ISSN 1346-602X

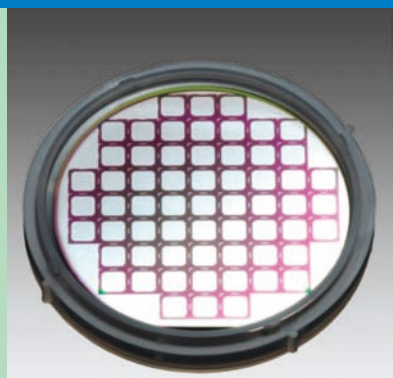
Full Research in Society, for Society

AIST TODAY

National Institute of Advanced Industrial Science and Technology

International Edition

2010
2010-3
No.37



MESSAGE

President's Message
Enhancing hub functions for open innovation

FEATURE

Bioinformatics
Toward the establishment of Japanese leadership in genome information analysis and utilization technologies

Research Hotline
UPDATE FROM THE CUTTING EDGE (April–June 2010)

In Brief

President's Message
Enhancing hub functions for open innovation

Tamotsu Nomakuchi

President
National Institute of Advanced Industrial Science and Technology (AIST)

1. Introduction

We are almost six months into the Third Term of AIST. "Solving 21st Century issues" and "Enhancing hub functions for open innovation" have been defined as the missions of the institute. With the former, a research strategy that takes into consideration the new growth strategy set forth by the government was developed, and it is now transitioning to the execution phase. For the latter, we have entered an era where close alliances that share more outcome awareness with industry and universities are required. Furthermore, I have personally visited various research institutes at home and abroad and have reconfirmed the important role that public research institutes play as hubs in global competition for innovation. I would like to discuss this further here.

2. New trends

There is a new way for industry and universities to tie up with AIST. This involves the new ability for AIST itself to participate in "technology research associations". A legal revision in 2009 removed restrictions on participation in technology research associations. Until now, AIST has contributed in the form of external contributions to associations that have been made by companies based on economic and industrial policy. From now on, however, AIST will be able to contribute as an association member from the stage of drafting projects, through research execution, to utilization of findings. Key human resources such as researchers, project leaders, directors, and managing directors from AIST are already participating in eight associations, and many of the principal research execution bases are located within AIST. I believe that a more strategic approach has become possible, encompassing not only commercialization but also incorporating intellectual property strategies and standardization strategies in order to maintain and enhance the industrial competitiveness of our country.

① Stereo Fabric Research Association (SFRA)
② Photovoltaic Power Generation Technology Research Association (PVTEC)
③ Bio Electro-mechanical Autonomous Nano Systems Laboratory Technology Research Association (BEANS Laboratory TRA)
④ Lithium Ion Battery Technology and Evaluation Center (LIBTEC)
⑤ Fuel Cell Cutting-Edge Center Technology Research Association (FC-Cubic TRA)
⑥ Advanced Laser and Process Laboratory Technology Research Association (ALPROT)
⑦ R&D Partnership for Future Power Electronics Technology (FUPET)
⑧ Technology Research Association for Single Wall Carbon Nanotubes (TASC)

Please refer to the AIST website for details of scope of research. There are a number of topics still under consideration, and the number of participating technology research associations will increase still further in the future.

The formation of the global nanotech research base, "Tsukuba Innovation Arena TIA−nano", through the joint efforts of the University of Tsukuba, the National Institute for Materials Science, industry, and AIST is a major challenge unseen in recent years. I think that this collaboration will give rise to green innovation that will lead the world in six areas: nanoelectronics, power electronics, MEMS, carbon nanotubes, nanogreen, and nanomaterial safety assessment. Researchers from several dozen companies and several universities will come together. The technology research associations ③, ⑦,

and ⑧ mentioned above are primarily carrying out research in this arena. Utilization of the arena framework as an education and training ground for post-graduate students and private sector technicians is also planned, and on this account I expect that this will contribute to the aspect of developing human resources.

Construction of the new AIST Chugoku was completed in the city of Higashi Hiroshima, and an opening ceremony attended by local residents and numerous other interested parties was held on June 25, 2010. Although the new site is more compact than the former site in the city of Kure, it occupies a corner of the Hiroshima Central Science Park. The AIST Chugoku takes pride in its status as a global research base for biomass fuels research, and Hiroshima University, which has concluded a partnership and cooperation agreement with AIST, is located nearby. For this reason, I believe that joint research will proceed efficiently. In addition, the Chugoku International Center of the Japan International Cooperation Agency (JICA) is located within the same park. The research unit director also works as the director for the Research Core for Asian Biomass Energy of the Economic Research Institute for ASEAN and East Asia (ERIA), and many visitors from Southeast Asia come for research and training. The JICA center is thus cooperating immensely in the intake of these visitors. The Chugoku region of Japan has traditionally been an area where there are great opportunities to utilize biomass, and on this account it can be said that the conditions for further development of the new AIST Chugoku as a global open innovation hub have been further enhanced.

AIST has always had in place various schemes for collaboration with industry, academia, and the public sector, such as one-to-one alliances with companies, consortium-based alliances with multiple companies, and such alliances that also incorporate the participation of universities and public research institutes. In the Third Term, however, we will endeavor to further strengthen and expand these ties. In no way can I introduce all of this in such few pages. Those who are interested should by all means refer to the Third Term Research Strategy (published in fiscal 2010).

3. Open innovation is an activation engine

Since the start of this year, I have visited Europe, Singapore, Vietnam, and neighboring South Korea, and have had the opportunity to exchange views with personnel at universities and public research institutes. In Japan, meanwhile, we have been visited by government leaders from numerous countries as well as leaders of public research institutes. Through these experiences, I strongly feel that the world is in the midst of intense competition in innovation, and with the diverse participation of industry, academia, and the public sector, open innovation itself is viewed as a growth engine and an activation engine for countries and regions.

The globalization trend is not only accelerating unification of a single global economic market but also giving rise to numerous challenges that cannot be overcome without global-scale alliances; among them is the issue of climate change. The trend of digitalization has to date promoted evolution and fusion of a diversity of sciences and technologies developed and accumulated over a great number of years in countries of so-called developed nations, but has not allowed being content with the existing sense of values. If there is further progress in the life sciences sector such as in genetic manipulation, I think that this need for advancement will become even further entrenched. This signifies opportunities for innovation regardless of whether the country is developed or developing, and this is becoming a factor in the global innovation battle. Japan must make all efforts to win this battle.

Recently, I visited Belgium's Interuniversity Microelectronics Centre (IMEC) and Singapore's A*STAR (Agency for Science, Technology and Research), both of which are world renowned innovation hubs. Since its establishment in 1984, IMEC has focused on research in the microelectronics and nanotechnology. A*STAR was formed through the reorganization of Singapore government ministries in 2002, and is the country's largest research and educational institute that leads the science and technology sector. What I see as common with both institutes is more than merely the objective of exploiting research findings for the prosperity of the host country. Rather, it is the more important objective of bringing together capable human resources from throughout the world and creating a large research park. The idea is to activate the country and the area through activating the exchange of knowledge and people. I believe that this is a new way of thinking not present in our own philosophy of trying to contribute to strengthening industrial competitiveness through creating good products and businesses. The most important objective of open innovation promoted by AIST is needless to say strengthening industrial competitiveness; nevertheless, I believe that we need to clearly recognize the significance of this new philosophy and strategically promote it in order to achieve a greater contribution to society.

Bioinformatics
Toward the establishment of Japanese leadership in genome information analysis and utilization technologies

Bioinformatics opening the door to life innovation

Introduction

The rapid progress of measuring technologies, as reflected in the recent dramatic improvement (more than 1,000-fold in 10 years) in the reading speed of DNA sequencers, is greatly changing not only genomic analysis but also the whole concept of the field of life sciences and its applications. Bioinformatics is an information technology for extracting meaningful knowledge from the flood of data that has gone far beyond individual manual analysis.

This special feature highlights the activities of AIST targeted at 'life innovation' through bioinformatics, focusing on database development, applied research for drug discovery, and activities for fostering human resources.

