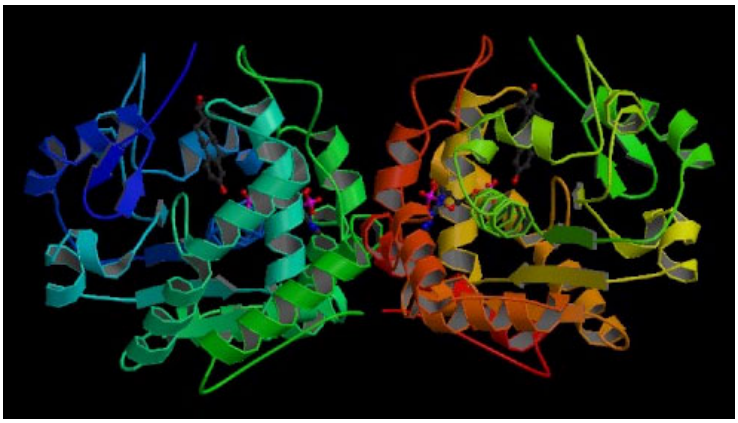
CBI 2006

Chem-Bio Informatics Society
情 報 計 算 化 学 生 物 学 会

CBI 2006 PROCEEDINGS

Integration of Chem-Bio-Pharma Informatics for Rational Drug Development and e-ADMET



Human SULT1A1 complexed with PAP and Estradiol

July 24-26, 2006

Komaba Eminence, Tokyo

CBI 2006 PROCEEDINGS

Integration of Chem-Bio-Pharma Informatics
for Rational Drug Development and e-ADMET

Chem-Bio Informatics Society

CONTENTS

What's CBI

The CBI Association, established in 1981, sponsored nearly two hundred meetings, such as lectures, seminars, and symposia on chemical computing, molecular biology, and informatics during the past two decades. All of these activities were financially supported by members of the CBI Company association, which included a wide range of leading Japanese companies in the fields of pharmaceuticals, chemicals and computers as well as informatics. Hitherto, this association stimulated Japanese academic and industrial activities related to computer-aided drug design, protein engineering, structure-based biology, biochemical (DNA and protein) chips, biocomputing, applications of pattern recognition, artificial intelligence and the Internet, computational toxicology, and genomic technologies. In April 2000, the CBI Association changed its structure and name to become the CBI Society with a new mandate to enhance its activities on the basis of the previous association and to evolve towards the new era of biotechnology in the 21st century. The executive team of the CBI Society comprises leading scientists from universities, industrial laboratories, and national research institutes.

CBI 2006 PROCEEDINGS

Edited by Takashi Mizuma
July 24, 2006

Chem-Bio Informatics Society(CBI)
Office: Iida Bldg., Room 301
4-3-16, Yoga, Setagaya-ku
Tokyo, 158-0097, JAPAN
Tel: +81-3-5491-5423 Fax: +81-3-5491-5462
Email: cbistaff@cbi.or.jp

Preface

Welcome to the 2006 CBI Annual Meeting entitled
Integration of Chem-Bio-Pharma Informatics for Rational Drug Development and e-ADMET

In May 2005, our CBI Society held a seminar entitled "Toward development of e-ADMET." The fact that so many attended the workshop made us realize again that CBI society members and non-members alike are very interested in this issue. ADMET, which represents both pharmacokinetics and toxicity, has had a key role in drug development and usage; and the recognition of its importance has been continuing to increase more and more. Nevertheless, there are many problems with ADMET, such as its cost, speed, and so on, to be solved to further expand its use for drug development.

To overcome these problems, we need to establish a novel method. One way to do this is to develop e-ADMET, by which we can enable advancements in drug development by relying less on "wet" approaches (experimentation), and more on "dry" approaches (informatics and computational science) that can efficiently utilize the knowledge accumulated by our predecessors. Once we establish this approach, the way of carrying out drug development becomes changed, resulting in monetary savings and more efficient use of personnel, materials, and time. Moreover, issues regarding the use of experimental animals and human tissues become less related to drug development.

I believe that for rational drug development to be successful, it is necessary to know chemistry and biology, and to understand phenomena and their mechanisms derived from interactions between chemicals and the body. Pharmaceutical science is a study integrating these issues. Medicines (drugs) are not just about chemicals. Medicines accompanied by information is "Medicine." The members of CBI society are multidisciplinary; that is, our fields cover chemistry, biology, informatics, computational science, pharmaceutical science, and their "wet" and "dry" studies. This means that our society can provide the best platform for a brainstorming project about rational drug development.

On the basis of this current background, the 2006 annual meeting of the CBI society is entitled "Integration of Chem-Bio-Pharma Informatics for Rational Drug Development and e-ADMET." I hope that the 2006 annual meeting, which consists of thematic sessions as well as regular sessions, becomes an indispensable opportunity for attendees to get something new and innovative for advancing rational drug development and e-ADMET.

Takashi Mizuma, Ph.D. Chairman of the 2006 annual meeting
(Tokyo University of Pharmacy and Life Science)

Host

Chem-Bio Informatics Society

Benefactors

The Chemical Society of Japan
The Pharmaceutical Society of Japan
The Japanese Society for Artificial Intelligence
The Biophysical Society of Japan
The Japanese Society for the Study of Xenobiotics
Molecular Biology Society of Japan
Japanese Society for Bioinformatics
Protein Science Society of Japan
The Japan Society of Drug Delivery System
The Academy of Pharmaceutical Science and Technology, Japan
Information Processing Society of Japan
Japan Association for Medical Informatics

Contributors

AdIn Research, Inc.
Daiichi Pharmaceutical Co.,Ltd.
Eisai Co.,Ltd.
Mitsui Chemicals Inc.
Mochida Pharmaceutical Co.,Ltd.
Taisho Pharmaceutical Co.,Ltd.
Taiho Pharmaceutical Co.,Ltd.
Tokyo University of Pharmacy and Life Science

Exhibition & Luncheon Seminor Companies

AdvanceSoft Corporation
Cerep Japan
CTC Laboratory Systems Corporation
Fujitsu Ltd.
Hitachi, Ltd.
IBM Japan, Ltd.
Infocom Corporation
Mathematical Systems, Inc.
NABE International Corporation
NalaPro Technologies, Inc.
PharmaDesign, Inc.
Ryoka Systems Inc.
Sumisho Computer Systems Corporation
SGI Japan, Ltd.

Talous Inc.
Cybernet Systems Co.,Ltd.
Wavefunction, Inc.

CBI Consortium Members

Accelrys Inc.
AdIn Research, Inc.
AdvanceSoft Corporation
Ajinomoto Co.,Inc.
Astellas Pharma Inc.
Banyu Pharmaceutical Co.,Ltd.
CAC Corp.
Chugai Pharmaceutical Co.,Ltd.
CTC Laboratory Systems Corporation
Daiichi Pharmaceutical Co.,Ltd.
Dainippon Sumitomo Pharma Co.,Ltd.
DNA Chip Research Inc.
Eisai Co. Ltd.
Fuji Photo Film Co. Ltd.
Fujitsu Ltd.
GlaxoSmithKline K.K.
Hitachi, Ltd.
IBM Japan, Ltd.
Infocom Corp.
Mitsui Chemicals Inc.
Mizuho Information & Research Institute, Inc.
Mochida Pharmaceutical Co.,Ltd.
NEC Corporation
Otsuka Pharmaceutical Co.,Ltd.
Sankyo Co.,Ltd
Shionogi & Co.,Ltd.
Sony Corp.
Sumisho Computer Systems Corporation
Taiho Pharmaceutical Co.,Ltd.
Taisho Pharmaceutical Co.,Ltd.
Tanabe Seiyaku Co.,Ltd.
Teijin Pharma Ltd.
Toray Industries Inc.
Ryoka Systems Inc.
Yokogawa Electric Corporation

Organizer

Chairman

Takashi Mizuma (Tokyo University of Pharmacy and Life Science)

Chairs

Yukio Tada (Taiho Pharmaceutical Co., Ltd)

Tsuguchika Kaminuma (Tokyo Medical and Dental University)

Masami Uebayasi (AIST)

Yuji Takaoka (Taisho Pharmaceutical Co., Ltd.)

Akihiko Konagaya (RIKEN Genomic Science Center)

Kazuro Shimokawa (RIKEN Genomic Science Center)

Akihiko Hirose (National Institute of Health Sciences)

Hiroshi Tanaka (Tokyo Medical and Dental University)

Yoshinori Harada (Life Science Group, Hitachi,Ltd.)

Yoshiro Nakata (Gunma University)

Program

Program for CBI2006

July 24th (Mon.)

10:00		Opening Remarks for Japanese Session <Let's Share and Solve our Common Problems> Takashi Mizuma (Chairman of The 2006 Annual Meeting)
		10:05<Problem A> Proposal：Hiroyuki Kusuhara((Univ. of Tokyo) Answer： Motohiro Kato(Chugai Pharmaceutical Co. Ltd.) 10:35<Problem B> Proposal：Eiichiro Ichiishi(Tohoku University) Answer： Shigeki Mitaku(Nagoya University) Answer： Tomokazu Konishi(Akita Prefectural University) 11:30<Problem C> Proposal：Makoto Tsuruoka(Tokyo University of Technology) 11:35<Problem D> Proposal：Yoshihiro Mori(Okayama University of Science)
11:50		Opening Remarks Yukio Tada (President of The CBI Society)
12:00		Luncheon Seminar by **AdvanceSoft Corporation**
13:15		Opening Speech Takashi Mizuma (Chairman of The 2006 Annual Meeting)
13:30		Regular Session 1: Area 1 & 2 <Keynote & Invited Talks>
	S01	13:30-14:15
		Kazuo Kitaura (Research Institute for Computational Sciences, AIST) "Quantum Chemical Calculations of Binding Energy between Protein and Ligand" Chairman: Tatsuya Nakano (National Institute of Health Sciences)
	S02	14:15-15:00
		Masaki Tomimoto (SGX Pharmaceuticals Inc.) "Fragment-Based Drug Discovery: Concept and Example" Chairman: Yuji Takaoka (Taisho pharmaceutical Co., Ltd.)
		15:00-15:30 Break
	S03	15:30-16:30
		Ruben Abagyan (The Scripps Research Institute) "Computational Structural Proteomics and Drug Discovery" Chairman: Mayuko Takeda-Shitaka (Kitasato Univ.)
16:30		Poster Oral Session (Poster ID: P1-03, P1-06, P2-01, P2-02, P3-03, P3-09)
17:00 18:00		Poster Session

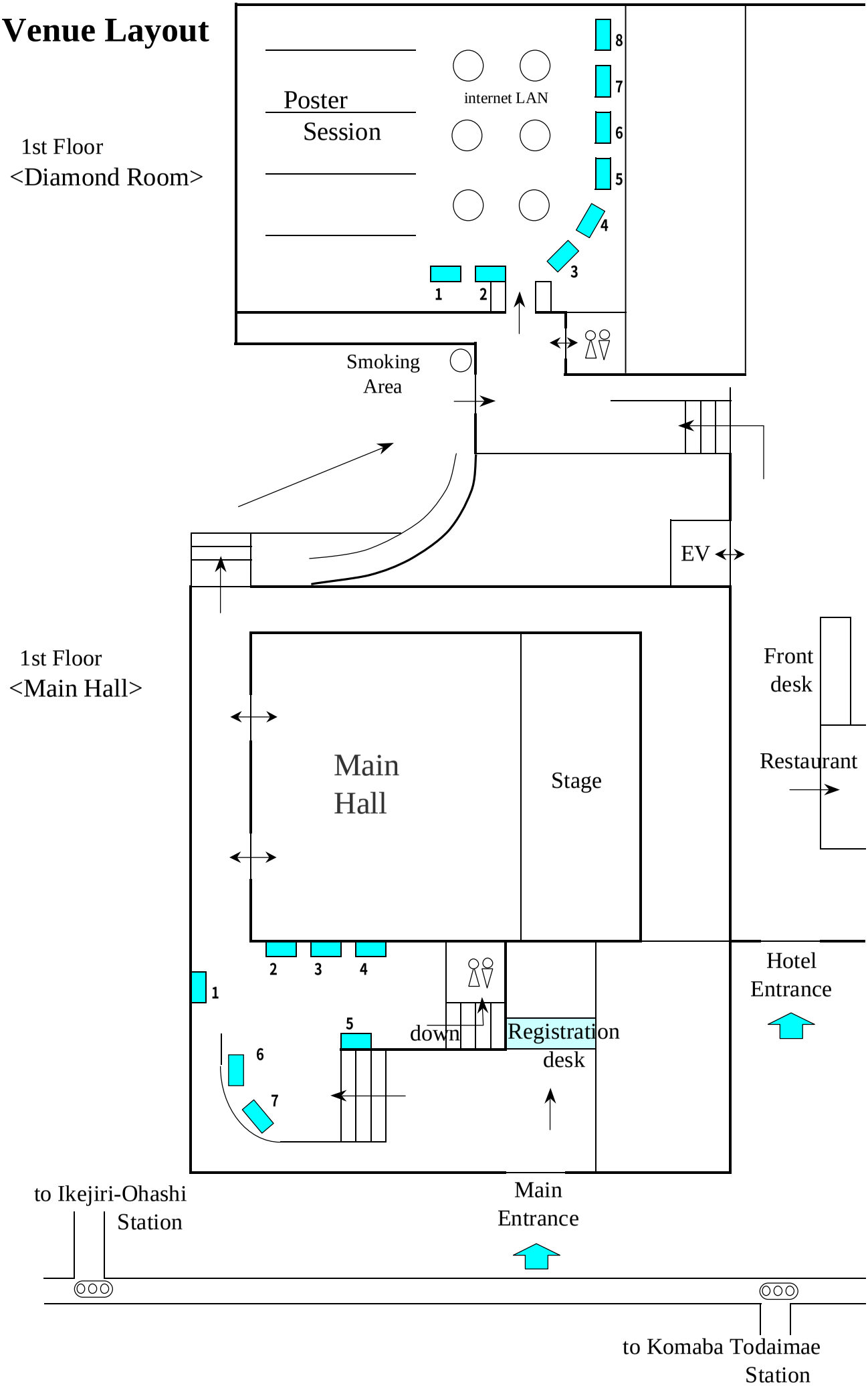
July 25th (Tue.)

9:30		Annual theme session 1 (Wet study of ADMET) <Keynote & Invited Talks>
	S04	9:30-10:30
		Svein Oie (College of Pharmacy The Univ. of Georgia) “Optimizing use of Pharmacokinetics and Toxicokinetics in Drug Discovery” Chairman: Takashi Mizuma (Tokyo Univ. of Pharmacy and Life Science)
	S05	10:30-11:15
		Mitsuru Hashida (Kyoto Univ.) "Development of DDS: Analysis and Control of Drug Disposition in the Body" Chairman: Toshiharu Horie (Chiba Univ.)
	S06	11:15-12:00
		Atsushi Ono (National Institute of Biomedical Innovation) "Current topics of toxicogenomics - Challenges on issues in Toxicogenomics Project -" Chairman: Akihiko Hirose (National Institute of Health Sciences)
12:00		Luncheon Seminar by **Fujitsu Ltd.**
13:30		Regular Session 2: Area 3 & 4 <Keynote & Invited Talks>
	S07	13:30-14:30
		Kyoung Tai No (Yonsei Univ.) "Introduction of Integrated e-Drug Discovery System, Lead Explore" Chairman: Akihiko Konagaya (RIKEN, Genomic Science Center)
	S08	14:30-15:15
		Shigehiko Kanaya (Nara Institute of Science and Technology) "KNApSAcK: Secondary metabolite database related between metabolites and species" Chairman: Akihiko Konagaya (RIKEN, Genomic Science Center)
		15:15-15:45 Break
	S09	15:45-16:30
		Sumi Yoshikawa (RIKEN, Genomic Science Center) "Pharmaco-Ontologies and their Roles" Chairman: Akihiko Konagaya (RIKEN, Genomic Science Center)
16:30		Poster Oral Session (Poster ID: P3-11, P4-02, P4-04, P5-07, P5-11, P5-16)
17:00		Poster Session
18:00		Poster Award

July 26th (Wed.)

9:30		Annual theme session 2 (Dry study of ADMET) <Keynote & Invited Talks>
	S10	9:30-10:30
		Sean Ekins (Arizona Commodity Traders, LLC&Univ. of Maryland) "Combining Quantitative Structure Activity Relationships and Systems Biology Approaches for *in silico* ADME/Tox" Chairman: Yasushi Okazaki (Saitama Med. Schl. Res. Ctr. for Genomic Medicine)
	S11	10:30-11:15
		Edward H. Kerns (Wyeth Res.) "Current Strategies for ADMET Integration to Enhance Efficiency and Quality in Drug Discovery" Chairman: Fumiyoshi Yamashita (Kyoto Univ.)
	S12	11:15-12:00
		Kazuya Maeda (Univ. of Tokyo) "*In silico* database, simulation and prediction for the optimization of pharmacokinetic properties of drug candidates" Chairman: Takashi Mizuma (Tokyo Univ. of Pharmacy and Life Science)
12:00		Luncheon Seminar by **Cerep Japan**
13:30		Regular Session 3: Area 6 & 7 <Keynote & Invited Talks>
	S13	13:30-14:20
		Akihiko Takano (National Institute of Informatics) "Information Access based on Association" Chairman: Yoshinori Harada (Hitachi, Ltd.)
	S14	14:20-15:10
		Asako Koike (Hitachi, Ltd.) "Biomedical text mining: state of the art and future direction" Chairman: Yoshinori Harada (Hitachi, Ltd.)
		15:10-15:25 Break
	S15	15:25-16:15
		Teruhiko Yoshida (National Cancer Center) "Cancer Genomics: Chip, SNP, Database" Chairman: Hiroshi Tanaka (Tokyo Med. Dent. Univ.)
	S16	16:15-17:05
		Kunio Shiota (Univ. of Tokyo) "Application of Epigenetics on Development of Drugs" Chairman: Takatoshi Kawai (Eisai Co.,Ltd.)
17:10 17:30		Closing Remarks

Venue Layout



Exhibition/ Luncheon Seminor/ Advertisement Companies

Companies	Exhibition	Luncheon Seminor	Advertisement
	Diamond Room		
Cybernet Systems Co.,Ltd.	1		○
Talous Inc.	2		○
CTC Laboratory Systems Corporation	3		○
Hitachi, Ltd.	4		○
PharmaDesign, Inc.	5		○
SGI Japan, Ltd.	6		○
Wavefunction, Inc.	7		○
Mathematical Systems Inc.	8		○
	Main Hall Foyer		
Infocom Corporation	1		○
Sumisho Computer Systems Corporation	2		○
Ryoka Systems Inc.	3		○
Fujitsu Ltd.	4	July 25	○
NABE International Corporation	5		○
AdvanceSoft Corporation	6	July 24	○
NalaPro Technologies, Inc.	7		○
Cerep Japan		July 26	○

Notice

Poster Awards

- Participants are requested to vote for the best poster.
- The Best Poster Awards will be given for five excellent presentations and posted on the display board on July 25th, 18:00 in Diamond Room.
- Winners of the Poster Prize will be decided by the CBI Journal Publishing Committee based on the votes of conference participants.
- The winners will receive the awards plus a privilege of exception from submission charges to the CBI Journal. The winners will also be invited to attend the Dining Party (only one person per poster).

Dining Party (Buffet style)

- Date: July 25th, 18:00 ~ 20:00
- Place: "Diamond Room"

Keynote & Invited Talks

Keynote and Invited Talks

S01 **Quantum Chemical Calculations of Binding Energy Between Protein and Ligand**
Kazuo Kitaura
Research Institute for Computational Sciences (RICS)
National Institute of Advanced Industrial Science and Technology (AIST)

S02 **Fragment-Based Drug Discovery: Concept and Example**
Masaki Tomimoto
Associate Director, Computational Chemistry, SGX Pharmaceuticals, Inc.

S03 **Computational Structural Proteomics and Drug Discovery**
Ruben Abagyan
The Scripps Research Institute

S04 **Optimizing use of Pharmacokinetics and Toxicokinetics in Drug Discovery**
Svein Øie
College of Pharmacy, University of Georgia

S05 **Development of DDS:Analysis and Control of Drug Disposition In the Body**
Mitsuru Hashida
Graduate School of Pharmaceutical Sciences, Kyoto University

S06 **Current topics of toxicogenomics– Challenges on issues in Toxicogenomics Project –**
Atsushi Ono
Toxicogenomics Project, National Institute of Biomedical Innovation

S07 **Introduction of Integrated e-Drug Discovery System, Lead Explore**
Kyoung Tai No
Department of Biotechnology, Yonsei University
Bioinformatics and Molecular Design Research Center (BMDRC)

S08 **KNApSAcK: Secondary metabolite database related between metabolites and species**
Shigehiko Kanaya
Graduate school of Information Science, Nara Institute of Science and Technology

S09 **Pharmaco-Ontologies and their Roles**
Sumi Yoshikawa
RIKEN Genomic Science Center

S10 Combining Quantitative Structure Activity Relationships and Systems Biology Approaches for in silico ADME/Tox

Sean Ekins
ACT LLC and Department of Pharmaceutical Sciences, University of Maryland

S11 Current Strategies for ADMET Integration to Enhance Efficiency and Quality in Drug Discovery

Edward H. Kerns
Wyeth Research

S12 *In silico* database, simulation and prediction for the optimization of pharmacokinetic properties of drug candidates

Kazuya Maeda[1], Hiroyuki Kusuhara[1], Yoshihisa Shitara[2], Motohiro Kato[3], Yuichi Sugiyama[1]
[1]Graduate School of Pharmaceutical Sciences, The University of Tokyo, [2]Graduate School of Pharmaceutical Sciences, Chiba University, [3]Chugai Pharmaceutical Co., Ltd.

S13 Information Access based on Association

Akihiko Takano
National Institute of Informatics, Tokyo

S14 Biomedical text mining: state of the art and future direction

Asako Koike
Central Research Laboratory, Hitachi Ltd.

S15 Cancer Genomics; Chip, SNP, Database

Teruhiko Yoshida, Kazuhiko Aoyagi, Hiroki Sasaki and Hiromi Sakamoto
Genetics Division, National Cancer Center Research Institute

S16 Application of Epigenetics on Development of Drugs

Kunio Shiota
Laboratory of Cellular Biochemistry, Animal Resource Sciences, The University of Tokyo

S01

Quantum Chemical Calculations of Binding Energy Between Protein and Ligand

Kazuo Kitaura
Research Institute for Computational Sciences (RICS)
National Institute of Advanced Industrial Science and Technology (AIST)

We have developed the fragment molecular orbital (FMO) method[1] for quantum chemical calculations of very large molecules. In the method a molecule is divided into fragments and *ab initio* MO calculations are performed on the fragmnets, their dimers and optionally trimers to obtain the total enegy and other properties of the whole molecule. The FMO method reproduces regular *ab initio* properties with good accuracy. Various FMO-based correlation methods have been developed combining with the density functional theory(DFT), the 2nd order Møller-Plesset perturbation theory(MP2), the coupled cluster theory(CC), and MCSCF. The multilayer FMO method[2] treats the important part of the system with high accuracy and the rest with low level of theory. Very recently the polarisable continuum model (PCM) was interfaced with the FMO method[3], allowing one to treat solvent effects of real size proteins. These methods have been interfaced with the GAMESS program package[4].

We applied the method to four FK506 binding protein (FKBP) complexes (denoted by their PDB codes 1fkb, 1fkf, 1fkg, and 1fki)[5]. The geometries of the ligands in the complexes were optimized at the restricted Hartree-Fock (FMO-RHF) level using the 3-21G basis set. The binding energies were refined at a higher level of theory (2nd order Møller-Plesset perturbation theory FMO-MP2 with the 6-31G* basis set). Solvent effects on the protein-ligand binding were estimated using the FMO/PCM method. The important interactions between the protein residues and the ligand were revealed based on the FMO pair interaction analysis.

[1] D. G. Fedorov and K. Kitaura, Theoretical development of the fragment molecular orbital (FMO) method, in “Modern methods for theoretical physical chemistry and biopolymers”, edited by E. Starikow, S. Tanaka and J. Lewis, Elsevier (2006).
[2] D.G. Fedorov, T. Ishida, K. Kitaura, *J. Phys Chem. A* , 109, 2638-2646 (2005).
[3] D.G. Fedorov and K. Kitaura, *J. Comp. Chem.*, 27, 976-985 (2006).
[4] GAMESS, http://www.msg.ameslab.gov/GAMESS/GAMESS.html
[5] I. Nakanishi, D.G. Fedorov, K. Kitaura, submitted.

S02

Fragment-Based Drug Discovery: Concept and Example

Masaki Tomimoto
Associate Director, Computational Chemistry, SGX Pharmaceuticals, Inc.

Fragment-based drug discovery (FBDD) is already generating a lot of attention. FBDD has been applied to various targets not only in biotech companies but big pharmas, and it generates lead compounds and drug candidates. However, there is no standard recipe for FBDD so far. Each company develops its own FBDD strategy suited for its technological advantage. SGX Pharmaceuticals, Inc. has developed a FBDD platform that utilizes high-throughput X-ray crystallography for lead identification/optimization. The proprietary FAST™ (Fragments of Active STructures) exploits crystallographic screening to detect, visualize, and identify fragments (molecular weight 150-200, typical hit rates 1-5%) that are bound to the target protein. Each member of the FAST™ fragment library is amenable to rapid chemical elaboration at two or three points of chemical diversity using high-throughput organic synthesis. Initial lead optimization involves using knowledge of the co-crystal structure of the target-fragment complex and advanced computational chemistry tools to guide synthesis of small focused linear (one-dimensional) libraries. These linearly elaborated fragments are then evaluated with *in vitro* biochemical assays and crystallography. Thereafter, optimal variations at each point of chemical diversity are combined to synthesize focused combinatorial (two- or three-dimensional) libraries that are again examined with biochemical assays and crystallography. Lead compound series are prioritized for further medicinal chemistry and compound development efforts using the results of cellular and animal model assays, *in vitro* and *in vivo* ADME and *in vitro* toxicology studies in concert with structural information. Basic concept and application of the FAST・ fragment-based lead generation process to various oncology targets will be presented.

S03

Computational Structural Proteomics and Drug Discovery

Ruben Abagyan
The Scripps Research Institute

Abstract

Computational prediction of protein structure, protein function, and ligand binding is an area of critical importance considering the rapidly growing number of protein structures in Protein Data Bank. We focused on several related computational tasks:

(i) predicting druggable binding sites from a single structure;
(ii) predicting protein-protein interaction patches from a single structure
(iii) predicting conformations of protein loops in models by homology with high accuracy
(iv) accurate docking of flexible ligands to flexible protein pockets and virtual screening
(v) protein docking and refinement of flexible protein interfaces

We demonstrated that the small molecule binding pockets can be predicted with a certain transformation of the Lennard Jones potential [1]. This algorithm is useful in predicting new or allosteric binding sites or the feasibility of inhibiting protein-protein interaction with a small molecule. Predicting transient protein-protein interaction interfaces without knowing its partner
was also proposed and validated on a large benchmark[2].

Including the receptor pocket flexibility into account in ligand docking remains to be a task highly specific to the nature of the receptor. We attempted to formulate a docking protocol which is relatively general and includes both the side-chain sampling and loop movements [3]. The ICM docking protocol led to the discovery of novel inhibitors against a number of biomedical targets, including de novo discovery of antagonists of RAR and thyroid hormone receptor ([4]), discovery of inhibitors of alpha 1-antitrypsin amyloid formation, discovery of new antimalarial agents, and novel antagonists of the androgen receptor.

The CAPRI protein docking experiment provides a good platform for evaluation of protein docking algorithms. While a fully reliable prediction of the association geometry given two uncomplexed structures remains unattainable, considerable progress has been made in the last few years [5]. The difficulty of the problem is a function of the scale of the induced conformational changes upon association. We demonstrate that at least in some cases these changes can be correctly predicted. A new way of exchanging and browsing the chemical and structural data is also presented [6].

References and Footnotes

1. An, J., Totrov, M. and Abagyan, R. Pocketome: Comprehensive Identification and Classification of Ligand-Binding Envelopes. Mol Cell Proteomics 4, 752-61 (2005).

2. Fernandez-Recio, J., Totrov, M., Skorodumov, C. and Abagyan, A.Optimal docking area: a new method for predicting protein-protein interaction sites. Proteins, 58,134-143 (2005)

3. Cavasotto, C.N. and Abagyan, R. Protein flexibility in ligand docking and virtual screening to protein kinases J Mol Biol, 337, 209-225, (2004).

4. Schapira, M., Raaka, B.M., Das, S., Fan, L., Totrov, M., Zhou, Z., Wilson, S.R., Abagyan, R., Samuels, H.H. Discovery of diverse thyroid hormone receptor antagonists by high-throughput docking. PNAS USA, 2003, 100,7354-9.

5. Fernandez-Recio, J., Totrov, M. and Abagyan, R. ICM-DISCO docking by global energy optimization with fully flexible side-chains Proteins, 52.113-117, (2003).

6. Abagyan, R., Lee, W.H., Raush, E., Budagyan, L., Totrov, M., Sundstrom, M., Marsden, B.D.
Disseminating structural genomics data to the public: from a data dump to an animated story. Trends Biochem Sci. (2006), 31, 76-8

S04

Optimizing use of Pharmacokinetics and Toxicokinetics in Drug Discovery

Svein Øie
College of Pharmacy, University of Georgia

Pharmacokinetics (absorption, distribution, metabolism and elimination – ADME) is extremely important for developing optimal dosing regimens. Predictable and desirable plasma concentrations, compliance and optimal response depend upon the pharmacokinetic characteristics of the drug.

As cost of discovering and developing new drug entities has spiraled out of control, the use of computer design and high throughput screening of chemical libraries has become an integral part drug discovery. Although a larger number of lead compounds can be generated, most compounds still fail during clinical studies. In 1998 it was reported that at least 50 % of the drugs failed in clinical trials due to pharmacokinetics and animal toxicity. To reduce the number of failings due to pharmacokinetics and toxicity, a number of cellular, in vitro and in silico screening techniques have been adopted. In silico screening has been particular successful in identifying "drugable" chemical entities with better bioavailability, and membrane permeability.

Although these approaches have reduced the number of clinical failures due to poor absorption, they come at a cost. The current methods do not adequately consider factors such as active transport, modification of ADME through formulation and chemical modification, and that "optimal" ADME is not necessary for all compounds to be effective.

