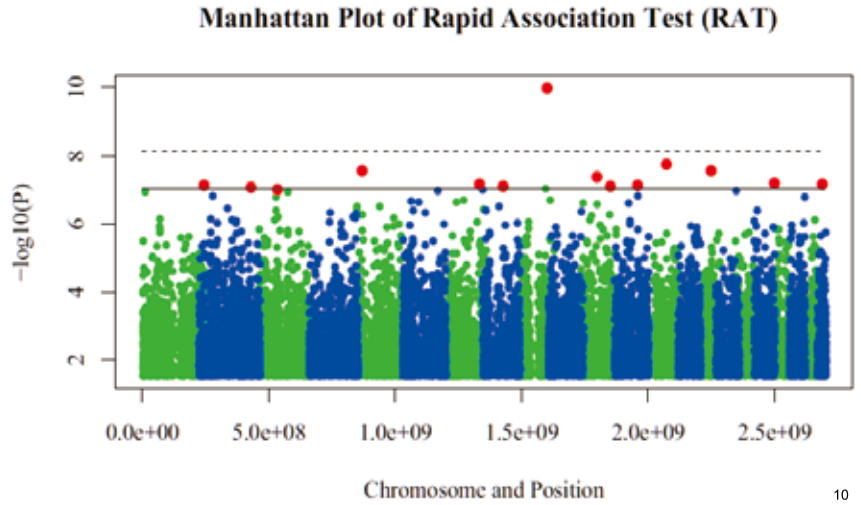
Haplotype phasing with Mach 1.0

↓

Massively-parallel processing of haplotype-based GWAS

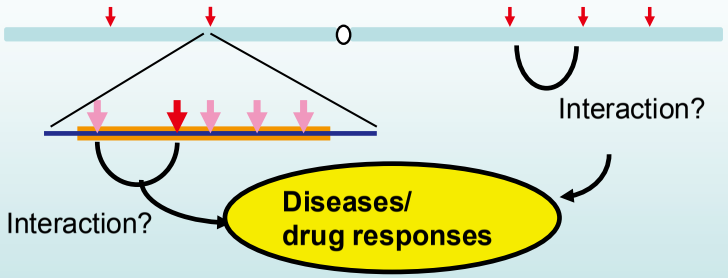


Example output of ParaHaplo



10

Interlocus interaction



Association between diseases/drug responses and combination of SNPs

11

Genome-wide gene-gene interaction

Genome-wide scan for gene-gene interaction

$\text{logit}(P) = a_1x_1 + a_2x_2 + a_3\text{int}(x_1, x_2)$

↓

Bonferroni correction: too conservative

↓

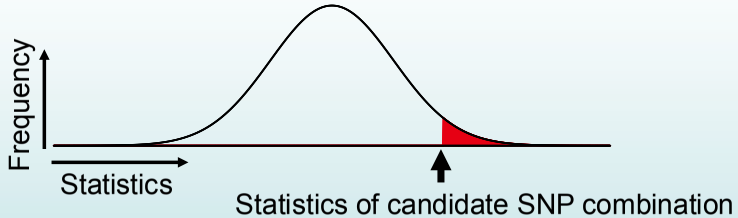
Permutation testing: too computationally demanding

↓

Rapid association test by importance sampling

Importance Sampling

Distribution of statistics based on permutation



Sampling probability of permutations from the red area

↓

Average ratio of sampling probability

Approximate p-value = red area / total

Extended Rapid Association Test (ExRAT)

