CBI2005 PROCEEDINGS

Chem-Bio Informatics in Post Genome Era

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. The participants of the previous meeting is approximately 400.

CBI2005 PROCEEDINGS

Edited by Akihiko Konagaya
August 24, 2005

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

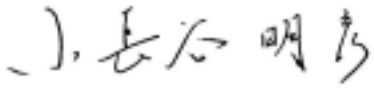
Preface

The scope of this 25th anniversary international conference is chem-bio informatics which is basis for pharmaceutical science, agriculture and environmental science. Established in 1981, the Chem-Bio Informatics Society (CBI) organized many meetings, seminars, training courses, and symposia for rational drug design targeted towards researchers in pharmaceutical companies and academia. Recently, the increasing amount of genome and post-genome data has raised many important issues in pathways and networks especially from the viewpoints of pharmacogenomics and personalized medicine.

This conference consists of three keynotes, four invited talks and nine oral presentations and poster presentations in the field of:
(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
The scientific program committee also provides two tutorial talks on bioinformatics, two adjunct workshops on “Immunoinformatics” and “Genome Medical Informatics”, and three lunch-on seminars organized by enterprise companies.

As for the keynotes, Richard Friesner will open the conference with his talk titled “Computational Modeling of Protein-Ligand Interactions”, Allen Roses will present “PGx in Early Development and Pharmacovigilance” and lastly, Yuichi Sugiyama will give his talk on “Role of drug transporter studies in the drug discovery and development; Molecular multiplicity, substrate specificity and drug-drug interaction”.

We hope all the participants will enjoy this conference by exchanging ideas and experiences through excellent keynotes, invited talks, oral presentations and poster presentations.



Akihiko Konagaya
Program Chair

Co-organization

RIKEN Genomic Sciences Center (GSC)
Japan Association of Medical Informatics

Benefactors

Information Processing Society of Japan
Japan Association for Medical Informatics
Japanese Society for Bioinformatics
The Biophysical Society of Japan
The Chemical Society of Japan
The Japanese Society for Artificial Intelligence
The Japanese Society for the Study of Xenobiotics
The Molecular Biology Society of Japan
The Pharmaceutical Society

Contributors

Accelrys Inc.
Daiichi Pharmaceutical Co.Ltd.
Eisai Co.Ltd.
IBM Japan, Ltd.
Initiative for Parallel Bioformatics (IPAB)
Mochida Pharmaceutical Co.,Ltd.
Taiho Pharmaceutical Co.,Ltd.
Taisho Pharmaceutical Co.,Ltd.

Exhibition & Lunch on Seminor Companies

AdvanceSoft Corporation
Affymetrix Japan K.K.
Cheminfonavi
CONFLEX Corporation
FUJITSU LIMITED
Infocom Corp.oration
Mathematical Systems, Inc.
NABE International Corporation
NEC Informatec Systems,Ltd.
Nittetsu Hitachi Systems Engineering,Inc.
Patcore Inc.
Ryoka Systems Inc.
SGI Japan, Ltd.
Sumisho Computer Systems Corporation
Tomy Digital Biology Co., Ltd.
Wavefunction, Inc.

CBI Members

Accelrys Inc.
AdIn Research Inc.
AdvanceSoft Corporation
Ajinomoto Co. Ltd.
Astellas Pharma Inc.
Banyu Pharmaceutical Co.,Ltd.
CAC Corp.
Chugai Pharmaceutical Co.,Ltd.
Daiichi Pharmaceutical Co.,Ltd.
Dainippon Pharmaceutical Co.,Ltd.
DNA Chip Research Inc.
Eisai Co. Ltd.
Fuji Film Co. Ltd.
Fujitsu Ltd.
GlaxoSmithKline K.K.
Hitachi Co. Ltd.
IBM Japan, Ltd.
Infocom Corp.
MediBIC
Mitsui Chemical Co.,Ltd.
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

Akihiko Konagaya (RIKEN GSC)

Chairs

Tsuguchika Kaminuma (Hiroshima University)

Yukio Tada (Taiho Pharmaceutical Co., Ltd)

Takatoshi Kawai (Eisai Co.Ltd.)

Hiroshi Tanaka (Tokyo Medical & Dental University)

Hiroshi Mizushima (National Cancer Research Institute)

Fumikazu Konishi (RIKEN GSC)

Masami Uebayasi (AIST)

Yoshiro Nakata(Gunma Univaersity)

Christian Schönbach (Riken GSC)

Jun Nakaya (Kobe University Graduate School of Medicine)

Tomokazu Matsusue (Mochida Pharmaceutical Co. Ltd.)

Masanori Arita (AIST)

Takashi Mizuma (Tokyo University)

Yoshihiro Takamoto (MediBIC)

24th(Wed.)

10:00		**Tutorial in Japanese**
	S1	New Opportunities and Challenging Topics in Informatics and Computing in Life Science **Tsuguchika Kaminuma (Hiroshima University)**
	S2	Bioinformatics for drug discovery in pharmaceutical company: current status and issues **Hiroki Shirai(Astellas Pharma Inc.)**
12:00		
		Luncheon Seminor **INFOCOM CORPORATION**
13:20		**Opening**
13:30		**Keynote Address**
	S3	Novel Methods for the Prediction of Structure and Binding Affinities of Protein-Ligand Complexes; Towards a New Theory and Practice of Molecular Recognition **Richard Friesner (Columbia Univ., USA)**
14:30		**Technical Session** Molecular Computation and Molecular Recognition
		Invited Talk
	S4	Computer-Assisted Drug Design (CADD) at Bristol-Myers Squibb **Roy S. Kimura (Bristol-Myers Squibb Company)**
		Invited Talk
	S5	From Genome Sequence to Protein Structure and Drug Discovery: computational pipeline with bioinformatics and supercomputing **Yutaka Akiyama (CBRC, AIST)**
	P7	Molecular Dynamics Simulations of HIV-1 Proteases - Insight into the Multi-drug Resistance conferred by Non-active Site Mutation L90M **Hirotaka Ode[1], Sabrou Neya[1], Masayuki Hata[1], Wataru Sugiura[2], ○Tyuji Hoshino[1,3]** Graduate School of Pharmaceutical Sciences, Chiba University[1], National Institute of Infectious Diseases[2] , JST[3]
	P51	Protein-protein interaction sites prediction and protein-protein docking by the methods of MD, grid scoring and the pair-wise interaction potential in CAPRI Rounds 3-5 **○ Genki Terashi, Mayuko Takeda-Shitaka, Daisuke Takaya, Katsuichiro Komatsu, Kazuhiko Kanou, Mitsuo Iwadate, Hideaki Umeyama** School of Pharmaceutical Sciences, Kitasato University
	P64	Structural analysis by molecular dynamics simulations:Protein-protein interactions of Ras-Raf and Ras-RalGDS complexes **○Noriyuki Futatsugi[1], Mikako Shirouzu [2], Atsushi Suenaga[1], Noriaki Okimoto[1], Tetsu Narumi[1], Toshikazu Ebisuzaki[3], Shigeyuki Yokoyama[2], Akihiko Konagaya[1], Makoto Taiji[1]** Computational and Experimental Systems Biology Group, RIKEN Genomic Sciences Center [1], Protein Research Group, RIKEN Genomic Sciences Center [2], Ebisuzaki Computational Astrophysics Laboratory, RIKEN [3]
17:10 18:30		**Poster Session**