Drugs can be classified as 1, "curing" diseases or conditions, 2, "controlling" diseases and conditions or 3, temporarily "relieving" symptoms. Because these classes of drugs vary in what stricture we set to their pharmacokinetics, toxicity, drug interactions and pharmacogenetic variability, our approach to "discovering" drugs in each of the classes will also vary.

For drugs that "cures" (e.g., antivirals, antibiotics, anticancer, genetic modifiers, etc.) the treatment period is usually short and effect long-term, allowing us to tolerate moderate side effects, accept less than optimal pharmacokinetics, and making cost usually less of an issue.

For drugs that control diseases or conditions (e.g., diabetes, high blood pressure, cholesterol, etc.) the treatment period is usually long-term requiring more optimal pharmacokinetics, low potential for drug-drug interactions and side-effects and low cost. This is an area where in silico screening methods are particular useful but is still in need of improvements.

For, drugs that temporarily relieve symptoms, the treatment period is usually short, but also frequently repetitive (e.g., seasonal allergies, migraines, etc.) and optimal pharmacokinetics is again less critical but requires low potential for drug-drug interactions and side-effects.

By understanding when pharmacokinetics and toxicity play major or minor roles in effective drug treatment and classifying drugs appropriately, we will be able to make further advances in rational drug design and drug discovery.

S05

Development of DDS: Analysis and Control of Drug Disposition in the Body

Mitsuru Hashida
Graduate School of Pharmaceutical Sciences, Kyoto University

Drug delivery system (DDS) is defined as a technology for a precise control of *in vivo* behavior of a drug aiming at optimization of its therapeutic efficacy and various applications such as controlled release, targeting, and absorption enhancement have been developed being indispensable especially for protein and gene medicines. Drug targeting, an approach to concentrate the drug on the specific site of the body through manipulation of its biodistribution profile attracts great interest and many macromolecular or particulate carriers have been developed: The use of polymer can be a formidable tool because of their high diversity and multiple functions, but successful application of a polymer carrier requires understanding of the pharmacokinetics of the polymer at whole body, organ, cellular and subcellular levels in relation to their physicochemical and biological characteristics.

In a series of our investigations, biodistribution of proteins and other types of macromolecules have been characterized based on pharmacokinetic analysis and a strategy for controlling their disposition profiles through chemical modification was proposed. The fundamental properties such as a molecular weight and electric charge are revealed to determine the disposition patterns of macromolecules. To achieve cell specific delivery of proteins, receptors existing on the surface of cells can be utilized and we have developed several methods such as galactosylation, mannosylation or succinylation, which lead proteins to be recognized by special cell types in the liver. Cell-specific delivery and resultant therapeutic effects of antioxidant enzymes, recombinant human superoxide dismutase (SOD) and bovine liver catalase (CAT), with chemical modifications were demonstrated. Similar approaches have been applied to gene delivery, and specific delivery of plasmid DNA and subsequent protein expression were observed in hepatocytes, macrophages, etc. Successful results have been observed in application of these systems to DNA vaccination.

In parallel to these approaches, we have interested in application of informatics in such as dosage form design and in silico ADME-Tox predication. With a recent progress in computer technology, we become able to apply sophisticated computation algorithms to QSAR and a better performance in QSAR modeling have been realized.

S06

Current topics of toxicogenomics
- Challenges on issues in Toxicogenomics Project -

Atsushi Ono
Toxicogenomics Project, National Institute of Biomedical Innovation

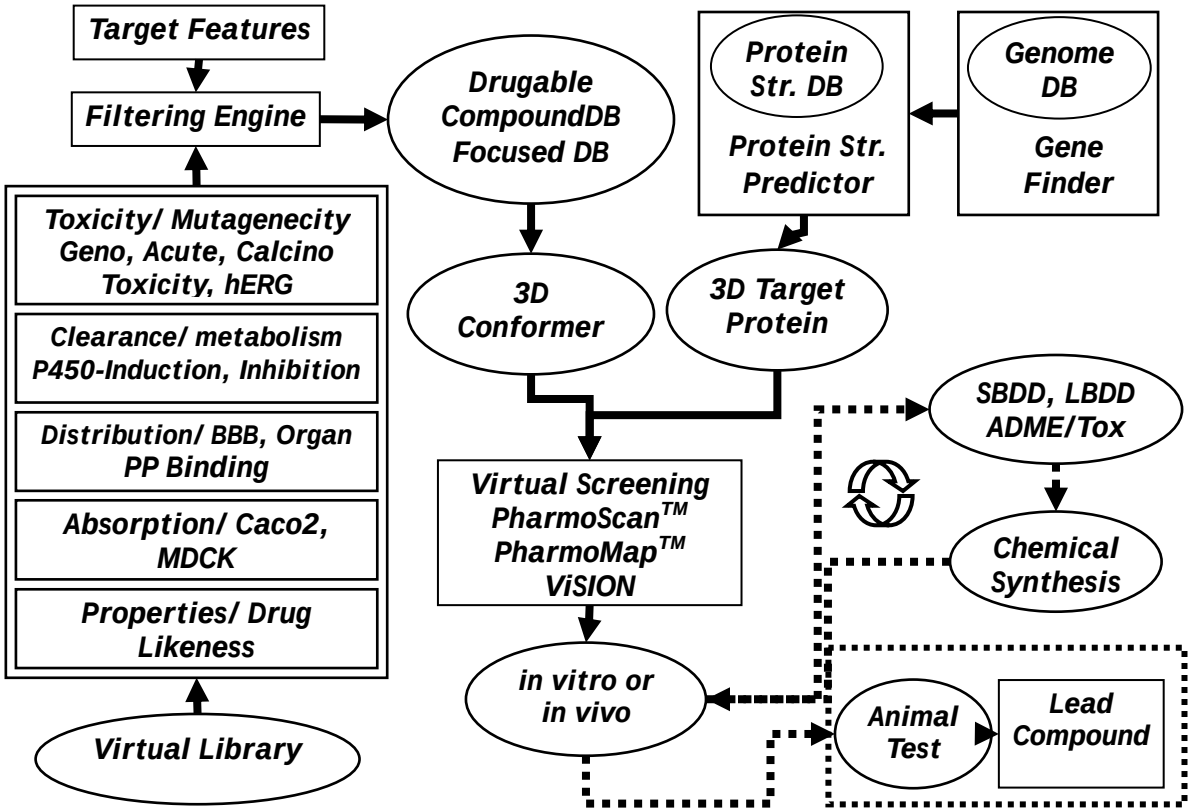
Gene expression analysis opens new perspectives in toxicology to identify mechanisms of toxicity and biological response and to improve toxicity assessment. However, there is difficulty in biological interpretation of the results obtained from the expression changes of thousands of genes because of the limitation of knowledge on relations of gene expression change and biological effect. Toxicogenomics Project (TGP) is a joint national project of National Institute of Health Sciences (NIHS), National Institute of Biomedical Innovation (NIBIO) and 15 pharmaceutical companies, which started at 2002. The goal of our project is to construct a large-scale database of over 150 drugs exposed to rats (*in vivo, in vitro*) and to human (*in vitro*), and develops a system that will forecast the toxicity of new chemical in the early stage, if possible *in vitro*, of drug development. Moreover, the bridging of the interspecies is considered using human and rat primary hepatocyte culture system in our project. TGP database integrate gene expression data analyzed by Affymetrix GeneChip with traditional toxicological endpoints including blood biochemical, hematological, and pathological data. Additional information about the toxicological, biological and medical effects of chemicals collected from the literature is stored. Strategic experimental design allows ensure high quality dataset. All stored data is fully searchable and informatics tools will aid mechanistic understanding of drug-induced toxicity, and lead to the identification of new biomarkers relate to toxic phenotypes. In order to develop a reliable systematic framework for phenotype classification and prediction from gene expression data, first critical step of analysis is to find gene sets whose expression is tightly anchored on to subset of the phenotype. We tried to use various unsupervised and supervised learning methods. In some *in vivo* case, such as glutathione depletion, nongenotoxic carcinogenesis and peroxisome proliferation, etc in the rat liver, we have succeeded in selection of biomarker genes and in classifying samples quantitatively with selected biomarker. On the other hand, prediction of *in vivo* toxicity from *in vitro* expression data as well as bridging of the interspecies has some limited by the fact that response of many genes is not coincidence in each system from our results. However, there is cluster of genes, which consistently changed *in vivo* and *in vitro*, and also in human and rat *in vitro*. Therefore, when attempting to use it for specific effects, *in vitro* data are useful to predict *in vivo* effects. The knowledge with our database will make predictive toxicology possible. The current version of our toxicity forecast system is prototype and further validation study and novel informatics to promote it is necessary. In case of its completion, a new era of toxicology "Systems Toxicology" will arrive.

S07

Introduction of Integrated e-Drug Discovery System, Lead Explore

Kyoung Tai No
Department of Biotechnology, Yonsei University
Bioinformatics and Molecular Design Research Center (BMDRC) www.bmdrc.org

An integrated e-Drug Discovery system, Lead Explore, have been developed since 1997 by BMDRC and it's collaborators in Korea and also foreign countries for providing systematic and cost effective drug discovery both for academic and industry researchers. The following diagram shows outline of Lead Explore and the components introduced.



Lead Explore consists of library design, chemical properties prediction, chemical management, ADME/Tox prediction, virtual screening, target protein analysis and pharmacophore define and some graphical interfaces components. At the presentation we will describe Lead Explore system in detail and will introduce some research results performed with the system.

S08

KNApSAcK:
Secondary metabolite database related between metabolites and species

Shigehiko Kanaya
Graduate school of Information Science, Nara Institute of Science and Technology

Many metabolites have been identified from plants, microbes and other organisms. Secondary metabolites are highly species specific and play a role for the survival of the producing organism within its natural habitat. The number of metabolites present in the plant kingdom is estimated to exceed 100,000—an enormous number indicating a large scale of structured diversity of compounds. Several databases have been made by collecting metabolite information of various organisms, and provide some chemical information and biological pathways on metabolites, however, they don't provide the relationships between metabolites and their biological origins. To systematically and comprehensively understand species-specific diversity of metabolites, we have designed a database system called KNApSAcK. This system is useful for obtaining information on metabolites and their corresponding species, chemical structure and biological activity. In addition, the database has a tool that can be used for analyzing datasets acquired using Fourier transform ion cyclotron mass spectrometry. We collected information on 24,604 metabolite-species pairs encompassing 10,181 metabolites and 7,362 species from published references (October 11th, 2005). These data have been stored on a server which is located in Nara Institute of Science and Technology. When the KNApSAcK system is started, these data are automatically downloaded from the server. This database system and online manual are freely available at http://kanaya.naist.jp/KNApSAcK/. This database system is available in Web- and Download-version. To use the Web-version, get into http://kanaya.naist.jp /KNApSAcK/KNApSAcK.php. If you want to use the Download-version, at the beginning, you have to install KNApSAcK and Java 1.4.2 on your local computer. In the CBI coference, some topics concerning species-specific metabolite will be discussed.

S09

Pharmaco-Ontologies and their Roles

Sumi Yoshikawa
Genomic Sciences Center (GSC), RIKEN

The increasing amount of digital information and the numerous opportunities of interdisciplinary activities are raising issues regarding the suitability of conventional information systems confined to certain domain communities. Ontology as a discipline is expected to become a solution since it allows to specify explicit semantics and relations in terms of both human and computer readable manners. “An ontology is a specification of a conceptualization, in the context of knowledge sharing” according to Gruber (1993). Historically, it has been investigated in Artificial Intelligent communities, providing knowledge representation framework theories and tools for inference systems, information retrieval, mining, and so on.

Today, ontologies are increasingly developed and adopted in various fields including biomedical domains. For example, Gene Ontology has been developed as a general gene annotation vocabulary system regardless of species differences. Biomedical Ontologies would be potentially associated with medical information and are expected to contribute to solve clinical problems. Pharmaco-ontologies overlap with biomedical ontologies in a wider sense, and deal with multi-dimensional aspects of drug functionalities. The application area of both ontologies will encompass drug discovery and R&D, decision making in treatment choices including application of pharmacogenetic information, communication of drug information between medical providers and patients / general public, etc. I will refer to some underlying informatics issues specific to the pharmaceutical / alternative medicine domains and how ontologies would contribute to the solutions.

The development and maintenance of domain-specific ontologies require knowledge input from domain experts and the user evaluation. I will also address the need for fostering application domain ontologists and/or ontology coordinators, and that for collaborative community activities for the development of solution-oriented ontology.

S10

Combining Quantitative Structure Activity Relationships and Systems Biology Approaches for in silico ADME/Tox

Sean Ekins
ACT LLC and Department of Pharmaceutical Sciences, University of Maryland,
20 Penn Street, Baltimore, MD 21201, USA

The challenge of predicting the ADME/Tox properties of a drug in humans has been approached independently using multiple experimental technologies and more recently computational approaches (e.g. computational quantitative structure activity relationships (QSAR)) and high content data. Understanding the complexity of biological systems requires a broader perspective rather than focusing on just one method in isolation for prediction. From experience of developing computational models for specific ADME/Tox targets such as the pregnane X receptor, human Ether-a-Go-Go-Related Gene potassium channel (hERG), drug transporters and cytochrome P450s to developing a software suite integrating many models, metabolite prediction, human cell signaling networks, metabolic pathways and and toxicogenomic data, the potential for understanding toxicity of small molecules computationally has greatly expanded. Recent work has included applying multiple computational methods to predict hERG inhibition and understanding the applicability of these methods. Further work has applied a systems approach, predicting potential small molecule interactions and visualizing these molecules as networks that can be compared to microarray or other high content data to show the networks of genes that are affected. This represents an example of how the combination of QSAR models, applicability methods and systems biology approaches may provide data mining opportunities for both pharmaceutical and environmental toxicology.

S11

Current Strategies for ADMET Integration to Enhance Efficiency and Quality in Drug Discovery

Edward H. Kerns
Wyeth Research, Princeton, NJ, USA

In the past decade, the increased attention to biopharmaceutical “drug-like” properties in drug discovery has enhanced success and efficiency. Higher quality clinical candidates are being advanced to development. Several strategies for integration of ADMET into discovery have proven to be successful. These strategies include the following:

- Diverse ADMET tools (rules, *in silico*, *in vitro*, *in vivo*) have been developed and have found their most appropriate stages of discovery.
- Chemical structures, and the role of medicinal chemists in selecting and modifying structures to optimize properties, are of crucial importance.
- Potency and properties are evaluated and optimized in parallel.
- ADMET experts provide consultation on selecting and optimizing lead candidates with optimal properties, as part of collaborative project teams.
- ADMET property knowledge has been successfully applied beyond *in vivo* drug delivery to optimizing *in vitro* biological assays.

In silico tools have the opportunity to enhance these ADMET strategies through: a) predicting properties and compound ranking, b) proactively guiding structure modifications for medicinal chemists, and c) modeling *in vivo* processes to assist visualization by project teams.

S12

In silico database, simulation and prediction for the optimization of pharmacokinetic properties of drug candidates

Kazuya Maeda[1], Hiroyuki Kusuhara[1], Yoshihisa Shitara[2], Motohiro Kato[3], Yuichi Sugiyama[1]
[1]Graduate School of Pharmaceutical Sciences, The University of Tokyo, [2]Graduate School of Pharmaceutical Sciences, Chiba University, [3]Chugai Pharmaceutical Co., Ltd.

Recently many metabolic enzymes and transporters responsible for the detoxification of xenobiotics including drugs have been identified and characterized, and clinical studies clearly indicated the importance of these molecules in the pharmacokinetics and subsequent pharmacological effect of many drugs. At the present, enormous amount of information has been accumulated in the field of pharmacokinetics. However, it is fairly difficult to utilize the whole information effectively for the drug development and many people are at a loss as to how to mine the valuable information from the piles of research articles and how to make use of the information to accelerate the drug development. In our laboratory, with the help of *in silico* technologies, we have developed the useful tools to support the acquisition of necessary information about drug transporters and the quantitative prediction of transport properties and pharmacokinetics of drugs as shown below.

1) **"TP-search" (http://www.TP-Search.jp/)**: This is a comprehensive Web-based database for all about drug transporters. We can easily obtain the information about substrates/inhibitors/inducers, tissue distribution, drug-drug interaction, genetic polymorphism, pathophysiology, knockout mice etc. with powerful search functions.
   The data are obtained from more than 2600 articles and routinely updated.
2) **Ligand-based prediction of the pharmacophore of transporters**: Based on the assumption that conformation of each substrate which can bind to the transporter is one of the energetically-stable conformers, we calculated the candidate conformers of each substrate, put "property sphere" representing the local physicochemical properties of chemical structure on each substrate and searched the common configulation of pharmacophores by superposing the structure of each substrate. Subsequent CoMFA analyses revealed the predicted 3D structure of binding site of transporters.
3) **Simulator for drug-drug interaction**: This simulator has also a database for pharmacokinetic parameters of substrate and inhibitor drugs. Using physiologically-based pharmacokinetic (PBPK) model, we can simulate the time-profiles of both substrate and inhibitor drugs with considering the time-dependent inhibitory effect of inhibitors on metabolic enzymes. This model was well validated by the reported clinical drug-drug interactions.
4) **Whole body PBPK modeling considering the transporter-mediated clearance**: Integrating the kinetic parameters for the transport properties of uptake and efflux transporters obtained from *in vitro* analyses into the PBPK model, we simulated the dose-dependent pharmacokinetics of pravastatin, which is taken up by OATP family transporters and excreted into bile by MRP2. Non-linear pharmacokinetics could be reproduced by this model.

In my presentation, I want to introduce the concept and overview of these *in silico* approaches and discuss their significance for optimizing the pharmacokinetic properties of drug candidates.

S13

Information Access based on Association

Akihiko Takano
National Institute of Informatics, Tokyo

GETA (**G**eneric **E**ngine for **T**ransposable **A**ssociation) is a software that provides efficient generic computation for association. It enables the quantitative analysis of various proposed methods based on association, such as measuring similarity among documents or words. Scalable implementation of GETA can handle large corpora of twenty million documents, and provides the implementation basis for the effective information access of next generation. Webcat Plus is the book information service for the union catalogue of one thousand research libraries in Japan, which is a successful example to show the power and the flexibility of GETA-based computation for association. It provides the associative search function which find related books based on similarities between books based on their TOC's and abstracs. It also provides users a set of Topic Words which characterizes the search result. This dual mode interface can intuitively inspire the users with further interactions.
The two-stage approach in the associative search, which is the key to its efficiency, also facilitates the content-based correlation among different databases.
In this talk we overview the basic features of GETA, and introduce the various existing services using it as their association engines.

Reference:

Webcat Plus: http://webcatplus.nii.ac.jp/en/
Cultural Heritage Online: http://bunka.nii.ac.jp/
Shinsho Map : http://shinshomap.info/
Book Town JIMBOU: http://jimbou.info/

S14

Biomedical text mining: state of the art and future direction

Asako Koike
Central Research Laboratory, Hitachi Ltd.
Dept. of Computational Biology, Graduate School of Frontier Science, The University of Tokyo

Due to the rapid progress of biomedical field and developments in high throughput technologies, an enormous amount of experimental results has been created and is stored mainly in text. Text-mining techniques have come into the spotlight for two main reasons. The first is because objective information, which is explicitly written in text, can be effectively retrieved using information extraction and information retrieval techniques. The second is because knowledge concealed within multiple sources can be discovered by connecting fragmented results described in the text. Extracting objective information is crucial to the interpretation of high-throughput analyses, while the discovery of knowledge hidden in text is advantageous in searching for unthought of but plausible implicit relationships between concepts.

To meet these needs, we have developed an information extraction system called "*PRIME*" and a text mining system called "*BioTermNet*". The *PRIME* has cumulated protein interactions, gene/protein functions, and gene/protein-disease relationships extracted by syntactic analysis from MEDLINE abstracts and full papers. Their protein networks and orthologous pathways can be drawn by a graphic viewer. The *BioTermNet* provides mainly two functions for knowledge discovery/hypothesis generation and interpretation of experimental results and the results are presented by a graphic viewer. The first function is to connect the explicit relationship and generate the conceptual network, and search the most appropriate implicit relationship for open (only start concept is given) and closed (start and end concepts are given) discoveries. The explicit relationship is calculated with a hybrid method of syntactic analysis and statistical analysis. This function is useful in the search for new diseases to apply approved drugs to, in the search for the dietary effects of foods and in the search for relationships between genes and phenotypes. The second function of *BioTermNet* is to cluster genes/concepts based on document similarities for the interpretation of high-throughput data. In this presentation, with introduction of the application of this system to the knowledge discovery and interpretation of experimental results, the future direction of biomedical text mining is also discussed. The *BioTermNet* and *PRIME* are respectively available at http://btn.ontology.ims.u-tokyo.ac.jp/ and http://prime.ontology.ims.u-tokyo.ac.jp:8081/ for non-commercial purposes.

S15

Cancer Genomics; Chip, SNP, Database

Teruhiko Yoshida, Kazuhiko Aoyagi, Hiroki Sasaki and Hiromi Sakamoto
Genetics Division, National Cancer Center Research Institute

This presentation tries to give an overview of the hopes and challenges which the speaker thinks that the cancer genomics is facing. The 2 words, Chip and SNP, may characterize the post sequencing era of the human genome research, and they have rapidly infiltrated many, if not most, fields of the cancer research. The increasing power and ease of the genome data acquisition have facilitated statistics-driven, hypothesis-free (or hypothesis-generating) research strategies based on the genome-wide screening of clinical samples. Cancer research has been one of the driving forces of the genomics, and its unique feature is that 2 types of genomes need to be analyzed in cancer cases: somatic and germline. In the somatic (cancer) cell genome analyses, 2 types of chips are now being used widely: array CGH (comparative genomic hybridization) and oligonucleotide microarray for expression profiling. Our ongoing experience will be presented regarding the application of the Affymetrix GeneChip technology to the class prediction problem of the chemoradiosensitivity of the esophageal cancer. In the germline analyses, SNP genome scan tool is now in the second generation, the post-HapMap, chip-based platform at least for the first screening. Our pre-HapMap experience during the Millennium Genome Project, helped by RIKEN, will be presented. With these floods of the genome data, the 3 obvious vital points are 1) high quality phenotype information, 2) sufficiently large number of the samples, and 3) functional database. We have started constructing GeMDBJ, Genome Medicine Database of Japan (http://gemdbj.nibio.go.jp), exploring a way for a synthetic analysis to correlate the phenotype data with the multiple types of the omics and their annotation data.

S16

Application of Epigenetics on Development of Drugs

Kunio Shiota
Laboratory of Cellular Biochemistry, Animal Resource Sciences,
The University of Tokyo

"Genome projects" of human being and various organisms have shown that mammals have a much bigger genome than less complex eukaryotes. However, the increased 'bio-complexity' is not reflected by an equivalent increase in the number of genes. Epigenetics appears to be one of the most important research fields in the post-genome sequence era. Epigenetics is defined as " the study of heritable changes in gene function that occur without a change in the sequence of DNA " or "the study of processes that produce a heritable phenotype that is not strictly dependent on the DNA sequence". DNA methylation is the only chemical modification in mammalian genome and it occurs in conjunction with histone modifications causing condensation of the chromatin packaging the genome, and results in the repression of genes. The study of epigenetics focuses on investigating the mechanism by which each cell expresses its own phenotype using the same genome. We have found tissue-dependent and differentially methylated regions (T-DMRs) in CpG-rich or poor unique sequences in rat and murine normal tissues and cell-types including embryonic stem cells. T-DMRs are widespread in the euchromatin of the mammalian genome. The complexity of mammals is likely to depend on the combinational control triggering a vast number of gene expression pattern. Thus, epigenetic research is anticipated to have implications for medicine and the understanding of the basic process of the cell fate determination. In this meeting, I will introduce "Epigenetics" for the various steps of development of drugs.

Reviews

1) Shiota K., and Yanagimachi R., Epigenetics by DNA methyaltion for development of normal and cloned animals. Differentiation 69:162-166 (2002).
2) Shiota K., DNA methylation profiles of CpG islands for cellular differentiation and development in mammals. Cytogenet Genome Res 105:325-334 (2004).
3) Lieb J. D., et al., Applying whole-genome studies of epigenetic regulation to study human disease. Cytogenet Genome Res 114:1-15 (2006).

Poster Presentations

Research Topics

1. Molecular Computing
2. Molecular Recognition
3. Bioinformatics and Bio Computing
4. Genome Wide Experimental Data Analyses
5. Information and Computing Infrastructure for Drug Design and Toxicology
6. Disease Mechanism and Control Models
7. Medical Genome Informatics and New Technology

Poster Presentations

Pn : n=Research Topic No.
P* : Oral Presentation

P1-01 Crystal structure and molecular dynamics simulation of ST0689 from *Sulfolobus tokodaii* strain7.

Yusuke Miyata[1], Takatoshi Arakawa[1], Okimasa Okada[2], Jyunichi Akutsu1[1, 3], Masafumi Yohda[1]
Department of Biotechnology and Life Sicence,Tokyo, University of Agriculture And Technology[1], Technology and Development, Fuji Xerox Co., Ltd., [2], AIST[3]

P1-02 Alkylation Mechanism of BPDE with DNA by Quantum Molecular Dynamics Simulation

Yoshiro Nakata, Wataru Ootani
Department of Biophysics, Faculty of Engineering, Gunma University

P1-03* Enantioselectivity *of Candida antarctica* Lipase TypeB toward Secondary Alcohols: *Ab initio* Fragment Molecular Orbital Study

Takahisa TANAKA, Yoshihiro MORI, Yoshinobu NAOSHIMA
Faculty of Informatics, Okayama University of Science

P1-04 The Low-Frequency Raman Modes of Crystals of Guanosine Dihydrate Analyzed from MD Simulation

Shigetaka Yoneda[1], Yoko Sugawara[1], and Hisako Urabe[2]
School of Science, Kitasato University[1], Tokyo Kasei Gakuin University[2]

P1-05 Intra- and intermolecular interactions between cyclic-AMP receptor protein and DNA: *Ab initio* fragment molecular orbital study

Kaori Fukuzawa[1], Yuto Komeiji[2], Yuji Mochizuki[3, 4], Takeshi Ishikawa[3, 4], Akifumi Kato[1], Tatsuya Nakano[4, 5], Shigenori Tanaka[4, 6]
Mizuho Information & Research Institute, Inc. [1], National Institute of Advanced Industrial Science and Technology[2], Rikkyo University[3], Japan Science and Technology Agency, CREST[4], National Institute of Health Sciences[5], Kobe University[6]

P1-06* Development of virtual screening method using Solvation Free Energy Density Model(SFED)

Chang Joon Lee, Se Han Lee, and Kyoung Tai No
Yonsei University

P1-07 Computational Study of Catalytic Reaction of Adenylate Kinase

Kenshu Kamiya, Shinya Morishita and Yoko Sugawara
School of Science, Kitasato University

P1-08 NFV-Resistant Mechanism due to Non-active Site Mutation N88S on Subtype AE HIV-1 PRs

Hirotaka Ode[1], Shou Matsuyama[1], Saburo Neya[1], Masayuki Hata[1], Wataru Sugiura[2], Tyuji Hoshino[1, 3]
Graduate School of Pharmaceutical Sciences, Chiba University[1],
AIDS Research Center, National Institute of Infectious Diseases[2], PRESTO, JST[3]

P1-09 Evaluation of the binding affinity of antibodies from antigen binding energy and V_H-V_L interaction energy calculated by the MM-GBSA method

Koji Iwamoto[1], Saburo Neya[1], Tyuji Hoshino[1, 2]
Graduate School of Phamaceutical Science,Chiba University[1], PRESTO, JST[2]

P1-10 Prediction of complex structure comprised of a protein and a glycosaminoglycan using docking simulation and cluster analysis

Tsubasa Takaoka[1], Noriaki Okimoto[2], Kenichi Mori[1], Saburo Neya[1], Tyuji Hoshino[1, 3]
Graduate School of Pharmaceutical Sciences, Chiba University[1]
RIKEN, GSC[2], PRESTO, Japan Science and Technology Agency[3]

P1-11 Catalytic reaction of the enzyme degrading biodegradable plastics

Y. Sakae[1], T. Matsubara[1], M. Aida[1], H. Kondou[2], K. Masaki[3], and H. Iefuji[3]
Hiroshima University[1], National Institute of Advanced Industrial Science[2],
Technology National Research Institute of Brewing[3]

P2-01* In silico drug screening by docking score index method using principalcomponent analysis

Sukumaran Murali[1], Yoshifumi Fukunishi[2], Haruki Nakamura[2, 3]
Japan Biological Information Research Center (JBIRC), Japan Biological Informatics Consortium (JBIC) [1], National Institute of Advanced Industrial Science and Technology (AIST), Japan[2], Institute for Protein Research, Osaka University[3]

P2-02* Solvent Site-Dipole Field Mediating Docking of Biomolecules

Norikazu Takano, Nobuyuki Hamasaki, Junichi Higo
Tokyo University of Pharmacy and Life Sciences

P2-03 Application of Amino Acid Descriptors for Prediction of MHC-Peptide Binding Affinity

Tomoko Niwa
Discovery Research Laboratories, Nippon Shinyaku Co., Ltd.