For linear regression models and likelihood ratio test

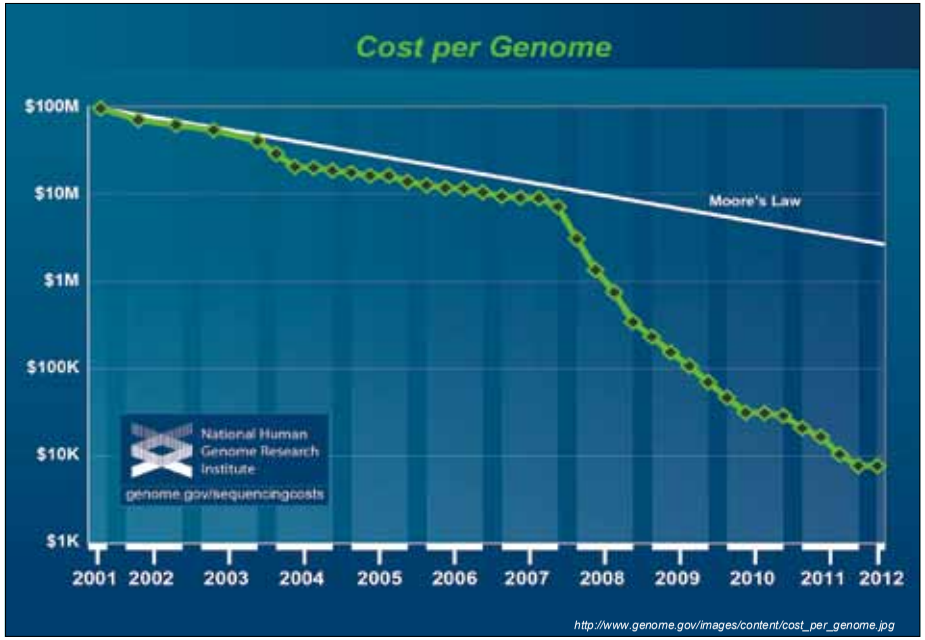
↓

Logistic regression model with a SNP-SNP interaction term

P: probability of being affected

$$\mathrm{logit}(P) = \mu + \alpha x_A + \beta x_B + \gamma x_A x_B$$

Effect of SNP *A* | Effect of SNP *B* | Interaction b/w SNPs



Whole genome sequencing

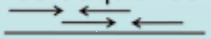
Sample DNA

↓ DNA fragmentation and library preparation

↓ Sequence

Adaptor
Sequenced region

Mapping to reference sequence



Reference

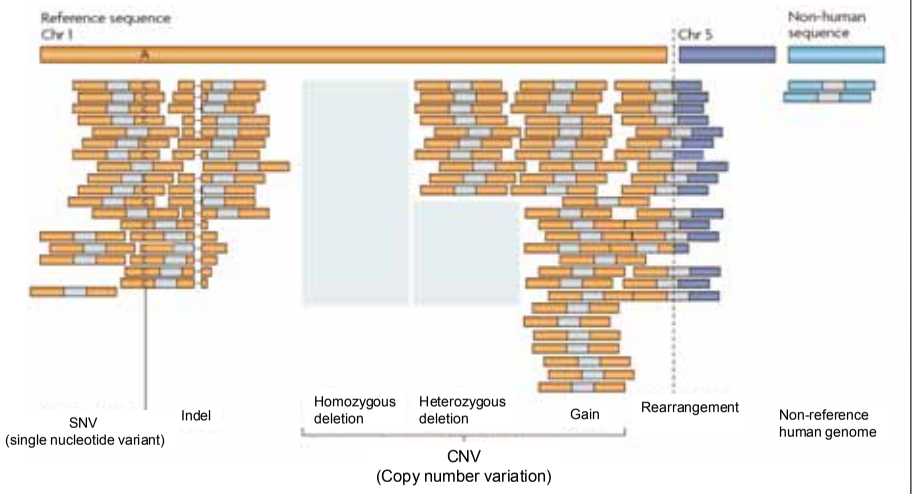
Identify polymorphism

NGSanalyzer

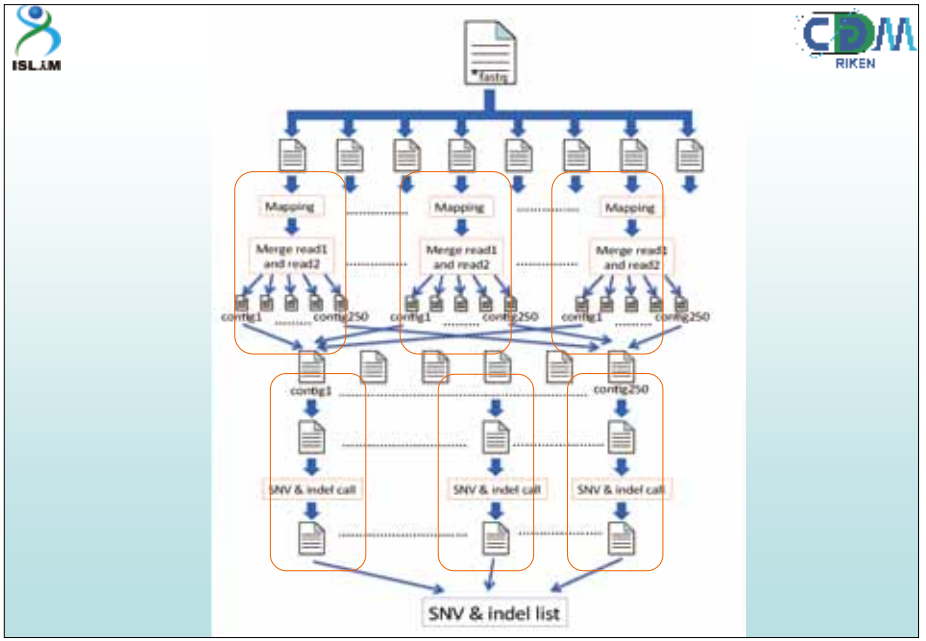
↓ *De novo* assemble

Construct novel sequence

Genetic variation detection



Meyerson, M. *et al., Nature Reviews Genetics*, **11**, 685- 696 (2010).



First WGS and analysis of a Japanese genome

ARTICLES

nature genetics

Nature Genetics, 42, 931-936(2010).

Whole-genome sequencing and comprehensive variant analysis of a Japanese individual using massively parallel sequencing

Akihiro Fujimoto[1,2], Hidewaki Nakagawa[1], Naoya Hosono[1], Kaoru Nakano[1], Tetsuo Abe[1], Keith A Boroevich[1], Masao Nagasaki[3], Rui Yamaguchi[3], Tetsuo Shibuya[3], Michiaki Kubo[1], Satoru Miyano[2,3], Yusuke Nakamura[1,3] & Tatsuhiko Tsunoda[1,2]

We report the analysis of a Japanese male using high-throughput sequencing to ×40 coverage. More than 99% of the sequence reads were mapped to the reference human genome. Using a Bayesian decision method, we identified 3,132,608 single nucleotide variations (SNVs). Comparison with six previously reported genomes revealed an excess of