25th(Thu.)

9:45		**Immunoinformatics Workshop**
	S6	Strategies for Improved Vaccine Design **Nikolai Petrovsky (Flinders Medical Centre, Australia)**
	S7	Beyond Epitope Mapping: Making Vaccines and Modifying Therapeutics **Anne S DeGroot (EpiVax Inc., USA)**
	S8	Automatic Construction of Specialized Databases, Exemplified by the VBASE2 Database, an Integrative Database for Immunoglobulin Variable Genes **Werner Müller (German Research Centre for Biotechnology, Germany)**
	S9	Transcriptional and Translational Variation from an Immunomics Point of View **Christian Schönbach (RIKEN GSC, Japan)**
12:00		
		Luncheon Seminor **RYOKA SYSTEMS INC.** Interactive Cheminformatics: Workflow meets Visualization
13:30		**Keynote Address**
	S10	Pipeline pharmacogenetics: Phase I, Phase II, & Phase III & Surveillance **Allen Roses (Glaxo Smith Kline)**
14:40		**Technical Session** Bioinformatics and Genome-wide Experimental Data Analysis
		Invited Talk
	S11	High throughput biology: from genome diversity to transcriptional regulation **Hiroyuki Aburatani (Univ. of Tokyo)**
		Invited Talk
	S12	System-level Identification of Mammalian Circadian Clocks **Hiroki Ueda (RIKEN CDB)**
	P28	Amino acid sequence analysis of *Saccharomyces cerevisiae* proteome in terms of electric charge distribution ○**Runcong Ke , Shigeki Mitaku** Department of Applied Physics, Graduate School of Engineering, Nagoya University
	P35	Semi-automatic interpretatio of experimental results based on conceptual network ○**Asako Koike [1,2] and Toshihisa Takagi [1]** Graduate School of Frontier Science, The University of Tokyo[1], Central Research Laboratory, Hitachi Ltd. [2]
	P48	*In silico* analyses of biosynthesis rate limiting factor in cell-free system ○**Naoya Fujita[1], Motoaki Seki[2], Kazuki Shinozaki[2], Kengo Kinoshita[1], Tatsuya Sawasaki[3], Yaeta Endo[3], Kenta Nakai[1]** Inst. of Med. Sci., Univ. of Tokyo[1], GSC, RIKEN[2], CSTRC, Ehime Univ.[3]
17:10 18:30		**Poster Session**

26th(Fri.)

10:00		**Medical Genome Informatics Workshop**
	S14	**Key Note** Omics-based systems approach to diseases -A new paradigm for clinical medicine in the post-genomic era- **Hiroshi Tanaka (Tokyo Medical & Dental University)**
	S15	**Invited Talk** Future Vision: A possibility of Genome Information toward the Personalized Medicine and Personalized Healther **Hiroshi Mizushima (National Cancer Research Institute)**
	S13	**Invited Talk** MC's Comment The Translational Research Informatics: A bridge toward the Genome Medicine **Jun Nakaya (Kobe Univ.)** Discussion: The Goal, the Hurdles, and the Path of the Genome Medical Informatics? **Hiroshi Tanaka, Hiroshi Mizushima, Jun Nakaya (MC)**
12:00		
		Luncheon Seminor **AFFYMETRIX JAPAN** Affymetrix Gene Chip® Technology and its Applications in Clinical Research
13:30		**Keynote Address**
	S16	Role of drug transporter studies in the drug discovery and development; Molecular multiplicity, substrate specificity and drug-drug interaction **Yuichi Sugiyama (Univ. of Tokyo)**
14:30		**Technical Session** Drug Design and Toxicology, Disease Mechanism and Control Model
	P19	Extraction of Information on Chemicals-CYP3A4 Interactions from Literature ○**Chunlai Feng, Fumiyoshi Yamashita, Mitsuru Hashida** Graduate School of Pharmaceutical Sciences Kyoto University
	P72	A Grid-based Information Integration System for Drug Discovery by using Freely Accessible Compound Databases ○**Gen Kawamura[1, 2], Masato Kitajima[1,3], Takahiro Kosaka[1], Susumu Date[1], Shinji Shimojo[4], Hideo Matsuda[1]** Graduate School of Information Science and Technology, Osaka University[1], Aztec-System Co., Ltd.[2], Fujitsu Kyushu System Engineering Limited[3], Cybermedia Center, Osaka University[4]
	P14	Evaluation of flexible molecular docking programs for virtual screening ○**Kenji Onodera, Kazuhito Satou, Hiroshi Hirota** Riken Genomic Sciences Center
15:30		**Closing** Ceremony for poster awards

Keynotes and Invited Talks

S1 **New Opportunities and Challenging Topics in Informatics and Computing in Life Science**

Tsuguchika Kaminuma

(Center for Quantum Life Science, Hiroshima University)

S2 **Bioinformatics for Drug Discovery in Pharmaceutical Company: current status and issues**

Hiroki Shirai

(Astellas Pharma Inc.)