P2-04 Molecular modeling and 3D-QSAR studies of benzothiazol-2-ylacetonitrile derivatives as c-Jun N-terminal kinase 3 inhibitors

Abdul Rajjak Shaikh[1], Mohamed Ismael[1], Hideyuki Tsuboi[1], Michihisa Koyama[1], Akira Endou[1], Hiromitsu Takaba[1], Momoji Kubo[1, 2], Carlos A. Del Carpio[1], Parasuraman Selvam[1], Kazumi Nishijima[3], Tetsuya Terasaki[1], Akira Miyamoto[1]
Tohoku University[1], JST-PRESTO[2], Mochida Pharmaceutical Company[3]

P2-05 Bootstrap-based consensus scoring method for high-performance and robust drug screening by single scoring function

Hiroaki Fukunishi, Reiji Teramoto, Jiro Shimada
Fundamental and Environmental Research Laboratories, NEC corporation

P2-06 Introduction of initial binding orientation in protein-protein docking

Jong Young Jung[1], Chang Joon Lee[1] and Kyoung Tai No[1, 2]
Yonsei University[1], Research Institute of Bioinformatics and Molecular design[2]

P2-07 Functions of the key residues in the ligand-binding pocket of vitamin D receptor : Fragment molecular orbital-interfragment interaction energy analysis

Kenji Yamagishi[1, 2], Keiko Yamamoto[3], Sachiko Yamada[3], and Hiroaki Tokiwa[4]
The University of Tokushima[1], CREST-JST[2], Tokyo Medical and Dental University[3], Rikkyo University[4]

P2-08 A Workbench for Selective Nuclear Receptor Modulators

Naomi Komiyama[1], Tatsuya Nakano[2], Kaori Fukuzawa[3], Junpei Komiyama[4], Masumi Yukawa[5], Hiroshi Tanaka[5], Kotoko Nakata[6], Tsuguchika Kaminuma[5]
Chem-Bio Informatics Society[1], National Institute of Health Sciences[2], Mizuho Information and Research Institute[3], The University of Tokyo[4], Tokyo Medical and Dental University[5], Advance Soft Co./The University of Tokyo[6]

P3-01 Discrimination of β-barrel Membrane Proteins: Comparison between Statistical Methods and Machine Learning Algorithms

M. Michael Gromiha, Makiko Suwa
Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST)

P3-02 SPLITS: an Expanding Program for Split and Intron-Containing tRNA Prediction

Junichi Sugahara[1, 2], Nozomu Yachie[1, 3], Yasuhiko Sekine[4], Akiko Soma[4], Motomu Matsui[1, 2], Masaru Tomita[1, 2, 3], Akio Kanai[1, 2]
Institute for Advanced Biosciences, Keio University[1], Department of Environmental Information, Keio University[2], Bioinformatics Program, Graduate School of Media and Governance, Keio University[3], Department of Life Science, College of Science, Rikkyo University[4]

P3-03* Phylogenetic profiling approach to generate RNA-protein networks in *E. coli*

Nozomu Yachie[1, 2], Koji Numata[1, 2], Yoshiteru Negishi[1, 2], Hiroyuki Nakamura[1, 3], Junichi Sugahara[1, 3], Rintaro Saito[1, 3], Masaru Tomita[1, 2, 3]
Institute for Advanced Biosciences, Keio University[1]
Bioinformatics Program, Graduate School of Media and Governance, Keio University[2]
Department of Environmental Information, Keio University[3]

P3-04 Development and application of a molecular dynamics simulation system : myPresto

Satoru Kubota[1], Yoshifumi Fukunishi[2], Ikuo Fukuda[1], Eiji Kanamori[1], Katsumi Omagari[1], Haruki Nakamura[3]
Japan Biological Information Research Center (JBIRC), Japan Biological Informatics Consortium (JBIC) [1], Biological Information Research Center (BIRC), National Institute of Advanced Industrial Science and Technology (AIST) [2], Institute for Protein Research, Osaka University[3]

P3-05 Automatic Procaryote ORF Finding by Markov Model

Daigo Nakahara, Hiroyuki Nakajima, Yuuichi Tokuda, Atsushi Hanada, Yuuki Shibata, Koichi Takahashi
Kinki University

P3-06 Prediction of nuclear localization proteins based on charge periodicity of 28 residues

Noriyuki SAKIYAMA, Runcong KE, Shigeki MITAKU
Department of Applied Physics, Graduate School of Engineering, Nagoya University

P3-07 WinBEST-KIT(Biochemical Reaction Simulator for Analyzing Metabolic Pathways): Application to the System Analysis of Acetone-Butanol-Ethanol Fermentation in *Clostridium saccharoperbutylacetonicum*

Masahiro Okamoto[1], Hideaki Shinto[2], Tatsuya Sekiguchi[3],Yukihiro Tashiro[2], Mayu Yamashita[2], Genta Kobayashi[4], Taizo Hanai[1], Kenji Sonomoto[3]
Graduate School of Systems Life Sciences, Kyushu University[1], Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University[2], Department of Information Engineering, Maebashi Institute of Technology[3], Ariake Sea Research Project, Saga University[4]

P3-08 An Integrated Database of Flavonoids

Takashi Kinoshita, Zsolt Lepp, Yoshichika Kawai, Junji Terao, Hiroshi Chuman
Institute of Health Biosciences, The University of Tokushima

P3-09* Estimation of Gene Regulatory Network by means of AdaBoost

Shinya Nabatame, Hitoshi Iba
School of Engineering The University of Tokyo

P3-10 Theoretical studies of the ATP hydrolysis mechanism of HisP protein

Qiang Pei[1], Hideyuki Tsuboi[1], Michihisa Koyama[1], Akira Endou[1], Hiromitsu Takaba[1], Momoji Kubo[1, 2], Carlos A. Del Carpio[1], Parasuraman Selvam[1], Kazumi Nishijima[1, 3], Tetsuya Terasaki[1], Akira Miyamoto[1]
Tohoku University[1], JST-PRESTO[2], Mochida Pharmaceutical Company[3]

P3-11* Proteins with charge periodicity of 28 residues which are coded in total genomes

Runcong KE, Shigeki MITAKU
Department of Applied Physics, Graduate School of Engineering, Nagoya University

P3-12 Systematic Errors Commonly Found in GeneChip Expression Data and Effective Compensation Methods thereof

Tomokazu Konishi[1], Keishi Hata[2], Shigeru Takasaki[3], Akihiko Konagaya[3]
Akita Pref. Univ.[1], Akita Research Institute for Food and Brewing[2], RIKEN[3]

P3-13 Protein functional site prediction -DNA binding proteins-

Yukako Sakatsuji[1], Ayumi Suzuki[2], Ichiro Yamato[2], Satoru Miyazaki[1]
Department of Pharmaceutical Science, Tokyo university of science[1],
Department of Biological Science and Technology, Tokyo university of science[2]

P3-14 HP-Lattice-model Structure Prediction of Proteins using Genetic Programming

Takahiro Yabuki, Hitoshi Iba
Frontier Informatics, Graduate School of Frontier Sciences, the University of Tokyo

P3-15 Prediction of Human Cytochrome P450 3A4 Substrates using Statistical analysis

Hyesun Han, Ji Hoon Jung, Kyoung Tai No
Yonsei University

P3-16 Branch And Bound Median Search algorithm for finding a substrate specific motif at human SLC22 transporters

Takayuki Arakawa, Satoru Miyazaki
Department of Pharmaceutical Science, Tokyo university of science

P3-17 A Molecular Modeling Study on Antigenic Drifts of Influenza virus Surface Glycoprotein

Manabu Igarashi[1], Kimihito Ito[1], Hiroshi Kida[1, 2], and Ayato Takada[1]
Department of Global Epidemiology, Research Center for Zoonosis Control[1],
Department of Disease Control, Graduate School of Veterinary Medicine[2],
Hokkaido University

P3-18 Analysis of amino acid sequences of motor protein by physical finger print method

Hideki Tanizawa and Shigeki Mitaku
Department of Applied Physics, Graduate School of Engineering, Nagoya University

P3-19 A New Method for Analysis of SNP's Effect on Protein Function and Structure

Carlos A. Del Carpio[1], Eichiro Ichiishi[2], Hideyuki Tsuboi[1], Michihisa Koyama[1], Akira Endou[1], Hiromitsu Takaba[1], Momoji Kubo[1, 3], Akira Miyamoto[1, 2]
Graduate School of Engineering, Tohoku University1, NICHe, Tohoku University[2], JST-PREST[3]

P3-20 Application of Connectivity Matrix Method to Atom-level Analysis of Metabolic Networks: Calculation of Paths Connecting Two Metabolites

Jun Ohta
Okayama University

P4-01 Microarray analysis based on Standard Deviation of Effecting of Biological Factors

Chiaki Handa[1], Kayoko Takashima[2], Hidetoshi Maruyama[2], Naoto Nagasawa[1], Toshiki Honma[2], Takeshi Nakabayashi[2], Shinji Kikuchi[2]
Kissei comtec Co., Ltd. [1], Kissei pharmaceutical Co., Ltd. [2]

P4-02* Detecting Cell Cycle Regulated genes in S. pombe with Non-metric Multidimensional Scaling without sinusoidal fittings

Y-h. Tagcuhi
Department of Physics, and Institute for Science and Technology, Chuo University

P4-03 Comprehensive analyses on relationships between alternative-splicing patterns and developmental stages in mouse

Yasuko Takahashi, Soichi Ogishima, Hiroshi Tanaka
Dept. of Computational Biology, Tokyo Medical and Dental University

P4-04* Genome-wide Prediction of Novel Small Peptides in *Escherichia coli* by Using Liquid Chromatography Mass Spectrometry

Motomu Matsui[1, 2], Nozomu Yachie[1, 3], Kosaku Shinoda[4], Naoyuki Sugiyama[4], Fumio Ando[1, 2], Yuki Okada[1, 3], Martin Robert[1], Rintaro Saito[1, 2]*, Masaru Tomita[1, 2, 3, 4]
Institute for Advanced Biosciences, Keio University[1], Department of Environmental Information, Keio University[2], Bioinformatics Program, Graduate School of Media and Governance, Keio University[3], Human Metabolome Technologies, Inc., c/o Insitute for Advanced Biosciences, Keio University[4]

P4-05 Transcriptomic approach for understanding the effect of cryoprotectant, trehalose and DMSO of Saccharomyces cerevisiae

Yuko Momose, Rena Matsumoto, Mine Odani, Masakazu Yamaoka
The National Institute of Advanced Industrial Science and Technology (AIST)

P4-06 GC contents of nucleotide sequences for proteins with charge periodicity of 28 residues

Ryusuke Sawada, Runcong Ke, Noriyuki Sakiyama, Masashi Sonoyama
Shigeki Mitaku
Department of Applied Physics, School of Engineering, Nagoya University

P5-01 PBPK-modeling as a tool for interpreting and understanding of pharmacokinetics

Hiroko Akiyama, CTC LABORATORY SYSTEMS Corporation [Presenter]
S. Willmann, W. Schmitt, Bayer Technology Services GmbH [Authors]

P5-02 Scaffold analysis of GPCR ligand datasets from literature

Masami Fujiki[1], Yanhai Fan[2], Naotaka Ohishi[3]
Patcore, Inc. [1], Phas Technology, Ltd. [2], Spotfire Japan, K.K. [3]

P5-03 Development of New *in silico* Method for Predicting Drug Metabolism

Michihisa Koyama[1], Kota Kasahara[1], Hideyuki Tsuboi[1], Akira Endou[1],
Hiromitsu Takaba[1], Momoji Kubo[1, 2], Carlos A. Del Carpio1, Kazumi Nishijima[1, 3], Tetsuya Terasaki[1], Akira Miyamoto[1]
Tohoku University[1], Japan Science and Technology Agency[2],
Mochida Pharmaceutical Co.[3]

P5-04 "CALTA", a novel structure-property-activity profiling method: Compound-scaffold profiling and *in silico* prediction for the drug-discovery compounds that target membrane proteins

Yoshimasa Kobayashi, Katsuya Akimoto, Hiroshi Kikuchi, and Kenichi Sudo.
Drug Metabolism & Physicochemistry Research Lab., Daiichi Pharmaceutical. Co., Ltd.

P5-05 Virtual Screening Models for Predicting Activities in Various Therapeutic Areas

Zsolt Lepp, Takashi Kinoshita and Hiroshi Chuman
Institute of Health Biosciences, The University of Tokushima

P5-06 Convergence properties of absolute binding free energy calculation using BAR method

Yoshiaki Tanida, Masakatsu Ito, Hideaki Fujitani
Fujitsu Laboratories Ltd., Fujitsu Limited

P5-07* Building up an original database of launched drugs for their ADME and physicochemical HTS profiling data and human PK data. Rating drugability based on the database and profiling data for in-house compounds.

Atsuko Kameda, Kyosuke Suzuki, Yoshimasa Kobayashi, Katsuya Akimoto, Hiroshi Kikuchi and Kenichi Sudo
Drug Metabolism & Physicochemistry Research lab., Daiichi Pharmaceutical. Co., Ltd.

P5-08 Conceptual Design of a Framework for ADME Ontology

Takeshi Arikuma[1], Sumi Yoshikawa[1, 2], Kentaro Watanabe[1]*, Kazumi Matsumura[1, 2], Akihiko Konagaya[1, 2]
Department of Computer Science, Tokyo Institute of Technology[1], Genomic Sciences Center, RIKEN[2]

P5-09 Acquisition of Extended Information Regarding Chemicals-CYPs Interactions by Natural Language Processing

Chunlai Feng, Fumiyoshi Yamashita, Mitsuru Hashida
Graduate School of Pharmaceutical Sciences, Kyoto University

P5-10 Ontology Development for RTK-Associated Information Retrieval System

Sumi Yoshikawa, Kazumi Matsumura, Akihiko Konagaya
Genomic Sciences Center (GSC), RIKEN

P5-11* Physiologically Based Pharmacokinetic (PBPK) model for Pravastatin - Analysis of transporter-related pharmacokinetics -

Takao Watanabe, Hiroyuki Kusuhara, Kazuya Maeda, Yuichi Sugiyama
The University of Tokyo

P5-12 A Web-Based Drug Safety Information Community

Kumiko Sakamoto[1], Eiko Nakata Doolin[2], Mikio Sasaki[3], A.Ammar Ghaibeh[3], Hiroshi Chuman[1], Aiko Yamauchi[1]
Institute of Health Biosciences, The University of Tokushima[1],
Department of Pharmacy, NTT-East Kanto Medical Center[2], Saila Systems Inc.[3]

P5-13 Instant Computing Technology for Molecular Design Environment for Non-Cluster Experts

Fumikazu KONISHI[1], Toru Yagi[2], Ryo Umetsu[1], Shingo Ohki[1], Akihiko KONAGAYA[1]
RIKEN GSC[1], BESTSYSTEMS, Inc.[2]

P5-14 Conformational Analysis and Docking Study of the c-Jun N-Terminal Kinase Inhibitors Having a Benzothiazol Moiety

Mohamed Ismael[1], Abdul Rajjak Shaikh[1], Hideyuki Tsuboi[1], Michihisa Koyama[1], Akira Endou[1], Hiromitsu Takaba[1], Momoji Kubo[1, 2], Carlos Del Carpio[1], Kazumi Nishijima[3], Tetsuya Terasaki[1], Akira Miyamoto[1]
Tohoku University[1], JST-PRESTO[2], Mochida Pharmaceutical Co.[3]

P5-15 Development of an Empirical Model for the Activation Energy of CYP Reaction

Doo Nam Kim[1], Won Seok Oh[2], Kyoung Tai No[1, 2]
Bioinformatics & Molecular Design Research Center[1], Yonsei University[2]

P5-16* Prediction of Skin Permeability for Chemical Penetration Enhancers using Quantitative Structure-Activity Relationship Analysis and Classification Methodology

Hyejin Kang[1], KyoungTai No[1, 2]
Yonsei University[1], BMDRC[2]

P5-17 *In Silico* Prediction of CYP1A2-mediated Metabolism

Jihoon Jung, Suyeon Kim, Kyoung Tai No
Yonsei University

P5-18 Classification of P-Glycoprotein Substrates and Inhibitors

Sokyoung Joo[1], Kyoung Tai No[1, 2]
Yonsei University[1], Research Institute of Bioinformatics and Molecular Design[2]

P5-19 NMR metabolomics 1/2 Multivariate analysis for diagnostic study of inherited metabolic disorders using human urine

Tadashi Nemoto[1], Masako Fujiwara[1], Itiro Ando[1, 3] Kazunori Arifuku[2], Tomiko Kuhara[4]
National Institute of Advanced Industrial Science and Technology[1], JEOL DATUM LTD[2], Environmental Research Center Ltd.[3], Kanazawa Medical University[4]

P5-20 NMR metabolomics 2/2 Multivariate analysis for diagnostic study of streptozotocin-induced diabetes using rat urine

Masako Fujiwara[1], Itiro Ando[2, 3], Kazunori Arifuku[1], Taeko Kataoka[2], Kenji Kanazawa[2], Yasuhiro Natori[4], Tadashi Nemoto[2]
JEOL DATUM LTD.[1], National Institute of Advanced Industrial Science and Technology[2], Environmental Research Center Ltd.[3], Research Institute International Medical Center of Japan[4]

P5-21 Development of a database with simulation function for quantitative prediction of drug-drug interactions

Yoshihisa Shitara[1], Toshiharu Horie[1], Toshihiko Ikeda[1], Masato Kitajima[1], Jose M. Ciloy[3], Yuichi Sugiyama[4]
Chiba University[1], Sankyo Co. Ltd.[2], Fujitsu Kyushu System Engineering Ltd.[3], The University of Tokyo[4]

P5-22 **In vitro screening system using a combination of suspensions and monolayers of Caco-2 cells for detecting drug-drug interactions in intestinal absorption and conjugative metabolism of drugs**

Marie Tanaka, Takashi Mizuma, Masahiro Hayashi
School of Pharmacy, Tokyo University of Pharmacy and Life Science

P5-23 **Transport and metabolism of ampicillin prodrugs and their Interactions with P-glycoprotein: *in vitro* study on successful prodrugs using Caco-2 cell monolayers**

Sachie Tanaami, Sayaka Sakaguchi, Takashi Mizuma, Masahiro Hayashi
School of Pharmacy, Tokyo University of Pharmacy and Life Science

P6-01 **Mathematical modeling of gene expression associated with Wnt signaling pathway in colorectal carcinoma**

Yoshimi Naruo, Kaoru Mogushi, Hiroshi Tanaka
Department of Bioinformatics, Tokyo Medical and Dental University

P6-02 **Modeling of extracellular matrix degradation processes in cancer metastasis using covariance structure analysis**

Akiko Hatano, Kaoru Mogushi, Hiroshi Tanaka
Department of Bioinformatics, Tokyo Medical and Dental University

P6-03 **Comprehensive analysis of drug-specific mutation patterns in protease gene of HIV-1 subtype B using decision tree**

Kaoru Mogushi, Fengrong Ren, Naoki Hasegawa, Hiroshi Tanaka
Department of Bioinformatics, Tokyo Medical and Dental University

P7-01 **Tutorial and Training Courses on Some Areas of Interest of the CBI Society**

Tsuguchika Kaminuma[1], Masumi Yukawa[1], Naomi Komiyama[2], Yoshitomo Tanaka[1], Takako Takai-Igarashi[1], Hiroshi Tanaka[1]
Tokyo Medical and Dental University[1], Chem-Bio Informatics Society[2]

Crystal structure and molecular dynamics simulation of ST0689 from *Sulfolobus tokodaii* strain7.

Yusuke Miyata[1], Takatoshi Arakawa[1], Okimasa Okada[2], Jyunichi Akutsu[1,3], Masafumi Yohda[1]

Department of Biotechnology and Life Sicence,Tokyo, University of Agriculture and Technology[1], Technology and Development, Fuji Xerox Co., Ltd., [2], AIST[3]

ST0689 of the thermoacidophilic archaeon, *Sulfolobus tokodaii*, is a protein of 122 amino acids which has not been functionally annotated. It exhibits high homology with HEPN domain (higher eukaryotes and prokaryotes nucleotide-binding domain). HEPN domain exists in various chaperons or enzymes, but its functional role has not been elucidated. In this study, we resolved the structure of ST0689 by X-ray crystallography, and analyzed the dynamic structural change by MD simulation.

The crystal structure has shown that ST0689 is composed of a bundle of four helixes with two loops exposing on one side of it. RYPD box, the most conservative sequence motif of HEPN proteins, located in the loop. We performed MD simulations of monomer and dimer of ST0689 to grasp the behavior of the loop containing the motif. During the MD simulation of the monomer, we observed large fluctuation of RYPD box at 300K. However, at higher temperature, it became small and large fluctuation region has shifted to the downstream in the loop. We also observed the change of the structure at each temperature. The loop around RYPD box and the above site were moving during 1 ns at 300 K and 1 ns at 350 K. However, the structure was stable during further 2 ns at 350K. On the other hand, in dimer, the fluctuation of not only RYPD box but also all residues was small even at high temperature. Because X-ray crystal structure determination has shown the B factor around RYPD box is large, we thought that ST0689 exists as monomer with the structure in the state after 1 ns at 350 K during the MD simulation.

Reference

Grynberg M, Erlandsen H, Godzik A, “HEPN: a common domain in bacterial drug resistance and human neurodegenerative proteins”, Trends Biochem Sci, 28, 224-226(2003).

Alkylation Mechanism of BPDE with DNA by Quantum Molecular Dynamics Simulation

Yoshiro Nakata , Wataru Ootani
Department of Biophysics, Faculty of Engineering, Gunma University

Benzo[a]pyrene and structurally-related polycyclic aromatic hydrocarbon environmental pollutants are metabolically activated in cellular environments into highly reactive and genotoxic diol epoxide derivatives. These molecules can attack DNA to form bulky covalently-linked molecules called carcinogen-DNA adducts[1].

Quantum Molecular Dynamics calculations were performed to simulate the alkylation process of benzo[a]pyrene diol epoxide(BPDE) with DNA. The starting structures of the simulation are the 1AVX in Protein Data Bank[2], and the intercalated model structure by MD simulation and several intermediate structures between the former and the latter.

Quantum-MD simulations were carried out with PEACH Ver.5.8 Windows Version[3] using the parallel computing with MPICH Ver1.2.5[4]. PEACH[5),6)] is capable of quantum simulation and minimization by use of Fragment Molecular Orbital method (FMO)[7].

The results of Quantum MD simulations show the process of alkylation and the activation barrier in the alkylation.

References

1) Meehan, T. and Straub, K., Nature, 277, 410-412(1979).
2) http://www.rcsb.org/pdb/
3) Nakata, Y., Takizawa, T., Ogino, M., Uno, M., J.Comput.Chem.Jpn, Vol.2, No.4, 149-154(2003).
4) http://www-unix.mcs.anl.gov/~ashton/mpich.nt/References
5) Komeiji, Y., Nakano, T., Fukuzawa, K., Ueno, Y., Inadomi, Y., Nemoto, T., Uebayasi, M., Fedorov, D. G., Kitaura, K., Chem. Phys. Lett., 372, 342-347 (2003).
6) Komeiji, Y., Haraguchi, M., Nagashima, U., Parallel Computing 27, 977-987(2001).
7) Nakano, T., Kaminuma, T., Sato, T., Fukuzawa, K. Akiyama, Y., Uebayasi, M., Kitaura, K., Chem. Phys. Lett. 351, 475-480 (2002).

Enantioselectivity of *Candida antarctica* Lipase TypeB toward Secondary Alcohols: *Ab initio* Fragment Molecular Orbital Study

Takahisa TANAKA, Yoshihiro MORI, Yoshinobu NAOSHIMA
Faculty of Informatics, Okayama University of Science

Lipases, the most widely studied enzymes in the fields of biotechnology, are now routinely used as an environmentally friendly and efficient biocatalyst for obtaining enantiomerically pure and biologically important compounds in the fields of organic synthesis. As part of our continuous works toward conducting biocatalytic reactions with lipases,[1)] we are aimed at understanding the enantioselectivity of *Pseudomonas cepacia* lipase (PCL) and *Candida antarctica* lipase typeB (CALB) toward non-natural organic compounds.[2)] Although a large amount of experimental data is available for these two popular lipases, the mechanism of their enantioselectivity is still not completely elucidated. In this work the *ab initio* fragment molecular orbital (FMO) calculations were carried out for the CALB complexed with chiral alcohol esters.

The PDB X-ray crystallographic structure of CALB (1LBS) was optimized by energy minimization using OPLS-AA force field implemented in MacroModel. Chiral ligands, including aromatic and aliphatic alcohol esters, were prepared by employing a combination of molecular mechanics and MD simulated annealing technique and the subsequent Hartree-Fock 6-31G** calculations. The flexible docking simulations of CALB and chiral esters were performed through the Glide four-stage minimization. The CALB-chiral ester complexes thus obtained were subjected to an *ab initio* FMO calculation at MP2/STO-3G level by using BioStation package, on a cluster system of Intel Itanium2 processors. Similar FMO calculations were carried out toward CALB enzyme and chiral esters, respectively. The FMO calculations indicated that the binding energy between CALB and chiral esters is larger for the fast reacting (*R*)-enantiomer than for the slow reacting (*S*)-enantiomer. We also found that the hydrogen bond network between Thr40 and chiral esters is significant in the formation of each CALB complex. It is expected that the difference of the binding energy makes it possible to predict the enantiopreference of CALB toward a variety of secondary alcohols. FMO calculations on the tetrahedron intermediate in the ester hydrolysis are under way.

REFERENCES

1) Y. Naoshima, T. Kimura, Y. Mori, and M. Kamezawa, *Recent Res. Devel. Org. Bioorg. Chem.*, **6**, 1-9 (2004). 2) Y. Mori, Y. Yagi, and Y. Naoshima, Proceedings of 25th Anniversary International CBI Conference, 53 (2005).

The Low-Frequency Raman Modes of Crystals of Guanosine Dihydrate Analyzed from MD Simulation

Shigetaka Yoneda[1], Yoko Sugawara[1], and Hisako Urabe[2]
School of Science, Kitasato University[1], Tokyo Kasei Gakuin University[2]

Molecular dynamics simulation was applied to a new region of calculational analysis of organic crystal. In the previous study, we reported MD simulation of a crystal of guanosine dihydrate and discussed the mobility of the crystal water molecules.[1] In this study, we present analysis of low-frequency Raman active phonon modes.

In low-frequency Raman spectra of DNA, one characteristic vibration has been observed around 20 cm^{-1},[2] which is called S-mode. There exist similar modes in Raman spectra of nucleoside and nucleotide crystals in which base moieties are stacked along one direction. Frequency shift of the S-mode was observed in the A-B transition of DNA and in the humidity-induced phase transition between the guanosine dihydrate and the anhydrous states.[3]

In order to determine the origin of the S-mode, we performed MD simulation for a crystal of guanosine dyhdrate made of 4×5×8 unit cells with the AMBER6 program, using the Particle Mesh Ewald method. The molecular trajectory of guanine bases in the MD simulation was Fourier transformed in time and in molecular number. Among optically active modes, one significant peak was found for the origin of the S-modes. Phase distribution of vibrational waves in the crystal of guanosine dihydrate is displayed with graphics.

References

1) S. Yoneda, Y. Sugawara, and H. Urabe, "Crystal Water Dynamics of Guanosine Dihydrate: Analysis of Atomic Displacement Parameters, Time Profile of Hydrogen-Bonding Probability, and Translocation of Water by MD Simulation", *J. Phys. Chem.* B, 109, 1304-1312 (2005).

2) H. Urabe and Y. Tominaga, " Low-lying collective modes of DNA double helix by Raman spectroscopy Biopolymers", *Biopolymers*, 21, 2477(1982)

3) H. Urabe, Y. Sugawara, and T. Kasuya, " Humidity-dependent structural transition of guanosine and disodium adenosine 5'-triphosphate crystals studied by low-frequency Raman spectroscopy", *Phys. Rev.* B, 51, 5666-5672(1995).

Intra- and intermolecular interactions between cyclic-AMP receptor protein and DNA: *Ab initio* fragment molecular orbital study

Kaori Fukuzawa[1], Yuto Komeiji[2], Yuji Mochizuki[3,4], Takeshi Ishikawa[3,4], Akifumi Kato[1], Tatsuya Nakano[4,5], Shigenori Tanaka[4,6]

Mizuho Information & Research Institute, Inc.[1], National Institute of Advanced Industrial Science and Technology [2], Rikkyo University[3], Japan Science and Technology Agency, CREST[4], National Institute of Health Sciences[5], Kobe University[6]

The ab initio fragment molecular orbital (FMO) calculations were performed for the cAMP receptor protein (CRP) complexed with a cAMP and DNA duplex to elucidate their sequence-specific binding and the stability of the DNA duplex, as determined by analysis of their inter- and intramolecular interactions. Calculations were performed with the AMBER94 force field and at the HF and MP2 levels with several basis sets. The interfragment interaction energies (IFIEs) were analyzed for interactions of CRP-cAMP with each base pair, DNA duplex with each amino acid residue, and each base pair with each residue. In addition, base-base interactions were analyzed including hydrogen bonding and stacking of DNA. In the interaction between DNA and CRP-cAMP, there was a significant charge transfer (CT) from the DNA to CRP, and this CT interaction played an important role as well as the electrostatic interactions. It is necessary to apply a quantum mechanical approach beyond the “classical” force-field approach in order to describe the sequence specificity. In the DNA intramolecular interaction, the dispersion interactions dominated the stabilization of the base-pair stacking interactions. Strong, attractive 1,2-stacking interactions and weak, repulsive 1,3-stacking interactions were observed. Comparison of the intramolecular interactions of free and complexed DNA revealed that the base-pairing interactions were stronger, and the stacking interactions were weaker, in the complexed structure. Therefore, the DNA duplex stability appears to change due to both the electrostatic and the CT interactions that take place under conditions of DNA-CRP binding.

Reference

1) K. Fukuzawa et.al., *J. Comp. Chem.* 27, 948-960 (2006).

Development of virtual screening method using Solvation Free Energy Density Model(SFED)

Chang Joon Lee[1], Se Han Lee[1], and Kyoung Tai No[1]
Yonsei University [1]

Solvation plays an important role in protein-ligand association and has a strong impact on comparisons binding energies for dissimilar molecules.

In Solvation Free Energy Density Model(SFED), the solvation free energy of a solute molecule can be calculated by summing the product of the linear combination coefficients and the basis interaction functions for the solute. Especially in solvated system, hydrophobicity is an important factor in molecular recognition and the accurate prediction of the binding modes of ligand to proteins in aqueous solvent, which is uesful for ligand docking and drug design. SFED maps of each protein and ligand could be calculated by using the Solvation Free Energy Density (SFED) model. The density map is understood that interaction field between a solute and solvent at specific three dimensional space.

The basic concept of density is to extract the information present in 3D molecular density map into 3D vectors which are easy to understand and to interpret. Thus, 3D vectors matching is applied to fast automate filtering of structure-based screening. Accordingly, we provide SFED model which is essential for successful virtual screening.

References

1) K.T No. and H.A Scheraga, “Determination of Net Atomic Charges Using a Modified Partial Equalization of Orbital Electronegativity Method. I~V” *J. phys. Chem*, 1990,1993,1995

2) YoungYong. In, Han Ha Chai and K.T No, “A Partition Coefficient Calculation Method with the SFED Model”, *J. ACS*, 2005

Computational Study of Catalytic Reaction of Adenylate Kinase

Kenshu Kamiya, Shinya Morishita and Yoko Sugawara
School of Science, Kitasato University

Adenylate kinase catalyzes the following reaction:

$$\mathrm{ATP} + \mathrm{AMP} + \mathrm{Mg}^{2+} \rightleftarrows \mathrm{ADP} + \mathrm{ADP} + \mathrm{Mg}^{2+}$$

In order to elucidate the nature of the catalytic reaction of adenylate kinase, we carried out model calculations using molecular mechanics and quantum chemistry.