singleton nonsense and nonsynonymous SNVs, as well as singleton SNVs in conserved non-coding regions. We also identified 5,319 deletions smaller than 10 kb with high accuracy, in addition to copy number variations and rearrangements. *De novo* assembly of the unmapped sequence reads generated around 3 Mb of novel sequence, which showed high similarity to non-reference human genomes and the human herpesvirus 4 genome. Our analysis suggests that considerable variation remains undiscovered in the human genome and that whole-genome sequencing is an invaluable tool for obtaining a complete understanding of human genetic variation.

WGS and analysis of liver cancers

LETTERS

Nature Genetics, 44, 760-764(2012).

Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators

Akihiro Fujimoto[1-16], Yasushi Totoki[2,16], Tetsuo Abe[1], Keith A Boroevich[1], Fumie Hosoda[2], Ha Hai Nguyen[1], Masayuki Aoki[1], Naoya Hosono[1], Michiaki Kubo[1], Fuyuki Miya[1], Yasuhito Arai[2], Hiroyuki Takahashi[2], Takuya Shirakihara[2], Masao Nagasaki[3], Tetsuo Shibuya[3], Kaoru Nakano[1], Kumiko Watanabe-Makino[1], Hiroko Tanaka[3], Hiromi Nakamura[2], Jun Kusuda[4], Hidenori Ojima[5], Kazuaki Shimada[6], Takuji Okusaka[7], Masaki Ueno[8], Yoshinobu Shigekawa[8], Yoshiiku Kawakami[9], Koji Arihiro[10], Hideki Ohdan[11], Kunihito Gotoh[12], Osamu Ishikawa[12], Shun-ichi Ariizumi[13], Masakazu Yamamoto[13], Terumasa Yamada[12], Kazuaki Chayama[1,9], Tomoo Kosuge[6], Hiroki Yamaue[8], Naoyuki Kamatani[1], Satoru Miyano[3], Hitoshi Nakagama[5,14], Yusuke Nakamura[1,15], Tatsuhiko Tsunoda[1], Tatsuhiro Shibata[2] & Hidewaki Nakagawa[1]