Software and databases

Backed by theories and algorithms, of information science bioinformatics is widely used in the life sciences in the form of software and databases. AIST has been developing software programs and databases for analyses of various targets, such as genome sequence information, protein structure information, gene expression information, and sugar chain information. This feature focuses on H-InvDB, a human gene database; MEDALS, a portal site serving as a guidepost for databases; and the functional RNA database, which is the representative database in AIST's life science research field. In order to utilize the immense amounts of data accumulated not only in our in-house databases but also in other domestic and overseas databases, it is necessary to build a framework for retrieving useful data rather than using databases merely as electronic encyclopedias. Whereas the most common use of databases is to obtain information by narrowing down search results, bioinformatics software is created with a view to providing more advanced information analysis. Following the description of databases, we introduce the integration of information infrastructures using workflow technology. This concept aims at extracting meaningful information by arithmetic operations on computers with an ensemble of software and databases, the fruit of our information analysis technology, working together. AIST has already developed a large number of world-class software tools for information analysis of sequence data for next-generation sequencers, structure prediction and information analysis for functional RNA, protein three-dimensional structure prediction, etc. We plan to actualize many practical workflows utilizing these tools.

Development of research associated with drug discovery support

One of the major goals of life innovation through bioinformatics is to develop technologies to support drug discovery and to return the results to society. Technologies for drug discovery support utilizing bioinformatics are required for the selection of biological substances (proteins, etc.) to be targeted by drugs, screening of drugs (compounds, etc.) acting on target substances, and prediction of side effects through the analysis of biological networks such as gene expression and metabolism. AIST Tokyo Waterfront has gathered together researchers who are conducting theoretical studies on bioinformatics, development of software and databases, practical studies to support drug discovery, and experimental research in collaboration with bioinformatics, etc., and a large number of outstanding results have been obtained. Protein structure-based drug screening, introduced by Senior Research Scientist Yoshifumi Fukunishi, is one of the most important technologies among AIST's major outcomes. Bioinformatics contributes to life science and, at the same time, is a technology based on leading-edge information theories. The high-speed neighborhood search method, introduced by Senior Research Scientist Koji Tsuda, has enabled large-scale comparison of ligand-binding sites that has never before been possible with conventional technologies. Life is a complex system composed of intricate interactions among all biological substances; no individual genes function independently of each other in the living body. Studies are essential to clarify the mechanisms of life through the modeling and analysis of life as a system, for such purposes as elucidation of disease mechanisms and improvement of the operability of pluripotent cells. Team Leader Katsuhisa Horimoto introduces biological network analysis using systems biology.

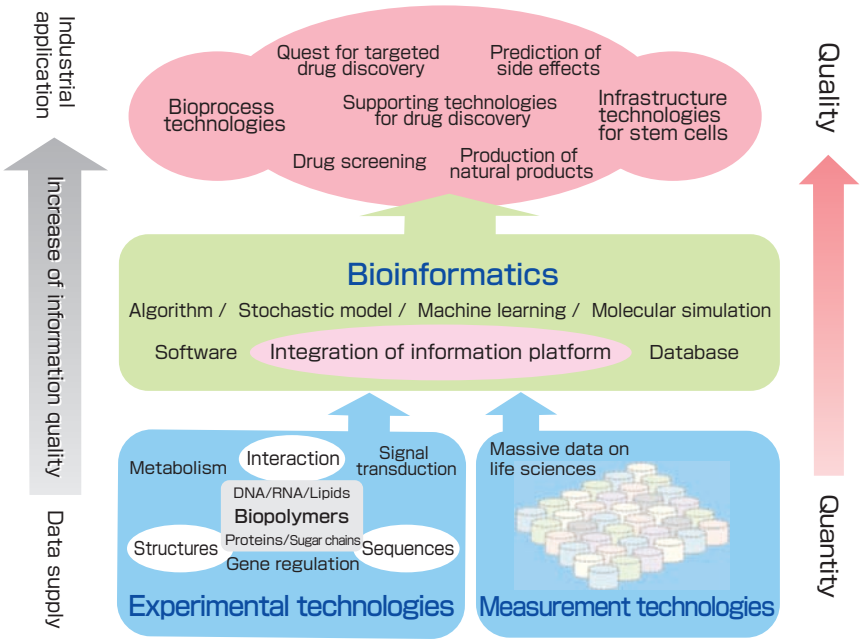
Fostering of human resources in life information science

Bioinformatics is faced with a shortage of human resources to promote education, research, and development, who will cater to

For inquiries about this article：Computational Biology Research Center http://www.cbrc.jp/index.eng.html

the rapidly growing needs of today's society. AIST has also been addressing the fostering of human resources who will assume an active role in bioinformatics by collaborative research; acceptance of technical trainees; training of R&D staff (focusing on corporate researchers) in the Bioinformatics Training Consortium, which is described later; and acceptance of graduate students utilizing the Cooperation Program between AIST and Graduate Schools.

Director
Computational Biology Research Center
Kiyoshi Asai



Life innovation through bioinformatics

Expectations for bioinformatics research in AIST

Database Center for Life Science
Research Organization of Information and Systems
Toshihisa Takagi

Research styles in the life sciences are dramatically changing from the hypothesis-driven style to a data-driven style because of the rapid progress in measuring devices typified by next-generation DNA sequencers. In response to the needs of this era, Japan has been creating a framework that will appropriately coordinate existing life science databases so that they can be utilized for research and industrial applications. Since AIST's bioinformatics research is expected to contribute to society through industrial technology, we expect AIST's bioinformatics researchers to not only set their sights on biological discoveries and their application, but also to develop new data-analysis technologies leading to life innovation. Strong expectations are being placed on the Computational Biology Research Center (CBRC) and Biomedicinal Information Research Center (BIRC) of AIST, which are among the largest core facilities dealing with bioinformatics in Japan and have been developing a large number of globally competitive software programs.

What is bioinformatics?

Arrival of the era of massive data

One of the major characteristics of recent molecular life science is the generation of massive amounts of data accompanied with the progress of measurement technologies. This trend began around the '70s and '80s, when sequence databases were constructed as the progress in DNA sequencing technologies was made, and has been accelerated with the subsequent advancement of genome projects and post-genome research. It is expected that the next-generation sequencers will drive this trend for the time being. It is now becoming difficult for the researcher in molecular life science to develop his or her work without these high volumes of data, even if the researcher is not directly involved in the genome-wide studies. The success of the research is highly dependent on whether the information required for the study is efficiently extracted from the ocean of data. Practically, it is impossible to access such extensive and diverse data without computers. One of the significant roles of bioinformatics in the life sciences is to help researchers by navigating this vast ocean of information and assist in steering them through the process of acquiring new knowledge in life sciences including medicine and agriculture.

Acceleration of studies based on construction of hypotheses

There are a large number of examples in which bioinformatics has played a critical role in the research of life science. Recently, Prof. Shinya Yamanaka and his colleagues of Kyoto University had tried to search for transcription factors involved in the maintenance of pluripotency in embryonic stem cells, but their experimental study had reached a deadlock In silico analysis with databases, however, broke the deadlock and led them to a new stage of the study[1]. Another example is found in the study of drug target discovery, in which novel cytokines and novel cytokine receptors have been obtained from genome or expressed sequence tag (EST) databases by sequence similarity search, as novel targets of drugs[2]. We can easily find many other examples. A common feature shared by these examples is that bioinformatics is used for the construction of hypothesis. Unlike the conventional study with the style of hypothesis testing, the study with bioinformatics can assist the researchers to step up their studies into new stages by proposing new hypotheses from massive volumes of data. A more familiar term "prediction" is a form of hypothesis construction. The drug target discovery in genome pharmaceuticals is regarded as a kind of hypothesis construction. The study with the style of hypothesis testing is performed by examining the hypothesis deduced from the experience and the knowledge of a researcher. However, the truth of nature is often far beyond such human guesses. In that case, hypothesis construction is more effective for the development of the study, and bioinformatics serves as an accelerator of serendipity.

Conversion from "quantity" into "quality"

The explosion of information in the field of life sciences has just begun, and more and more massive volumes of data will be generated in the future. In such a situation, it is important to develop the methods to convert "quantity" to "quality." Bioinformatics transforms information from quantity to quality by constructing hypothesis, as described above. Then, the quality of the hypothesis is critical for further development of the studies. Inaccurate prediction (or hypothesis) actually delays the research and industrialization. The time required to convert quantity into quality also matters. Even if an accurate prediction is obatined, the prediction is meaningless if it requires, say, 100 years for the computation. Furthermore, the integration of diverse biological information is significant for life science. The biological data being currently generated are not only enormous in size, but also highly diverse. For example, such data include genome sequences, expression profiles and protein networks. Each of such data represents a different aspect of living organisms, and the integration of the data will deepen the insight into the living organisms as hierarchical and complex systems. Such an integration requires a flexible platform on which the diverse data can be comprehensively processed. The explosion of the biological information has changed the style of research of living organisms, in which the computer has become an essential tool of biology and the role of bioinformatics in life science will be further expanded.