We constructed the complex structure of enzyme and substrates, ATP and AMP with Mg ion, based on the PDB data 2AKY, complex of enzyme and inhibitor AP5. After an extra PO3 group was removed from AP5 to make ATP and AMP molecules, the whole molecular structure was optimized except for the main chain atoms with the force field of AMBER. The molecular dynamics calculation in 300K was performed to investigate the role of water molecules and for the structural refinement to the model structure trancated within 10 A from substrate molecules. The quantum chemical calculations were carried out for the model structure of the enzyme, the phosphate groups of substrates, Mg^{2+} and water molecules coordinating the ion, and some molecules which model the side chain atoms of surrounding amino acids. The reactant, product and transition state structures were optimized within partial degrees of freedom of the model.

The transition structure accompanied with the phosphate group inversion like the SN2 reaction was determined using partial optimization of model structure using ONIOM (HF/6-31G(d):PM3) method without the surrounding protein molecule. However, more quantitative calculations using DFT method or extended basis sets, or with surrounding amino acid residues, showed that these effects were quite large.

Also, molecular dynamics calculations revealed the possibility that some water molecules participated in the reaction mechanism.

NFV-Resistant Mechanism due to Non-active Site Mutation N88S on Subtype AE HIV-1 PRs

Hirotaka Ode[1], Shou Matsuyama[1], Saburo Neya[1], Masayuki Hata[1],
Wataru Sugiura[2], Tyuji Hoshino[1,3]
Graduate School of Pharmaceutical Sciences, Chiba University[1],
AIDS Research Center, National Institute of Infectious Diseases[2], PRESTO, JST[3]

Abstract

Human immunodeficiency virus type 1 protease (HIV-1 PR) is one of the proteins that several anti-HIV-1 drugs are targeting on. NFV is a currently available HIV-1 PR inhibitor (PI). However, the drug efficacy of NFV is reduced because of the appearance of resistant mutations in HIV-1 PRs. Ariyoshi et al. reported that patients failing in treatment with NFV predominantly carry D30N mutant of HIV-1 PRs if they were infected with the subtype B virus[1]. On the other hand, N88S mutant of HIV-1 PRs predominantly emerges in patients of treatment failure with NFV and of carrier of the subtype AE virus[1]. Both D30N and N88S are related to the resistance against NFV. We previously reported that an active site mutation D30N on subtype B PR conferred resistance by canceling direct hydrogen bonds between the 30th residue and NFV[2]. In contrast, there still remains a question why N88S mutation on subtype AE confers resistance against NFV, because N88S mutation occurs at the non-active site of PR. The non-active site residues do not have any direct contacts with PIs. In this study, we have executed molecular dynamics (MD) simulations of subtype AE reference PR and N88S mutant in complex with NFV. The simulations suggest that N88S mutation on subtype AE PR affects binding with NFV indirectly. N88S mutation creates a new hydrogen bond between mutated side chain of S88 and side chain of D30. The formation of this hydrogen bond dislocates D30 to make a distance from NFV and NFV and reduces interaction between D30 and NFV. As shown in our previous report[2], D30 is one of the most important residues with which NFV interacts. This is the reason why N88S mutation confers resistance against NFV.

References

1) K. Ariyoshi et al., “Patterns of Point Mutations Associated With Antiretroviral Drug Treatment Failure in CRF01_AE (Subtype E) Infection Differ From Subtype B Infection”, *JAIDS*, 33, 336-342 (2003).
2) H. Ode et al., “Resistant Mechanism against Nelfinavir of Human Immunodeficiency Virus Type 1 Proteases”, *J. Phys. Chem. B.* 109, 565-575 (2005).

Evaluation of the binding affinity of antibodies from antigen binding energy and V_H-V_L interaction energy calculated by the MM-GBSA method

Koji Iwamoto[1], Saburo Neya[1], Tyuji Hoshino[1,2] ,
Graduate School of Phamaceutical Science,Chiba University[1], PRESTO,JST[2],

Abstract

Because antibodies are a kind of proteins having high affinity and specificity for antigen, they are expected to be suitable candidates for drugs which have more efficient and less side effect than conventional chemical drugs. In fact, the increasing number of biopharmaceutical drugs are now available or on trial. To support the developments of these drugs, rapid and highly accurate screening is indispensable. In silico screening is one of powerful solutions to this purpose.

In this study, we performed molecular dynamics simulation of antibody HyHEL-10 and its antigen white hen egg lysozyme(HEL) to collect the trajectories of binding structures. After MD simulation, we have executed the molecular mechanics-Generalized Born surface area(MM-GBSA) method to evaluate the binding affinity between HyHEL-10 and HEL. Antigen binding energy calculated with the MM-GBSA method is partially incompatible with the changes in free energy measured by isothermal titration calorimetry(ITC)[1,2]. The variants mutated in the light chain (LS91A,LS93A,LY50F) show less satisfactory result than the variants mutated in the heavy chain(HD32A,HD96A,HD32AD96A). To resolve the discrepancy between the experimental and the computational results, we have also calculated interaction energy between antibody heavy chain variable region flagment(VH)/antibody light chain variable region flagment(VL) with MM-GBSA. Combination of these two calculated energy shows better correlation with the experimental results than the respective energy.

References

1) M. Shiroishi, A. Yokota, K. Tsumoto, H. Kondo, Y. Nishiyama, K. Horii, M. Matsushima, K. Ogasahara, K. Yutani, I. Kumagai, *J. Biol. Chem.*, **276**, 23042-23050 (2001)
2) A. Yokota, K. Tsumoto, M. Shiroishi, H. Kondo, and I. Kumagai, *J. Biol. Chem.*, **278** 5410-5418 (2003)

Prediction of complex structure comprised of a protein and a glycosaminoglycan using docking simulation and cluster analysis

Tsubasa Takaoka[1], Noriaki Okimoto[2] , Kenichi Mori[1], Saburo Neya[1], Tyuji Hoshino[1,3]
[1]*Graduate School of Pharmaceutical Sciences, Chiba University*
[2]*RIKEN, GSC* [3]*PRESTO ,Japan Science and Technology Agency*

One of the docking programs, AutoDock3.0, predict the information on the structure of ligand-recepter complexes and their stability as a docking energy. We have developed an original method combining AutoDock3.0 with a cluster analysis, and succeeded in remodeling the docking structure to be more accurate than the original output of AutoDock3.0.

At first, we tried to predict a complex structure of basic Fibroblast Growth Factor (bFGF) with heparin, using only AutoDock3.0. Two X-ray crystal structures were already obtained for bFGF. One is a complex of the protein and a heparin, which is a kind of the glycosaminoglycans, and the other is only for a protein, which will be called simplex structure hereafter. Considering induced fitting, we docked a heparin taken from the complex crystal structure onto the simplex protein, and the energy of the docked complexes was minimized by AMBER8.

Although the docking energy calculated by AutoDock3.0 could not satisfactorily predict the correct complex structure, we found out the majority of the models generated by AutoDock3.0 were similar to the crystal structure. Hence, we utilized only the structures for the evaluation, and carried out a cluster analysis with them. This procedure has successfully selected a quite similar structure to the crystal one.

We have tried to predict other heparin-binding proteins, 1E03&1E04, 1G5N&1A8A using the above approach. Two crystal structures, complex structure and simplex one, are also available for these proteins as well as bFGF. The trials show fine prediction of binding structures of heparin and the proteins.

The approach proposed in this study will be effective for docking of any other biomolecules with proteins.

References

1) Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. “Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function”, *J Comput Chem.* 1998;**19**:1639–1662.

Catalytic reaction of the enzyme degrading biodegradable plastics

Y. Sakae[1], T. Matsubara[1], M. Aida[1], H. Kondou[2], K. Masaki[3], and H. Iefuji[3]
Hiroshima University[1], National Institute of Advanced Industrial Science[2],
Technology National Research Institute of Brewing[3]

The environmental pollution caused by the disposal of the plastics is quite serious. Biodegradable plastics that can be decomposed to water and carbon dioxide by microorganisms widely distributed in the nature have attracted much attention as “Green Plastics”. Recently, it has been reported that an enzyme, which belongs to the class of serine proteases, degrades the biodegradable plastics very efficiently through the hydrolytic reaction[1]. The origin of the high catalytic activity of this enzyme derived from the yeast *Cryptococcus* sp. Strain S-2 has not been clarified yet. We therefore theoretically examine the mechanism of the catalytic reaction of this enzyme using the model molecule by the density functional theory method (B3LYP) with the basis set of 6-31G**.

The catalytic reaction consists of four elementary steps, and the rate-determining step is the third step, the nucleophilic attack of an H_2O molecule to the substrate. We examine in detail the third step using a larger molecular system of the active site taking account of the environmental effect of the neighboring amino acid residues. We replaced Thr17 by Ala17 and examined the contribution of Thr17 (-OH) to the catalysis. As a result, the activation barrier of the rate-determining step was increased by 3.4 kcal/mol by this replacement, because the transition state is destabilized in energy by the loss of the hydrogen bond between Thr17 (-OH) and the H_2O that attacks the substrate. Thus, it is probable that the neighboring amino acid residue plays an important role on the catalysis.

References

1) K. Masaki et al., “Cutinase-Like Enzyme from the Yeast *Cryptococcus* sp. Strain S-2 Hydrolyzes Polylactic Acid and Other Biodegradable Plastics”, *Applied and Environmental Microbiology*, 71, 7458-7550(2005).

In silico drug screening by docking score index method using principal component analysis

Sukumaran Murali,[1] Yoshifumi Fukunishi[2] and Haruki Nakamura[2,3]

[1] Japan Biological Information Research Center (JBIRC), Japan Biological Informatics Consortium (JBIC) , [2] National Institute of Advanced Industrial Science and Technology (AIST), Japan, and [3] Institute for Protein Research, Osaka Universit*y*

The docking score index (DSI) method is one of *in silico* drug screening methods and is a sort of similarity search method based on a protein-compound docking affinity calculated by a docking software. By the DSI method, the protein-compound docking scores are converted to the docking score indexes by the principal component analysis (PCA) method and each compound is projected into a PCA space. The DSI method projects all the compounds onto a subspace of the PCA space to reduce the computational noise and selects the compound, which is close to the known active compound. The database enrichment by the DSI method strongly depends on how to select the principal components to form the subspace. In this study, we propose an automatic method to select a set of suitable principal component axes and evaluate the database enrichment by this method for 12 target proteins. This work was supported by a grant from NEDO.

References

1) Y. Fukunishi et al., “Similarity among receptor pockets and among compounds: Analysis and application to *in silico* ligand screening”, J. Mol Graph & Modelling, 24, 34-45 (2005).

2) Y. Fukunishi et al.,“Classification of chemical compounds by protein-compound docking for use in designing a focused library”, J. Med. Chem., 49, 523-533 (2006).

3) Y. Fukunishi et al.,“Multiple target screening method for robust and accurate in silico screening”, J. Mol Graph & Modelling,25, 61-70 (2005).

4) Y. Fukunishi et al., “A virtual active compound produced from the negative image of a ligand-binding pocket, and its application to in-silico drug screening”, J. Compt-Aided Mol Design, (2006) accepted

5) Y. Fukunishi et al., “Finding ligands for G-protein coupled receptors based on the protein-compound affinity matrix”, J. Mol Graph & Modelling, (2006) accepted

Solvent Site-Dipole Field Mediating Docking of Biomolecules

NorikazuTakano[1], Nobuyuki Hamasaki[1], Junichi Higo[1]
Tokyo University of Pharmacy and Life Sciences[1]

Biomolecular interaction occurs in solvent. According to a continuum-solvent theory (e.g. Poisson-Boltzmann equation), the electrostatic interaction between biomolecules decays rapidly with increasing the inter-biomolecular distance. Thus, we naturally imagine that the biomolecules interact effectively only when they come close to each other by chance through diffusive motions.

On the other hand, we have computationally demonstrated that the orientation of water molecules is ordering around biomolecules [1-3]. We called the patterns of ordering 'a solvent site-dipole filed'. Furthermore, we have computationally observed the solvent site-dipole field formed in an inter-molecular zone between two small biomolecules, where the minimum inter-biomolecule distance is 14 Å [4]. The mechanism of the field formation is as follows: atomic charges on protein surfaces restrain the orientation of the first-layer water molecules [3]. This first-layer ordering restrains the orientation of the second-layer water molecules, and so on. The water molecules visiting the crystal-water sites anchored the solvent site-dipole field. We have shown the fluctuations of the solvent site-dipole field around a protein, which can be a trigger of biomolecular interaction [5].

In this study, we did a molecular dynamics simulation of two biomolecules (a protein and a peptide), which were located apart to each other in an explicit solvent environment in the initial state of simulation. We observed a process where the two biomolecules approach close to each other. Then, we analyze the solvent site-dipole field formed in the inter-biomolecule zone and discuss the biological significance of the solvent site-dipole field.

References

1) J. Higo, H. Kono, H. Nakamura, A. Sarai, Proteins 40 (2000) 193.
2) J. Higo, M. Sasai, H. Shirai, H. Nakamura, T. Kugimiya, Proc. Natl. Acad. Sci. USA 98 (2001) 5961.
3) J. Higo, M. Nakasako, J. Comp. Chem. 23 (2002) 1323.
4) J. Higo, M. Sasai, H. Shirai, H. Nakamura, T. Kugimiya, Proc. Natl. Acad. Sci. USA 98 (2001) 5961.
5) T. Yokomizo, S. Yagihara, J. Higo, Chem. Phys. Lett. 374 (2003) 453.

Application of Amino Acid Descriptors for Prediction of MHC-Peptide Binding Affinity

Tomoko Niwa[1]
Discovery Research Laboratories, Nippon Shinyaku Co., Ltd.[1]

Prediction of the binding affinities of antigen peptides to major histocompatibility complex (MHC) molecules is important for not only developing vaccines but also understanding peptide-protein interactions. Various models have been developed to predict these binding affinities. However, most models were so complicated that it was difficult to understand the specificity of peptide-protein interactions.

Amino acid descriptors such as electronic, steric and hydrophobic ones have been successfully applied to quantitative structure-activity relationship (QSAR) studies on bioactive peptides.[1,2] The objective of this study was to develop “interpretable” QSAR models for prediction of MHC-peptide binding affinities by using amino acid descriptors and multiple regression analysis (MRA) method, where relevant descriptors were selected with a genetic algorithm. Models were derived for HLA-A*0201,[3] H2-Db,[4] and H2-Kk.[4] The predictabilities of our models were comparable to those of previously reported models. Use of amino acid descriptors and MRA enabled us to clarify which amino acids and which properties were important for the binding and to understand the specificity of peptide-protein interactions.

References

1) Sotomatsu-Niwa, T.; Ogino, A. Evaluation of the Hydrophobic Parameters of the Amino Acid Side Chains of Peptides and their Application in QSAR and Conformational Studies. *THEOCHEM* **1997**, *392*, 43-54.
2) Niwa, S. T.; Ogino, A. Multiple Regression Analysis of the Beta-Sheet Propensity of Amino Acids. *THEOCHEM* **1997**, *419*, 155-160.
3) Doytchinova I. A.; Walshe V.; Borrow P.; Flower D. R. J. Towards the Chemometric Dissection of peptide--HLA-A*0201 binding affinity: comparison of local and global QSAR models. *Comput. Aided Mol. Des.* **2005**, *19*, 203-12.
4) Hattotuwagama C. K.; Doytchinova I. A.; Flower D. R. In Silico Prediction of Peptide Binding Affinity to Class I Mouse Major Histocompatibility Complexes: A Comparative Molecular Similarity Index Analysis (CoMSIA) Study. *J. Chem. Inf. Model.* **2005**, *45*, 1415-23.

Molecular modeling and 3D-QSAR studies of benzothiazol-2-ylacetonitrile derivatives as c-Jun N-terminal kinase 3 inhibitors

Abdul Rajjak Shaikh[1], Mohamed Ismael[1], Hideyuki Tsuboi[1], Michihisa Koyama[1], Akira Endou[1], Hiromitsu Takaba[1], Momoji Kubo[1,2], Carlos A. Del Carpio[1], Parasuraman Selvam[1], Kazumi Nishijima[3], Tetsuya Terasaki[1], Akira Miyamoto[1]

Tohoku University, [2]JST-PRESTO, [3]Mochida Pharmaceutical Company

Abstract

The c-Jun N-terminal kinases (JNKs), also known as stress-activated protein kinases, are members of mitogen activated kinase family. They are generally associated with apoptotic and inflammatory responses induced by a variety of chemical and physical stresses including oxidative stress, UV light, inflammatory cytokines, and osmotic shock. In the present study, for the first time, we report molecular modeling and three-dimensional quantitative structure activity relationship (3D-QSAR) studies performed on 44 JNK 3 inhibitors.[1,2] The 3D-QSAR models, which include molecular field analysis (MFA) and receptor surface analysis (RSA), was employed using 34 compounds as training set and its predictive ability was assessed using a test set of 10 compounds. The predictive 3D-QSAR models have r^2 values of 0.849 and 0.766 while the r^2_{cv} values of 0.616 and 0.605 for MFA and RSA, respectively. Fig.1 depicts the actual and predicted activity of all the studied compounds. The results of the 3D-QSAR models were further compared with docking studies. On the one hand, the MFA 3D-QSAR model demonstrate the importance of several electrostatic and steric properties required for the activity. On the other hand, the RSM 3D-QSAR necessities the importance of hydrophobic and hydrogen bonding property required for JNK3 activity. These observations further validated with docking studies and correlated well with 3D-QSAR models.

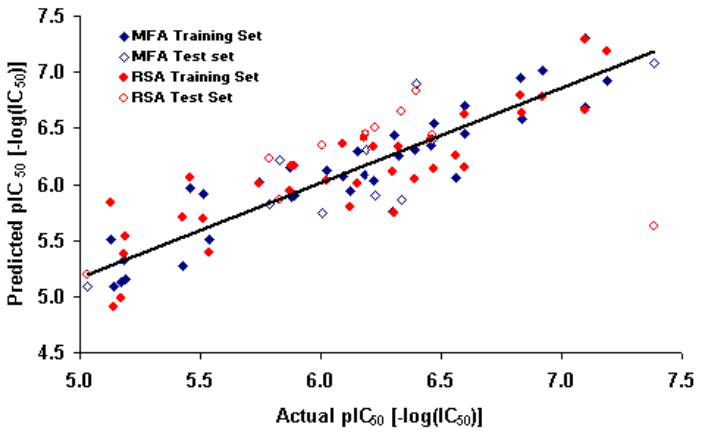


Fig. 1: The predicted and actual activities of JNK-3.

References

1) P. Gaillard et al., "Design and synthesis of the first generation of novel potent, selective, and in vivo active (Benzothiazol-2-yl) acetonitrile inhibitors of the c-Jun N-terminal kinase", *J. Med. Chem.*, 48, 4596-4607 (2005).
2) A. R. Shaikh et al., "Three-dimensional quantitative structure activity relationship (3D-QSAR) and docking studies on (benzothiazole-2-yl) acetonitrile derivatives as a c-Jun N-terminal kinase 3 (JNK3) inhibitors", *Bioorg. Med. Chem. Lett.* (in press).

Bootstrap-based consensus scoring method for high-performance and robust drug screening by single scoring function

Hiroaki Fukunishi[1], Reiji Teramoto[1], Jiro Shimada[1]
Fundamental and Environmental Research Laboratories, NEC corporation[1]

We developed a bootstrap-based consensus scoring (BBCS), which is inspired from bagging of ensemble learning, to improve performance of the existing scoring function. In the BBCS, the bootstrap method [1] is employed to increase a training set from one to two more, which are used for optimizing energy parameter set of the scoring function. Each training set includes protein-ligand complexes; each complex has X-ray structure and decoy ones. The optimization method, which is employed in each training set, explores energy parameter set maximizing an average z-score of X-ray structure for all complexes. Once two or more different energy parameter sets are generated, by combining scoring with these parameter sets, consensus scoring can be implemented. The reason why we reached the idea of the BBCS is that it is difficult to uniquely determine energy parameter set give global minimum for binding free energy from one training data, in particular, when properties of complexes in training set are biased. On the other hand, a conventional consensus scoring (CS) [2] gets consensus among a few scoring functions. Judging from published reports [3,4], it seems that it is not easy to find the best combination of scoring functions because of heuristic search. In contrast, for the BBCS, the bothersome work such as the heuristic search is not necessary. In this study, we applied BBCS to FlexX scoring function. From training data including 50 complexes, 100 energy parameter sets were generated. Using these parameter sets, we could get consensus of much more scorings than CS. To estimate performance of the BBCS, against each of 48 complexes as test data, X-ray structure and 100 decoy ones were ranked. Compared to original FlexX scoring function and CS by CScore(Sybyl), improvements of rank of X-ray structure were shown in 61.9% and in 62.5% of test data, respectively. In particular, for the complexes binding with mainly hydrophobic interactions, the BBCS were quite effective.

References

1) Breiman, L., Machine Learning., 24 ,123-140(1996).
2) Charifson, P. S. et al., *J. Med. Chem.*, 42, 5100-5109(1999).
3) Stahl, M., Rarey, M., *J. Med. Chem.*, 44, 1035-1042(2001).
4) Wang, R. et al. *J. Med. Chem.*, 46, 2287-2303(2003).

Introduction of initial binding orientation in protein-protein docking

Jong Young Jung[1], Chang Joon Lee[1] and Kyoung Tai No[1,2]
Yonsei University[1], Research Institute of Bioinformatics and Molecular design[2]

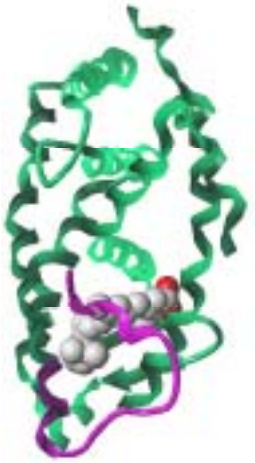
In protein-protein complex, the interface of the complex has different characteristics from other area of surface. Because the phenomena of protein-protein docking occur in the cellular environment, solvent effect has to be considered in docking. Surely salvation affects on binding interfaces in each protein to orient each other. In that sense, hydrophobicity is considered to be the one of principal factors to bind proteins in the correct sites. We generate hydrophobic field around each protein using SFED (Solvation Free Energy Density) model. And then, through the process to analyze hydrophobic fields near protein-protein interface, we determine the hydrophobic standard to guide proteins to attach each other. But hydrophobicity is not sufficient to determine interfaces in protein pairs which form a complex. Interface residue information, chemical characteristics on interface and shape complementarity between interfaces are also important factors. Using all of these informations, the vector model is constructed, which represent approximate orientation of protein components that constitute protein complexes. In comparison with most of the existing methods to simulate protein-protein docking, initially introduced orientation reduce the computation and scanning time to generate putative binding complex candidates in initial stage.

Reference.

1) Youngyong In, Han Ha Chai, and Kyoung Tai No, “A partition coefficient calculation method with the SFED model”, *J. Chem, Inf, Model.* 45, 254-263(2005)

Functions of the key residues in the ligand-binding pocket of vitamin D receptor: Fragment molecular orbital-interfragment interaction energy analysis

Kenji Yamagishi[1,2], Keiko Yamamoto[3], Sachiko Yamada[3], and Hiroaki Tokiwa[4]
The University of Tokushima[1], CREST-JST [2],Tokyo Medical and Dental University[3], Rikkyo University [4]



All electron calculations based on the *ab initio* Fragment Molecular Orbital (FMO) method enable to evaluate correctly the interactions between ligand/substrate and residues of receptor/enzyme.

In this study, FMO-interfragment interaction energy calculations of the vitamin D receptor (VDR)/1α,25-dihydroxyvitamin D_3 complex were utilized to assign functions of key residues in the ligand-binding pocket (LBP) of the VDR. We have reached the following conclusions. (1) The important residues that form hydrogen bonds to the 1-, 3- and 25-OH groups are Arg274, Tyr143 and His397, respectively. Only one residue forms a significant interaction with the corresponding hydroxy group of the ligand, although two residues are located around each hydroxy group. (2)The degradation of binding affinity for derivatives upon removal of a hydroxy group is closely related to the trend in the strength of the hydrogen bonds. (3)Arg274 forms the strongest hydrogen bond with the ligand in the LBP of the VDR. Type II hereditary rickets due to an Arg274 point mutation is caused by the elimination of the strongest hydrogen bond between the ligand and the residues in the LBP of the VDR.

References

1) K. Yamamoto, D. Abe, N. Yoshimoto, M. Choi, K. Yamagishi, H. Tokiwa, M. Shimizu, M. Makishima, S. Yamada, J. Med. Chem. 49 (2006) 1313.
2) K. Yamagishi, K. Yamamoto, S. Yamada, H. Tokiwa, Chem. Phys. Lett. 420 (2006) 465.

A Workbench for Selective Nuclear Receptor Modulators

Naomi Komiyama[1], Tatsuya Nakano[2], Kaori Fukuzawa[3], Junpei Komiyama[4], Masumi Yukawa[5], Hiroshi Tanaka[5], Kotoko Nakata[6], Tsuguchika Kaminuma[5]

Chem-Bio Informatics Society[1], National Institute of Health Sciences[2], Mizuho Information and Research Institute[3], The University of Tokyo[4], Tokyo Medical and Dental University[5], Advance Soft Co./The University of Tokyo[6]

The Selective Nuclear Receptor Modulators (SNuRMs) are the chemicals (lignds) that bind to nuclear receptors to modulate transcriptional and post-transcriptional signal transduction pathways and networks. Some of these chemicals are drugs already in the market used for wide range of diseases including classical endocrine hormonal diseases, inflammation, breast and prostate cancer, and metabolic syndrome. Well-known examples of SNuRMs are breast cancer drugs, tamoxifen and raloxifene, which are classical hormone estrogen receptor agonists/antagonists, and is the PPARγ agonist thiazolidinediones, which is one of the most well known therapeutic agent for diabetes. Ligands for LXRs, and FXRs are also under intensive investigation from anti-obesity and reverse insulin resistance viewpoint for improving metabolic syndrome.

However these existing drugs in the market show some adverse effects, and fine-tuning of structures of the existing drugs or finding new structural scaffold drew attention of drug designers in this field. Identifying ligands for receptors, classified as true orphan receptors at present, is another challenge, for these ligands may be transformed to lead chemicals for new drugs. Except few cases SNuRMs research has been carried out focusing on particular target or a class of targets in mind rather than on the whole superfamily members. Eventual goal of our workbench development is to provide essential data, docking study tools, and graphical analyzing tools for all members of the family as the infrastructure of SNuRMs research.

We have employed the FMO (fragment molecular orbital) method for energy calculation, and chose ZINC, PubChem, PDB, and *KiBank* as the basic molecular databases, and BioStation as model builder and viewer. Since these component resources are distributed as free software for nonprofit use, our workbench will also be able to be used freely. The workbench has been developed in parallel with carrying several docking studies for selective receptors such as FXR.

References

1) T. Kaminuma, *CBI Journal*, **3**, 130-156 (2003)

2) K. Fukuzawa, et al., *J. Comp. Chem.*, **26**, 1-10 (2005)

Discrimination of β-barrel Membrane Proteins: Comparison between Statistical Methods and Machine Learning Algorithms

M. Michael Gromiha and Makiko Suwa

Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-42 Aomi, Koto-ku, Tokyo 135-0064, Japan

Discriminating β-barrel membrane proteins from other folding types of globular and membrane proteins is an important task both for identifying β-barrel membrane proteins from genomic sequences and for the successful prediction of their secondary and tertiary structures. We have developed statistical methods and machine learning techniques for discriminating β-barrel membrane proteins using amino acid composition, residue pair preference and motifs.The information about amino acid composition could correctly identify the β-barrel membraneproteins at an accuracy of 89% and exclude globular and α-helical membrane proteins at the accuracy level of 80% (1,2). The residue pair preferences and motifs have more information than amino acid composition and these methods improved the accuracy of more than 95% in detecting β-barrel membrane proteins (3).

On the other hand, we have used support vector machines and neural networks for discriminating β-barrel membrane proteins. These machine learning techniques improved the overall accuracy to 92%. The sensitivity and specificity are, respectively, 89% and 94%, which indicate that the machine learning techniques excluded the globular and α-helical membrane proteins at better accuracy than identifying the β-barrel membrane proteins (4). From the comparison of statistical methods and machine learning techniques we observed that the statistical methods could identify the β-barrel membrane proteins at high accuracy while an opposite trend is observed for machine learning techniques, which correctly excluded other folding types of globular and membrane proteins.

References:

1. M.M. Gromiha and M. Suwa (2005) *Bioinformatics* 21, 961-968.
2. M.M. Gromiha et al. (2005) *Nucl. Acids Res.* 33, W164-167.
3. M.M. Gromiha et al. (2005) *Comp. Biol. Chem.* 29, 135-142.
4. M.M. Gromiha and M. Suwa (2006) *PROTEINS: Structure Function. Bioinf.* (**in press**).

SPLITS: an Expanding Program for Split and Intron-Containing tRNA Prediction

Junichi Sugahara[1,2], Nozomu Yachie[1,3], Yasuhiko Sekine[4], Akiko Soma[4], Motomu Matsui[1,2], Masaru Tomita[1,2,3], Akio Kanai[1,2]
Institute for Advanced Biosciences, Keio University[1], Department of Environmental Information, Keio University[2], Bioinformatics Program, Graduate School of Media and Governance, Keio University[3], Department of Life Science, College of Science, Rikkyo University[4]

Transfer RNA (tRNA), one of the large non-coding RNA families, is a small RNA that transfers a specific amino acid to synthesize polypeptide chain at the ribosomal site during translation. In eukaryotes and archaea, approximately 15% of tRNA precursors contain intron(s) not only in anticodon loop region but also in diverse sites of the gene (intron-containing tRNA or *cis*-spliced tRNA). Also the parasite *Nanoarchaeum equitans*, a member of the Nanoarchaeota kingdom creates functional tRNA from separate genes; one encoding the 5′-half and the other the 3′-half (split tRNA or *trans*-spliced tRNA). Although recent researches reported that there are various types of tRNA genes, computational methodology for comprehensive search is yet to be established.