Towards efficient improvement of transcriptional circuit models by Life Science Data Assimilation System (LiSDAS)

Tomoyuki Higuchi

Director-General, Institute of Statistical Mathematics

Profile:

Tomoyuki Higuchi is currently Director-General of the Institute of Statistical Mathematics (ISM), Japan. He is also Professor of ISM and of the Graduate University for Advanced Studies. He obtained the B.S. degree in 1984, the M.S. degree in 1986, and the Ph.D. degree in 1989 from the University of Tokyo. His primary research interests are in a Bayesian modeling and sequential Monte Carlo computation, in particular, data assimilation.

Abstract

1. Introduction

It has been about 40 years since a computer simulation is called as the third method of science. A recent explosion of data, the so-called appearance of big data, strengthens the research domain to study a method of tools for analyzing big data such as statistics, machine learning, data mining, and visualization technologies. This phenomenon is called the fourth paradigm after a publication. The data assimilation (DA) is a synthesis technique based on the Bayesian filtering method by embedding observation/experiment data in a numerical simulation. DA is an emerging area in earth sciences, particularly oceanography. Its research motivation is easily understood simply if we notice that there are too many uncertainties in the model such as the boundary condition, initial condition, unknown parameters, and unknown dynamics. It yields an accommodation ability to make a simulation real, and the better initial and boundary conditions can be automatically obtained.

Major objectives of DA are classified into the following five aspects. The first is to find the best or better initial condition for forecasting. It is actually realized in the real weather forecast. The second is to find the best or better boundary condition in constructing a simulation model. This procedure includes a setting of appropriate boundary conditions necessary for dealing with the coupled phenomena. The third is to optimize the configuration of macroscopic parameters which substitute a description of complicated real phenomena involving different temporal and spatial scales. A validation of the empirically given values is included in this problem. The fourth is to inter/extrapolate physical quantity with some numerical simulation model at times and locations where no observation data is available. This procedure is called “a generation of re-analysis dataset (product)”. The fifth is to perform a sensitivity analysis of virtual observation network in an attempt to construct an effective observation network system with less budgetary cost and less consuming time.

2. Applications

We are studying the sequential DA methods based on the Ensemble Kalman and particle filters, and conducting the DA experiments in several specific areas. We will give a brief explanation for the sequential DA and demonstrate a part of applications carried out by our DA research group.

The understanding of transcription circuit is not enough from both the experimental and the theoretical views. Life Science Data Assimilation System (LiSDAS) has been developed as an application dedicated for K computer, with the aim of proposing improved models by assimilating observation data to a conventional model. LiSDAS describes transcriptional circuits in terms of differential equations, which include processes of protein-protein interactions, the activation or the repression of mRNAs by transcription factors, and the degradation of mRNAs and proteins. We will make a demonstration in our talk, employing two network models: a circadian rhythm network model and a statistically constructed model for network of mRNAs in lung cancer cells.

Data assimilation works also in intracellular fluid dynamics. A large flow is observed in a fertile egg right before a cell division its mechanism is yet uncertain. One hypothesis is that proteins “myosin” in the cell wall drive physically such intracellular flow. We have performed data assimilation in order to obtain dynamical evidence supporting this hypothesis. Our model obtained by data assimilation successfully explains the observed cytoplasmic flow.

We have also applied DA to influenza epidemics and estimated connection of cities in the context of 2009 influenza pandemic. It is important to know how degree cities are coupled since such preventive activities are conducted basically by individual autonomies.