Deputy Director
Computational Biology Research Center
Hiroyuki Toh

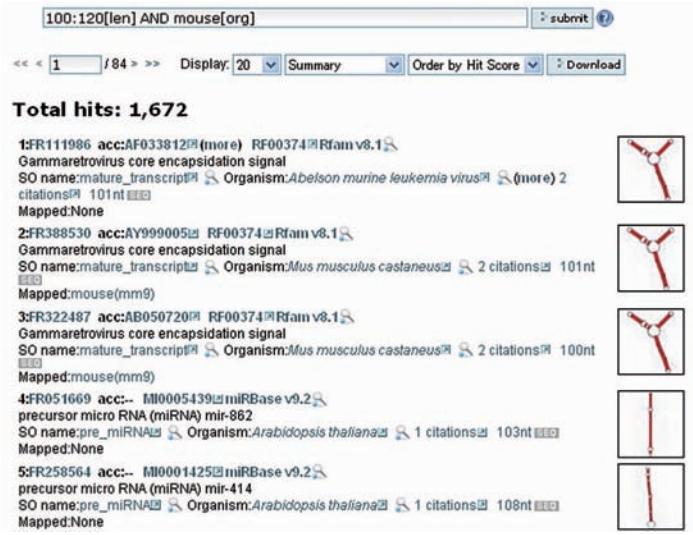
References
[1] M. Tanaka: *iPS cells-To what extent can humans be regenerated?*, Nippon Jitsugyo Publishing (2008)(in Japanese).
[2] H. Nomura: *Genome-based drug discovery-Personalized medicine and genome data mining*, Saiensu-Sha (2005) (in Japanese).

For inquiries about this article：Computational Biology Research Center http://www.cbrc.jp/index.eng.html

Functional RNA database

What are functional RNAs?

A ribonucleic acid (RNA) having a certain function is referred to as a functional RNA. It might be said that messenger RNAs (mRNAs), which carry genetic information for protein synthesis, are a typical example of functional RNAs. However, it has recently been found that a wide variety of RNAs are present in cells. Thus, the term functional RNA has come to denote RNAs having a function, excluding mRNAs. Most of what have been reported as functional RNAs are either still in the process of functional analysis or have not even begun to be analyzed. Expectations are therefore being placed on future development in order to find functional RNAs that are worth using from the viewpoint of medical science and industrial applications.



Screen displaying search results on the fRNAdb sequence database
Thumbnails of secondary structures are displayed in the list.

Functional RNA database

The objective of the Functional RNA Project funded by the New Energy and Industrial Technology Development Organization (NEDO) and implemented over a five-year period since 2005 is the development of infrastructure technologies for functional RNAs. Database construction is one of the pivotal infrastructure technologies. The functional RNA database plays a greater role in the discovery of novel functional RNAs and as an information tool for supporting functionality analysis than in the catalog development of known functional RNAs. The major challenge has been to identify promising functional RNAs using sequence information present in large volumes, rather than to newly measure RNAs. This database is a combination of two databases working in coordination with each other: the fRNAdb sequence database and the UCSC Genome Browser for Functional RNA. The functional RNA database hosts a large collection of information about known functional RNAs as well as predicted functional RNA candidates, allowing users to search and browse a diverse range of information such as sequence information, genome mapping information, information about neighboring genes, literature information, expression information based on uniquely designed microarrays, relevant disease names, etc. To further enhance its convenience as a practical information tool, a feature has been incorporated into the database that authorizes certain users to obtain unpublished information designated in the database. This has proven to be extremely useful in joint research activities handling large-scale sequence information. The database is available for public use at http://www.ncrna.org/.

Future development

There is growing demand for genome-wide analysis of RNA nucleotide sequences accompanying the widespread use of next-generation sequencers. For example, by reading RNA sequences contained in tissue samples from a patient on a next-generation sequencer and comparing the RNA sequences with the sequences of the overall genes (annotation), the gene expression level can be determined with high sensitivity. The application of this technique to clinical tests will provide more useful information for pathological diagnosis than ever before. The functional RNA database technology is thus contributing to improvements in the precision of annotation to identify expression levels of functional RNAs.

Computational Biology Research Center
Toutai Mitsuyama

Construction of infrastructure technologies for bioinformatics information

Infrastructure technologies for integration of life science information

The Computational Biology Research Center (CBRC) of AIST has been making efforts using state-of-the-art information technology to construct a bioinformatics information infrastructure for the seamless integration of internally and externally developed and maintained life-science-related software programs and databases. With the increasing complexity and subdivision of life science research, it has become necessary to diversify analysis tools and databases. However, the coexistence of a large number of individual databases, analysis tools, and systems necessitates a great deal of time and effort for their management and operation, making it difficult for researchers to cope with rapid changes and novel technologies in their research field. With this as a background, CBRC has been engaged in developing technologies related to workflows that enable users to perform efficient analyses in a shorter time by defining a series of job flows (Fig. 1), rather than utilizing dispersed analysis tools one by one. This technical development is being carried out through participation in the Life Science Integrated Database Project implemented by the Ministry of Education, Culture, Sports, Science and Technology (MEXT).

Development of workflows

The workflows we are developing are classfied into two types: web workflows and active workflows. A web workflow is a combination of fixed processes, which allow users to specify parameters, etc. only within a limited scope. Thus, web workflows are easily executable web services. In contrast, an active workflow enables users to specify a combination of processes. It has a user-friendly, flexible design for defining the main flow, as required by users, featuring a platform environment known as KNIME with excellent operability using an intuitive graphic user interface. On this platform, users select independent programs needed for their analysis among a group of programs for the analysis tools developed at CBRC or externally developed tools and databases, and combine them as they wish. This realizes the users' original analysis and a series of integrated analysis flows (Fig. 2). In addition, the analysis tools we are developing under this platform environment use the Simple Object Application Protocol (SOAP), a protocol enabling communication with remote PC programs and data. Processes by local PCs are separated from the computing processes requiring computing power, and the user can access computer servers at CBRC via SOAP communication. This makes seamless workflow analysis possible.

Computational Biology Research Center
Kazuhiko Fukui

Fig. 1 Screen image of CBRC Integrated DB Information Infrastructure Website (http://togo.cbrc.jp/)

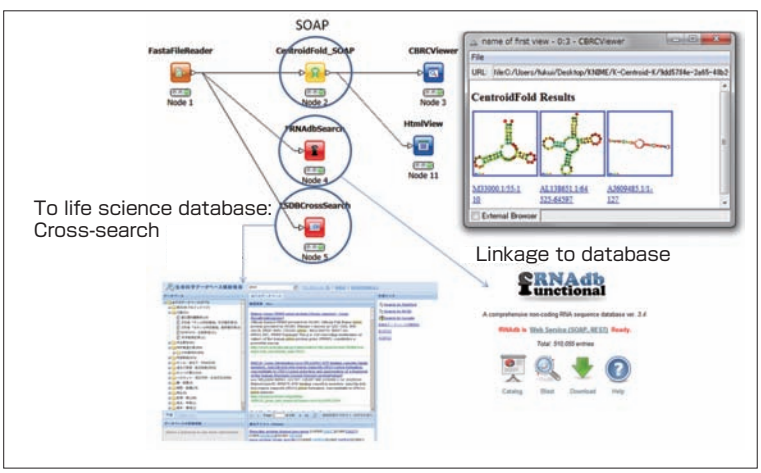


Fig. 2 Example of analysis in an active workflow using the platform

For inquiries about this article：Computational Biology Research Center http://www.cbrc.jp/index.eng.html

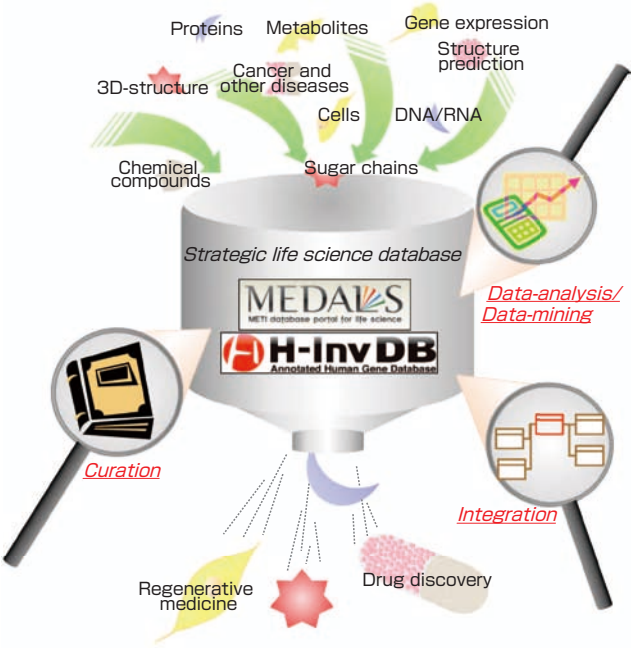
Toward the construction of strategic databases in the life science field