S3 **Novel Methods for the Prediction of Structure and Binding Affinities of Protein-Ligand Complexes; Towards a New Theory and Practice of Molecular Recognition**

Richard Friesner

(Columbia University, USA)

S4 **Computer-Assisted Drug Design (CADD) at Bristol-Myers Squibb**

S. Roy Kimura

(Bristol-Myers Squibb Company)

S5 **From Genome Sequence to Protein Structure and Drug Discovery: computational pipeline with bioinformatics and supercomputing**

Yutaka Akiyama

(Director, Computational Biology Research Center (CBRC) National Institute of Advanced Industrial Science and Technology (AIST), Japan)

S6 **Strategies for Improved Vaccine Design**

Nikolai Petrovsky

(Australian Centre for Adjuvant Research, Flinders Medical Centre)

S7 **Beyond Epitope Mapping: Making Vaccines and Modifying Therapeutics**

Annie De Groot, Bill Martin, Luisa Marcon, Paul Knopf, Betty Bishop, Julie Mcmurry, Dan Rivera, Michelle Klutzer, Christine Boren, Claire Rodriguez-annon, David Weiner

(TB/HIV Research Laboratory, University of Pennsylvania, and EpiVax, Inc.)

S8 **Automatic construction of specialized databases, exemplified by the VBASE2 database, an integrative database for Immunoglobulin Variable Genes**

Werner Müller

(Department of Experimental Immunology, GBF German Research Centre for Biotechnology, Braunschweig, Germany)

S9 **Transcriptional and translational variation from an immunomics point of view**

Christian Schönbach

(Immunoinformatics Team, Advanced Genome Information Technology Group RIKEN Genomic Sciences Center, Yokohama, Kanagawa 230-0045, Japan)

S10 **Pipeline pharmacogenetics: Phase I, Phase II, & Phase III & Surveillance**

Allen D. Roses

(GlaxoSmithKline)

S11 **High throughput biology: from genome diversity to transcriptional regulation**

Hiroyuki Aburatani

(Genome Science Division, Laboratory for Systems Biology and Medicine, Research Center for Advanced Science and Technology, University of Tokyo)

S12 **System-level identification of Mammalian Circadian Clocks.**

Hiroki R. Ueda[1,2,3]

(Team Leader, Laboratory for Systems Biology and Subunit Leader, Functional Genomics Subunit, Center for Developmental Biology, RIKEN, Kobe Japan. [1], Visiting Professor, Tohoku University.[2], Visiting Professor, Tokushima University.[3])

S13 **MC's Comment: The Translational Research Informatics: A bridge toward the Genome Medicine**

Jun Nakaya

(Associate Professor, Clinical Genome Informatics Center, School of Medicine, Kobe University)

S14 **Omics-based systems approach to diseases - A new paradigm for clinical medicine in the post-genomic era -**

Hiroshi Tanaka

(Department of Bioinformatics, Tokyo Medical and Dental University)

S15 **Future Vision: A possibility of Genome Information toward the Personalized Medicine and Personalized Healthcare.**

Hiroshi Mizushima

(Head of Bioinformatics Section. Center for Medical Genomics. National Cancer Center Research Institute.)

S16 **Role of drug transporter studies in the drug discovery and development; Molecular multiplicity, substrate specificity and drug-drug interaction**

Yuichi Sugiyama

(Department of Molecular Pharmacokinetics, Graduate School of Pharmaceutical Sciences, The University of Tokyo)

Poster Presentations

Odd Number: 8/24 17:30 – 19:00 Even Number: 8/25 17:30 – 19:00

P6 Structural Analysis and Discrimination of Outer Membrane Proteins

M. Michael Gromiha, Makiko Suwa

(Computational Biology Research Center (CBRC), AIST, 2-42, Aomi, Koto-ku, Tokyo 135-0064, Japan)

P7 Molecular Dynamics Simulations of HIV-1 Proteases - Insight into the Multi-drug Resistance conferred by Non-active Site Mutation L90M -

Hirotaka Ode[1], Saburou Neya[1], Masayuki Hata[1], Wataru Sugiura[2], Tyuji Hoshino[1,3]

(Graduate School of Pharmaceutical Sciences, Chiba University[1], National Institute of Infectious Diseases[2], JST[3])

P8 GLYMM: a program for modeling a 3D structure of a carbohydrate chain and a lipid membrane: the application to molecular dynamics simulations of glycolipid (GM1) membranes

Kenichi Mori[1], Tsubasa Takaoka[1], Koji Iwamoto[1], Masayuki Hata[1], Saburou Neya[1], Tyuji Hoshino[1,2]

(Graduate School of Pharmaceutical Science, Chiba University[1], PRESTO JST[2])

P9 Computational Chemistry Platform for Drug Metabolomics

Michihisa Koyama[1], Satoshi Kawagoe[1], Kota Kasahara[1], Hideyuki Tsuboi[1], Akira Endou[2], Momoji Kubo[1,3], Carlos A. Del Carpio[1], Ewa Broclawik[1], Kazumi Nishijima[4,5], Tetsuya Terasaki[4,6], Akira Miyamoto[4,1]

(Department of Applied Chemistry, Tohoku University[1], Institute of Fluid Science, Tohoku University[2], PRESTO, Japan Science and Technology Agency[3], New Industrial Creation Hatchery Center[4], Development Division, Mochida Pharmaceutical Co. Ltd.[5], Graduate School of Pharmaceutical Sciences, Tohoku University[6])

P10 Theoretical Study on the Mechanism of Theophylline Metabolism by Compound I of Cytochrome P450

Mohamed Ismael[1], Abdul Rajjak Shaikh[1], Hideyuki Tsuboi[1], Michihisa Koyama[1], Akira Endou[2], Momoji Kubo[1,3], Carlos A. Del Carpio[1], Ewa Broclawik[4], Kazumi Nishijima[4,5], Tetsuya Terasaki[4,6], Akira Miyamoto[1,4]