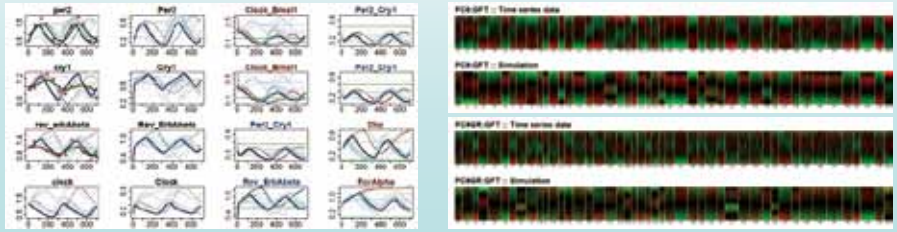
References

1. T. Higuchi, Proceedings of 14th International Conference Fusion (2011).

Data assimilation in life-science

The Institute Statistical Mathematics
Tomoyuki Higuchi

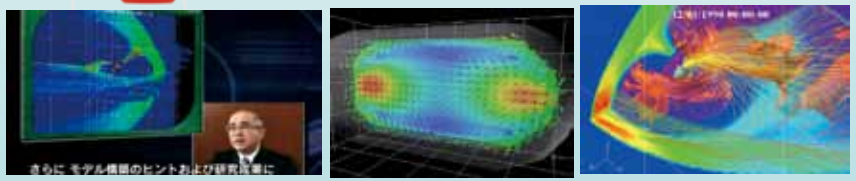
Masaya M. Saito Ryo Yoshida Hiromichi Nagao Shin'ya Nakano



Data Assimilation – Method of Integration for Simulation and Data

- Solutions for differetn types of inversion problems
  - Estimation of unknown parameters (*e.g.* boundary conditions)
  - Reconstruction of latent physical process
  - Remodeling
- Practiced widely in weather forecasts - Typhoon, El Nino, etc
- Towards Life Science

Watch YouTube "Data Assimilation R&D Center at ISM"

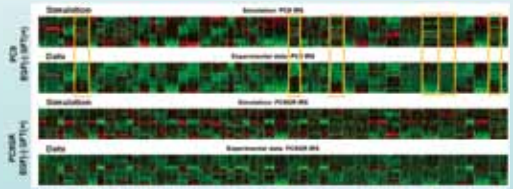


LiSDAS - Life Science Data Assimilation Systems

Members: Saito, M, Yoshida, R, Nagao, H, Nakano, S, Higuchi, T (ISM)
Developments: Data assimilation software implemented on *K*-computer
Applications: Two types of inversion problems using biochemical and biophysics simulators

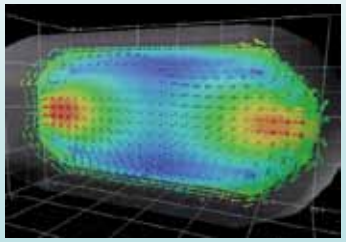
Gene regulation models for lung cancer cells with acquired resistance to anticancer drugs

Collaboration with Miyano, S and Gotoh, N (HGC, Univ.Tokyo)



Cell mechanics for cytoplasmic flow in the *C. elegans* embryo

Niwayama, R and Kimura, A (NIG)



State Space Representation

- System model – Your simulator is represented as a Markov process

$$x_n = f_G(x_{n-1}, q) + \text{noise}$$

$$G \;\rightarrow\; f_G(\cdot\,|q) \;\rightarrow\; x = \{x_n\} \;\rightarrow\; D = \{y_n\}$$

Model / Parameter / State variables / Data

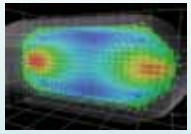
- Measurement model
  - Relate the model output to given data

$$y_n = g(x_n) + \text{noise}$$

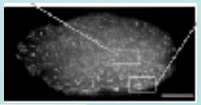
Bayesian Inference – Solving the Inversion Problem

Posterior distribution – *Likeliness* of ω evaluated after observing D

$$p(w|D) \propto p(D|w)p(w)$$



Estimation of unknown objects ω with *given* data D



ω can be any of G, θ, x

G — Model

$f_G(\cdot|q)$ — Parameter

$x = \{x_n\}$ — State variables

$D = \{y_n\}$ — Data

Provided by Dr. Kimura, A (National Institute of Genetics)