Here, we developed SPLITS, an expanding program for tRNAscan-SE (1) and Split-tRNA-Search (2) to improved detection of split and intron-containing tRNA genes. SPLITS initially predicts bulge-helix-bulge splicing motifs, which is well known structure of pre-tRNA introns, to determine and remove intronic regions of tRNA genes. The resulting DNA sequences are automatically queried to tRNAscan-SE. SPLITS can predict known tRNAs with single introns located at unconventional sites on the genes (100%), tRNAs with double introns (87%), and known split tRNAs (100%). The repeated trials by using random DNA sequence demonstrated high accuracy of our prediction. Our program will be very useful for identifying novel tRNA genes after genome projects. SPLITS source code is freely downloadable at http://splits.iab.keio.ac.jp/.

References

(1) Todd M. Lowe and Sean R. Eddy. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* **25**, 955-964 (1997).
(2) Lennart Randau, Richard Munch, Michael J. Hohn, Dieter Jahn and Dieter Soll. *Nanoarchaeum equitans* creates functional tRNAs from separate genes for their 5'- and 3'- halves. *Nature* **433**, 537-541 (2005).

Phylogenetic profiling approach to generate RNA-protein networks in *E. coli*

Nozomu Yachie[1,2], Koji Numata[1,2], Yoshiteru Negishi[1,2], Hiroyuki Nakamura[1,3], Junichi Sugahara[1,3], Rintaro Saito[1,3], Masaru Tomita[1,2,3]

[1]Institute for Advanced Biosciences, Keio University
[2]Bioinformatics Program, Graduate School of Media and Governance, Keio University
[3]Department of Environmental Information, Keio University

Abstract

Recent genome and transcriptome projects have identified many examples of non-coding RNA (ncRNA) genes in various species. NcRNA do not encode protein products but rather encode structural, regulatory or catalytic RNA molecules. However, cellular functions of these molecules are still largely unclear except well-known RNA families such as tRNA, rRNA and micro RNA. Although cell-wide examination of functional linkages between RNAs and proteins are required in order to understand cellular roles of RNAs, no bioinformatics or experiments based methodologies to discover them have been hitherto established.

Here, we developed a computational system to generate whole-cell network of RNA-protein functional linkages in *Escherichia coli* K12 mainly based on phylogenetic profiling. In the system, RNA- and protein-encoded regions within *E. coli* genomic DNA were initially compared with genome sequences of the other 307 species (264 prokaryotes, 15 eukaryotes and 28 archaeas) using NCBI BLAST program, and homology profiles which consists of 307 homology scores for each RNA and protein was created. Then every pairs of two molecules were evaluated on the basis of mutual information calculated using corresponding pair of E-values. The mutual information for each pair was statistically assessed according to those calculated from shuffled homology profile and statistically significant pairs were assumed to have a functional linkage.

The functional linkage network (RNA-protein and protein-protein) of *E. coli* was generated full-automatically by using ~100 CPUs. The map contained 295 RNA-protein and 2,615 protein-protein linkages among 92 RNA and 748 protein nodes. Previously reported protein complexes such as those related to transport systems, transcriptional regulators and enzyme complexes were clustered in the map. In addition, some previously reported RNA-protein linkages such those of ribosomal components (composed of large subunit and small subunit), tRNA and their corresponding aminoacyl-tRNA synthetase, and RNaseP complex were detected. The generated linkages were further validated by comparing these with the documented protein-protein interactions in EcoCyc and DIP databases. The coverage and accuracy were significantly different from random network and also higher than those generated by previous works using phylogenetic profiling. We also mapped these links into 111 unique metabolic pathways in KEGG database and found that the linked proteins and RNAs are likely to participate in the same pathway. Based on our results and the other genome-scale data, e.g. microarray expression profiles and results of comprehensive gene knockout experiments, possible function and metabolisms of ncRNAs are discussed.

Development and application of a molecular dynamics simulation system : myPresto

Satoru Kubota [1], Yoshifumi Fukunishi [2], Ikuo Fukuda [1], Eiji Kanamori [1], Katsumi Omagari [1], Haruki Nakamura [3]

Japan Biological Information Research Center (JBIRC), Japan Biological Informatics Consortium (JBIC) [1], Biological Information Research Center (BIRC), National Institute of Advanced Industrial Science and Technology (AIST) [2], Institute for Protein Research, Osaka University [3]

We have developed a molecular dynamics simulation system: myPresto. The aim of this system is *in silico* drug screening with general simulation for protein-ligand complexes. The current version, version 3 was released in March, 2006. In this version 3, several new facilities are provided: generalized ensemble methods, in particular, a simulated tempering method, protein-ligand docking method, methods for construction of focused libraries, and a 3-D structural database having more than three million compounds for *in silico* drug screening.

We have enhanced our original *in silico* screening methods, multiple targets screening (MTS) method and docking score index (DSI) method, based on a multi-receptor vs. multi-ligand docking affinity matrix. According to these methods, compounds in a database were docked to multiple proteins. In the MTS method, the compounds, which likely bind to the target protein among these proteins, are selected as the members of the candidate-hit compound group. The DSI method is one of similarity search methods and utilizes the docking scores of each compound as indexes instead of 1D/2D descriptors by the conventional similarity search method.

In addition, the component programs of myPresto were designed for parallel computers.

This work was supported by grants from NEDO and METI.

References

1) Y. Fukunishi et al., "The filling potential method: A method for estimating the free energy surface for protein-ligand docking",J. Phys. Chem. B. (2003) 107, 13201-13210.

2) J.G. Kim et al., "Multicanonical molecular dynamics algorithm employing adaptive force-biased iteration scheme", Phys. Rev. E 70, 057103 (2004).

3) Y.Fukunishi et al.,“Similarity among receptor pockets and among compounds: Analysis and application to *in silico* ligand screening”, J. Mol Graph & Model, 24, 34-45 (2005).

Automatic Procaryote ORF Finding by Markov Model

Daigo Nakahara, Hiroyuki Nakajima, Yuuichi Tokuda,
Atsushi Hanada, Yuuki Shibata, Koichi Takahashi
Kinki University

Gene codes of organisms have acquired their own specific stochastic biases in codons. Escherichia coli are widely adopted to build parameters of Markov model[1]. In the cases which belong to the other families than E-coli or newly found ones such as Nanoarchaeum equitans, ordinary E-coli Markov model is of no use. An automatic model building method is needed for gene findings for various kinds of genomes.

Audic and Claverier[2] have presented a new method for predicting coding regions in prokaryote genomes. In genomes there are seven kinds of stochastically different regions. They are three direct classes which are based on reading frames, three reverse (complementary) coding classes and a non-coding region. A Markov model with seven kinds of transition probabilities is adopted to find gene candidates.

With the initial randomly selected sequences, repeated improvements of transition probabilities of Markov model have discriminated gene candidates as shown in Fig.1. We searched the coding regions for 14 prokaryotes[3], in the performance of less than ten cycles (Fig.1.). After refinements of gene candidates with stop and start codons, i.e., ORFs, ca 70% of genes were found.

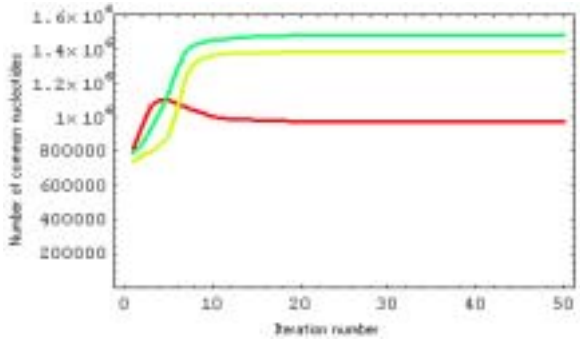


Fig. 1. Convergence of the iterative homogeneous Markov model for Escherichia coli K12

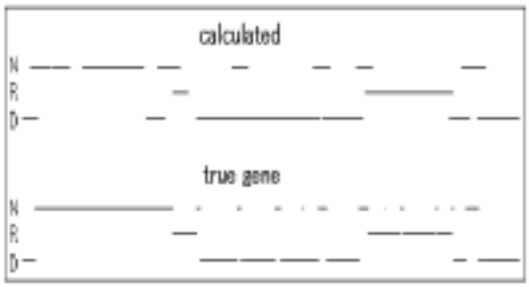


Fig.2. Direct(D), reverse(complementary) coding regions(R), and non-coding regions(N), of Bacillus subtilis by inhomogeneous Markov model

References

1) A.V.Lukashin and M.Borodovsky, *Nucleic Acids Research*, 26, 1107–1115(1998).
2) S.Audic and J.M.Claverie, *Proc. Natl. Acad. Sci. USA*, 95, 10026-10031(1998).
3) NCBI, http://www.ncbi.nlm.nih.gov/

Prediction of nuclear localization proteins based on charge periodicity of 28 residues

Noriyuki SAKIYAMA, Runcong KE, Shigeki MITAKU
Department of Applied Physics, Graduate School of Engineering, Nagoya University

There is strong correlation between the localization of proteins in a cell and their function, and the accurate prediction of the cellular localization of proteins is very interesting problem from the viewpoint of the annotation of unknown sequences. Nucleus is the largest organelle and nuclear proteins have very important biological functions such as the transcription factors. It has been detected that some DNA binding protein as Zinc Finger Proteins have charge periodicity of 28 residues by research of Ke *et al.*. Moreover, myosin and tubulin other than DNA binding proteins also had charge periodicity of 28 residues. Therefore, in this study, our aim is to focus on charge periodicity of 28 residues and discriminate of nuclear proteins from other ones.

The dataset of nuclear proteins and other proteins having charge periodicity of 28 residues were prepared from the SWISS-PROT database release 48.2. We analyzed these proteins by two steps. The first step was to analyze the average values of eight physicochemical properties of whole proteins sequences. The eight properties are the densities of positively charged residues, negatively charged residues, proline, glycine, small polar amino acid residues and aromatic residues, the amphiphilicity [1] and the hydorophobicity. The second step was the analysis of local sequences around the peek of charge symmetry [2]. The discriminant analysis was performed by these two modules. As the result, nuclear proteins having charge periodicity of 28 residues could be predicted in the specificity of about 94%.

References

1) S. Mitaku et al., “Amphiphilcity index of polar amino acids as an aid in the characterization of amino acid preference at membrane-water interfaces”, *Bioinformatics*, **18**, 608-616(2002).

2) K. Imai et al., “Secondary structure breakers and hairpin structures in myoglobin and hemoglobin”, *Chem-Bio Informatics Journal*, **5**, 65-77(2005)

WinBEST-KIT(Biochemical Reaction Simulator for Analyzing Metabolic Pathways): Application to the System Analysis of Acetone-Butanol-Ethanol Fermentation in *Clostridium saccharoperbutylacetonicum*

Masahiro Okamoto[1], Hideaki Shinto[2], Tatsuya Sekiguchi[3],Yukihiro Tashiro[2], Mayu Yamashita[2], Genta Kobayashi[4], Taizo Hanai[1], Kenji Sonomoto[3]

Graduate School of Systems Life Sciences, Kyushu University, Japan[1], Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University, Japan[2], Department of Information Engineering, Maebashi Institute of Technology, Japan[3], Ariake Sea Research Project, Saga University, Japan[4]

We have implemented an efficient, user-friendly biochemical reaction simulator named web-based BEST-KIT (http://www.best-kit.org) for analyzing large scale nonlinear networks such as metabolic pathways. The users can easily design and analyze an arbitrary reaction scheme through the Internet and the efficient graphical user interface without considering the mathematical equations. The rate velocity of reactant (dx/dt, where x and t represent reactant and time, respectively) involves several kinds of reaction terms; some are represented by mass action law and some are by approximated velocity functions of enzyme kinetics at steady state such as Michaelis-Menten, competitive inhibition and so on. The problem here is that all the reaction mechanisms of the actual reaction schemes in vivo have not always been elucidated. It is desirable for the users to speculate reaction mechanisms and to construct the corresponding mathematical equations with trial and error in the simulator. For this purpose, we have developed new version of BEST-KIT (Microsoft Windows® version named WinBEST-KIT[1]) in order for the users to define and register original mathematical equations and to customize them in user-defined library. In this study, we have applied WinBEST-KIT to the system analysis of Acetone-Butanol-Ethanol fermentation in *Clostridium saccharoperbutylacetonicum* N1-4. By using WinBEST-KIT, we made a suitable mathematical model and estimated the kinetic parameters followed by the system analysis, based on several experimentally observed time-course datasets of glucose, acetone, butanol, ethanol, acetate, butyrate and biomass. The first thing we have to do is to identify “bottleneck reaction process” in order for the production of the target metabolite (butanol in this case) to be maximized. Using the established mathematical kinetic model, we can examine the transient change of flux of each reaction process time by time (time-sliced metabolic flux analysis) after adding some perturbation to the system. In this study, we shall describe on the system analysis of mathematical kinetic model for Acetone-Butanol-Ethanol fermentation in *Clostridium saccharoperbutylacetonicum* N1-4.

References

1) T. Sekiguchi and M. Okamoto, “WinBEST-KIT: Windows-based Biochemical Reaction Simulator for Metabolic Pathways”, J. Bioinformatics and Computational. Biology, (in press)

An Integrated Database of Flavonoids

Takashi Kinoshita, Zsolt Lepp, Yoshichika Kawai Junji Terao and Hiroshi Chuman

Institute of Health Biosciences, The University of Tokushima

Abstract

Flavonoids are polyphenolic compounds that exist ubiquitously in foods of plant origin. Flavonoids show various interesting biological activities, such as removal of oxygen radical, anti-cancer action, improvement of high blood pressure, antibacterial, antibiotic and anti-allergy actions. So far, over 4000 structurally unique flavonoids have been isolated from plant sources. Recently, databases of chemicals have been utilized in to help chemical and biological researches, however the comprehensive database of flavonoids with information about structural, biological and physicochemical properties not yet available. We have constructed the integrated database of flavonoids for nutrition research. Moreover, prediction of activity against enzyme was performed using a virtual screening procedure to demonstrate a possible application of the database for pharmaceutical research.

References

1) Lepp Z, Kinoshita T, Chuman H., Screening for new antidepressant leads of multiple activities by support vector machines. J. Chem. Inf. Model, 46, 158-67, 2006.

2) Kinoshita T, Lepp Z, Kawai Y, Terao J Chuman H., An Integrated Database of Flavonoids. Biofactors, in press.

Estimation of Gene Regulatory Network by means of AdaBoost

Shinya Nabatame[1], Hitoshi Iba[1]
School of Engineering The University of Tokyo[1]

The major step in inferring a Gene Regulatory Network (GRN) from the time series data of gene expression is the estimation of the values of different parameters of a system of differential equations. In this context, various models such as S-system, linear model, SDE model have been proposed. To measure the goodness of a model, criteria such as Akaike Information Criteria (AIC) and Mean Square Error with heuristics have been used. However, the inferred GRN by these models do not agree well with the real network confirmed with biological experiments. In other words, accuracy is not high. Additionally, the performance depends on the criteria.

Ensembles of multiple models and information criteria are expected to improve the estimation accuracy of a GRN. Consequently, we use AdaBoost, which is a very powerful algorithm for integrating multiple estimation methods, in order to improve the accuracy. In AdaBoost, the decision function uses a weighted majority voting technique. However, the success of AdaBoost depends on the way of treating the training data and on the base algorithm used as learning machine.

In GRN estimation, each data point of gene expression time series is used as training data. To be more precise, the loss function is calculated from the difference between the fitting result of learning machine and the observation data. Learning is conducted by updating weight of each data point from the value of the loss function.

Models and information criteria used as learning machine are as follows: (1) Models: linear model, SDE model, weighted matrix model. (2)Information criteria: AIC, BIC, Mallows' statistics, determination coefficient adjusted for the degrees of freedom, Hannan-Quinn IC, penalized likelihood. Thus, the number of learning machine is 18 (i.e. 3models * 6 criteria).

To validate the effectiveness of proposed method, we estimated a part of S.O.S DNA repair network of E.coli. The expressed data are the kinetics of main 8 genes, i.e. lexA, umuD, recA, uvrA, uvrY, ruvA, ruvA, and polB are monitored with 50 instants evenly spaced every 6 minutes. Our proposal method got the accuracy of 78% Sensitivity and 51% Specificity which are higher than those of almost learning machines.

As a future work, we will apply our method to other networks such as cell-cycle network of S.cerevisiae.

Theoretical studies of the ATP hydrolysis mechanism of HisP protein

Qiang Pei[1], Hideyuki Tsuboi[1], Michihisa Koyama[1], Akira Endou[1], Hiromitsu Takaba[1], Momoji Kubo[1,2], Carlos A. Del Carpio[1], Parasuraman Selvam[1], Kazumi Nishijima[1,3], Tetsuya Terasaki[1], Akira Miyamoto[1]

[1]Tohoku University; [2]JST-PRESTO; [3]Mochida Pharmaceutical Company

Abstract

The reported high-resolution crystal structures of several nucleotide-binding domains (NBD), e.g., HisP protein - a subunit of ATP-binding-cassette (ABC or traffic ATPase) of the histidine transporter complex ($HisQMP_2J$), not only offered the first overall view of NBD but also gave a significant breakthrough toward the elucidation of the binding of ATP to the protein.[1] On the other hand, how the ATP is exactly hydrolyzed to ADP is currently a matter of considerable debate, and mainly different models have been proposed.[2] In the present investigation, density functional theory (DFT) method was used to study the ATP hydrolysis reaction in HisP by considering the ATP binding site, especially the γ-phosphate group, surrounding residues and water molecules. Since the hydrolysis is favored by the presence of cations, we have substituted Wat407 with Mg^{2+}. Based on the gas phase DFT calculations, we propose that ATP hydrolysis in an aqueous environment is initiated by nucleophilic attack of a single water molecule (Wat437) on the γ-phosphate; the hydrolysis product ADP acts as leaving group. Further, we describe the proton transfer through water surrounding γ-phosphate. Transition state structure was determined by approximate saddle-point search as shown in Fig. 1. Gln100, Asp178 and Glu179 formed a coordinated complex with both β- and γ-phosphates through the interaction of Mg^{2+}. The bond distances between P_γ -$O_{\beta\text{-}\gamma}$ and between P_γ –O_w are 2.00 and 1.84 Å, respectively. The angle O_w-P_γ -$O_{\beta\text{-}\gamma}$ and the dihedral angle formed by four atoms of γ-phosphate group were also analyzed.

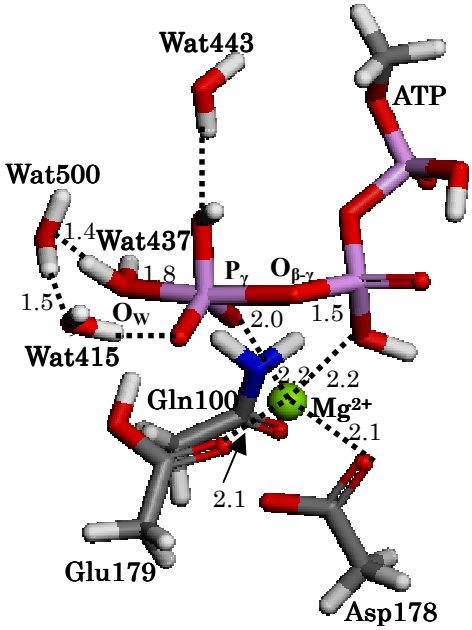


Fig. 1: Transition state structure of ATP binding site hydrolysis process of HisP protein.

References

1) L. Hung et al., "Crystal structure of the ATP-binding subunit of an ABC transporter", *Nature*, 396, 703-707 (1998).
2) C. van der Does et al., "How do ABC transporters drive transport?", *Bio. Chem.*, 385, 927-933 (2004).

Proteins with charge periodicity of 28 residues which are coded in total genomes

Runcong KE and Shigeki MITAKU
Department of Applied Physics, Graduate School of Engineering, Nagoya University

The electric charge distribution was recently analyzed for all proteins from total genomes of several biological organisms (1, 2). We calculated the autocorrelation function of electric charges of amino acid sequences in eukaryota and prokayota. The results indicated that the autocorrelation function of electric charges of all amino acid sequences showed significant positive correlation for eukaryota and no correlation for prokaryota. The characteristic length of the positive correlation in the case of eukaryota was as long as several tens residues. Furthermore, we found recently significant charge periodicity with 28 residues in the charge autocorrelation function of amino acid sequences from human genome.

In this work, we extracted the proteins, which have 28-residues periodicity, and classified their functions for revealing the relationship between the charge periodicity and the protein functions. Many of those proteins are not annotated but almost 80% of those proteins are similar to known nuclear proteins such as zinc-finger proteins. The cytoplasmic proteins with same characteristics are mainly contractile and cytoskeleton proteins such as myosin and tubulin. The nuclear proteins with charge periodicity of 28 residues are about 10% of all nuclear proteins, when data from Swissprot are analyzed. These findings indicated the relationship between charge periodicity in amino acid sequence and protein function. The charge periodicity of 28 residues may be used for the prediction of nuclear proteins, combining with the physical finger print method. In fact, the physical properties of amino acid sequences are different between the nuclear proteins and the cytoplasmic ones.

References

1) R. Ke and S. Mitaku, “Gaussian analysis of net electric charge of proteins in the *Drosophila melanogaster* genome”, *Chem-Bio Informatics J*., 4, 101-109(2004)
2) R. Ke and S. Mitaku, “Local repulsion in protein structures as revealed by charge distribution analysis of all amino acid sequences from the *Saccharomyces Cerevisiae* genome”, *Journal of Physics: Condensed Matter*, 17, 2825-2831(2005)

Systematic Errors Commonly Found in GeneChip Expression Data and Effective Compensation Methods thereof

Tomokazu Konishi[1], Keishi Hata[2], Shigeru Takasaki[3] and Akihiko Konagaya[3]
Akita Pref. Univ.[1], Akita Research Institute for Food and Brewing [2], RIKEN [3]

Measurements can never be free from errors. However, errors that have been derived by systematic reasons can be removed or reduced by calculations, if the causes of the errors are identified. In this poster, it is introduced that the systematic errors, their possible causes and compensation methods frequently found in GeneChip expression analysis arrays (Affymetrix Inc.). By using the methods, one can reduce the effects of noises and improve reproducibility of measurements.

GeneChip is a well used microarray; it is equipped with the most comprehensible format available and user-friendly manipulating system. In the meantime, calculation methods for GeneChip seem to base on arbitral ideas, which spoil objectivity of data analysis. Indeed, criteria of data evaluations have been altered in many updating of the software. Without rigid criteria, it is difficult to systematize measurement errors, since stability in analytical results cannot be expected.

By using a suitable analytical technique, the problems caused by the arbitral data handling methods for GeneChip have been dissolved. Moreover, the systematic errors are discovered through data analyzing against a parametric model. The model, which describes statistical distribution of transcriptome data, has been confirmed over many experimental evidences [3)]. Additionally, this model bases on a theory that explains formation of transcriptome in a cell, in the terms of thermodynamics [4)]. The systematic errors are found as frequently occurring tendencies in sets of data. Estimations on the causes of the errors and correction methods are discussed on experimental evidences.

References

1) T. Konishi, “A thermodynamic model of transcriptome formation” *Nucleic Acids Research*, 33, 6587-6592 (2005).
2) T. Konishi, “Three-parameter lognormal distribution ubiquitously found in cDNA microarray data and its application to parametric data treatment”, *BMC Bioinformatics*, 5, 5(2004).

Protein functional site prediction -DNA binding proteins-

Yukako Sakatsuji[1], Ayumi Suzuki[2], Ichiro Yamato[3], Satoru Miyazaki[1]

Department of Pharmaceutical Science, Tokyo university of science,
2641 Yamazaki, Noda-shi, Chiba 278-8510, Japan[1]
Department of Biological Science and Technology, Tokyo university of science, 2641 Yamazaki, Noda-shi, Chiba 278-8510, Japan[3]

E-mail：j3106635@ed.noda.tus.ac.jp

Abstract

It is difficult to define functional sites in any proteins, because classification, characterization or definition of all the functional sites are not yet clear. Recently useful methods such as fold recognition search and 3D template search are thought out, but it is not yet easy to detect a functional site in a protein even if we can find the similar structures in PDB. To overcome these problems, we developed a novel algorithm and program to predict a functional site of a protein. Our method, "FCANAL", is a program used for functional site prediction using the score matrix obtained from the distances between CA atoms and frequencies of appearance of any amino acids.

The FCANAL was applied to characterize DNA binding proteins. For ZINC FINGER C2H2 protein analysis, we constructed the score matrix to identify functional site with high accuracy and characterize the functional site. As a matter of course, this function prediction algorithm, FCANAL, worked well for other DNA binding proteins such as LEUCINEZIPPER, HOMEOBOX, and HELIX-TURN-HELIX. From these results, we are going to search for a commonality of functional sites of DNA binding proteins.

We will report these results and will discuss about the commonality of functional sites of DNA binding proteins.

References

1) A.Suzuki, T.Asaoka, I.Yamato and S.Miyazaki ,CBIJ,3,39-55(2005)
2) T.Asaoka, T.Ando, T.Meguro and I.Yamato, CBIJ,3,96-113(2003)

HP-Lattice-model Structure Prediction of Proteins using Genetic Programming

Takahiro Yabuki[1], Hitoshi Iba[1]
Frontier Informatics[1], Graduate School of Frontier Sciences, the University of Tokyo

Since folding real-life proteins is extremely complex and demands extensive resources, a simpler model is often used for benchmark computational purposes. One of the most commonly used model is Hydrophobic-Polar model (**HP-model**) , which involves folding a binary-based genome on a two or three dimensional regular lattice. In this paper, HP-model in two dimensional regular lattice (**2DHP**) and three dimensional regular lattice (**3DHP**) are used.

HP-model is widely used to simulate protein folding. Several benchmark sequences and minimum-energy solutions have been described in the literature. HP-model was conceived by researchers in order to gain insight into the folding process of a protein, and, albeit simple, it exhibits many features of real-life protein folding. HP-model considers only one force driving protein folding, namely, the hydrophobicity of amino acids. In HP-model, All amino-acids (20 kinds) which constitute protein sequences are classified into H(hydrophibic, water-haters) or P (hydrophilic, water-lovers). H amino acids try form bonds(H-H bonds) with other H amino acids. The more H-H bonds there are, the better the Structure (result) is.

In this poster, the technique based on **GP** (Genetic Programming) is introduced concerning the prediction of structure of proteins. We have extended the traditional methods through the usage of exhaustive search (**LookN**) and Counting Backward Tree (**CBT**) . LookN performs an exhaustive search of depth N and follows the best path found. For example, Look2 examines all possible paths of length 2 and then takes the one that results in maximal energy improvement. CBT creates Trees without LookN from good results (counts backward tree from results) and raises the average fitness.

In 2DHP Our method managed to find the best-known free energy value in several cases. Next, we try to extrapolate the great results in 3DHP.

References

1) K.Yue et al., “A test of lattice protein folding algorithms” *Chemistry* Vol. 92, pp. 325-329, 1995.
2) R.Jakovski et al., “Applying Genetic Programming to the HP Protein folding Problem” in Proc. of *Problem Solving from Nature* (PPSN2004), 2004.

Prediction of Human Cytochrome P450 3A4 Substrates using Statistical analysis

Hyesun Han[1], Ji Hoon Jung[1], and Kyoung Tai No[1]
Yonsei University, Seoul, Korea[1]

Human Cytochrome P450s are considered to be the most important enzyme family in drug metabolism. CYP3A4 is one of the major P450s and plays an important role in metabolic degradation of approximately 50% of all marketed drugs. Consequently, it is necessary to develop system for predicting of CYP3A4 substrates. This system will be used in predicting metabolism profile of novel compounds for drug discovery and development. To separate CYP3A4 substrates or non-substrates, we used ensembled statistical methods using logistic regression, linear discriminant analysis (LDA), support vector machine (SVM) and so on. In present study, we discussed the factors relating specificity of CYP3A4 enzyme.

References

1) Jan M. Kriegl et al., “A support vector machine approach to classify human cytochrome P450 3A4 inhibitors”, *J. Comp-Aid. Mol. Design*, 19, 189-201(2005).

Branch And Bound Median Search algorithm for finding a substrate specific motif at human SLC22 transporters

Takayuki Arakawa[1], Satoru Miyazaki[1]

Department of Pharmaceutical Science, Tokyo university of science, 2641 Yamazaki, Noda-shi, Chiba 278-8510, Japan[1]

Email：j3106601@ed.noda.tus.ac.jp

The solute carrier 22 (SLC22) family comprises organic ion and drug transporters of various tissue cells. The study of the correlation between their functions and their active sites is helpful for pharmacokinetics and a drug development. However, few SLC22 active sites have been identified because their three-dimensional structure analyses are difficult technically. Therefore, the development of an algorithm for predicting their active sites is expected.

We thought that a conserved region of protein sequences whould correlate with an active site. To find an active site of SLC22 family proteins, as well as other proteins, we developed an amino acid sequence motif finding algorithm which is based on the Branch And Bound Median Search algorithm (BABMS) for DNA motif findings. This algorithm allows us to evaluate the Hamming distance as an alignment score and finds the conserved region with a local multiple alignment of amino acid sequences. This method avoids the insertion of involuntary gaps often caused by length differences among sequences. By using this, we have analysed orthologous proteins of SLC22 in several organisms and these paralogous proteins of human respectively.

We focused on SLC22 family proteins and have searched a substrate specific motif about these proteins. In this presentation, we will discuss about the derived results.

References

1) NEAL C.JONES AND PAVEL A.PEVZNER, AN INTRODUCTION TO BIOINFORMATICS ALGORITHMS, 91-115(2000)
2) 遠藤　仁ら, ”有機イオントランスポーターファミリー”, *PNE*, vol46, No5, 583-604-611(2001)
3) 菅原　秀明ら, あなたにも役立つバイオインフォマティクス, 21-31(2004)

A Molecular Modeling Study on Antigenic Drifts of Influenza Virus Surface Glycoprotein

Manabu Igarashi[1], Kimihito Ito[1], Hiroshi Kida[1,2], and Ayato Takada[1]
Department of Global Epidemiology, Research Center for Zoonosis Control[1] and
Department of Disease Control, Graduate School of Veterinary Medicine[2],
Hokkaido University

The hemagglutinin (HA) is the major surface glycoprotein of influenza viruses and plays an important role in virus entry into host cells. The HA undergoes antigenic drifts which occur by accumulation of a series of amino acid substitutions. Influenza viruses that escape from antibody-mediated immune pressure of human population acquire a new antigenic structure and continuously cause epidemic in the world. Since the HA is a favorable target for the development of vaccines and antivirals to control influenza, it is important to understand the structural changes caused by amino acid mutations made in the evolution of influenza viruses.