(Department of Applied Chemistry, Tohoku University[1], Institute of Fluid Science, Tohoku University[2], JST-PRESTO[3], New Industry Creation Hatchery Center, Tohoku University[4], Development Division, Mochida Pharmaceutical Co., Ltd.[5], Graduate School of Pharmaceutical Sciences, Tohoku University[6])

P11 An *in Silico* discriminant model constructed for HERG potassium channel inhibitors

Motoi Tobita[1,2], Tetsuo Nishikawa[1]
(Reverse proteomics research institute, 2-6-7 Kazusa-kamatari, Kisarazu-si, Chiba 292-0818, Japan,[1], Advanced Research Laboratory, Hitachi, Ltd. 1-280 Higashi-Koigakubo, Kokubunji-shi, Tokyo, 185-8601, Japan[2])

P14 Evaluation of flexible molecular docking programs for virtual screening

Kenji Onodera, Kazuhito Satou, Hiroshi Hirota
(Riken Genomic Sciences Center)

P15 Implementation of a Haplotype Reconstruction Algorithm and a tagSNP Selection Algorithm for Large-scale SNP Information Processing

Sang Jun Kim, Sang-soo Yeo, Sung Kwon Kim
(Chung-Ang University)

P16 Computational study of conformational preference in the tetrahedral intermediates of the acylation step of ester hydrolysis catalyzed by lipase

Yu Takano[1,2], K. N. Houk[2]
(Institute for Protein Research, Osaka University[1], Department of Chemistry and Biochemistry, University of California Los Angeles[2])

P17 Computational Analysis of CYP3A4 Mediated Metabolism of KCNQ2 Potassium Channel Opener

Shaikh Abdul Rajjak[1], Mohamed Ismael[1], Hideyuki Tsuboi[1], Michihisa Koyama[1], Akira Endou[2], Momoji Kubo[1,3], Carlos A. Del Carpio[1], Ewa Broclawik[4], Kazumi Nishijima[4,5], Tetsuya Terasaki[4,6], Akira Miyamoto[1,4]
(Department of Applied Chemistry, Tohoku University[1], Institute of Fluid Science, Tohoku University[2], JST-PRESTO[3], NICHe, Tohoku University[4], Mochida Pharma Co.,Japan[5], Graduate school of Pharmaceutical Sciences, Tohoku University, Japan[6])

P18 Computational Study on the Enantioselectivity of Lipase Enzyme toward Non-Natural Organic Compounds

Yoshihiro Mori, Yoichiro Yagi, Yoshinobu Naoshima
(Faculty of Informatics, Okayama University of Science)

P19 Extraction of Information on Chemicals-CYP3A4 Interactions from Literature

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

P20 Spectral Analysis of the Daily Prescription Variations at a Pharmacy for Infectious and Non-Infectious Diseases

Katsunori Segawa[1], Tatsuya Nakano[1], Kotoko Nakata[1], Kazushige Ijuin[2], Noriko Hatanaka[3], Yuzuru Hayashi[1]

(National Institute of Health Sciences[1], Tanashi Yakuhin Co. Ltd.[2], Kakunoki Pharmacy[3])

P21 Computational study on ATP hydrolysis by ABC transporter subunit HisP

Qiang Pei[1], Hideyuki Tsuboi[1], Michihisa Koyama[1], Akira Endou[2], Momoji Kubo[1,3], Carlos A. Del Carpio[1], Ewa Broclawik[4], Kazumi Nishijima[4,5], Tetsuya Terasaki[4,6], Akira Miyamoto[1,4]

(Department of Applied Chemistry, Graduate School of Engineering, Tohoku University[1], Institute of Fluid Science, Tohoku University[2], JST-PRESTO[3], NICHe, Tohoku University[4], Developmen Division, Mochida Pharmaceutical Co. Ltd.[5], Graduate School of Pharmaceutical science, Tohoku University[6])

P22 *In silico* screening for the human P2Y12 receptor

Yosuke Nonaka, Takeshi Hiramoto, Kiyoshi Baba, Norihisa Fujita

(Laboratory of Pharmcoinformatics, Department of Bioinformatics, Ritsumeikan University, Kusatsu, Shiga 525-8577)

P23 Systematic comparisons of consensus scores for computational ligand-docking

Akifumi Oda[1], Keiichi Tsuchida[1], Tadakazu Takakura[1], Noriyuki Yamaotsu[2], Shuichi Hirono[2]

(Toyama Chemical Co. Ltd.[1], Kitasato University[2])

P24 A Multilocus Simulation Study to Multiple Testing for Association Test

Woosung Yang, Jun Nakaya, Norihiro Sakamoto

(Kobe University of Clinical Genome Informatics Center)

P25 Kinetic assessment of lenampicillin absorption using Caco-2 cells: in vitro study aiming for rational design of prodrug

Takashi Mizuma, Sayaka Sakaguchi, Sachie Tanaami, Masahiro Hayashi

(Department of Drug Absorption and Pharmacokinetics, School of Pharmacy, Tokyo University of Pharmacy and Life Science (TUPLS))

P26 Database and analysis of binding site residues and ligands in membrane protein-ligand complexes

M. Xavier Suresh, M. Michael Gromiha, Makiko Suwa

(Computational Biology Research Center (CBRC), AIST, 2-42, Aomi, Koto-ku, Tokyo 135-0064, Japan)

P27 GeneT2D: A database for the genetic variation in type II diabetes

Yu Lin, Jun Nakaya, Norihiro Sakamoto
(Kobe University Graduate School of Medicine, 7-5-2 Kusunokicho, Chuo-ku, Kobe, Hyogo 650-0017, JAPAN1)

P28 Amino acid sequence analysis of *Saccharomyces cerevisiae* proteome in terms of electric charge distribution

Runcong Ke, Shigeki Mitaku
(Department of Applied Physics, Graduate School of Engineering, Nagoya University)

P29 Folding dynamics of 10 residues protein, chignolin

Atsushi Suenaga, Tetsu Narumi, Noriyuki Futatsugi, Ryoko Yanai, Yousuke Ohno, Noriaki Okimoto, Makoto Taiji
(Computational and Experimental System Biology Group, RIKEN Genomic Sciences Center)