MEDALS: A guidepost through the flood of information

MEDALS (http://medals.jp/) is a portal site providing a collection of databases and analysis tools that have been produced by various life science projects led by the Ministry of Economy, Trade and Industry of Japan (METI). In the past, the management of project results was left up to the individual researchers implementing each project, and there was no means of obtaining a bird's-eye view of all projects for effective use of the resulting data. MEDALS has been developed to overcome this problem and to unify the management of project outcomes. The MEDALS catalogs are effectively listed by classifying the outcomes obtained by METI according to data type, biological species, etc., so that each researcher can easily select outcomes meeting his or her objectives. Moreover, an environment for utilizing this accumulated information has been provided by taking advantage of tools such as "MEDALS cross-search" in 250 databases, developed in collaboration with the MEXT Integrated Database Project (http://lifesciencedb.jp/) and "PubMedScan," which reports information on related articles to users by e-mail.

H-InvDB: Infrastructure information specializing in human genes

H-InvDB (http://hinv.jp/) is an integrated database of human genes and transcripts. It started as a large-scale project in which researchers manually annotated human full-length complementary DNAs (cDNAs) under a policy of maintaining both comprehensibility and high reliability based on observed facts.



Information useful for life science research is scattered throughout the Internet; the outcomes of METI projects are one example. We add our own analysis data to the raw information or delete less reliable data by curation, and publish such information as strategic integrated databases, aimed at application to drug discovery, regenerative medicine, etc.

For the past six years, having hosted a wide variety of data including those on diseases, expression levels, protein interaction, etc., H-InvDB now functions like a data center for human genes. Information in H-InvDB is linked to information on materials for experiments, such as human full-length cDNA clones and antibodies, allowing a smooth flow from searches on the PC to experimental studies in the laboratory. Moreover, data for downloading are organized according to purpose, serving as ideal resources for information and theoretical research.

In recent years, to support the acquisition of new knowledge we have released "Navigation search," an extended search system for human genes based on complicated search criteria; and "HEAT," a data mining tool for automatically identifying features common to given gene sets. Armed with these tools that add a fresh dimension to data, we plan to contribute to frontier research activities such as regenerative medicine and drug discovery.

Future prospects

We believe that henceforth it is desirable to build up strategic databases by investing resources intensively in pivotal research fields, such as updating of information on transcription networks that is useful for such purposes as the acceleration of regenerative medicine and disease research, to list two examples. We have very high expectations that our databases will facilitate collaboration of experimental researchers and bioinformatics researchers to produce significant outcomes.

Biomedicinal Information Research Center
Tadashi Imanishi

Bioinformatics Training Consortium

Fostering of human resources with bioinformatics capabilities

Bioinformatics cannot dispense with biological information itself nor with hardware infrastructures such as large-scale computers for processing information, not to mention the essential human resources who can make effective use of this software and hardware. Now that next-generation DNA sequencers are in widespread use, product development using bioinformatics, such as genome-based drug discovery, is an extremely promising area. Nevertheless, the reality is that companies have yet to provide an environment for producing epoch-making results, with only a few bioinformatics engineers dotted among many experimental technicians (general computer programmers, etc., in the case of information technology companies).

Our efforts up to now and strong track record

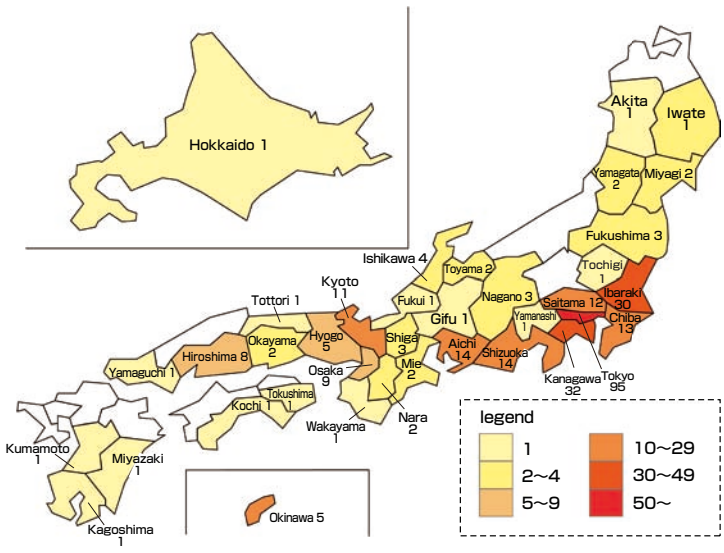
The Computational Biology Research Center (CBRC) has been making efforts to foster human resources since the reorganization of AIST. In fiscal 2001 we established the Bioinformatics Personnel Training Course funded by the Special Coordination Fund for Promoting Science and Technology of MEXT, jointly with the Biological Information Research Center (the current Biomedicinal Information Research Center). The training of advanced-level specialists under this course, particularly researchers, expert engineers and annotators, has gained a high reputation. Subsequently, since fiscal 2005, we have implemented the Bioinformatics Training Course for Engineers intended for people working in companies etc., which has also been supported by the same MEXT fund. This course, which includes short-term learning and e-learning, has successfully trained 446 people from a wide range of industries and areas over the past five years. Of special note is the Pharmacoinformatics Engineer Training Course with its unique, highly practical curriculum that is conscious of the actual forefront of drug development, in which a large number of people from pharmaceutical companies have received training. Another significant achievement has been the active exchanges among numerous corporate researchers that have taken place throughout the course.

Toward the further development of bioinformatics

The human resource fostering activities of CBRC, whose horizons have been successfully expanding, have been a focus of very high expectations among the people concerned in industrial, academic, and government circles. To meet these expectations, in 2010 CBRC established the Bioinformatics Training Consortium as an AIST official consortium for paid members only. The time has come when all researchers and engineers involved in the life sciences should be able to use bioinformatics. One objective of the Consortium is to build up efficient, needs-oriented curricula. We plan to continue making contributions to the development of bioinformatics-related industries from the standpoint of the supply of human resources.

Computational Biology Research Center
Takatsugu Hirokawa
Hiroko Sakai

Courses approx. 30 days per year are held at the CBRC Collaboration Corner which is equipped with 40 PCs for hands-on training, a teleconferencing system, etc.



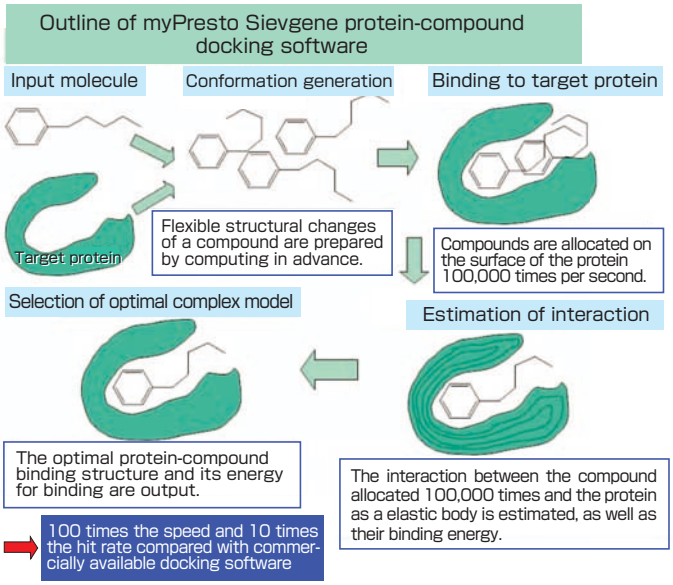
Distribution of applicants for the Bioinformatics e-Learning Course, which has attracted applicants from throughout Japan (273 in total in fiscal 2009)
The e-learning system enables anyone, including those busy with daytime work, to learn about bioinformatics or pharmacoinformatics at any time via the Internet.