Three-dimensional (3D) molecular models were built for 15 antigenically different HAs of the human H3N2 virus strains isolated during 1968 to 2006. Based on these predicted models, we analyzed the change of electrostatic potential (ESP) on the surface of HA molecule, which may associate with antigenic property. The 3D distribution of ESP showed that the charge distribution of HA surface, especially in the globular head which interact with the receptor, have gradually changed with increased positive regions. To further examine the biological significance of ESP change, we performed a retrospective and comprehensive analysis of amino acid sequences of H3N2 subtypes derived from human, swine, and avian hosts. Amino acid sequence data of about 1,900 virus strains were collected from the NCBI influenza virus resources, and used for the calculation of the isoelectric points (pI) of HA molecules. We found that the pI of human virus HAs gradually moved to basic pH during 1968 to 1984 and, since then, has been fairly constant. In contrast, avian virus HAs have retained their pI since 1963. These results suggest that these structural changes of the HA molecule were required for the adaptation of avian virus to human populations.

Analysis of amino acid sequences of motor protein by physical finger print method

Hideki Tanizawa and Shigeki Mitaku
Department of Applied Physics, Graduate School of Engineering, Nagoya University

The movement of cells and subcellular structures of eukaryotic cells are generated by the activities of motor proteins that interact with rigid cytoskeletal fibers. There are three superfamilies of motor proteins: myosin, kinesin and dynein. The sequence homology between motor proteins within a superfamily is considerably high, but the sequence homology among three superfamilies is low. They commonly have the ATP-binding sites as well as the domain which bind the cytoskeletal fibers. If the physical mechanism of the coupling between the chemical reaction and the movement of proteins is common among the superfamilies, there have to be common features also in the physical properties of sequences despite of their very low homology. Therefore, the purpose of this work is to find out the common clusters of particular physical properties around the domains which bind with the skeletal fibers, leading to the prediction of any motor proteins from all amino acid sequences coded in genome sequences. Here, we adopted the physical fingerprint approach, in which various sequences of the physical properties are used for the prediction on the assumption that the motor proteins function by the common mechanism.

First, we prepared dataset of proteins of three superfamilies. Then, the plots of the moving average of various physical properties were compared around the binding regions. We found that there are three significant clusters of negative charges. The amino acid sequences of those regions are highly conserved among superfamilies but there are no homologous sequences with known function. We developed a method to discriminate motor proteins using physical fingerprints of three regions including the clusters of negative charges. We used motor proteins as positive dataset and other kinds of ATP-binding proteins as negative dataset. As a result, the accuracy was as high as 80%. The physical meaning of the three regions with negative clusters will be discussed.

A New Method for Analysis of SNP's Effect on Protein Function and Structure

Carlos A. Del Carpio[1], Eichiro Ichiishi[2], Hideyuki Tsuboi[1], Michihisa Koyama[1], Akira Endou[1], Hiromitsu Takaba1, Momoji Kubo[1,3], Akira Miyamoto[1,2]

Graduate School of Engineering, Tohoku University[1], NICHe, Tohoku University[2], JST-PREST[3]

【Introduction】

Single nucleotide polymorphisms (SNPs) are the most frequent type of DNA variation and they are being used in association studies of gene and protein function distortion as well as genetic diseases in general. To understand the relationship between genetic and phenotypic variation in particular, it is essential to evaluate the consequences of the specific non-synonymous mutations in the structure as well as function of the respective proteins they encode.

【Method】

Here, we have mapped known distorting mutations onto the known three dimensional structure of the wild type p53 which is inactivated in the majority of cancers, mostly through single mutations that cause conformational changes in the DNA binding core domain of the protein

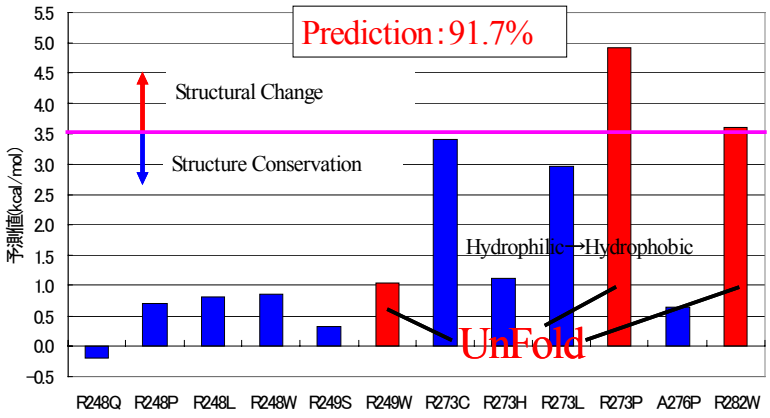


Fig. 1 Prediction of effects of SNP's in P53

【Results and Discussion】.

Propensities to changes in folding characteristics of proteins caused by cSNP's on the wild type structure can be calculated in terms of the variation of the hydrophobicity patterns of the protein surface using the above exposed scheme. To perform these calculations we have utilized the hydrophobic patch inference module of the system for bio-macromolecular interaction MIAX[1,2]. The study is performed by mapping mutations dictated by common SNP's on the wild type structure of P53 and analyzing the changes arisen on the surface of the protein.The methodology applied to the P53 protein results in a high prediction rate as shown in Fig. 1 for twelve of the most common SNP's observed experimentally.

【References】

1)Del Carpio et al. PROTEINS: Structure, Function and Genetics, 48:696-732, 2002

2)Del Carpio et al. Genome Informatics 16-2:148-160,2005

Application of Connectivity Matrix Method to Atom-level Analysis of Metabolic Networks: Calculation of Paths Connecting Two Metabolites

Jun Ohta[1]
Okayama University[1]
E-mail: jo25@md.okayama-u.ac.jp

Connectivity matrix method is an approach for analyses of biological networks. In this approach, all the connectivities of interest are expressed in a matrix, connectivity matrix. Each row of connectivity matrix, which corresponds to one connectivity, is formed by combination of three row vectors or numerals indicating two different nodes and one edge, respectively. Connectivity matrix carries information defining or characterizing nodes or connectivities. Analyses of networks are performed based on the information described in connectivity matrix on Octave and Matlab. Octave and Matlab are software suitable for manipulation of matrices and simulation. Both can also manipulate strings and be easily customized using script files and functions, indicating their utility and flexibility as analytical tools.

Atom-level analysis of metabolic networks is the first application of connectivity matrix method. Atom is the smallest node in metabolic networks. The structure of metabolic networks is defined at atom-level by inter-metabolite connectivity through enzymatic reactions and intra-metabolite connectivity through chemical bonds. Calculation of atom-level paths or cycles can be performed by connectivity matrix method. Whereas atom-level consideration is important, metabolite-level consideration is also important, because metabolic network works through consumption and production of metabolites. This presentation shows 2 algorithms for calculation of paths connecting metabolites. Both algorithms are based on connectivity matrix method. In the algorithm 1, calculation of metabolite-level path is performed based on substrate-product relationship extracted from connectivity matrix, whereas, in the algorithm 2, metabolite-level paths are formed by rearrangement of the calculated atom-level paths.

References

1) J. Ohta, “Connectivity matrix method for analyses of biological networks and its application to atom-level analysis of a model network of carbohydrate metabolism”, submitted.
2) http://www.metabo-info.org

Microarray analysis based on Standard Deviation of Effecting of Biological Factors

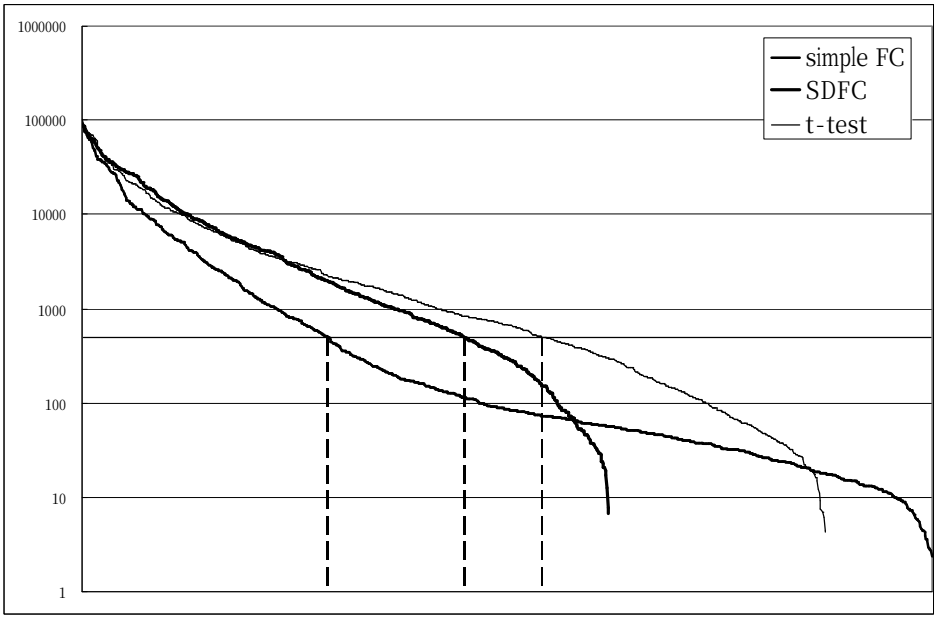
Chiaki Handa[1], Kayoko Takashima[2], Hidetoshi Maruyama[2], Naoto Nagasawa[1], Toshiki Honma[2], Takeshi Nakabayashi[2], Shinji Kikuchi[2]

Kissei comtec Co., Ltd.[1], Kissei pharmaceutical Co., Ltd.[2]

Background: The DNA microarray technology is a powerful tool for large-scale gene expression analysis. The most important step in the analysis of microarray expression profiling data is the identification of genes that display significant changes in expression signals between two biological conditions (e.g. vehicle vs treatment). However, no universal standard methods exist for the identification of a significant change in expression signals between two conditions.

Methods: In this study, we propose a method using a computer to identify genes changed significantly out of the large number of experimental data. We called this method SDFC(Fold Change analysis using Standard Deviation), which is combination of simple fold-change and t-test. The feature of SDFC is adopting not only control data from same experiment but also many control data from other experimental data as normal conditions.

Results: The result of SDFC compared with simple fold-change or t-test indicated that SDFC is more useful to display genes changed significantly than two methods. This result suggests that the effect of biological factors in the analysis of DNA microarray is important. We will report this result in details.



Detecting Cell Cycle Regulated genes in S. pombe with Non-metric Multidimensional Scaling without sinusoidal fittings

Y-h. Tagcuhi[1]

1Department of Physics, and Institute for Science and Technology, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan,[1]

Microarray is versatile method to depict relative amount of transcripted material under several experimental conditions. It can measure relative amounts for several thousands of genes simultaneously. However, this advantage can be a weak point on the other hand. Even if only limited number of genes are significant under some experimental conditions, measurements are performed for almost all genes. Since we do not know which genes are significant since this is what we would like to know, this forbids us to observe only the significant genes. Thus, it is hoped that some method can depict what the important genes are without any preknowledge.

In this poster, we would like to propose non-metric multidimensional scaling method (nMDS)[1-3] in order to check which genes are worth while investigated. *Schizosaccharomyces* ***pombe***'s cell division cycle experiments by several groups are analyzed by nMDS and compared with each other. By requiring consistency among those performed by distinct research groups, we can pick up two sets of so-called "cell cycle regulated genes" without assuming sinusoidal behavior which are frequently assumed. Polar angle obtained by the circular distributions of genes in the embedded space turns out to be cell cycle phases estimated by sinusoidal fittings in the proceeding studies in spite that we do not assume sinusoidal fittings. Biological meanings of two sets of genes (by gene ontology) as well as the problem in assuming sinusoidal behavior will be discussed.

References

1) Y.-h. Taguchi and Y. Oono, "Relational patterns of gene expression via non-metric multidimensional scaling analysis" Bioinformatics, 21, 730-740 (2005) .
2) Y-h. Taguchi, "Detecting Cell Cycle Regulated genes in S. pombe with Non-metric Multidimensional Scaling without sinusoidal fittings" IPSJ SIG Technical Report, 2005-BIO-3, 59-66 (2005).
3) Y-h. Taguchi and Satwik Rajaram, "Noise reduction procedure for microarray experiments with non-metric multidimensional scaling method" IPSJ SIG Technical Report, 2006-BIO-4, 9-16 (2006).

Comprehensive analyses on relationships between alternative-splicing patterns and developmental stages in mouse

Yasuko Takahashi, Soichi Ogishima, Hiroshi Tanaka
Dept. of Computational Biology, Tokyo Medical and Dental University.

"Human genome has almost 22,000 genes (, whereas, for example, worm genomes have almost 19,000 genes)". It is an astonishing finding of the human genome project. On the other hand, recent studies showed that RNAs from approximately 40-79% of multi-exon human genes undergo alternative splicing. These findings suggest that alternative splicing is considered to be necessary for achieving complexity of higher eukaryotes. Preceding studies have demonstrated that alternative splicing has key roles in temporal and spatial regulation of development and differentiation in higher-multicellular organisms; tissue-specific splicing variants have been reported, for example, mitochondorial ATP synthase gamma-subunit (F1 gamma) gene undergo alternative splicing in mouse muscle development.

However, the overall picture of alternative splicing in development has not been obtained. Here we show the overall relationships between alternative-splicing patterns and developmental stages in mouse by employing comprehensive analyses on mouse full-length cDNAs. We built a splicing-variant dataset which consists of 4,850 transcriptional units (TUs), based on over 102,801 FANTOM3 full-length cDNA clones. To reveal the occurrences of alternative splicing and their patterns at each developmental stage, we classified the splicing variants by their patterns and examined the relationships between these patterns and developmental stages. As a result, we found that alternative splicing rarely occurs in early developmental stage (from stage 4 to stage 9), though it might be a result of bias in a splicing-variant dataset. We also found that rates of sharing common-exons are distributed centered at 0.5. A rate of sharing common-exons were defined as a rate of numbers of transcribed exons to numbers of total exons of a transcriptional unit. In this poster presentation, we will also report the relationships between splicing variants and domain-compositional changes which result in functional complexity.

References

1) Nakai K, et al., Large-scale analyses of human alternative protein isoforms: pattern classification and correlation with subcellular localization signals, *Nucleic Acids Research*, 33, 2355-2363 (2005).
2) Tomita M, et al., Inferring alternative splicing patterns in mouse from a full-length cDNA library and microarray data, *Genome Research*, 12, 1286-1293 (2002).

Genome-wide Prediction of Novel Small Peptides in *Escherichia coli* by Using Liquid Chromatography Mass Spectrometry

Motomu Matsui[1,2], Nozomu Yachie[1,3], Kosaku Shinoda[4], Naoyuki Sugiyama[4], Fumio Ando[1,2], Yuki Okada[1,3], Martin Robert[1], Rintaro Saito[1,2,*], Masaru Tomita[1,2,3,4]

[1]Institute for Advanced Biosciences, Keio University, Tsuruoka, 997-0035, *Japan*, [2]Department of Environmental Information, Keio University, Fujisawa, 252-8520, *Japan*, [3]Bioinformatics Program, Graduate School of Media and Governance, Keio University, Fujisawa, 252-8520, *Japan*, [4]Human Metabolome Technologies, Inc., c/o Insitute for Advanced Biosciences, Keio University, Tsuruoka, 997-0017, *Japan*

MOTIVATION: uORFs (upstream Open Reading Frames) are small open reading frames which are located upstream of protein coding region (CDS) and some of them encode small peptides which act as important translational regulators in several eukaryotes. In prokaryotes, there are similar small open reading frames upstream of main CDS, which encode leader peptides although most of them are yet to be identified and their functions are unclear. Genome-wide predictions and verifications of novel small ORF candidates are all-important to investigate their cellular roles, which may provide additional insights into the prokaryotic metabolisms.

RESULTS: Here we introduce a novel method to predict small ORFs from the *Escherichia coli* (*E. coli*) genomic DNA with the combination of Liquid Chromatography Mass Spectrometry (LC/MS) and MS/MS ion search to verify their expressions. All possible longest ORFs in *E. coli* strain K12 genome were initially screened by computational approaches, and total protein extractions were analyzed by LC/MS. Then the computationally predicted ORFs are compared with the peptide signals by MS/MS ion search. As a result, 201 small peptide candidates with high specificity and sensitivity were finally extracted. Among them, six had promoter motifs and two were overlapped with downstream ORFs, and suggested as sORFs. These results suggest that there are many small peptides than previously thought and some of them may have important functions.

Transcriptomic approach for understanding the effect of cryoprotectant, trehalose and DMSO of Saccharomyces cerevisiae

Yuko Momose [1], Rena Matsumoto[1], Mine Odani[1] , Masakazu Yamaoka[1]
The National Institute of Advanced Industrial Science and Technology (AIST)[1],

Background: In our institute, International patent organism depositary (IPOD), we should keep various organisms in good states of the preservation. Freezing is a good and convenient method for the preservation, but some organisms are susceptible to freeze-thaw stress, which get serious damage and decrease viability after freeze-thaw treatment. Therefore we tried to improve more efficient cryoprotectants for those freeze-sensitive organisms. To understand of molecular mechanism of various cryoprotectants, and choose more effective combination of those, we analyzed global gene expression profiles just before freeze using the model eukaryotic organism, Saccharomyces cerevisiae.

Methods: The yeast Saccharomyces cerevisiae was frozen at -80 ◦C (cooling rate, 1 ◦C/min) for 1 day with or without cryoprotectant and thawed at 30 ◦C for 6min. After freeze-thaw, we checked viability by flow cytometry using Vybrant Apoptosis Assay #4 YO-PRO-1 kits (Molecular Probes). We analyzed the gene expression profile of yeast before freeze-thawing with and without the cryoprotectants, DMSO (dimethyl sulfoxide) and trehalose and their combination.

Results: From the aspect of the viability after freeze, DMSO (dimethyl sulfoxide) and trehalose had almost the same level of protect freeze-stress at proper concentration respectively in yeast, and after we compare the profiles of gene expression obtained from the yeast treated by DMSO or Trehalose, or their combination (DT), the profiles of those treatments were similar to induce the genes related to protein synthesis, transporter, energy and stress response categories. But there were some difference in gene expression. DMSO induced genes of protein synthesis and trehalose induced genes related to protein fate more than other treatments. From those results, we tried to pick up the gene related to survival after freeze in both cases and suggest the more effective combination of cryoprotectants.

GC contents of nucleotide sequences for proteins with charge periodicity of 28 residues

Ryusuke Sawada, Runcong Ke, Noriyuki Sakiyama, Masashi Sonoyama
and Shigeki Mitaku
Department of Applied Physics, School of Engineering, Nagoya University

The eukaryotic genomes have different regions with high and low GC contents. Length of each type of region is in the order of 100kbp. It is considered that AT-rich regions are more compact than GC-rich regions and that the housekeeping genes are mainly in GC-rich region. These facts indicate that GC content is one of the most important factors for considering the meaning of genes in terms of the expression in cells. Recently, we found that proteins with the charge periodicity of 28 residues (PCP28) are abundant in vertebrate genomes, most of which are nuclear proteins. In this work, we found that the GC content for PCP28 is very low.

We first studied the histogram of GC content of the total ORFs from 9 eukaryotic genomes, as the control data of the particular type of proteins. The histograms had two peaks and were fitted well with the combination of two Gaussian distribution. For human genome, the GC contents of the two peaks were about 0.45 and 0.6. Then, the histogram of GC content for PCP28 from human genome was plotted, the shape of which was very different from that for total ORFs, having a single peak at the GC content of 0.42.

This characteristic distribution of GC content for PCP28 is formed in the process of evolution under some pressure from the environment. Most codons are degenerated into smaller number of amino acids, in which the 3rd letter of codon is variable. Therefore, the GC content of a gene corresponding a protein can vary, even when the amino acid sequence is kept the same. In order to elucidate whether the 3rd letters of codons in PCP28 occur randomly, we carried out the simulation of random occurrence of the 3rd letters with the amino acid sequence maintained. The result showed that the distribution of the GC content for PCP28 is far from the random occurrence of nucleotides at the 3rd position of codons. The meaning of the AT-rich character of nucleotide sequences for PCP28 will be discussed.

References

1) T. Ikemura et al., "Giant G+C% Mosaic structures of the Human genome found by arrangement of GenBank human DNA sequences according to genetic positions", *Genomics*, 8, 207-216(1990)
2) B. Alberts et al., *molecular biology of the cell*, 4th ed., chapter 4, Garland Science.

PBPK-modeling as a tool for interpreting and understanding of pharmacokinetics

Hiroko Akiyama, CTC LABORATORY SYSTEMS Corporation [Presenter]
S. Willmann, W. Schmitt, Bayer Technology Services GmbH [Authors]

In-silico methods in ADME are mainly seen as tools for property prediction. However, mechanistic computer based simulation techniques as physiology based pharmacokinetic (PBPK) modeling have a much larger potential and can deal as valuable tools to support the interpretation and understanding of experimental data. By means of different exemplary case studies it is shown how initially hidden information important for optimization and development of drug candidates can be revealed from pharmacokinetic data using PBPK-simulations. Together with the gained knowledge the simulations build then the basis for reliable predictions of the ADME behavior in experimentally not investigated scenarios. Thus PBPK-modeling bears a high potential for accelerating optimization and development of new drug molecules.
In particular the following topics are discussed:

Bioavailability

- Distinguishing between solubility and permeability limitation of oral absorption.

Analysis of the plasma curve form allows to distinguish between the to possible causes for reduced fraction absorbed.

- Disclosing of active transport in gut wall.

p-Gp-transport of a compound results in a characteristic change of curve form compared to passively absorbed compound.

Distribution in the body

- Organ specific accumulation behavior

Depending on the physicochemical properties of a compound an accumulation in certain organs may appear that is not apparently reflected by the blood concentration curve.

- Species specifity of distribution kinetics:

Under certain circumstances the distribution kinetics depends on the size of the organism, leading to significantly different plasma levels in different species.

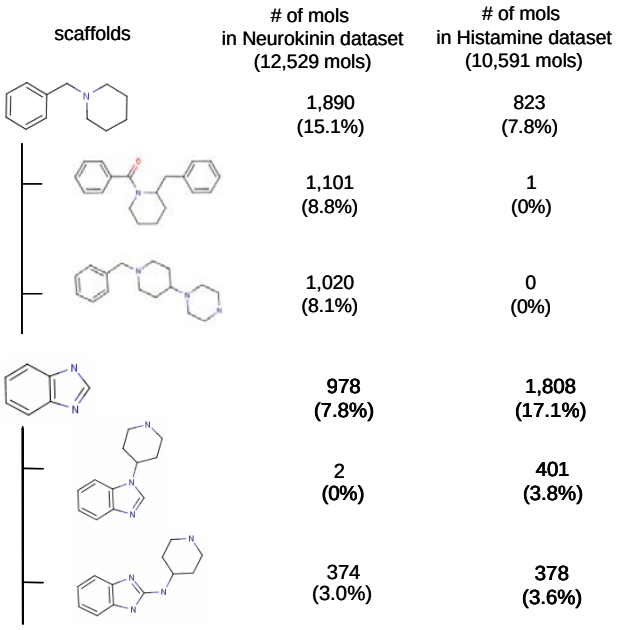
Scaffold analysis of GPCR ligand datasets from literature.

Masami Fujiki[1], Yanhai Fan[2], Naotaka Ohishi[3]
Patcore, Inc.[1], Phas Technology, Ltd.[2], Spotfire Japan, K.K.[3]

Maximum Common Substructure (MCS)-based clustering is a highly intuitive method to categorize chemical datasets by common chemical motifs that exist within a chemical dataset. Because molecules are subdivided and grouped by chemical substructure or scaffolds, small changes in molecular structure can be easily identified, thereby facilitating interpretation of large datasets in a chemically intuitive manner. However, MCS-based clustering methods are computationally more demanding than classical clustering techniques that exploit chemical fingerprints or molecular keys. In this study the benefits of conducting MCS-based clustering analysis versus classical clustering methods is addressed.

For this evaluation, 1) the ability to correlate activity to chemical motifs between MCS clustering and other classical clustering methods (Tanimoto and Ward's based clustering) are evaluated, and 2) scaffolds (substructures) are compared by different target from datasets collected from patents and journal articles.

The results show that MCS clustering has comparable or better detection power as methods that use fingerprint based clustering, and it has the potential to find unique substructures larger than the known privileged structures in the set.



Development of New *in silico* Method for Predicting Drug Metabolism

Michihisa Koyama[1], Kota Kasahara[1], Hideyuki Tsuboi[1], Akira Endou[1],
Hiromitsu Takaba[1], Momoji Kubo[1,2], Carlos A. Del Carpio[1], Kazumi Nishijima[3,1],
Tetsuya Terasaki[1], Akira Miyamoto[1]
Tohoku University[1], Japan Science and Technology Agency[2],
Mochida Pharmaceutical Co.[3]

Predicting drug metabolism *in silico* is of siginificant importance and a great challenge. Conventionally, "local" or "global" method was applied to predict drug metabolism *in silico*[1]. "Local" method is typically based on molecular simulation while "global" method utilizes database or knowledgebase. The former method is not capable of predicting macroscopic metabolic characteristics of living body; on the other hand the latter often faces the predictability issue for newly developed drug molecules. For these issues, we proposed a new concept for the prediction of metabolism[2]. In this study, we developed a new *in silico* method that bridges "local" and "global" methods.

In the developed method, first binding of drug molecule was assessed by using Monte Carlo method as schematically shown in Fig. 1. Based on the binding energy, macroscopic metabolic characteristics are predicted.

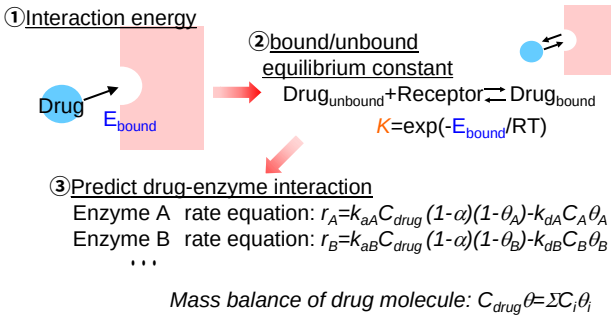


Fig. 1 Scheme of new *in silico* method

We applied the developed method for the metabolism of several drug molecules by CYP3A4 and CYP2C9. Fig. 2 shows results for Phenytoin, which is metabolized by both enzymes. Details are discussed at the conference.

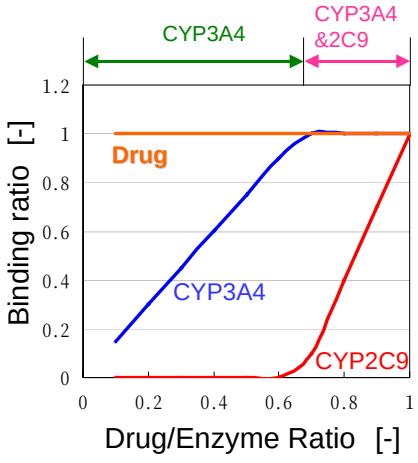


Fig. 2 Calculated results for Phenytoin

References

1) B. Testa et al., "Predicting Drug Metabolism-An Evaluation of the Expert System *METEOR*", *Chemical and Biodiversity*, 2, 872-885(2005).
2) M. Koyama et al., "Computational Chemistry Platform for Drug Metabolomics", *CBI2005 Proceedings*, p.46(2005).

"CALTA", a novel structure-property-activity profiling method: Compound-scaffold profiling and *in silico* prediction for the drug-discovery compounds that target membrane proteins

Yoshimasa Kobayashi, Katsuya Akimoto, Hiroshi Kikuchi, and Kenichi Sudo.
Drug Metabolism & Physicochemistry Research Lab., Daiichi Pharmaceutical. Co., Ltd.

Purpose: When drug-targeted proteins are located near biomembranes, drugs with high log D are concentrated in the membrane and a high apparent affinity (Ka_{app}) can be observed. From the viewpoint of the ADME, however, nonessential high log D must be avoided for drug-discovery compounds. In this presentation, a novel structure-property-activity profiling method, here termed "Correlation Analysis for Lipophilicity and Target Affinity (CALTA)", has been proposed. This method allows researchers to "intuitively" find out the intrinsic SAR using only the dataset of Ka_{app} and log D, thus avoiding undesirable ADME properties caused by nonessential high lipophilicity of the compounds.

Data & Theory: (1) Certain activity values (e.g. $IC_{50,app}$) and the octanol/water (pH 7.4) partition coefficients ($D_{7.4}$) were measured for the in-house compounds of several research projects. (2) The compounds for a project were classified into clusters based on their substructures. Their $IC_{50,int}$ values were estimated from CALTA based on the following theory: The formula in Fig.1 is derived from the assumption that nonspecific interaction is positively correlated with drug lipophilicity. When a series of active compounds is plotted on a scatter diagram using their D and $IC_{50,app}$ values as shown in Fig.1, producing a "CALTA-map", plots of compounds with the same $IC_{50,int}$ are located around an identical line along the formula. The y-intercept corresponds to their $IC_{50,int}$.

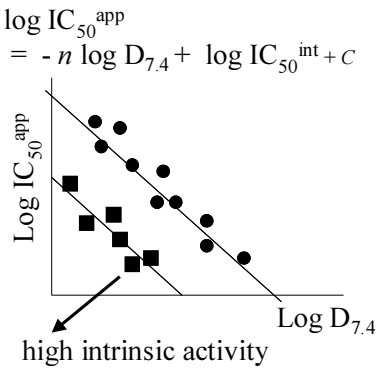


Fig. 1. CALTA-map

Results: (1) Each cluster of compounds was located around the corresponding down-sloping line in their CALTA-maps (Fig.2). The compounds with the lowest $IC_{50,int}$, around the left-lowest line, were found to be the most effective in the *in vivo* studies. (2) The pharmacophores generated from the $IC_{50,int}$ *in silico* predicted the activities of other new compounds more accurately than those from the $IC_{50,app}$.