P30 Massively parallel calculation of binding free energy for biomolecules

Masakatsu Ito[1], Yoshiaki Tanida[1], Hideaki Fujitani[1], Michael R. Shirts[2], Guha Jayachandran[2], Christopher D. Snow[2], Vijay S. Pande[2]
(Fujitsu Laboratories Ltd.[1], Chemistry Department, Stanford University[2])

P31 Construction of the Web-Accessible comprehensive transporter database for drug discovery and development, "TP-Search" (Part II)

Takafumi Ochiai[1], Naoki Ozawa[1,2], Yoshimasa Hama[1], Katsuhiko Taki[3], Toshiyuki Ueda[4], Noriko Sugawara[4], Yukiko Abe[4], Hitoshi Matsui[4], Kiyomi Ito[5], Kazuya Maeda[6], Hiroyuki Kusuhara[6], Yuichi Sugiyama[6]
(Advanced Research Institute for Science and Engineering, Waseda University[1], Wyeth K.K.[2], Nihon Visual Science.Inc[3], CAC Corporation[4], Hoshi University[5], Graduate School of Pharmaceutical Sciences, the University of Tokyo[6])

P32 Particle simulation approach for subcellular spatio-temporal events

Ryuzo Azuma[1], Tetsuji Kitagawa[2], Tomoyuki Yamamoto[3], Takeshi Arikuma[2], Hiroshi Kobayashi[4], Akihiko Konagaya[1,2]
(Genomic Sciences Center, RIKEN[1], Mathematics and Computing Sciences, Tokyo Institute of Technology[2], School of knowledge Sci. JAIST[3], Graduate Sch. of pharmaceutical Sciences, Chiba Univ.[4])

P33 Spontaneous Docking MD Simulations of Aspartate Racemase - Roles of Amino Acid Residues in Docking Process -

Okimasa Okada[1], Tomohiro Seko[1], Masafumi Yohda[2]
(Fuji Xerox Co., Ltd.[1], Tokyo University of Agriculture and Technology[2])

P34 Transplantation of the PEACH5.8 from UNIX to Windows System, and Practice in the Simulation of TNA

Yoshiro Nakata, Wataru Ootani, Noriaki Saitoh
(Department of Biophysics, Faculty of Engineering, Gunma University)

P35 Semi-automatic interpretation of experimental results based on conceptual network

Asako Koike[1,2], Toshihisa Takagi[1]
(Dept. of Computational Biology, Graduate School of Frontier Science, The University of Tokyo, Kiban-3A1(CB01) 5-1-5, Kashiwanoha Kashiwa, Chiba, 277-8561, Japan[1], Central Research Laboratory, Hitachi Ltd. 1-280 Higashi-Koigakubo Kokubunji, Tokyo, 185-8601, Japan[2])

P36 Development of Effective Computer-Aided Drug Design Strategy

Noriaki Okimoto, Atsushi Suenaga, Noriyuki Futatsugi, Ryoko Yanai, Tetsu Narumi, Yosuke Ohno, Makoto Taiji
(Computational and Experimental Systems Biology Group, RIKEN Genomic Sciences Center)

P37 Novel Type of Secondary Structure Breaker in Soluble Proteins

Kenichiro Imai, Shigeki Mitaku
(Nagoya University, Graduate School of Engineering, Department of Applied Physics)

P38 A solvent site-dipole field mediating DNA-protein binding

Nobuyuki Hamasaki[1], Daisuke Mitomo[1], Junichi Higo[1], Akihiko Yamagishi[1], Hiroh Miyagawa[2]
(School of Life Science, Tokyo University of Pharmacy and Life Science[1], Taisyo Pharmaceutical co., Ltd[2])

P39 Development of a Three-Dimensional-Structure Database of Natural Metabolites (3DMET) and Problems about Conversion from 2D to 3DStructures.

Miki Maeda, Hisataka Numa
(Genome Research Department, National Institute of Agrobiological Sciences)

P40 How effective for fold recognition is a potential of mean force that includes relative orientations between contacting residues in proteins?

Sanzo Miyazawa[1], Robert L. Jernigan[2]
(Faculty of Technology, Gunma University, Japan[1], L. H. Baker Center for Bioinformatics and Biological Statistics, Iowa State University, USA[2])

P41 MOLWORKS+G: Integrated Workbench for the Chimo-Bio Molecular Design by Grid Computing

Fumikazu Konishi[1], Toru Yagi[2], Akihiko Konagaya[1]
(RIKEN GSC, Advanced Genome Information Technology Research Group[1], BestSystems Inc.[2])

P42 A mutation in a cuticle collagen causes hypersensitivity to bisphenol A in *Caenorhabditis elegans*

Masahito Watanabe[1], Nanako Mitani[1], Naoaki Ishii[2], Keizaburo Miki[3]
(Japanese Institute of Pearl Science.[1], Department of Molecular Life Science, Tokai University School of Medicine[2], BIOS Inc.[3])

P43 *Ki*Bank, major new developments and status in 2005

Junwei Zhang[1], Masahiro Aizawa[1], Shinji Amari[1], Yoshio Iwasawa[2], Tatsuya Nakano[3], Kotoko Nakata[3]
(Collaborative Research Center of Frontier Simulation Software for Industrial Science, Institute of Industrial Science, University of Tokyo[1], AdvanceSoft Corporation[2], Division of Safety Information on Drug, Food and Chemicals, National Institute of Health Sciences[3])

P45 Software system for the prediction of mitochondria localization on the basis of physicochemical profiles in amino terminal segments

Toshiyuki Tsuji[1,2], Shigeki Mitaku[1]
(Department of Applied Physics Graduate School of Engineering Nagoya University[1], 21st Century COE program "Frontiers of Computational Science" Nagoya University[2])

P46 Comparative Analyses for Selecting Effective siRNA Target Sequences

Shigeru Takasaki, Akihiko Konagaya
(Advanced Genome Information Technology Group, RIKEN Genomic Sciences Center)