Bayesian Supercomputing

Major difficulty arises in the posterior computation

$$w \sim p(w|D) \propto p(D|w)p(w)$$

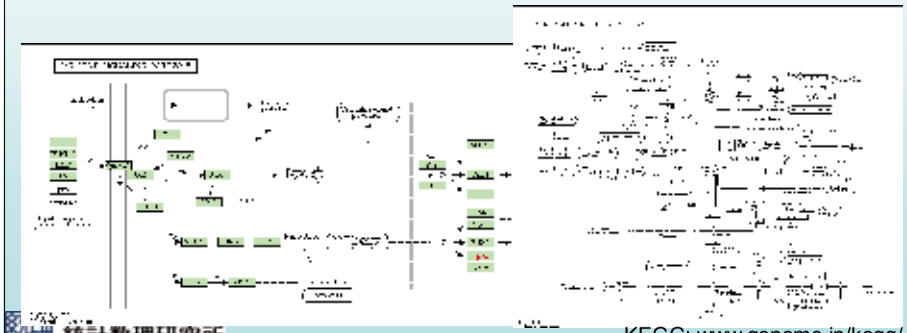
- Monte Carlo simulation from the posterior distribution
- Search over exceedingly high-dimensional Ω
- Hard computation is necessary for biophysics simulation

K computer

Massively-parallel computation for the Markov chain Monte Carlo simulation

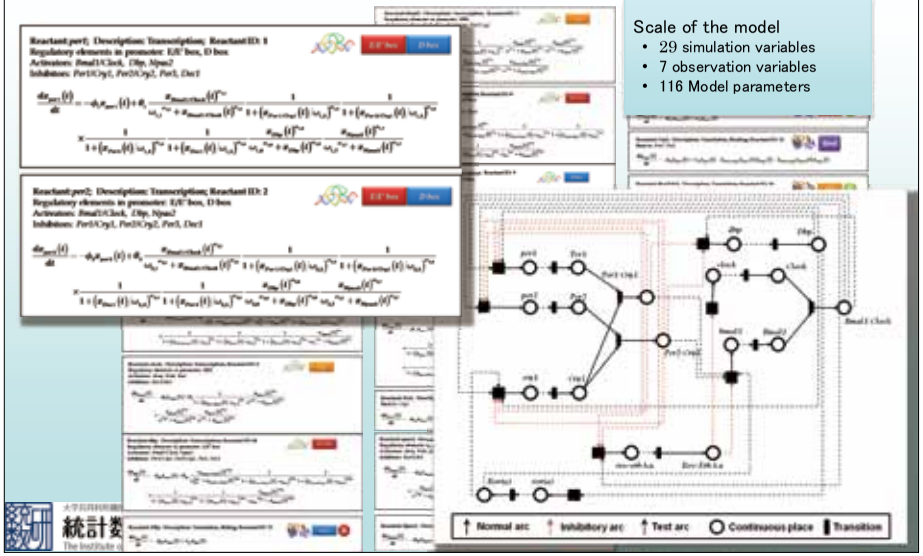
Biological Networks

- Cell - an integrated device made of several types of biochemical reactions involving diverse molecules
- Functions are programmed into *Networks*
  - Variable: RNA, protein, metabolite, carbohydrate
  - Process: Production, binding, degradation, diffusion