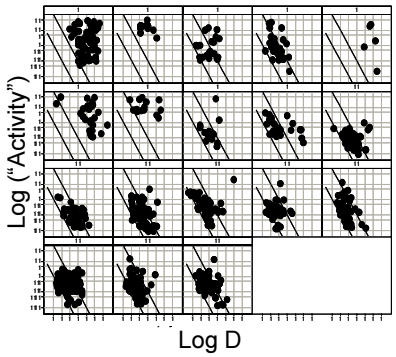


Fig.2. CALTA-maps for some compound-groups of a project classified by their substructures

Virtual Screening Models for Predicting Activities in Various Therapeutic Areas

Zsolt Lepp, Takashi Kinoshita and Hiroshi Chuman
Institute of Health Biosciences, The University of Tokushima.

A very important aim for drug research is to identify the possible side effects of drug candidates. A straightforward way to do it would be to predict those therapeutic areas for which the compounds might be effective, directly from the molecular structures. However there are relatively few software tools that could do this task.

The aim of the work was develop virtual screening models for therapeutic areas, in order to help the process of new lead design in high throughput fashion. The virtual screening was done by the means of support vector machines using a cluster of personal computers. Simple atom type descriptors were used to characterize the molecules.

Altogether 30 general therapeutic areas were included in the study. The datasets were selected from the MDDR database from MDL. To improve the applicability of the method large numbers of molecules were chosen to build the models.

About 50-80% of the positive, and more than 96% of the negative samples of the large external test sets could be correctly identified. The method was suitable to assign therapeutic area-likeness properties to large virtual libraries.

Various statistical analyses were also carried out in order to understand the relationship between therapeutic areas.

References

1. Screening for New Antidepressant Leads of Multiple Activities by Support Vector Machines. Z. Lepp, T. Kinoshita, H. Chuman. *J. Chem. Inf and Model* **2006**; *46*(1); 158-167
2. 33rd QSAR Symposium of Chemical Society of Japan, 2005

Convergence properties of absolute binding free energy calculation using BAR method

Yoshiaki Tanida, Masakatsu Ito, Hideaki Fujitani
Fujitsu Laboratories Ltd., Fujitsu Limited

Recent development of the BAR method[1, 2] provides us the capability of the quantitative estimation of the binding free energy for biomolecular docking problem. In this paper, we examine the influence of the convergence of the interaction energies for obtaining the absolute binding energy. The procedure of estimating the binding free energy of ligand-receptor system is as follows. First, the free energy of the annihilation of ligand from the complex system is calculated, hereafter called $\Delta G_{complex}$. Next, the free energy of the annihilation of ligand from solvated ligand system is calculated, hereafter called ΔG_{ligand}. Finally, we obtain the absolute binding energy as the former value minus the latter one, $\Delta G = \Delta G_{complex} - \Delta G_{ligand}$. PME is used to compute the electrostatic interaction. We also use the switched LJ potential with long range correction for describing the van der Waals interaction. Figure 1 shows the cutoff dependency of the electrostatic energy and the van der Waals one for ΔG_{ligand} and for $\Delta G_{complex}$ in theophylline-RNA system. It is found that ΔG_{ligand} gradually decreases as the cutoff radius is increased as shown in Fig.2. It can be seen that $\Delta G_{complex}$ has also the same tendency. Consequently, we find that the cutoff radius of 0.9 nm is enough to obtain the absolute binding free energy.

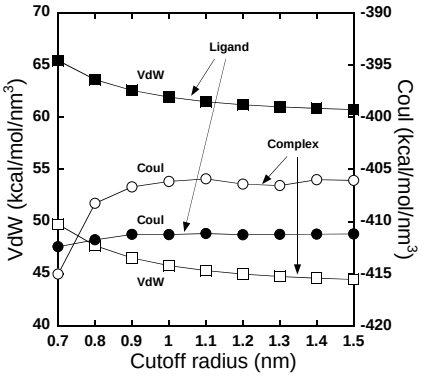


Fig.1 Convergence of interaction energies

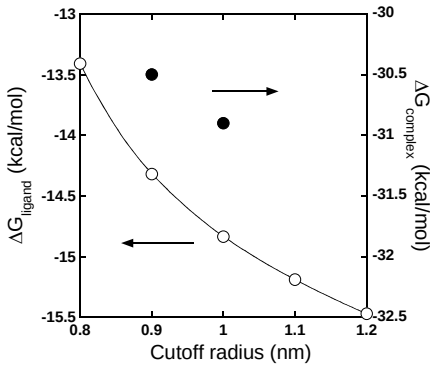


Fig.2 Free energy convergence as a function of the cutoff radius

References

1) M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande, Phys. Rev. Lett. **91**, 140601 (2003).

2) H. Fujitani, Y. Tanida, M. Ito, G. Jayachandran, C. D. Snow, M. R. Shirts, E. J. Sorin, and V. S. Pande, J. Chem. Phys. 123, 084108 (2005).

Building up an original database of launched drugs for their ADME and physicochemical HTS profiling data and human PK data. Rating drugability based on the database and profiling data for in-house compounds.

Atsuko Kameda, Kyosuke Suzuki, Yoshimasa Kobayashi, Katsuya Akimoto,
Hiroshi Kikuchi and Kenichi Sudo
Drug Metabolism & Physicochemistry Research lab., Daiichi Pharmaceutical. Co., Ltd.

Purpose: An original database of launched drugs was built up to identify the profiles of these drugs. The database consists of a wet part; ADME and physicochemical HTS profiling data, and a dry parts; literature-based human pharmacokinetics data. With this database, the profiles of in-house compounds were compared with those of the launched drugs, and the drugability was evaluated.

Methods: The literature-based information for approximately 1400 launched drugs, mainly FDA-approved drugs, is compiled with data from Prous Science's Integrity database, IMSworld drug launches database, GVK's drug database, Technomics's "Asu-no-sin'yaku," and Goodman & Gilman's "The pharmacological basis of therapeutics: Tenth Edition." The commercially available drugs in that database were purchased and their ADME and physicochemical parameters were measured. Those values are linked to the literature-based information above. The key statistical values for the launched drugs and in-house compounds were compared to rate drugablity.

Results: The collected PK data were reviewed from the point of the confirmation of healthy human data and the uniformity of units. The original hybrid database, called KameDB, which consists of both the dry and wet data, was built up. The approximately 170 compounds that had wet data within KameDB and approximately 7000 in-house compounds were statistically analyzed for the group of drugs and non-drugs, respectively. Consequently, the following conclusions were obtained. (1) The distribution coefficient correlates with many ADME parameters. (2) Increasing the molecular weight by 10 and increasing the distribution coefficient by 1 decreases drugability in the ratios of 0.82 and 0.66, respectively.

Conceptual Design of a Framework for ADME Ontology

Takeshi Arikuma[1], Sumi Yoshikawa[1, 2], Kentaro Watanabe[1*], Kazumi Matsumura[1, 2], Akihiko Konagaya[1, 2]

Department of Computer Science, Tokyo Institute of Technology[1],
Genomic Sciences Center, RIKEN[2]

Ontology is a science that studies the formal representation of domain knowledge. An ontological issue in ADME knowledge is the representation of hierarchical biological interactions between drugs and human beings from molecular level to body level. In addition, we should consider the fact that biological interactions are dynamic and depend on context such as health conditions and multiple drug administration. In order to solve the above issues, we propose a framework for ADME Ontology based on the Triadic Relations of continuants [1].

The framework consists of a class hierarchy of continuants and processes, and sets of instances of biological interactions represented by Triadic Relations (Fig.1). In a Triadic Relation, a process is represented by three kinds of participants, *i.e.* 'trigger', 'situator' and 'resultant' [1]. Instances of primitive processes are represented by Triadic Relations accompanied by instances of continuants as the participants. The originality of the framework is in its use of 'instance constructors' and 'anonymous classes'. Instance constructors are inference rules applied to connect Triadic Relations to generate pathways. Anonymous classes are classes that intentionally have no formal definitions and do not exist in the class hierarchy of the Ontology. Generated pathways are considered as instances of anonymous classes to avoid combinatorial explosion of class definitions and instance variations.

We applied this framework to the ADME Ontology of Irinotecan, an anti cancer drug. 162 classes and 65 instances of organs, molecules and molecular level primitive processes are implemented with OWL-DL and prolog.

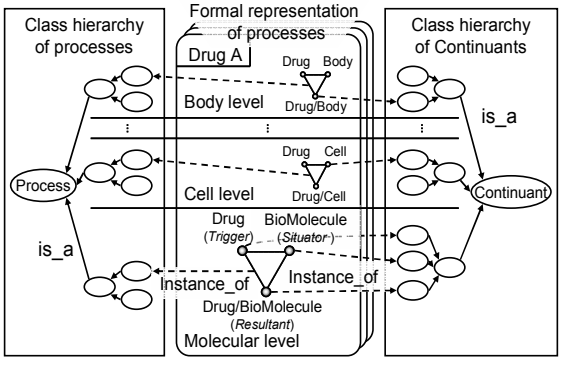


Fig.1 ADME Ontology

References

1) Sumi Yoshikawa, Satou Kenji, Akihiko Konagaya. Medinfo, 2004; 11, pp454-458

* Current affiliation: System Development Laboratory, Hitachi Ltd.

Acquisition of Extended Information Regarding Chemicals-CYPs Interactions by Natural Language Processing

Chunlai Feng [1], Fumiyoshi Yamashita [1], Mitsuru Hashida [1]
Graduate School of Pharmaceutical Sciences, Kyoto University[1]

Cytochrome P450s (CYPs) are a superfamily of enzymes that are responsible for oxidative metabolism of a wide variety of drugs. Drug-drug interaction that arises due to CYP metabolism becomes a critical problem in clinical usage of many drugs. Therefore, it is an important issue to predict the interaction of new chemical entities with CYPs in drug discovery and development. In the CBI 2005 meeting, we reported a natural language processing system for extracting information on chemicals-CYP3A4 interactions from literature. This system implements the rules that identify chemical names and extract interaction in a context-based manner, in addition to a chemical name dictionary. Information extraction is carried by the following steps: identification of chemical and enzyme names in the text, transformation of sentences into multiple simple clauses containing single event, and pattern matching of keyword sequences inside the clauses. It gave a high performance on extraction of chemical names (0.842 recall, 0.939 precision) and chemical-enzyme interaction relationships (0.868 recall, 0.877 precision). In the present study, we applied the system to extract interaction information between chemicals and other metabolic enzymes, including CYP1A2, CYP2C9, CYP2C19, CYP2D6 and CYP2E1. Using PubMed database, 1600 records on interaction with 570 chemicals for CYP1A2, 900 records (300 chemicals) for CYP2C9, 850 records (260 chemicals) for CYP2C19, 1670 records (620 chemicals) for CYP2E1, and 1790 records (450 chemicals) for CYP2D6 were obtained. A corpus of randomly selected one hundred CYP2D6 abstracts was used for validation of the text mining system used. It achieved 83.6% recall and 89.1% precision in information extraction of chemicals-CYP2D6 interactions. The recall and precision were comparable to those with CYP3A4.

References

1) Guengerich FP, “Role of cytochrome P450 enzymes in drug-drug interactions”, *Advances in pharmacology*, 43:7-35(1997)
2) Ono T. et al., “Automated extraction of information on protein-protein interactions from the biological literature”, *Bioinformatics*, 17(2), 155-61(2001)

Ontology Development for RTK-Associated Information Retrieval System

Sumi Yoshikawa, Kazumi Matsumura, Akihiko Konagaya
Genomic Sciences Center (GSC), RIKEN

RTK families have a relatively long history of investigation and are known to be involved in various basic biological processes such as cellular proliferation, differentiation and tumor progression. RTK-related literature information is continuously increasing in number and an efficient information retrieval system is demanded. Although literature search through PubMed is convenient and efficient when the MeSH terminology system is relevant to users' searching scope and purposes, this general purpose biomedical terminology system is not always sufficient when users intend to focus on a specific biological domain in depth. In case of RTK families, for example, search results are often so large in number that users tend to spend too much time screening out really needed ones (specificity problem). On the other hand, they may be overlooked due to the fact that relevant gene names and synonyms are not fully incorporated (sensitivity problem).

In order to solve these problems, we developed a prototype of RTK-specific literature retrieval system, which is a kind of filtering tool for PubMed search. Our approach is focused on the following two methods:

1) Collecting and structuring RTK-specific terminologies, including names of biological substances, roles and events. These RTK-terminology systems are incorporated as an ontology. The terms are scanned on the result of user's free term search on PubMed, followed by showing the distribution pattern in a hierarchy tree, as well as showing markings on the abstract.

2) Preparation of a gene-MeSH terms matrix which provides numbers of relevant papers in each cell and a direct (one-click) access to the abstracts through PubMed.

To collect RTK-specific terms, we examined terms in Gene Ontologies and other existing terminology systems, and currently, we have collected 1, 063 concept entries (no duplicate of synonyms), comprising 804 gene entries and 259 process oriented terms. Among these 804, 143 were extracted from Gene Ontology annotation, and others were manually collected with the aid of a commercially available tool for extracting co-occurrence of gene names in the literature. Future work will be done to further develop the terminology system, accompanied by the development of an efficiency evaluation method, as well as constructing links to another RTK-information system under development in GSC, RIKEN.

Physiologically Based Pharmacokinetic (PBPK) model for pravastatin - Analysis of transporter-related pharmacokinetics -

Takao Watanabe, Hiroyuki Kusuhara, Kazuya Maeda, Yuichi Sugiyama
The University of Tokyo

Cumulative studies have elucidated importance of drug transporters in the pharmacokinetics of drugs (hepatobiliary and renal transport, intestinal absorption, and tissue barriers). Recently, many reports on genetic polymorphisms in drug transporters and transporter-related drug-drug interactions have been published, and these are part of mechanisms of interindividual difference in drug response. It is important to predict such interindividual difference at early stage of drug development. The purpose of the present study is to establish a PBPK model that includes the transporter-mediated saturable membrane transport processes and to investigate the effect of changes in transporter function on the pharmacokinetics and, ultimately, the pharmacological and/or toxicological effects.

In this study, pravastatin, the HMG-CoA reductase inhibitor, was used as a model compound, because cumulative studies have demonstrated the importance of transporters in the pharmacokinetics (hepatobiliary and renal transport) of pravastatin. We constructed a PBPK model which consists of compartments such as the liver, kidney, muscle and brain and includes physiological and biochemical parameters, and simulated the plasma and tissue concentrations in rats at various doses. The parameters of membrane transport were taken from previous studies.

The simulated data were comparable to the observed data that exhibited non-linear pharmacokinetics because of the saturation of transporter-mediated membrane transport and/or metabolism. We also performed sensitivity analysis to understand the effect of the changes in hepatic transporter function on the plasma and liver concentration. The changes of transporter activity affected the plasma and liver concentration in a different manner. This PBPK model can be thus applied to predict changes of drug concentration in plasma and target organs caused by changes in transporter function and expression level.

References

1) M. Yamazaki et al., “Na^+-independent multispecific anion transporter mediates active transport of pravastatin into rat liver”, *American Journal of Physiology*, 264, G36-G44 (1993).
2) M. Yamazaki et al., “Biliary excretion of pravastatin in rats: contribution of the excretion pathway mediated by canalicular multispecific organic anion transporter (cMOAT)”, *Drug Metabolism and Disposition*, 25, 1123-1129 (1997).

A Web-Based Drug Safety Information Community

Kumiko Sakamoto[1], Eiko Nakata Doolin[2], Mikio Sasaki[3], A.Ammar Ghaibeh[3],
Hiroshi Chuman[1], Aiko Yamauchi[1]
Institute of Health Biosciences, The University of Tokushima[1],
Department of Pharmacy, NTT-East Kanto Medical Center[2], Saila Systems Inc. [3]

Progress of the recent science and technology are revolutionary. It is important that such technologies are used to support the safe society and safe citizen's life. And conversely, the information from society and citizens can be fed back to the progress of Science and Technology. Since the thalidomide catastrophe occurred in the beginning of the 1960's, many people have assumed that all medicines have generally some teratogenicity. Thus the drug-induced birth defect seems to be over-assessed and feared excessively by the pregnant woman who had taken medicines without noticing the pregnancy. In addition, the information about drug teratogenicity is beneficial for the pharmaceutical scientists in the development of new drugs. We have pursued a research/development project, "Construction of a Drug Safety Information Community". The objective of this research is to develop a drug safety information community for teratogenic agents as a social system. This information community system allows an effective circulation of drug safety information among health care consumers, health care professionals, and drug discovery researchers through a shared knowledge on the Internet.

The integrated bi-directional database system has three classes of information;

1) Documentary information of the medicinal drugs on teratogenicity from package inserts, scientific papers and the summary basis for approval of new drugs

2) Prediction of teratogenicity for a given compound by the structural similarity search of drugs with teratogenicity.

3) Information of the clinical case reports registered by the general public.

The information provided by each community member will contribute to the practice of evidence-based medicine, an efficient drug research and development, and training for "the informational pharmacist" who will play important roles in risk communication in the area of clinical pharmacy. The database is going to be partially released soon.

Acknowledgement: This research is supported by Research Institute of Science and Technology for Society (RISTEX), Japan Science and Technology Agency.

Instant Computing Technology for Molecular Design Environment for Non-Cluster Experts

Fumikazu KONISHI[1], Toru Yagi[2], Ryo Umetsu[1],
Shingo Ohki[1], Akihiko KONAGAYA[1]
RIKEN GSC[1], BESTSYSTEMS, Inc [2]

Abstract

We have designed and implemented a new portable system (Knoppix for Molworks) that can rapidly construct a computer environment where high-throughput research applications can be performed instantly. The Knoppix is a collection of GNU/Linux and features one CD live file system (iso9660) that can be customized as a full-featured portable computer system. This paper presents an approach for instant computing using Knoppix technology that can allow even a non-computer specialist to easily construct and operate a Beowulf cluster.

In order to show an actual example using a reference application, we have chosen MolWorks while is an integrated software tool for molecular design. The MolWorks supports a molecular builder and pre/post interface to quantum mechanics calculation software, such as Gaussian, GAMESS, Q-Chem and MOPAC. MolWorks also estimates properties of the molecules.

Build the Structure of Molecules

It is very easy to build up your own molecules in MolWorks. You can draw the molecular structure within "Molecule Window". MolWorks also provides "Optimizer" for the purpose to relax the structure. It is so easy to change element type and bond order with the selection of each atoms and bonds. Currently, MolWorks has the capability to display the molecular structure with wire, and ball & stick.

Quantum Mechanics Calculation Assistant

BSI migrate the interface for well-known quantum mechanics programs, like Gaussian, GAMESS, Q-Chem and MOPAC. You can select the keywords and options for calculation from the menu. After you save the input file for calculation, it is very easy to submit the calculation on your own PC. MolWorks also provide CNDO/2 calculation engine, so you can calculate charge distribution and display molecular orbital with the graph of energy level.

Instant Build High Throughput Computing Environment

Knoppix for MolWorks can be booting up from CD-ROM image and be automatically setup up for a high-performance cluster computing system. Therefore, all of the necessary troublesome installation work isn't necessary by researcher.

Conformational Analysis and Docking Study of the c-Jun N-Terminal Kinase Inhibitors Having a Benzothiazol Moiety

Mohamed Ismael[1], Abdul Rajjak Shaikh[1], Hideyuki Tsuboi[1], Michihisa Koyama[1], Akira Endou[1], Hiromitsu Takaba[1], Momoji Kubo[1,2], Carlos Del Carpio[1], Kazumi Nishijima[3], Tetsuya Terasaki[1], Akira Miyamoto[1]
Tohoku University[1], JST-PRESTO[2], Mochida Pharmaceutical Co.[3]

Stress activated protein kinases or c-Jun N-terminal kinases (JNKs) belong to the mitogen-activated protein kinase (MAPK) family, which regulate signal transduction in response to environmental stress. Several evidences support the hypothesis that JNKs plays a critical role in a wide range of diseases (e.g. cell death). In this study we applied computational calculation methods to investigate the activities of well characterized JNKs inhibitors using docking and molecular dynamics simulation techniques.

Selected (Benzothiazol-2-yl)acetonitrile inhibitors were docked to the active site of the c-JUN protein[1]. According to these results the ligands showed three main binding modes depending in the inhibitor size. Our classical molecular dynamics simulation (New-Ryudo) program was used to perform a dynamics equilibration for the (protein/ligand) complex models at constant temperature (300K) for 100 ps to provide additional relaxation. All the geometrical properties converged after 50 ps of simulation. Our results show that the ligands bind to the active site pocket and form two types of hydrogen bonds with the carbonyl of Met149 and another with the amine of the Gln155. Fig. 1 shows the conformation of the most active inhibitor (ligand) in the active site pocket.

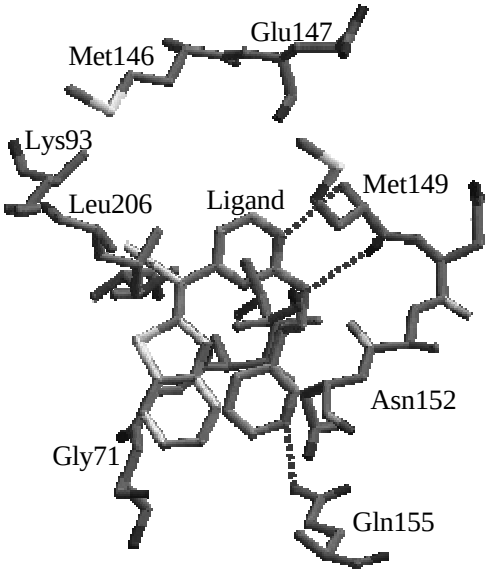


Fig. 1. Overlay of the docked orientation for the most active compound (ligand) in the active site of c-jun kinase.

Reference

1) P. Gaillard et al., "Design and synthesis of the first generation of novel potent, selective, and in vivo active (benzothiazol-2-yl)acetonitrile inhibitors of the c-jun n-terminal kinase" J. Med. Chem., 48 (14), 4596 -4607(2005).

Development of an Empirical Model for the Activation Energy of CYP Reaction

Doo Nam Kim[1], Won Seok Oh[2], Kyoung Tai No[1,2]

Bioinformatics & Molecular Design Research Center[1], Yonsei University[2]

Human cytochrome P450 (CYP450) is the most abundant enzyme in the human liver and is responsible for the metabolism of nearly over 90% of the drugs on the market. Accordingly, it is necessary to determine regioselectivity which shows what site is metabolized preferentially for the purpose of reducing time and resources of drug discovery and development. However, the lack of quantitative experimental metabolic data and the fact that CYP system depends on both steric effect and electronic reactivity indicate that a determining regioselectivity is difficult.

Here, we present newly developed empirical model for prediction of activation energy of CYP reactions for the purpose of estimating electronic reactivity. We used AM1 quantum mechanics calculation to collect values of activation energies of various molecules with para-nitrosophenoxy radical as a surrogate oxygenating species of Compound I which is responsible for CYP catalytic property, Modified Partial Electron Orbital Equalization (MPEOE) program to see atomic charge, polarizibility, electron negativity of each atom, and derived empirical model for activation energies of each atoms of a given molecule by statistical method. This model explains aliphatic hydroxylation, aromatic hydroxylation, and dealkylation which altogether constitute most reactions mediated by CYP and also is valuable regarding that further research will enable to know activation energy of not only metabolism reactions but also of general molecules.

Prediction of Skin Permeability for Chemical Penetration Enhancers using Quantitative Structure-Activity Relationship Analysis and Classification Methodology

Hyejin Kang[1], KyoungTai No[1,2]
Yonsei Univ[1], BMDRC [2]

Skin penetration enhancers are chemicals used to help drug sufficiently permeable to skin. Making mathematical model for skin penetration enhancers is required to predict new compounds. Recursive partitioning (RP) and Quantitative-Structure Activity Relationship are useful tools to classify enhancers into fluidizers and extractors which have been considered main types of mechanism of stratum corneum perturbation, and to correlate structure of each categorical enhancers with permeability. In this study, almost all kinds of enhancer containing anionic, cationic, zwitterionic and nonionic surfactant, fatty acid, ester and amine, terpene and terpenoid, pyrrolidone derivatives and azones have been studied. The resulting RP models show classification rate is sufficiently high, which is about 93.7% . Regression results show that fluidizers with more hydrophobic part and carbonyl group tend to have high potency. In case of extractors we further classified them into two groups based on the sphericity. Molecules with round shape shows correlation with 3 topological distance. On the other hand, permeability of long shape extractors have linear relationship with molecular weight and long distance topological distances. This classification based on the shape is proved appropriate by unsupervised learning, SOM (self organizing map), which shows similar results that extractors can be grouped into further classes based on the shape. Accordingly, this study supports more accurate means of predicting the permeability potential of classified enhancers.

References

1) P. Karande et al., “Design principles of chemical penetration enhancers for transdermal drug delivery ”, *PNAS*, 102, 4688-4693(2005).

In Silico Prediction of CYP1A2-mediated Metabolism

Jihoon Jung[1], Suyeon Kim[1], Kyoungtai No[1]
Yonsei University, Seoul, Korea[1]

Human Cytochrome P450 (CYP) enzymes are membrane-associated heme proteins involved in phase I metabolism of many drugs. Cytochrome P450 1A2 is a member of the CYP1 family that is responsible for the metabolism of a series of planar conjugated compounds, such as caffeine. To develop more efficient and effective drugs, it will be important to estimate quantitatively drug metabolism. Therefore, a model for the prediction of the regioselectivity of substrates for CYP1A2 enzyme was constructed through various computational methods. It mainly consists of homology modeling of human CYP1A2 enzyme and docking trials of selected chemicals into the binding site of protein. In this study, we will discuss the details of enzyme-substrate binding complex and their binding affinity.

References

1) D.F.V. Lewis et al., “Homology modelling of human CYP1A2 based on the CYP2C5 crystallographic template structure”, *Xenobiotica*, 33, 239-254(2003).
2) J.J. Lozano et al., “Three-dimensional modelling of human cytochrome P450 1A2 and its interaction with caffeine and MeIQ”, *Journal of Computer-Aided Molecular Design*, 11, 395–408(1997).

Classification of P-Glycoprotein Substrates and Inhibitors

Sokyoung Joo[1], KyoungTai No[1, 2]
Yonsei University[1], Research Institute of Bioinformatics and Molecular Design [2]

P-glycoprotein (P-gp), which belongs to a superfamily of ATP-binding cassette transporters, has been discovered in various resistant tumor cells as well as many normal tissues. The effect of P-gp-mediated drug efflux limits intestinal absorption and oral bioavailability of drugs. There is a great interest in anticipating whether drug candidates are P-gp substrates or inhibitors. Extensive research has been conducted to uncover the molecular features required for the substrates and inhibitors of P-gp. However, accurate classification is difficult due to broad substrate specificity, multiple binding sites and different modulator mechanisms.

Herein, a data set of diverse compounds taken from ADME database and experimental literatures were explored. Various different computational modeling methods have been set up using a set of predefined descriptors encoding the key molecular properties capable of discerning a substrate from an inhibitor. The performance was compared using various different descriptor sets based on a wide range of different properties and physicochemical descriptors.

References

1) Wang et al., “Classification of Substrates and Inhibitors of P-Glycoprotein Using Unsupervised Machine Learning Approach”, *J. Chem. Inf. Model,* 45, 750-757(2005).

NMR metabolomics 1/2
Multivariate analysis for diagnostic study of inherited metabolic disorders using human urine

Tadashi Nemoto[1], Masako Fujiwara[2], Itiro Ando[1,3] Kazunori Arifuku[2], and Tomiko Kuhara[4]

National Institute of Advanced Industrial Science and Technology [1],
JEOL DATUM LTD [2], Environmental Research Center Ltd.[3],
Kanazawa Medical University[4]

Metabolic disorders in new born babies are serious when it is overlooked, however adequate treatment cure at each early stage can mask the disease onset and ensures normal life for the patient. In Japan, only 6 kinds have been selected and does mass screening test with blood. On the other hand, gas chromatographic-mass spectrometry (GC/MS) technique using urine has been established for clinical diagnosis and revealed over 130 kinds of metabolic disorders[1)]. Using Urine is, non-invasive, merit having potential to detect wide range of such disorders including multiple complexities.

We have proposed NMR metabolic profiling (NMR-MP) method using ^{1}H NMR spectra pattern recognition analysis. Seventeen urine samples including several disorders and controls already diagnosed by GC-MS, were prepared by our method previously reported[2)]. ECA-500 (500MHz, JEOL) was used for 1D-^{1}H-NMR measurements and FIDs were directly processed by Alice2 for Matabolome software (JEOL). SIMCA (Soft Independent modeling of Class Analogy) method, which is an advanced multivariate analysis based on Principal component analysis (PCA) resulted in clear discrimination between disease and controls and in addition, detection of disease status, such as onset or remission, or so on. Our method of NMR-MP using urine would have promising scope for clinical diagnosis.

References:

1) T. Kuhara. *Mass Spec. Rev*. **2005**, *24*, 814

2) M. Fujiwara, K. Arifuku, I. Ando, and T. Nemoto, *Anal. Sci*. **2005**, *21*, 1259.

NMR metabolomics 2/2
Multivariate analysis for diagnostic study of streptozotocin-induced diabetes using rat urine

Masako Fujiwara[1], Itiro Ando[2, 3], Kazunori Arifuku[1], Taeko Kataoka[2], Kenji Kanazawa[2], Yasuhiro Natori[4] and Tadashi Nemoto[2]
JEOL DATUM LTD.[1], National Institute of Advanced Industrial Science and Technology [2], Environmental Research Center Ltd [3], Research Institute International Medical Center of Japan[4]

Assessment and evaluation of physical condition by non-invasive way is highly required to detect complications in a diabetes patient. We tried very early detection of nephritis in diabetes model rats by using NMR metabolic profiling method (NMR-MP).

Twenty male Wistar rats (11 week old) were used for the study and their left kidneys were surgically removed and then critical amount of streptozotocin was injected as citrate buffer solution and fed separately in metabolic cages under controlled condition.

Urine samples were collected every week and kept in a freezer. For NMR-MP measurements, urine samples were prepared by our method previously reported[1)]. ECA-500 (500MHz, JEOL) with automated sample changer was used for 1D-^{1}H-NMR measurements and FIDs were directly processed by Alice2 for Matabolome software (JEOL). After first stage analysis by principal component analysis (PCA), we found four outliers, those were found to be insufficient rats to be a diabetes model. In second stage analysis, we removed outliers and combine additional diagnostic and biochemical information and re-run multivariate analysis with the method named SIMCA to find a potential to very early detection of albuminuria before its starts.