P47 Development of a Workbench for Selective Nuclear Receptor Modulator

Naomi Komiyama[1], Tatsuya Nakano[2], Kaori Fukuzawa[3], Junpei Komiyama[4], Masumi Yukawa[5], Kotoko Nakata[6], Tsuguchika Kaminuma[7]
(Chem-Bio Informatics Society[1], The National Institute of Health Sciences[2], Mizuho Research Institute[3], The University of Tokyo[4], Tokyo Medical and Dental University[5], IIS, University of Tokyo / Advance Soft Co[6], Hiroshima University[7])

P48 *In silico* analysis of biosynthesis rate limiting factor in cell-free system

Naoya Fujita[1], Motoaki Seki[2], Kazuki Shinozaki[2], Kengo Kinoshita[1], Tatsuya Sawasaki[3], Yaeta Endo[3], Kenta Nakai[1]
(Inst. of Med. Sci., Univ. of Tokyo[1], GSC, RIKEN[2], CSTRC, Ehime Univ.[3])

P49 Cis-Regulatory Machinery Unit (CiRMU) Model: A portable system for studying gene transcriptional processes

Masumi Yukawa[1], Yoshitomo Tanaka[1], Hiroshi Tanaka[1], Kotoko Nakata[2], Naomi Komiyama[3], Tsuguchika Kaminuma[4]
(Tokyo Medical and Dental University[1], IIS, University of Tokyo / Advance Soft Co[2], Chem-Bioinformatics Society[3], Hiroshima University[4])

P50 A Knowledge Environment for Adipocyte and Adipokine Research

Tsuguchika Kaminuma[1], Masumi Yukawa[3], Naomi Komiyama[2], Yoshitomo Tanaka[3], Hiroshi Tanaka[3]

(Hiroshima University[1], Chem-Bio Informatics Society[2], Tokyo Medical and Dental University[3])

P51 Protein-protein interaction sites prediction and protein-protein docking by the methods of MD, grid scoring and the pair-wise interaction potential in CAPRI Rounds 3-5

Genki Terashi, Mayuko Takeda-shitaka, Daisuke Takaya, Katsuichiro Komatsu, Kazuhiko Kanou, Mitsuo Iwadate, Hideaki Umeyama

(School of Pharmaceutical Sciences, Kitasato University)

P52 Protein Structure Prediction in the 6th round of Critical Assessment of Techniques for Protein Structure Prediction (CASP6) using CHIMERA and FAMS

Mayuko Takeda-shitaka, Genki Terashi, Daisuke Takaya, Kazuhiko Kanou, Mitsuo Iwadate, Hideaki Umeyama

(School of Pharmaceutical Sciences, Kitasato University)

P53 Performance of multi-stage docking simulation with a fast Poisson-Boltzmann solver and re-scoring methods

Hiroaki Fukunishi[1], Jiro Shimada[1], Daisuke Tokushima[1], Takumi Washio[1], Hiroshi Kuramochi[2]

(NEC Corporation[1], NIPPON KAYAKU CO., LTD.[2])

P54 FAMS Complex: A fully automated homology modeling system for protein complex structures

Mayuko Takeda-shitaka[1], Genki Terashi[1], Daisuke Takaya[1], Kazuhiko Kanou[1], Mitsuo Iwadate[1], Katsuichiro Komatsu[2], Kinji Fuchikami[2], Hideaki Umeyama[1]

(School of Pharmaceutical Sciences, Kitasato University[1], In-Silico Sciences, Inc.[2])

P55 Fully automated protein structure predictions at CASP6 using homology-modeling server SKE-FAMSD

Kazuhiko Kanou, Mitsuo Iwadate, Genki Terashi, Daisuke Takaya, Mayuko Takeda-shitaka, Hideaki Umeyama

(Department of Biomolecular Desig School of Pharmaceutical Sciences, Kitasato University)

P56 Development and application of a molecular dynamics simulation system : prestoX

Satoru Kubota[1], Yoshifumi Fukunishi[2], Ikuo Fukuda[1], Masaru Horie[1], Katumi 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])

P57 Transcriptome according to a thermodynamic view

Tomokazu Konishi

(Bioresource Sciences, Akita Prefectural University)

P58 Automated genotyping of human CYP2C19 SNPs by a novel SNP typing system and development of a database of it

Yukari Kurebayashi[1,2], Naoko Matsumoto[1,2], Fumiko Kakihara[1,2], Mitsuharu Hirai[3], Setsuo Hasegawa[1], Masafumi Yohda[2]

(Sekino Clinical Pharmacology Clinic[1], Tokyo University of Agriculture and Technology[2], ARKRAY, Inc.[3])

P59 Gene expression profiles of the effect of cryoprotectants on the freeze-thaw damage in yeast

Yuko Momose, Masakazu Yamaoka

(The National Institute of Advanced Industrial Science and Technology)

P60 Analysis of drug-resistant mutation patterns of HIV-1 protease under highly active antiretroviral therapy using Bayesian network modeling

Kaoru Mogushi, Fengrong Ren, Naoki Hasegawa, Hiroshi Tanaka

(Department of Bioinformatics, Tokyo Medical and Dental University)

P61 Comparison of the validation methods with linear and non-linear classifier

Chihoko Yoshimura[1], Taizo Hanai[2], Masahiro Okamoto[2]

(Drug Discovery Lab., Hanno Research Center, Taiho Pharmaceutical Co., LTD[1], Laboratory for Bioinformatics, Graduate School of Systems Life Sciences, Kyushu University[2])

P62 Computational Proteomic Differential Analysis: *i-OPAL*

Mitsuhiro Kanazawa[1,3], Hisae Anyoji[1], Hsiao-kun Tu[1], Takao Kawakami[1,2], Toshihide Nishimura[1,2], Atsushi Ogiwara[1,2]

(Medical ProteoScope Co., Ltd.[1], Clinical Proteome Center, Tokyo Medical University[2], Graduate School of Pure and Applied Sciences, University of Tsukuba[3])

P64 Structural analysis by molecular dynamics simulations: Protein-protein interactions of Ras-Raf and Ras-RalGDS complexes