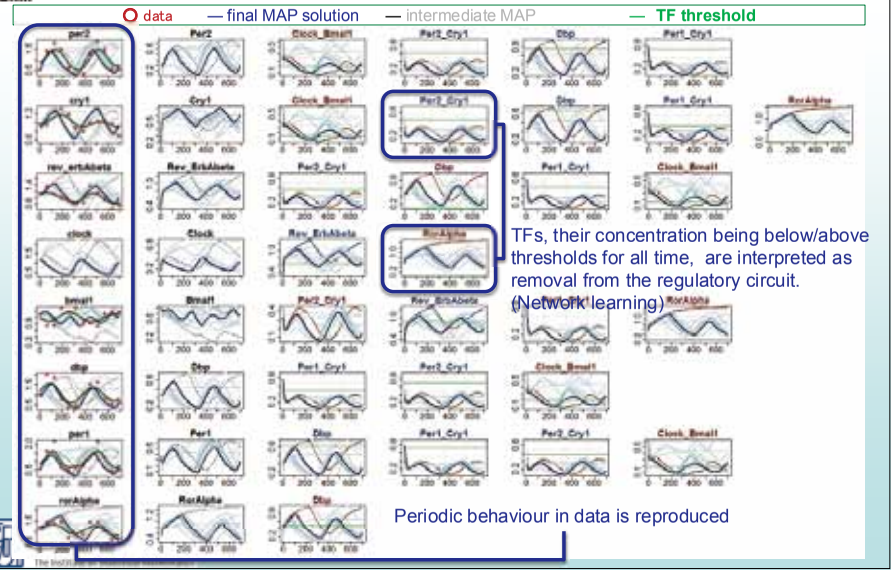
KEGG: www.genome.jp/kegg/

Example1: Gene regulatory systems for circadian clock controls



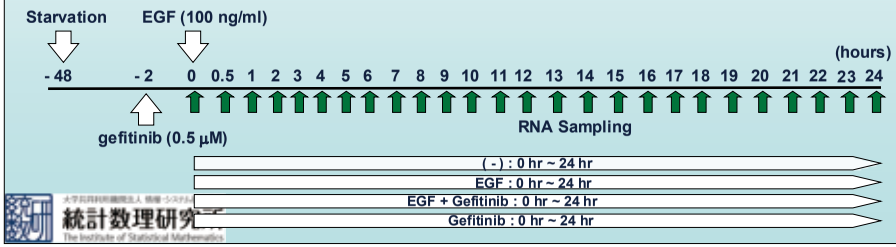
Scale of the model
- 29 simulation variables
- 7 observation variables
- 116 Model parameters

Circadian clock gene regulatory simulation



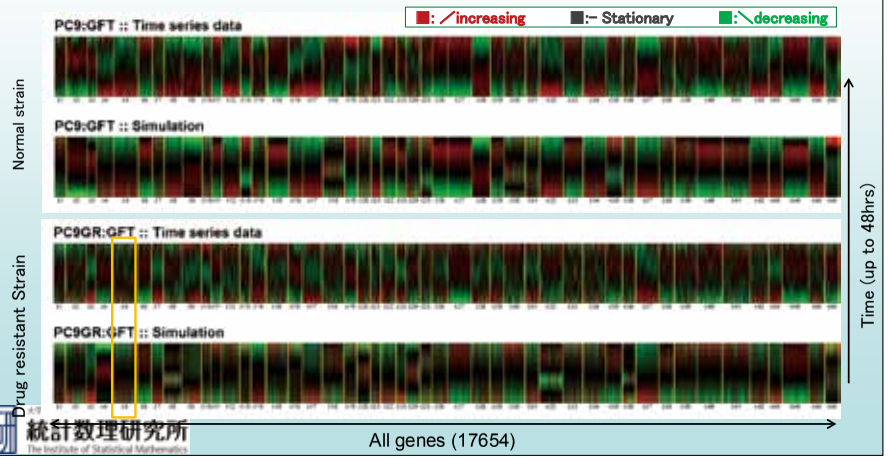
Example 2: Drug response prediction for non-small cell lung cancer

Human Genome Center

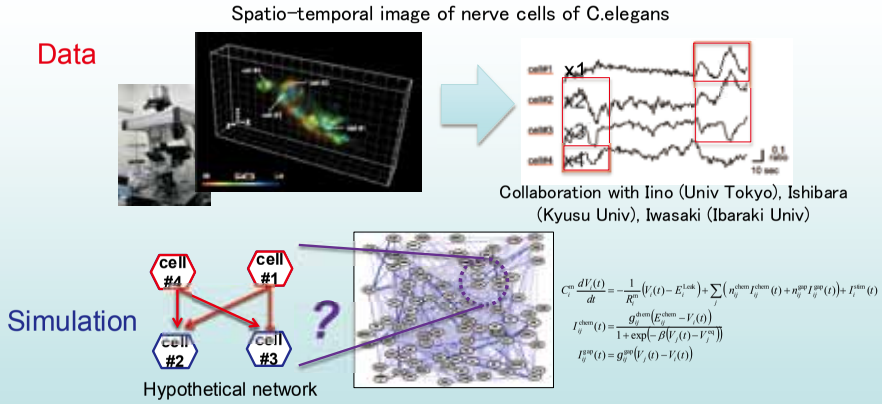


Example 2: Whole genes' transcription simulators

- Transcription dynamics of 17654 genes are modeled with 46 state variables.
- Observed gene expression patterns are well reproduced, though some of genes not.