For inquiries about this article：Computational Biology Research Center http://www.cbrc.jp/index.eng.html

Computational drug discovery based on protein 3D structures

What is protein-compound docking software?

Drug discovery based on protein structures is performed using a computer screening process in which chemical compounds that are likely to bind to the target protein are selected among commercial chemical compounds. Here, the key elements are the protein-compound docking software and the screening method used. The former predicts how compounds bind to proteins, while the latter organizes and integrates information, corrects computational errors, and predicts active compounds. Commercially available protein-compound docking software from other countries has been gaining the major share in this field. Such products, however, have extremely high royalties, amounting to several million yen per license annually. This problem prompted us to develop the "myPresto" in-house protein-compound docking software program.



Sievgene is a flexible docking software program that considers structural changes in compound conformations. To reduce the amount of computation, the atomic coordinates of the protein are not moved but approximated by a soft elastic body.

Principle of Sievgene protein-compound docking software

Sievgene allocates a compound such that it rolls over the surface of the target protein, estimates interactions such as the electrostatic force between the protein and the compound, and determines the configuration in which the compound is the most energetically stable. Generally, a docking software program treats a protein as a group of atoms. If atoms of a protein and those of a compound come too close and almost overlap each other, a repulsive force occurs; it is extremely difficult to locate a compound close enough to a protein without the atoms of both colliding with each other. In Sievgene, we have solved this problem by replacing protein with a soft elastic body (Figure). This makes it possible to perform docking at a rate of about one second per compound, about 100 times faster than that of standard commercial software products. Moreover, by combining Sievgene with our originally developed screening technique, we have achieved more than 10 times the hit rate in compound prediction compared with the standard commercial software.

Drug screening

A powerful computing screening technique has thus emerged. However, if the target protein has multiple structures due to large fluctuations etc., as many screening results as the number of structures are obtained. The success rate in computational prediction depends on which structures and results are adopted. We have overcome this issue using the concept of "universal active probes", based on the understanding that drug-binding pockets of proteins have a tendency to bind molecules having a "drug-like" structure. Thus, by adding "drug-like"-structured compounds to the screening database as active probes and selecting the screening results that rank these probes at high positions, predictions with high hit rate can be obtained. In this way, our unique approach has made it possible to apply computer screening to diverse target proteins.

Drug search software, myPresto is freeware, available at the following website:
http://MEDALS.jp/myPresto/

Biomedicinal Information Research Center
Yoshifumi Fukunishi

High-accuracy network analysis by systems biology approach

Systems biology

The recent development of epoch-making experimental techniques has enabled measurement of almost all characteristics of cellular molecules. In the past, biological studies went through a process from the formulation of a hypothesis to its testing. In systems biology, however, where the cell is treated as a system built up of molecules, the progress of experimental techniques has made it possible to conduct research using actual measured data. Moreover, the emphasis of research objectives has moved from exploring the characteristics of molecules to exploring the associations (networks) among molecules making up a system. However, this trend toward measuring the characteristics of huge numbers of molecules constituting cells requires tremendous expenditures of time and money. It is therefore necessary to seriously consider for what purpose the measurements are being made before starting a study, and to ensure high computational accuracy in analyzing the measurement data to obtain robust results. Here, we introduce two analysis methods that we have developed to realize the latter requirement; namely, high computational accuracy.

Identification of activated control networks

The first method is a technique for estimating control networks that are activated under specific conditions[1]. First, for networks having a particular structure, the statistical quantity (log-likelihood) is calculated using the values of measured data. Next, by preparing numerous graphs artificially and calculating the log-likelihoods of these graphs, the distribution of log-likelihoods is obtained. Finally, how rarely the log-likelihood of a given network appears in the distribution, i.e., the graph consistency probability, is calculated. Using this technique, the functions possessed by cells in a specific condition can be shown not merely in a list of the names of genes, but in the form of the networks consisting of the genes (Fig. 1). For this purpose, first the known networks and the TF (transcriptional factor)-gene binding data obtained from experiments or databases are prepared. Next, using the data measured in specific conditions, calculations are performed to determine which of the prepared networks have graph consistency probability at a significance level; i.e., which networks' forms rarely fit the data under specific conditions. We have named this procedure "network screening," because it selects the networks that appear only in specific conditions from among a huge number of networks.

High-accuracy estimation of parameters

The second method is a technique for enhancing the accuracy of parameter estimation in network dynamics[2]. The essence of this technique is to obtain, from a system of differential equations presenting network dynamics, another but equivalent system of differential equations, using mathematics referred to as differential elimination. The resulting system of differential equations is adopted as new constraints, together with the error function typically used in parameter estimation.

For the error function, the difference between the time-course values and estimated values is considered. On the other hand, since the new constraints include differentiated values, the form of the curve provided by the time-course values such as the intercept, inflection point can be further considered (Fig. 2). In fact, this is an overwhelming improvement compared with the approach

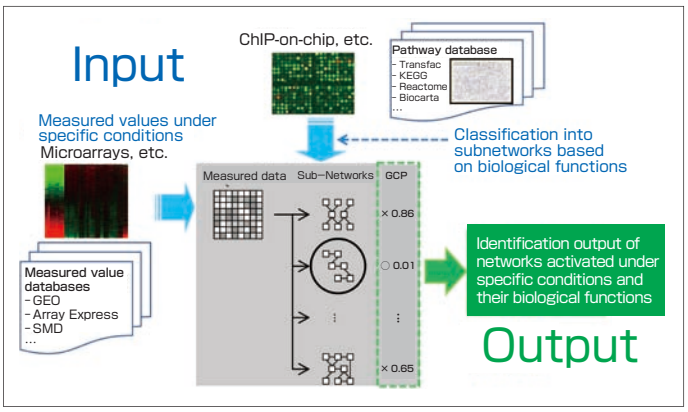


Fig. 1 Activated network detection method (network screening)

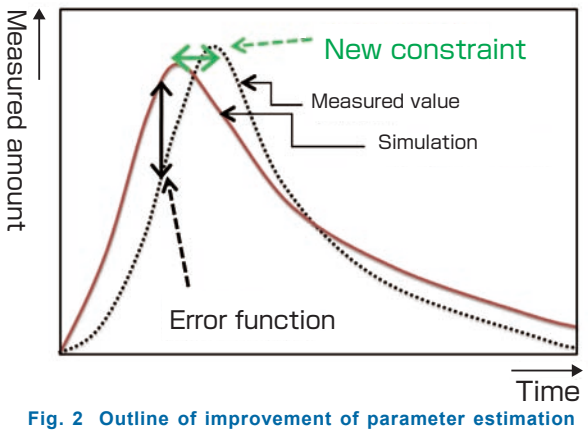


Fig. 2 Outline of improvement of parameter estimation accuracy using high-speed neighborhood search

For inquiries about this article : Computational Biology Research Center http://www.cbrc.jp/index.eng.html

using the conventional error function method. This is a general-purpose technique in various respects; we aim to apply it not only to the dynamics of biological phenomena but also to engineering issues in which parameter estimation is important.

Computational Biology Research Center
Katsuhisa Horimoto

References
[1] S. Saito *et al*. : *BMC Sys. Biol*., 2, 84, (2008).
[2] M. Nakatsui *et al*.: *BMC Sys. Biol*. (in press).

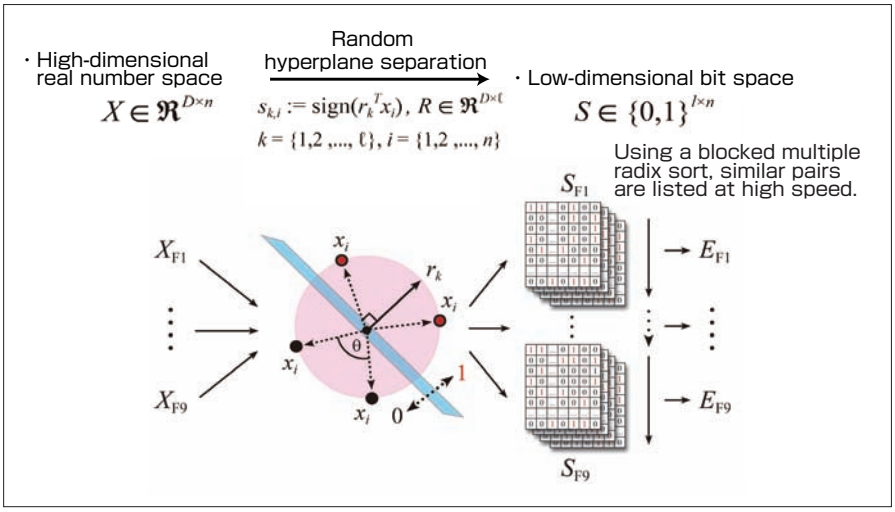
Large-scale comprehensive comparison of known and potential ligand-binding sites

Ligand-binding sites

It is known that a protein expresses its functions via interaction with other molecules (ligands) in many cases. That is, the domain (ligand-binding site) in a protein for binding to other molecules is a critical domain that characterizes the protein functions. Currently (as of June 2010), the Protein Data Bank (PDB) archives contain more than 60,000 protein three-dimensional structures, with the number of ligand-binding sites in the database amounting to hundreds of thousands. The recent development of techniques for ligand-binding site prediction has made it possible to efficiently detect local domains to which ligands can potentially bind. We believe that comprehensive comparison and classification of these enormous numbers of known and potential ligand-binding sites will greatly facilitate elucidation of the binding nature of ligands, and of the generality and evolution route of protein functions.