Noriyuki Futatsugi[1], Mikako Shirouzu[2], Atsushi Suenaga[1], Noriaki Okimoto[1], Tetsu Narumi[1], Toshikazu Ebisuzaki[3], Shigeyuki Yokoyama[2], Akihiko Konagaya[1], Makoto Taiji[1]

(Computational and Experimental Systems Biology Group, RIKEN Genomic Sciences Center[1], Protein Research Group, RIKEN Genomic Sciences Center[2], Ebisuzaki Computational Astrophysics Laboratory, RIKEN[3])

P66 NMR-based Metabolomics.1/3 1) Newly developed software; "Alice2 for Metabolome"; Integrated NMR spectroscopic and chemomerirc analysis for mixture

Kazunori Arifuku[1], Itiro Ando[2], Masako Fujiwara[1]

(JEOL DATUM LTD[1], Environmental Research Center LTD[2])

P63 NMR-based Metabolomics.2/3 2) How to analyze NMR-spectra of mixture samples by chemometric method; in examples of vegetable metabolome

Masako Fujiwara[1], Itiro Ando[2], Kazunori Arifuku[1]
(JEOL DATUM LTD[1], Envioronmental Research Center LTD[2])

P65 NMR-based Metabolomics.3/3 3) High-field NMR-spectroscopy for metabolomic analysis; An application to urine samples from model rats

Tadashi Nemoto[1], Taeko Kataoka[1], Itiro Ando[2], Katsuo Asakura[3], Kazunori Arifuku[4], Kenji Kanazawa[1], Masako Fujiwara[4]
(National Institute of Advanced Industrial Science and Technology (AIST)[1], Environmental Research Center LTD[2], JEOL LTD[3], JEOL DATUM LTD[4])

P67 Controlling dissociation and association of amyloid protofibrils using high pressure

Yuji O. Kamatari[1,2,3], Shigeyuki Yokoyama[2,3,4], Hideki Tachibana[5], Kazuyuki Akasaka[2,6]
(Gifu University[1], RIKEN Harima Institute[2], Riken Genome Science Center[3], The University of Tokyo[4], Kobe University[5], Kinki University[6])

P68 The evolutionary mechanism of protein-protein interaction network

Takeshi Hase, Soichi Ogishima, So Nakagawa, Hiroshi Tanaka
(Department of Bioinformatics, Tokyo Medical and Dental University)

P69 Is there a code for protein-DNA interaction? Yes, there exists the *analog* code for protein-DNA interaction.

Tomoki Yoshida[1], Misako Aida[1], Akinori Sarai[2]
(Center for Quantum Life Sciences (QuLiS), and Department of Chemistry, Graduate School of Science, Hiroshima University[1], Department of Bioscience and Bioinformatics, Kyushu Institute of Technology (KIT)[2])

P70 Investigation of Structural and Dynamic Properties of Importin-β Based on a Molecular Dynamics Study

Yoshinori Hirano[1], Noriaki Okimoto[1,2], Atsushi Suenaga[2], Naoko Imamoto[3], Toshikazu Ebisuzaki[1]
(Computational Astrophysics Laboratory, RIKEN[1], Bioinformatics Group, Genomic Sciences Center, RIKEN[2], Cellular Dynamics Laboratory, RIKEN[3])

P71 Matlab/GNU Octave as a Platform for Analyses of Metabolic Networks

Jun Ohta
(Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences)

P72 A Grid-based Information Integration System for Drug Discovery by using Freely Accessible Compound Databases

Gen Kawamura[1,2], Masato Kitajima[1,3], Takahiro Kosaka[1], Susumu Date[1], Shinji Shimojo[4], Hideo Matsuda[1]

(Graduate School of Information Science and Technology, Osaka University.[1], Aztec-System Co., Ltd.[2], Fujitsu Kyushu System Engineering Limited.[3], Cybermedia Center, Osaka University.[4])

P73 Conformational analysis of an antimicrobial peptide, Pediocin PA-1

Hitoshi Goto[1], Tatsunori Matsui[1], Toshiyuki Kamakura[1], Naofumi Nakayama[1], Tatsuya Tominaga[2], Hideo Nakajima[2]

(Toyohashi University of Technology[1], Saitama Industrial Technology Center North Institute[2])

P74 Normal mode analysis of the complex of HIV-1 protease with inhibitor vibrational behavior

Toshiyuki Kamakura, Hitoshi Goto

(Toyohashi University of Technology)

P75 Genome Functional Analysis based on Systematic Map of Protein Interactions

Carlos A. Del Carpio[1], Pei Qiang[1], Eichiro Ichiishi[2], Michihisa Koyama[1], Momoji Kubo[1,3], Akira Miyamoto[1,2]

(Department of Applied Chemistry, Tohoku University[1], New Industry Creation Hatchery Center, Tohoku University[2], JST-PRESTO[3])

P76 GenoSearch: High Performance Inter-Genomic Sequence Retrieval System using Suffix Array Technique

Aki Hasegawa, Akihiko Konagaya

(Informatics Infrastructure Team, Genome Core Technology Facilities RIKEN Genomic Sciences Center)

P77 Evaluation of polymorphism for an organic molecular crystal

Shigeaki Obata, Hitoshi Goto, Naofumi Nakayama

(Toyohashi University of Technology)

P78 Workflow for Open Genome Grid Service

Ryo Umetsu, Shingo Ohki, Akihiko Konagaya

(RIKEN GSC, Advanced Genome Information Technology Research Group)

P79 Drug Interaction Ontology (DIO) - OWL formalizations of interaction causal logic -

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

(Genomics Science Center (GSC), RIKEN (The Institute of Physical and Chemical Research)[1], Graduate School of Information Science and Technology, Tokyo Institute of Technology[2])

P80 Transition state of a SH3 domain detected with principle component analysis and a charge-neutralized all-atom protein model

Daisuke Mitomo, Junichi Higo
(School of Life Science, Tokyo University of Pharmacy and Life Science)

P81 CBDB: Cell Behavior Database - Toward understanding high order level biological phenomena -