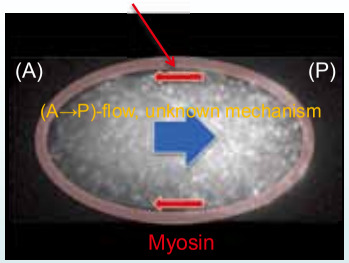


Post GC Application 1: Neuronal circuit model



Nerve cell activities are measured along time via the 4D calcium imaging data.
LiSDAS would promote the deep understanding of the complex network system behind.

Post GC Application 2: Intercellular fluid dynamics

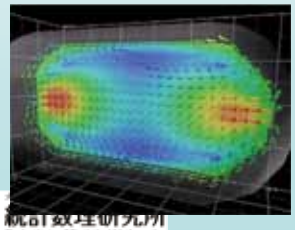


Data: Cytoplasmic flow from the anterior to the posterior

Hypothesis: Myosin on the membrane is a driving force of the flow.

Question

Is the observed flow (velocity field) reproduced under the hypothesis?



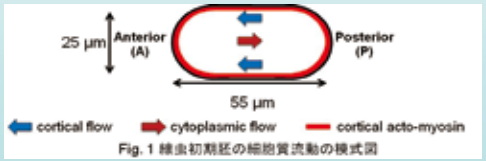
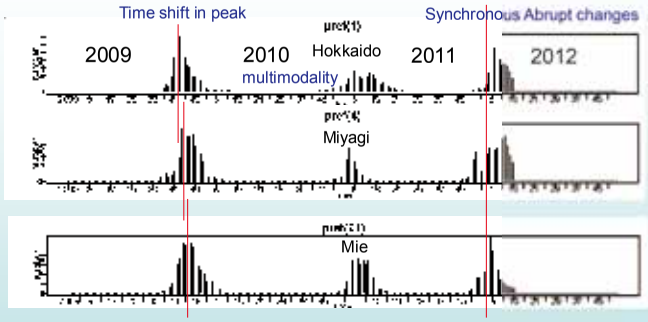


Fig. 1 線虫初期胚の細胞質流動の模式図

Schematic illustration of intercelluar fluid of C.elegans embryos
Collaboration with Dr Kimura (National Institute of Genetics)

Post GC Application 3: Infectious transmission network

The number of Influenza patients in recent three seasons



- Goal: Learn connection strength ε_{ij} between prefectures from features seen in time courses.
- Such a network model help us to incorporate transmission dynamics into the forecast of influenza epidemic.

14

Post GC Application 4:
Data-driven *de novo* drug design using kernel machine learning

Members: Yamashita, H (GUAS), Iba, Y (ISM), Higuchi, T (ISM), Yoshida, R (ISM)

Developments: Computational method for generating *de novo* chemical compounds

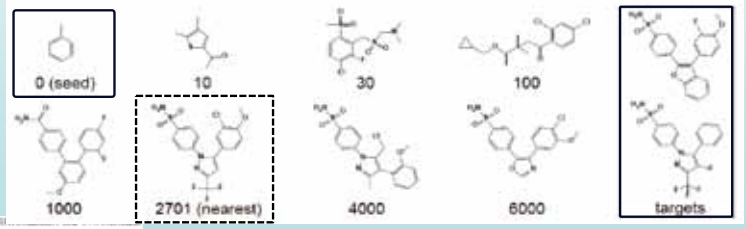
- Random sampling of chemical compound *graph* using a kernel machine and Bayesian computational techniques
- Collaboration of biophysics and statistical science

Pattern recognition and feature extraction

Alter the chemical structure by 'mixing' the identified key molecules using the *fragment-flip* MCMC method

Initial compound - toluene

Targets - COX2 activating compounds



15

Summary

- Data assimilation system «LiSDAS» has been developed for network system analyses for cells
- Potential applications
  - Neural cell network
  - Intercellular fluid
  - Infectious transmission network
  - (*Any whenever you have data and model*)
- Contact us

  http://daweb.ism.ac.jp/

4th Biosupercomputing Symposium 16

Contact:
RIKEN
Computational Science Research Program
2-1 Hirosawa, Wako, Saitama 351-0198, Japan
http://www.csrp.riken.jp