References:

1) M. Fujiwara, K. Arifuku, I. Ando, and T. Nemoto, *Anal. Sci*. **2005**, *21*, 1259.

Development of a database with simulation function for quantitative prediction of drug-drug interactions

Yoshihisa Shitara[1], Toshiharu Horie[1], Toshihiko Ikeda[2] , Masato Kitajima[3], Jose M. Ciloy[3], Yuichi Sugiyama[4]
Chiba University[1], Sankyo Co. Ltd.[2], Fujitsu Kyushu System Engineering Ltd.[3], The University of Tokyo[4]

Abstract

Most drugs are cleared from the body by P450-mediated metabolism and transporter-mediated hepatobiliary elimination. In clinical situations, a number of drugs are administered concomitantly. A coadministered drug (inhibitor) often inhibits the metabolism and/or hepatobiliary transport of another drug (substrate), resulting in a drug-drug interaction. The ratio of the unbound inhibitor concentration (I_u) to the inhibition constants (K_i) is used as an index for predicting drug-drug interactions. However, this often makes a false positive prediction. Thus, in the present study, we predicted pharmacokinetic alterations by simulation analysis to aim a quantitatively accurate prediction. For the accurate prediction, we estimated K_i values based on *in vivo* data (*in vivo* K_i). To estimate it, the plasma concentration-time profiles of substrates with coadministration of inhibitors were fitted to physiologically-based pharmacokinetic (PBPK) model. These *in vivo* K_i values were compared with the reported *in vitro* K_i values. The accuracy for predicting drug-drug interactions with the use of *in vivo* K_i values was compared with that obtained by *in vitro* $1 + I_u/K_i$. In the present study, we obtained *in vivo* K_i values of 11 inhibitors. The prediction from the PBPK model using *in vivo* K_i values was more accurate than the prediction obtained by $1 + I_u/K_i$. But there were false negative predictions of cytochrome P450 3A4 substrates possibly due to inhibition of intestinal metabolism. By taking the inhibition of intestinal metabolism into consideration, more accurate predictions were obtained. We are now trying to make accurate predictions of drug-drug interactions caused by transporter inhibition and mechanism-based inhibition of metabolic enzymes.

References

1) K. Ito et al., *Pharmacol. Rev.* **50**, 387-411, 1998
2) M. Kato et al., *Drug Metab. Pharmacokinet.* **18**, 121-127, 2003
3) Y. Shitara et al., *Annu. Rev. Pharmacol. Toxicol.* **45**, 689-723, 2005
4) Y. Shitara and Y. Sugiyama, *Pharmacol. Ther.* in press

In vitro screening system using a combination of suspensions and monolayers of Caco-2 cells for detecting drug-drug interactions in intestinal absorption and conjugative metabolism of drugs

Marie Tanaka, Takashi Mizuma, Masahiro Hayashi
School of Pharmacy, Tokyo University of Pharmacy and Life Science

We have previously reported on the kinetic impact of intestinal conjugative metabolism on oral bioavailability in humans using reported data [1]. For example, sequential analysis showed that the intestinal availability of salbutamol, which is metabolized by sulfate conjugation, was 0.700, whereas hepatic availability and fraction of drug absorbed were 0.893 and 0.802, respectively. The objective of this study was to detect drug-drug interactions in conjugative metabolism and the transport of salbutamol using Caco-2 cells.

Caco-2 cells were seeded onto Transwell inserts, and used for transport (absorption) and metabolism experiments. Metabolism experiments were done in Caco-2 cell suspensions as well. Salbutamol and salbutamol sulfate conjugate were determined by HPLC.

Salbutamol was metabolized to sulfate conjugate in Caco-2 cell suspensions. The sulfation metabolism of salbutamol was inhibited by the presence of fenoterol, isoproterenol, or acetaminophen. After salbutamol was added onto the apical side of Caco-2 cell monolayers, salbutamol appeared on the basal side and its amount increased over time. Salbutamol sulfate conjugate also appeared on the basal side and its amount increased over time. The sulfation metabolism rate of salbutamol across Caco-2 cell monolayers was decreased by the presence of fenoterol or isoproterenol. In contrast, the appearance rate of salbutamol on the basal side was unchanged by the presence of fenoterol or isoproterenol, suggesting that salbutamol interacts with fenoterol or isoproterenol not only in metabolism but also in membrane transport.

In conclusion, Caco-2 cell monolayers and suspensions are useful for the study of intestinal conjugative metabolism and transport. Even when assessing the intestinal metabolic impact on absorption, transport studies are indispensable for the prevention of misunderstandings about drug-drug interactions in intestinal absorption.

Reference

1. T. Mizuma *et al*., *J. Pharm. Sci*., 94, 571-575 (2005).

Transport and metabolism of ampicillin prodrugs and their interactions with P-glycoprotein: *in vitro* study on successful prodrugs using Caco-2 cell monolayers

Sachie Tanaami, Sayaka Sakaguchi, Takashi Mizuma, Masahiro Hayashi
School of Pharmacy, Tokyo University of Pharmacy and Life Science

Prodrugs have been developed to overcome the faults of active drugs. One typical example is to improve intestinal absorption by increasing lipophilicity of active drugs. Lenampicillin, bacampicillin, and talampicillin are prodrugs of antibiotics such as ampicillin. Nevertheless, the absorption process of these prodrugs has not been characterized in detail. In addition, it has been reported that a prodrug failed to improve drug absorption [1]. These tell us that it is important to verify the absorption process of prodrug. Therefore, we have previously characterized transport and metabolism of lenampicillin using Caco-2 cells [2]. In the present paper we report on the transport and metabolism of bacampicillin and talampicillin.

Bacampicillin and ampicillin appeared on the basal side after apical dosing of bacampicillin (AtoB transport). The amount of ampicillin on the basal side was much larger than that of bacampicillin. After basal dosing of bacampicillin, it and ampicillin also appeared on the apical side (BtoA transport). The amount of ampicillin on the apical side was much larger than that of bacampicillin as well. The sum of the transport clearances of ampicillin and bacampicillin in a direction of the AtoB transport was similar to that of the BtoA transport, and these clearances were much higher than the transport clearance of ampicillin, suggesting a high permeability of bacampicillin by passive transport. Similar results were observed in the transport of talampicillin. The transport of rhodamine 123, a substrate for efflux transporter of P-glycoprotein, from the basal to the apical sides was significantly inhibited by the presence of bacampicillin or talampicillin. The inhibitory effect of lenampicillin on the rhodamine 123 transport was observed in our previous study as well [2]. In contrast, ampicillin did not inhibit the rhodamine 123 transport. Therefore, the reason for the success of these ampicillin prodrugs is that the high membrane permeability of these prodrugs and no involvement of P-glycoprotein in ampicillin transport resulted in improvement of intestinal absorption, although these ampicillin prodrugs interacted with P-glycoprotein.

References. 1) N. Okudaira et al., J. Pharmacol. Exp. Therp., 295, 717 (2000). 2) T. Mizuma et al., Proceedings of the 2005 annual meeting of CBI society, Yokohama, 24-26 Aug., pp. 60 (2005).

Mathematical modeling of gene expression associated with Wnt signaling pathway in colorectal carcinoma

Yoshimi Naruo, Kaoru Mogushi, Hiroshi Tanaka
Department of Bioinformatics, Tokyo Medical and Dental University

Colorectal carcinoma (CRC) is one of the most common causes of cancer death in Europe and the United States. In Japan, according to dietary habit change to Western style food, CRC is thought to increase in the future. In cancer cells of CRC, β-catenin is known to accumulate in nucleus and cytoplasm because of aberration of Wnt signaling pathway. The well-known target genes of β-catenin are cyclin D1 and c-myc; the former is a cell cycle-related gene, the latter is an oncogene which controlls cell proliferation and apoptosis. Expression levels of these genes are upregulated by accumulation of β-catenin. Thus, β-catenin is one of the most important molecules of Wnt signaling.

Futhermore, recent researches in CRC show that CpG islands within the AXIN2 (axis inhibitor 2) promoter are methylated in CRCs with microsatellite instability (MSI) and that CPG islands within the APC (adenomatous polyposis coli gene) promoter are methylated in CRCs [1)]. Such epigenetic mutations decrease the mRNA expression and cause a downregulation of the protein expression.

In this research, we developed mathematical model about the relation of selected genes in Wnt signaling pathway. Using genes such as APC, AXIN2, and CTNNB1 (encoding β-catenin) associated with Wnt signaling pathway, we expressed the relation of these mRNA expression with differential equations based on reaction kinetics. Moreover, we evaluated their dynamics using mRNA expression obtained by microarray experiments. We used public microarray dataset of CRC specimen materials containing 10 MSI+ samples and 10 MSI- samples. The dataset is published on Gene Expression Omnibus with accession number GSE2138 [1,2)]. We set default values at constant offset from equilibrium point and ran a simulation by MATLAB software. As a result, MSI+ samples exhibited more β-catenin remained without ubiquitination, compared to MSI- samples. This is consistent with the report that β-catenin accumulated in nucleus and cytoplasm in cancer cells in MSI+ CRC [3)].

References:

1) K Koinuma et al., “Epigenetic silencing of AXIN2 in colorectal carcinoma with microsatellite instability”, *Oncogene*, 25(1), 139-46(2005).
2) Gene Expression Omnibus website (URL:http://www.nubi.nlm.nih.gov/geo)
3) Y Shimizu et al, “Frequent alterations in the Wnt signaling pathway in colorectal cancer with microsatellite instability”, *Genes, Chromosomes & Cancer* 33:73-81(2002).

Modeling of extracellular matrix degradation processes in cancer metastasis using covariance structure analysis

Akiko Hatano[1], Kaoru Mogushi[1], Hiroshi Tanaka[1]
Department of Bioinformatics, Tokyo Medical and Dental University[1]

Cancer metastasis is comprised of subsequent processes such as release of cancer cells from a primary focus, invasion in blood vessels and adhesion to target tissues. Various molecules are involved in each process including overexpression of extracellular matrix (ECM) degrading proteases and abnormality of cell adhesion molecules. Because cancer metastasis greatly affects prognosis of patients, it would be useful to clarify the mechanism of metastasis using expression information at a molecular level.

Matrix metalloproteinases (MMPs) and a disintegrin and metalloproteinases (ADAMs) are ECM degrading enzymes which relate to cancer metastasis, while tissue inhibitor of metalloproteinases (TIMPs) are their natural inhibitors.[1,2] Excessive degradation of ECM by loosing balance between such molecules is a critical cause for cancer metastasis. Cancer cells obtain metastatic ability through consecutive transformations, but it is difficult to observe temporal processes of such transformation events.

Covariance structure analysis (CSA) is a sophisticated method to estimate unobservable phenomena. CSA is one of the graphical modeling methods and has some properties of traditional multivariate analysis such as causal analysis, regression analysis and pass analysis. CSA can also be used to validate the hypothesis models. It is able to extract hidden characteristics behind observable phenomena, and to measure the degree of causality between such features.

In this study, we focused on the process of ECM degradation and vascular / lymphatic invasion. We analyzed a microarray dataset with CSA using mRNA expression levels of MMPs, ADAMs and TIMPs. We used an oral cancer dataset registered in Gene Expression Omnibus (GEO) with accession NO.GDS1062[3]. The dataset contains 22 samples including 14 samples with lymph node metastasis and 8 samples with no metastasis. We hypothesized a model of the balance between these enzymes and various factors in its destruction under metastasis progression.

References

1)M.D.Sternlicht et al., *Annu. Rev. Cell Dev. Biol,* 17, 463-516 (2001)
2)K.Reiss et al., *Pharmacol. Ther,* (2006)
3)R.K.O'Donnell et al., *Oncogene,* 24,1244-1251 (2005)

Comprehensive analysis of drug-specific mutation patterns in protease gene of HIV-1 subtype B using decision tree

Kaoru Mogushi [1], Fengrong Ren[1], Naoki Hasegawa[1], Hiroshi Tanaka[1]
Department of Bioinformatics, Tokyo Medical and Dental University[1]

The emergence of anti-HIV drug resistance has been the largest obstacle in the treatment of AIDS in recent decades. Characterization of drug-specific mutation patterns of HIV is crucial to the challenge of anti-HIV drug resistance acquisition. In this study, we employed C4.5 decision tree[1], a classification method which represents successive rules to assign each sample to a particular class, for characterizing the patterns of resistant mutations induced by protease inhibitors (PIs) and applied it to a large dataset with more than 10,100 HIV-1 protease genes downloaded from HIV Drug Resistance Database[2].

Prior to the analysis, we chose 3,902 isolates from the dataset with following criteria: (1) a specific drug regimen (including no PI treatment) had been received by a patient over 15 weeks, (2) the patient had never experienced any terminations of PIs other than those in the last regimen, (3) a complete PI drug history was known for the patient, and (4) protease mutation had no ambiguity.

Using these isolates, we constructed a decision tree to characterize specific mutation patterns in amino acid sequence of protease caused by PIs treatment. The part of amprenavir (APV) and atazanavir (ATV) treated isolates clearly separated by mutations of I50V and I50L, respectively. Furthermore, D30N mutation was typically found for nelfinavir (NFV) treatment. Occurrence of these mutations for APV, ATV and NFV was consistent with previous reports[3, 4]. We also found that mutation of V82A in conjunction with I84V was particular for isolates treated by ritonavir (RTV) and saquinavir (SQV) in combination. Such drug-specific mutation patterns extracted by decision tree can be helpful to understand the mechanisms of the drug resistance acquisition of HIV-1 under PIs treatment.

References:

1. J. R. Quinlan, C4.5: Programs for Machine Learning, San Francisco, Morgan Kaufmann Publishers, 1993.
2. HIV Drug Resistance Database. URL: http://hivdb.stanford.edu/
3. R. W. Shafer., *Clin. Microbiol. Rev.* 15:247-277 (2002).
4. R. J. Colonno et al., *Antimicrob. Agents Chemother.*, 47:1324-1333 (2003).

Tutorial and Training Courses on Some Areas of Interest of the CBI Society

Tsuguchika Kaminuma[1], Masumi Yukawa[1], Naomi Komiyama[2],
Yoshitomo Tanaka[1], Takako Takai-Igarashi[1], Hiroshi Tanaka[1]
Tokyo Medical and Dental University[1], Chem-Bio Informatics Society[2]

Quarter century have passed since the Chem-Bio Informatics (CBI) Society had started its activities. In these years the on-line journal and annual meetings were launched in addition to the monthly-based seminars. However now the Society is under strong pressure to change in order to adapt to the rapid transformation of biomedical science in the (post) genomic era. This transformation is embodied by the new axis of research of biomedicine namely Genome-Omics-Pathway/Network. Chemical and biological informatics and computing, the basic knowledge body and GOP/N methodologies of the Society, must grow and expand so that they can be relevant for the transforming biomedicine. Biomedical research in the post genomic era requires new type of informatics and computing specialists who not only understand basic molecular biology for structure and function of genes and proteins but also physiologies and pathologies. We considered that developing tutorial and skill up training courses that can convert present informatics and computing specialists to the new type of interdisciplinary researchers is one of the most urgent activities addressed to the Society.

As a starting effort we chose several cutting edge subjects and have designed courses and materials collaborating particularly with those who are operating special graduate level educational programs sponsored by Japan Science and Technology Agency. These subjects are (1) computer-aided drug design, (2) genome-omics-pathway/network, (3) pathway/network to disease (systems pathology), and (4) nuclear receptors and metabolic syndrome (NR-MS). Each course consists of classroom lectures, reading assignments, web-based skill-up problems, and report writings. The courses rely heavily on the Web technology, open resources on the Web, full text search system, computational modeling and simulation. Prototype system is build on the CBI and Tokyo Medical and Dental University servers and have been used by the assigned course participants.

In biomedical field discrepancy between rapidly advancing basic research and slow to change medical education is ever enlarging. We expect that our lectures and the materials or at least part of them will be put on open use and/or be adapted to regular graduate level courses at other academic institutions and even industry within few years.

Let's Share and Solve our Common Problems

Problem A: Proposal

Prediction of the distribution of plasma concentration profile of drugs in Japanese population : interindividual variation in drug disposition

Hiroyuki Kusuhara
Graduate School of Pharmaceutical Sciences, the University of Tokyo

Drug response is governed by pharmacokinetic factors as well as pharmacodynamic ones. And thus, it is necessary to select candidate compounds with proper pharmacokinetic properties to obtain sufficient pharmacological effect in vivo. Physiologically based pharmacokinetic model allows prediction of the average plasma ant tissue concentration time profiles in animals and even in human from *in vitro* studies. It is well known that the pharmacokinetic profiles of drugs in human exhibit interindividual variation, eventually leading to the interindividual variation in the pharmacological and toxicological action of drugs. Such interindividual variation in drug disposition is ascribed to the difference in the functional activity of enzymes and transporters responsible as well as physiologic parameters, such as blood-flow rate and tissue size. For example, CYP3A4, one of major drug metabolic enzymes involved in the metabolism of a variety of drugs, is well known enzyme the activity of which varies greatly. In addition, many reports on genetic polymorphisms in the promoter and coding region of drug metabolizing enzymes and transporters have been published, and the genetic polymorphisms affecting the enzyme- and transporter function have been found, resulting in a prolongation of plasma elimination half lives of drugs, or increased oral bioavailability. The frequencies of genetic polymorphisms depend on the ethnic group, and varied from rare one to frequent one. Prediction of the distribution of pharmacokinetic profiles of compounds at early stage of drug development will be important to avert the failure due to ineffectiveness and adverse effect at the clinical stage. Question is, when metabolic enzymes and transporters responsible for disposition of selected compounds are identified, how the distribution of plasma concentration in human can be predicted?, and what kind software is available for this prediction?

Problem A: Answer

Prediction of interindividual differences in pharmacokinetics

Motohiro Kato[1], Takashi Ito[2], Toshiko Koue[3], Kouji Chiba[4], Yuichi Sugiyama[5]
Chugai Pharmaceutical Co. Ltd.[1], Daiichi Pharmaceutical Co. Ltd[2]., Otsuka Pharmaceutical Co. Ltd.[3], Pfizer Japan Inc.[4], The University of Tokyo[5]

Abstract

Interindividual difference in the efficacy and toxicity of drugs is generally known. Sensitivity and exposure of a drug for a target tissue in each person varies. Factors including genetic polymorphism of metabolizing enzymes, body weight, sex, age, lifestyle habits such as smoking, and so on produce individual difference in the pharmacokinetics of drugs. We have been constructing a method for prediction of individual difference in pharmacokinetics using a Monte Carlo simulation of such factors.

The mean and standard deviation values of various physiological and biochemical parameters that affect pharmacokinetics were collected from the literature. A tube model for liver was used as the mathematical model. The coefficient of variation (CV) of the amount of CYP3A4 in human microsomes ranged from 33% to 99%. AUC data of CYP3A4 substrates were collected to compare actual CV values of AUC with the simulated results. Simulation results using a CV of 33% reflected the actual values in cases of both intravenous and oral administration. These results suggest that 33% is a suitable CV value for CYP3A4. The simulation showed that low bioavailability drugs exhibited large interindividual difference and reproduced results similar to actual results. We predicted the mean and SD values of clearance for midazolam in Japanese and European Americans. In all cases, the simulated mean and SD values reflected the observed values in human. We predicted the clearance of midazolam in each person from his or her body weight. Ninety-five percent (38/40) were in the 95% confidence interval. The predicted probability agreed with the observed probability. Individual differences in the AUC of CYP3A4 substrates may be predictable. This model would be most helpful in the design of Phase II and III studies. The ultimate goal is the prediction of clinical efficacy from virtual clinical trials, which would reduce the scale of clinical studies, especially in Phases II and III and thus reduce cost and time.

Problem B: Proposal

The status quo and problems of OMICS study

Eiichiro Ichiishi
Team Leader, Biomedical Engineering Research Organization, Tohoku University

In a situation of a doctor in the front hospital, a situation of a wet researcher, a situation of bio-informatician, I think that there are several problems about the status quo and problems of OMICS study in each field. I am going to have discussions about future medical care and the problem by OMICS study that we should get over in concerning the most wisdom "Two heads are better than one".

Even for several diseases, there are considerably real differences in approaching methods and problems in its own way on OMICS study. My approaches in OMICS study about "pollinosis" and "artificial dialysis" attract attention of the news in Nikkei biotechnology and winning each society prizes, a special feature magazine. In this session, I have a presentation about " The status quo and problems of an OMICS study" from my experience and situation as a proposer

As a clinical doctor of internal medicine, I want to aim at "clinical OMICS study". In comparison with standard "OMICS study", "clinical OMICS study" will be a little different in a purpose and a method and an item; may be approached. In other words I want to challenge "OMICS study from an exit". I think that it is to be important to push forward a study by "an inductive method" from the domain that is near to an exit, a clinical result.

I assume that various "OMICS study" is exhibited this time, but want to do various arguments to realize practical, useful "OMICS study" together.

Problem B: Answer 1

Analysis of Genome Scale Amino Acid Sequences with the "Physical Fingerprint" Method

Shigeki Mitaku
Nagoya University, Graduate School of Engineering, Department of Applied Physics

A protein is folded by the balance of several types of physical interactions. Thus, it would be desirable to predict the structure of all proteins with the use of molecular dynamics simulations that take these physical interactions into account. However, structure prediction by this approach is not possible even for a single protein of moderate size because of the problems of the computational time required and the complicated energy landscape. Therefore, a completely different approach is necessary for the analysis of amino acid sequences on the genome scale. We developed a 'physical fingerprint' method in which all amino acid sequences are classified by physicochemical parameters such as hydrophobicity, electric charge, amphiphilicity, and density of proline. It is expected that, when the appropriate parameters are selected for classification, the prediction of unknown sequences will be as accurate as that of known sequences.

The basic idea behind this method can be summarized in two points. (1) Each amino acid has various kinds of physicochemical properties, and we obtain various sequences of the indices of these properties from a single amino acid sequence. The idea here is that a phenomenon occurring on a protein is in principle caused by the combination of the sequences of various physical parameters. (2) As a transmembrane helix is formed by a cluster of hydrophobic residues, various other characteristics of protein structures may be formed by the particular distribution of physical properties. This idea leads to the coarse graining approach to amino acid sequence analysis.

The first system that employed the 'physical fingerprint' method was the membrane protein prediction system, SOSUI, in which the characteristic length of the coarse graining was the thickness of the lipid membrane, and the essential factors for prediction were the indices of amphiphilicity and hydrophobicity. Other high performance systems are SOSUIsignal (a signal peptide prediction system), SOSUIdumbbell (a dumbbell type protein prediction system) and SOSUIbreaker (a system for predicting the breaking points of secondary structures), which are available at the Internet site: http://bp.nuap.nagoya-u.ac.jp/sosui/. Several systems for cellular localization to organella and classification of GPCR are now under construction.

Problem B: Answer 2

Estimating Transcriptome from Genome: a Thermodynamic Approach

Tomokazu Konishi
Akita Prefectural University

Decoding genomic information is routine work for a cell, but is a challenging theme for researchers. Indeed, although genome sequences have been determined for many organisms, and transcriptomes has been measured under various conditions, genomes cannot be interpreted by means of transcriptomes, and transcriptomes are still not predictable from genome sequences. These problems base on our ignorance on the "grammar" of the genome; i.e., how quantitative information is embodied in it.

Introduced is a theory that describes the grammar. This theory is derived from an interpretation of biochemical studies using formulas of statistical thermodynamics. The concentration of each transcript is described according to three energies; each of the energy indicates the interaction between a protein factor and nucleic acid of the cell. The appropriateness of this model is verified through tests on hypotheses that have been deduced from the theory.

Using this theory, we will be able to analyze transcriptome data, find out heritable quantitative information and simulate transcriptome with minimum deflection between analysts. Additionally, minimum amounts of calculations will be required to perform those analyses, since the model is derived from linear equations of thermodynamics. Furthermore, the approach will have total fidelity to a cell; the theory describes the functions of protein factors that transcribe the genome and degrade the transcripts. Clearly, the theory used in this approach is contrast markedly with previous non-linear models, which often have been applied adhocery. With the stable basis on the theory, we can expect objectivity and high reproducibility in the analytical results.

References

1) T. Konishi, "A thermodynamic model of transcriptome formation" *Nucleic Acids Research*, 33, 6587-6592 (2005).

2) T. Konishi, "Three-parameter lognormal distribution ubiquitously found in cDNA microarray data and its application to parametric data treatment", *BMC Bioinformatics*, 5, 5(2004).

Problem C: Proposal

Weak Interaction between Inhibition Peptides and a Soluble Receptor for the Docking Analysis Using Software

Makoto Tsuruoka
School of Bionics & Engineering, Graduate School,
Tokyo University of Technology

A rapid screening method of ligands or receptors in the liquid phase has been examined in our laboratory. Interaction with the peptide ligand (SNP-1) to the receptor (BST-1: bone marrow stromal antigen 1) was measured in liquid phase, and the correlation with data by a solid phase method was obtained. The affinity of the ligand's three derivatives with the receptor was also examined by the single amino acid substitution. The receptor was stable in liquid phase because it was fused with the Fc region of human immunoglobulin G (IgG). SNP-1 is the peptide of 15 amino-acids. BST-1 is one of "GPI anchored membrane receptors", and the amino acid sequence of the active site is understood.

We have obtained the data of very weak affinity (Dissociation constant: 7000 nM or more) between one of the ligand derivatives and the receptor. It would be difficult to evaluate such weak interaction in the experiment using a solid phase method. For instance, it is considered that the ligand derivative which shows such weak affinity will give some suggestions to study reducing the side effect of medicines or controlling their levels of effect. We would like to compare the experimental data with those of the docking analysis using software, because it will be more interesting and helpful, if it becomes possible to explain cause or mechanism of the weak interaction from the viewpoint of molecular structures by the analysis.

[1] Sato A, et al. Inhibitor peptide SNP-1 binds to a soluble form of BST-1/CD157 at a 2:2 stoichiometry. Eur J Biochem. 1999 Sep; 264(2): 439-45. Erratum in: Eur J Biochem 1999, Dec; 266(3): 1219.

[2] Sato A, et al. Novel peptide inhibitor of ecto-ADP-ribosyl cyclase of bone marrow stromal cell antigen-1 (BST-1/CD157). Biochem J. 1999 Feb 1; 337 (Pt 3): 491-6.

[3] Yamamoto-K. S, et al. Site-directed removal of N-glycosylation sites in BST-1/CD157: effects on molecular and functional heterogeneity. Biochem J. 2001 Jul 15; 357 (Pt 2): 385-92.

[4] Yamamoto-K. S, et al. Crystallographic studies on human BST-1/CD157 with ADP-ribosyl cyclase and NAD glycohydrolase activities. J Mol Biol. 2002 Feb 22; 316(3): 711-23. Shimizu M, et al. Weak Interaction between Inhibition Peptides and a Soluble Receptor of Fusion Protein in Liquid Phase., submitted, 2006.

* Katakura 1404-1, Hachiouji, Tokyo, 192-0981 Japan. E-mail: tsuruoka@bs.teu.ac.jp

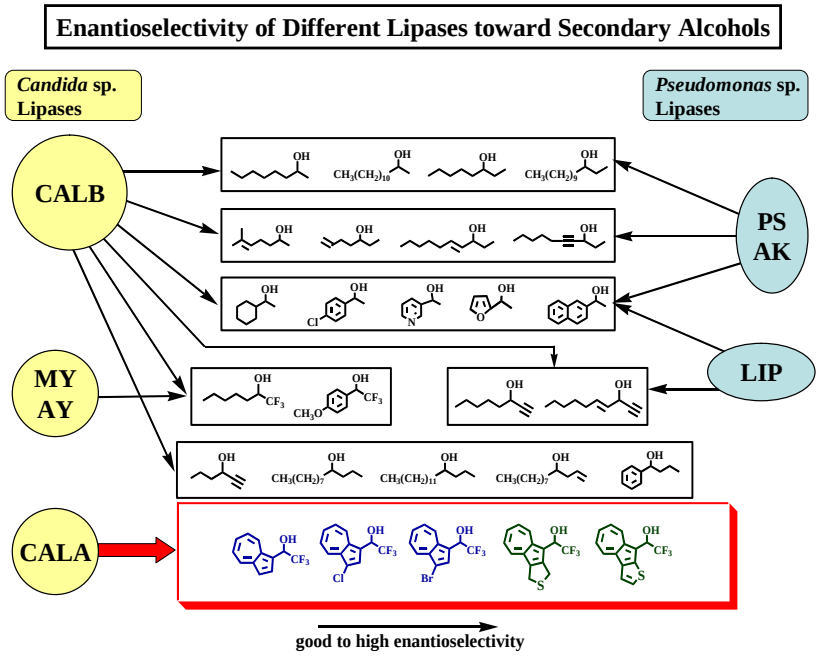
Problem D: Proposal

Computations on the Enantioselectivity of Lipase Enzymes

Yoshihiro MORI and Yoshinobu NAOSHIMA
Faculty of Informatics, Okayama University of Science

Protein lipases are enzymes hydrolyzing a variety of alcohol esters and are one of the most widely used biocatalysts in the synthesis of enantiomerically pure chiral compounds of biological relevance. A number of experimental data (including our results) are available for the stereopreference and enantioselectivity of *Pseudomonas cepacia* and *Candida antarctica* lipases toward secondary alcohols.[1)] Recently we have carried out the docking simulations (Glide) between these two lipases and chiral ligands and the molecular dynamics calculations (MacroModel) and the quantum chemical calculations (*ab initio* fragment molecular orbital method, BioStation) of the lipase-ligand complexes.[2)] Although we are neither specialists on theoretical chemistry nor on computational chemistry, we believe that computer simulation can be a useful technique for the elucidation of the enantioselectivity of the enzymes and related biomolecules toward chiral organic compounds.

We do hope that, through this kind of plan, we will be acquiring the knowledge and skill of molecular modeling and simulation more.



REFERENCES

1) Y. Naoshima, M. Kamezawa, T. Kimura, F. Okimoto, M. Watanabe, H. Tachibana, and T. Ohtani, *Recent Res. Devel. Org. Bioorg. Chem.*, **4**, 1-16 (2001) ; Y. Naoshima, T. Kimura, Y. Mori, M. Kamezawa, H. Tachibana, K. Kohara, and T. Ohtani, *Recent Res. Devel. Org. Bioorg. Chem.*, **6**, 1-9 (2004).

2) Y. Mori, Y. Yagi, and Y. Naoshima, Proceedings of 25th Anniversary International CBI Conference, 53 (2005).

Author Index