High-speed similar site search

However, an immense amount of computation time is required to comprehensively compare massive numbers of binding sites using the current approach based on alignment, graph matching, etc. To overcome this problem, we have proposed a novel comparison approach: by embedding ligand-binding sites to feature spaces according to their various properties (e.g., geometric properties such as inter-residue distance, amino acid attributes, and secondary structure attributes) and then applying "SketchSort," our originally developed extremely high-speed neighborhood search method for high-dimensional features, it is possible to perform high-speed and efficient detection of combinations (pairs) of ligand-binding sites that are similar in nature in the feature space.



High-speed neighborhood search by the SketchSort method

Result

As a result of comprehensive comparison of a combined total of about 1,200,000 sites consisting of known ligand-binding sites registered in the PDB and potential ligand-binding sites obtained by prediction, we confirmed that our method could comprehensively list pairs of similar sites with high accuracy, even using general desktop computers, in just over 10 hours. The result included pairs of not only ligand-binding sites common to homologous proteins but also large numbers of hidden active sites unknown in the past. We expect SketchSort to greatly contribute to the estimation of functions from protein three-dimensional structures found in structural genome projects, etc., and to screening in the field of drug design.

Computational Biology Research Center
Koji Tsuda

UPDATE FROM THE CUTTING EDGE

Apr.-Jun. 2010

**The abstracts of the recent research information appearing in Vol.10 No.4-6 of "AIST TODAY" are introduced here, classified by research areas.
For inquiry about the full article, please contact the author via e-mail.**

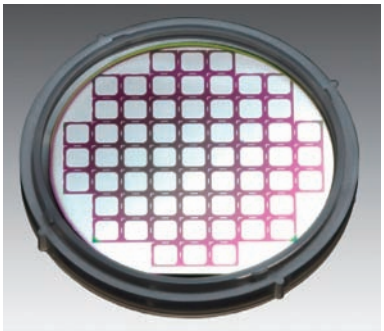
Environment and Energy

High-speed switching of high-voltage, high-power converter achieved by using SiC diodes

Size reduction of large-scale power converter systems in social infrastructure

We have developed a high-voltage, high-power converter prototype (300 kVA single-phase three-level power converter) using switching modules with SiC-PiN diodes and Si-IEGTs. The switching frequency of the modules is 2 kHz, which is four times higher than the conventional switching module.

The new switching module uses 4 × 4 mm SiC diodes (6 kV-class), developed through AIST's technology for large-area SiC devices, and Toshiba's Si-IEGTs. In switching modules using Si-IEGTs and Si diodes, the switching frequency is limited to about 500 Hz because of the limitations of Si diodes performance. The use of SiC diodes, with their excellent switching characteristics, results in a higher switching frequency of the switching modules (2 kHz). Due to the high switching frequency, the three-level converter can be adopted and the insulating transformer can be eliminated. In addition, the filter capacity can be reduced. Great size reduction of power converter systems (one-fifth compared to the conventional system) becomes possible with this technology.



SiC-PiN diodes (4×4 mm^2) fabricated in AIST
SiC-PiN diodes fabricated on 2 inch 4H-SiC wafer. A silver square area corresponds to one diode chip (4×4 mm^2).

Hiromichi Oohashi

Advanced Power Electronics Research Center (current affiliation: Energy Technology Research Institute)

h.oohashi@aist.go.jp

AIST TODAY Vol.10 No.4 p.20 (2010)

Structure and conduction mechanism of lithium ion in ionic liquid

Design of electrolyte materials for lithium secondary batteries with high charge rate

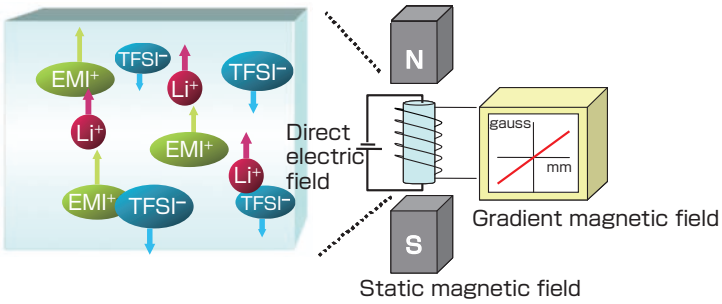
Diffusion coefficients and mobilities of the ionic species of lithium electrolytes were measured using the electric field applying pulsed gradient spin-echo NMR technique. Analyses of the dynamic values showed that the lithium ion electrolyte was coordinated by four bis(trifluoromethylsulfonyl)amide anions ($TFSI^-$s) and formed a cluster structure as $Li(TFSI)_4^{3-}$ in an equilibrium state. The electrolyte is lithium bis(trifluoromethylsulfonyl)amide dissolved in ionic liquid, 1-ethyl-3-methylimidazolium (EMI) bis(trifluoromethylsulfonyl)amide. High mobility was observed only when the electric field larger than a certain strength was applied to the electrolyte. This indicates that the ions are orientated in response to the electric field and the orientation may be due to the dielectric polarization of the ions. High mobility reveals the formation of pathways for efficient ion migration by the aligned ions. These results lead to the systematic designing of electrolyte materials applicable to the lithium secondary batteries with high charge rate.

Yuria Saito

Research Institute for Ubiquitous Energy Devices

yuria-saitou@aist.go.jp

AIST TODAY Vol.10 No.5 p.14 (2010)



Sketch of electric field applying pulsed gradient spin-echo NMR

A novel easy-to-recycle Li-Cu secondary battery

A large-capacity and low-cost rechargeable battery with metallic electrodes

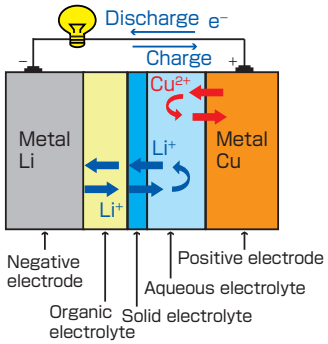
A novel secondary battery with metal Cu positive electrode and metal Li negative electrode has been developed. Aqueous electrolyte used for Cu electrode and non-aqueous electrolyte used for Li electrode are connected together by a glassy state electrolyte film through which only lithium ions can pass. During the charge and discharge processes, the dissolution-deposition of Cu (or Li) electrode and lithium ions transfer between aqueous electrolyte solution and non-aqueous solution occur. The highly reversible dissolution-deposition process of Cu metal positive electrode results in a capacity of 843 mAh g^{-1}, which is much higher than those of conventional positive electrodes. The active electrode materials of this new type of Li-Cu secondary battery are recyclable.

Haoshen Zhou

Energy Technology Research Institute

hs.zhou@aist.go.jp

AIST TODAY Vol.10 No.4 p.19 (2010)



Rechargeable battery	Capacity of positive electrode (Based on active material)	Recycle type
Lithium-ion battery	About 120–150 mAh/g	×
Lithium-copper rechargeable battery	840 mAh/g	○

Left: Schematic illustration of the novel "Li-Cu secondary battery", Right: Comparison of performance of the Li-Cu secondary battery and a conventional lithium-ion battery
Capacity of positive electrode is shown as capacity per unit weight of active material used in the positive electrode (mAh/g).

The driving force to generate new neurons in adult mammalian brain

Regulatory cues on both coding and non-coding genomic regions to promote adult neuron production and the diversity

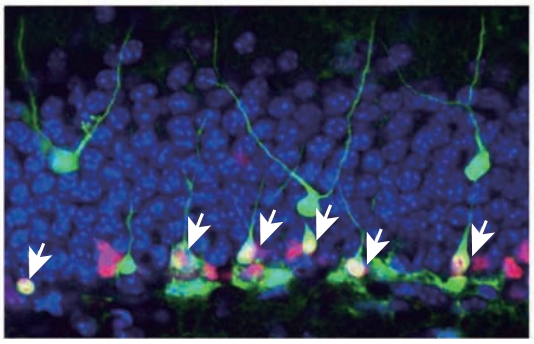
The direct relevance of neurogenesis in the adult mammalian brain to neural function and plasticity, and potentially to some neurological diseases, including Alzheimer's desease, depression and neurodegenerative conditions, is becoming well-established. Therefore, the molecular mechanisms underlying the control of adult neurogenesis have major implications for neurobiology. We have identified an important and surprisingly simple mechanism that triggers adult neurogenesis. The mechanism links intriguingly between genomic coding and non-coding regions including LINE-1 retro-elements, a family of mobile DNA elements that might contribute to neuronal diversification in the adult brain.

Tomoko Kuwabara

Research Center for Stem Cell Engineering

t.warashina@aist.go.jp

AIST TODAY Vol.10 No.5 p.12 (2010)



In control mice, newborn cells (green) are present among the more differentiated neurons deeper in the hippocampal granule cell layer.

Development of an array analysis system that assesses optimum ligands rapidly

The system is readily applicable to the purification of various antibody drugs

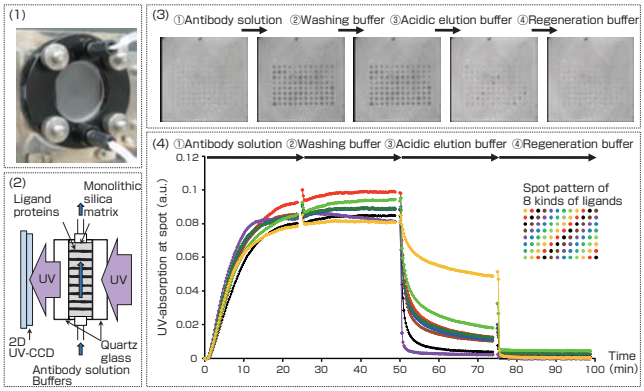
NEDO, Shimadzu Co., AIST, Kyoto Monotech Co., and JBA have jointly developed a protein array system which assesses optimum ligands for purification of various antibody drugs. The protein array is set into a flow-cell [(1)], so that solutions can pass through the inside of the array-matrix. Then UV light is applied to the protein array and transmitted light is monitored by a CCD camera to observe antibody directly without labeling [(2)]. It is observed that UV absorption at each spot increases when an antibody solution is applied, and decreases when an acidic elution buffer is applied, showing antibody binding to ligands and dissociation from ligands [(3)]. Quantified UV absorption at each spot shows that dissociation speed of antibody greatly varies among the ligands [(4)]. So, this array system enables high-throughput analysis of ligands' properties and assessment of optimum ligands for purification. Using this array system, we are attempting to establish methods to improve purification processes for the safe and cost-effective antibody drugs.

Kiyonori Hirota

Biomedical Research Institute

kiyo.hirota@aist.go.jp

AIST TODAY Vol.10 No.6 p.10 (2010)



(1) The array set in the flow-cell
(2) Mechanism of measurement (side view of array's cross section)
(3) An example of measurement for 8 kinds of ligands
(4) Quantified UV absorption at each spot

Singing synthesis technology by mimicking user's singing

VocaListener: Synthesis of more natural singing by mimicking pitch and dynamics of a user's singing voice

We have developed a singing synthesis system named *VocaListener* that automatically estimates parameters (pitch and dynamics) for singing synthesis by mimicking a user's singing voice with the help of song lyrics. Since a natural voice is provided by the user, the synthesized singing voice mimicking it can be human-like and natural without time-consuming manual adjustments.

VocaListener iteratively estimates singing synthesis parameters so that the synthesized singing can become more similar to the user's singing in terms of pitch and dynamics. The iterative estimation provides robustness with respect to different singing synthesis systems and their singer databases. Moreover, VocaListener has a highly accurate lyrics-to-singing synchronization function, and we also provide an interface that lets a user easily correct synchronization errors just by pointing them out. In addition, VocaListener also has a function to improve synthesized singing as if the user's singing skills were improved.

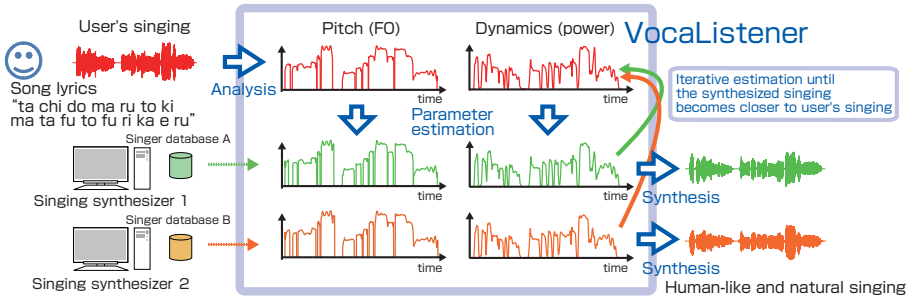
Demonstration videos including examples of synthesized singing are available at http://staff.aist.go.jp/t.nakano/VocaListener/.

Tomoyasu Nakano
t.nakano@aist.go.jp

Masataka Goto
m.goto@aist.go.jp

Information Technology Research Institute

AIST TODAY Vol.10 No.6 p.12 (2010)



Overview of VocaListener that automatically estimates parameters for singing synthesis from user's singing voice and its song lyrics

Hydrophilic plastic film with nanostructured surface made by nano-imprinting process

The use of nanostructured dies realizes improved light transmission and wettability of transparent panels

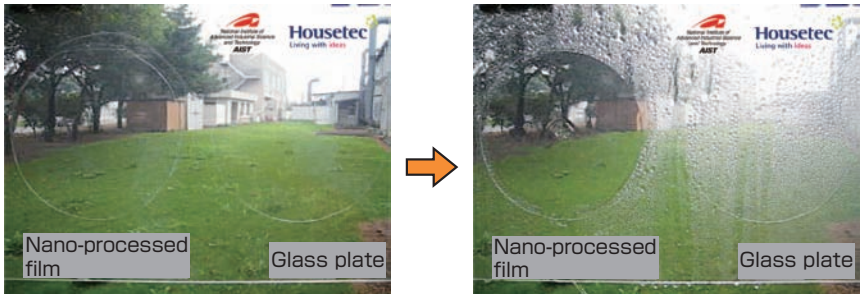
We have developed a large-area hydrophilic plastic film using a nanostructure. Using a transparent poly(ethylene terephthalate) (PET) film with the pillarlike nanostructure, the super-hydrophilic effect on the plastic surface emerged when the surface was covered with huge amounts of water droplets. The super-hydrophilic effect continued for over 60 days which was longer than that of other hydrophilic materials. In addition, the PET film with the pillarlike nanostrcture was produced by an imprinting process only. Thus, it was confirmed that super-hydrophilic PET film can be prepared by a low-cost process.

Kazuma Kurihara

Research Center for Ubiquitous MEMS and Micro Engineering

k.kurihara@aist.go.jp

AIST TODAY Vol.10 No.5 p.13 (2010)



Wettability of nanostructure-transferred film

Development of low operating temperature SOFCs using a zirconia-based electrolyte

Achieved 1 W/cm^2 at the operating temperature as low as 600 °C

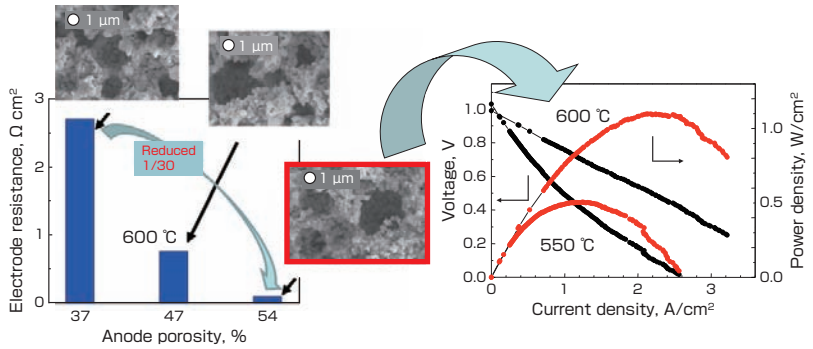
We investigated a correlation between the microstructure of the anode of a tubular solid oxide fuel cell (SOFC) and its electrochemical property. It was found that the electrochemical property of the cell was extensively improved when the size of constituent particles were reduced in a highly porous microstructure. Based on the results, an improved tubular SOFC was prepared using a conventional zirconia-based electrolyte, Ni cermet and (La, Sr)(Co, Fe) O_3 for anode and cathode materials, respectively and it showed outstanding power density of over 1 W/cm^2 at as low as 600 °C operating temperature. Thus, a zirconia-based cell could be utilized for low temperature SOFC systems under 600 °C just by optimizing the microstructure of the anode and operating conditions.

Toshio Suzuki

Advanced Manufacturing Research Institute

toshio.suzuki@aist.go.jp

AIST TODAY Vol.10 No.6 p.11 (2010)



Relation between electrode resistance and anode porosity, and cell performance of the cell with 54 % anode porosity (SEM images show reduced anode microstructure.)

Development of a stabilized laser source with a sub-Hertz linewidth

Towards an ideal oscillator for the steady "ticks"

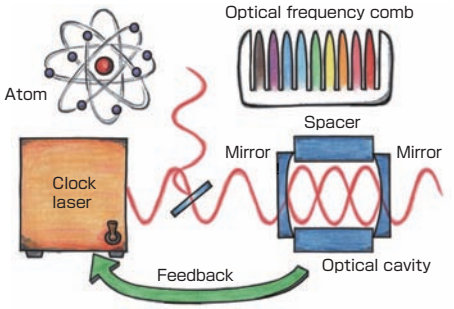
All clocks based on periodic phenomena in nature are composed of three elements: oscillator, reference, and counter. In the case of optical clocks, an ultra-stable laser is used as the local oscillator. (The reference is a narrow-linewidth transition in atoms, and the counter is an optical frequency comb.) The ultra-stable laser with a narrow linewidth and high frequency stability is essential to the operation of optical clocks, and the heart of a stable laser system is a high finesse optical cavity which must be situated in a good environment to isolate it from perturbations and should require an excellent design to be insensitive to perturbations. At the National Metrology Institute of Japan, AIST, an ultra-stable laser has been developed to drive a narrow linewidth clock transition $^1S_0$-$^3P_0$ in Yb atoms. Using very cold Yb atoms bound in optical lattices and optical frequency combs, we have been able to measure the frequency of the clock transition very precisely. The determined frequency was adopted as one of 'recommended standard frequencies' by the International Committee for Weights and Measures (CIPM).

Kazumoto Hosaka

Metrology Institute of Japan

kazu.hosaka@aist.go.jp

AIST TODAY Vol.10 No.4 p.21 (2010)



Frequency stabilization of a clock laser

A technique for quantitative measurement of acoustic cavitation generated by high pressure ultrasound

Measurement technology of generated acoustic cavitation for development of ultrasound medicine

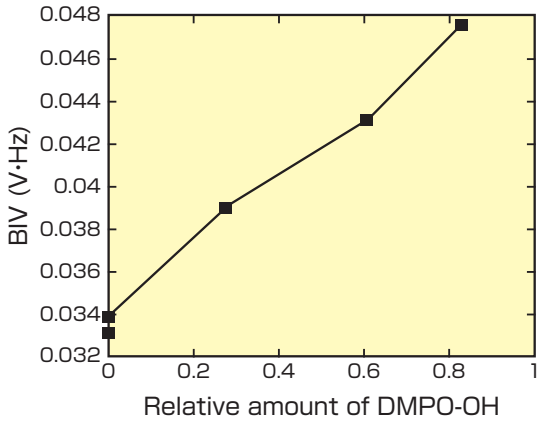
We have been studying quantitative measurement technique for acoustic cavitation in water. We have considered the use of broadband integrated voltage (BIV). BIV was calculated from the high frequency components of signals generated from the cavitation measured by a hollow cylinder type cavitation sensor. The result showed that BIV was dependent on the dissolved oxygen level in distilled water that was used as one of the parameters of the cavitation generation. Also, there was correlation between BIV and the amount of active oxygen species generated by the cavitation. This indicates that BIV has the potential to be used as an index of the amount of generated cavitation in water. In the future, this technique will be applicable to the ultrasound equipment such as an ultrasonic washing machine.

Takeyoshi Uchida

Metrology Institute of Japan

takeyoshi.uchida@aist.go.jp

AIST TODAY Vol.10 No.6 p.13 (2010)



Relationship between broadband integrated voltage (BIV) and the generated active oxygen species.

In Brief

Deputy Chairman of the Federation Council of Russia Visits AIST Tsukuba

On April 23, 2010, AIST Tsukuba was visited by Mr. Mikhail Efimovich Nikolaev, Deputy Chairman of the Federation Council of Russia, three members of the Russian parliament, and people of business in a party which totaled 14 people.

Deputy Chairman Nikolaev visited Japan on the invitation of Mr. Satsuki Eda, President of the House of Councillors of Japan. Apart from talks with Japanese politicians, he had interest in science and technology especially in robots for medical care. Concerning exchange of science and technology between AIST and Russia, he was highly interested in the exchange with the Russian Academy of Sciences, and Russian universities.

After hearing an overview presentation of AIST, he toured, with Vice-President Masakazu Yamazaki, Science Square TSUKUBA where he observed mainly robots, the therapeutic robot, Paro, the human nasal model for endoscopic surgery, and the myoelectric sensor.

During the tour, he tried the myoelectric sensor himself, and received explanation on how it can be applied to prosthetic arms. During his viewing of the "robotic arm for the disabled" of the Intelligent Systems Research Institute, the video presentation and the explanation by the researcher attracted his attention, and he expressed interest in practical implementations such as when it would be commercialized.

Deputy Chairman Nikolaev (second from right) and Vice-President Yamazaki

Secretary of State for the Development of the Capital Region of France Visits AIST Tsukuba

On March 9, 2010, Mr. Christian Blanc, Secretary of State for the Development of the Capital Region of France visited AIST Tsukuba.

Secretary of State Blanc is in charge of the "Grand Paris (a large capital region around Paris)" project which aims to enhance the attractiveness and development of the city, thus maintaining its position as a major metropolis. The main objective of his visit to Japan is to understand how the capital region around Tokyo is developed. He was especially interested in the development strategies of districts which play core roles in specific domains and regions that specialize in innovation, creation and research, and the economic contribution of the transportation network. He came to Tsukuba to verify points of his project theme, and he experienced firsthand the distance between the center of Tokyo and Tsukuba and the transportation infrastructure in a very short time.

Although he used to the utmost Tsukuba Express, the new infrastructure that joins Tokyo and Tsukuba, his schedule was tight and he could only stay for one hour. During his visit to Tsukuba, he toured in a chartered bus, and the only stop he made was at AIST.

Dr. Akira Ono, Senior Vice-President of AIST, welcomed Secretary of State Blanc at AIST and showed him places of his request, namely the Humanoid Research Group of the Intelligent Systems Research Institute and the AIST-CNRS Joint Robotics Laboratory. At this laboratory, AIST and Centre National de la Recherche Scientifique (CNRS) of France are jointly conducting research on intelligent robotic systems. 12 researchers mainly from CNRS are working there, and students from prestigious schools such as École Centrale Paris are receiving training.

Though his stay was short, he watched the demonstration of the humanoid robots and listened attentively to the lectures on its mechanism and ways of utilization. He enthusiastically questioned the researchers and showed high interest in the subject.

Secretary of State Blanc (right) and the humanoid robot HRP-2 under operation

Secretary of State Blanc (second from left) watching the female-type humanoid robot HRP-4C

Cover Photos
Above: SiC-PiN diodes fabricated on 2 inch 4H-SiC wafer (p. 14)
Below: Wettability of hydrophilic plastic film with nanostructured surface (p. 17)

AIST TODAY
International Edition
2010-3 No.37
Published in September, 2010

AIST NATIONAL INSTITUTE OF ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY (AIST)

Publication Office, Public Relations Department
National Institute of Advanced Industrial Science and Technology (AIST)

AIST Tsukuba Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan
TEL: +81-29-862-6217 FAX: +81-29-862-6212 Email: prpub@m.aist.go.jp URL: http://www.aist.go.jp/