Michiaki Iwazume[1], Hiroyuki Kaneko[2], Marina Dan[3]
(Brain Science Institute, RIKEN[1], Keio University[2], Hierarchical Biology Lab.[3])

P82 Drug discovery based on the structural dynamics of prion

Kazuo Kuwata
(Division of Prion Research, Center for Emerging Infectious Diseases, Gifu University, Gifu 501-1194)

P83 Parametric Normalization for GeneChip and Tiling Array Data

Tomokazu Konishi[1], Midori Katakami[2], Kenya Shibahara[2], Masahiro Hayashi[3], Takushi Nakajima[3]
(Bioresource Sciences, Akita Prefectural University[1], INTENSITY Co. Ltd.[2], Skylight Biotech Inc.[3])

P84 Consensus scores for efficient selection of computational protein-ligand docking models

Tadakazu Takakura[1], Akifumi Oda[1], Keiichi Tsuchida[1], Noriyuki Yamaotsu[2], Shuichi Hirono[2]
(Toyama Chemical Co. Ltd.[1], Kitasato University[2])

P85 Molecular dynamics simulation of prion using a coarse-grained model

Hironori K. Nakamura[1], Mitsunori Takano[2], Kazuo Kuwata[1]
(Division of Prion Research, Center for Emerging Infectious Diseases, Gifu University[1], Department of Physics, School of Science and Engineering, Waseda University[2])

P86 Molecular Interaction Between Estrogen Receptor and Their Ligands

Kaori Fukuzawa[1,2], Yuji Mochizuki[2,3], Shigenori Tanaka[2,4], Kazuo Kitaura[2,5], Tatsuya Nakano[2,6]
(Mizuho Information & Research Institute, Inc.[1], JST-CREST[2], Advancesoft[3], Kobe University[4], National Institute for Advanced Industrial Science and Technology[5], National Institute of Health Sciences[6])

S1

New Opportunities and Challenging Topics in Informatics and Computing in Life Science

Tsuguchika Kaminuma

Center for Quantum Life Science, Hiroshima University

Biology and related applied sciences such as medicine is getting into the so called post genomic era. The role of informatics and computing in these fields is dramatically increasing. This new trend is offering new challenging opportunities for researchers specialized in theory and computation, including mathematics, theoretical physics, computational chemistry, information science, and computer science. In this seminar what are the new opportunities and how to cope with the new challenges are discussed. Examples of new research topics are introduced, and necessity of educational reformation is touched. Questions and comments from the participants are most welcomed.

S2

Bioinformatics for drug discovery in pharmaceutical company: current status and issues

Hiroki Shirai

Astellas Pharma Inc.

Bioinformatics is being effectively used for “genomics / proteomics-based” drug discovery. There are four different roles; I) infrastructure construction <introduction and maintenance of hardware and software and construction of bioinformatics system and download of biological information> and management of routine operation, II) support for biological experiment <construction of database to be utilized by experimental researcher> III) analysis <analysis of sequence including genetic statistics / medical statistics method and a large amount of biological experimental data such as microarray data> IV) “*in silico*-driven” research <prediction of gene function and computational physics>. Current status and issues of each role is discussed.

S3

Novel Methods for the Prediction of Structure and Binding Affinities of Protein-Ligand Complexes; Towards a New Theory and Practice of Molecular Recognition

Richard Friesner

Columbia University, USA

The problem of predicting the structure and binding affinities of protein-ligand complexes is central to the effective application of computational methods in structure-based drug design. While significant progress has been made in both areas over the past decade, current approaches are lacking in the accuracy and robustness required to create a compelling technology that will qualitatively impact the drug discovery process. At the same time, the advent of inexpensive Linux clusters, and practical grid computing platforms, have offered the possibility of dramatically increasing the computing power that is available for these tasks. Thus, the crucial goal is to create new computational methods and models which exploit the available hardware platforms to deliver reliable answers that can drive both lead discovery and lead optimization.

Over the past several years, via collaborations between Schrodinger, Inc. and several academic research laboratories, we have developed new methods which represent a significant advance in docking and scoring technology. The major components of these advances can be enumerated as follows: (1) a scoring function for rigid receptor docking, based on a new theory of molecular recognition that identifies the key motifs capable of creating highly potent binding affinity with a relatively small number of ligand atoms; (2) sampling methods capable of accurately exploring the potential energy surface defined by this scoring function, which requires substantially higher resolution conformational search methods than standard empirical models for binding affinity; (3) coupling of docking and protein structure prediction algorithms to construct an efficient and robust methodology for binding mode prediction in cases where protein flexibility must be incorporated into the docking procedure, i.e. when induced fit effects are important.

Our new theory of molecular recognition is based on a concept which we term hydrophobic enclosure. The central idea is that a group of lipophilic ligand atoms which is surrounded on two sides by lipophilic protein atoms will yield a significantly larger contribution to binding affinity than is predicted by simply atom-atom pair hydrophobicity models. Examples from the medicinal chemistry literature will be used to graphically illustrate this idea. Furthermore, validation of the model has been carried out by calculating the binding affinity of a large number of protein-ligand complexes in the PDB, as well as the generation of enrichment curves for a substantial number of diverse, pharmaceutically interesting receptors. A similar analysis can be applied to hydrogen bonding, and "special" hydrogen bonds, which make exceptionally large contributions to binding affinity, can be identified. The hydrogen bonding terms require hydrophobic enclosure of ligand atoms surrounding the hydrogen bonding group. Finally, a set of empirical rules have been developed which assess when electrostatic interactions favor binding, based on multiple examples of ligands binding to pharmaceutically relevant targets with important charge-charge interactions with ligands in the active site.

Finally, our new induced fit methodology will be described and illustrated with a critical mass of examples, and its deployment in the workflow for computationally driven structure based drug design will be discussed. A number of advanced examples of application, e.g. to modeling of the HERG ion channel and to cytochrome P450, will be presented.

S4

Computer-Assisted Drug Design (CADD) at Bristol-Myers Squibb

S. Roy Kimura

Bristol-Myers Squibb Company

An overview of CADD methodologies applied to active drug development programs at BMS will be presented. High-throughput methods routinely used include ligand- and structure-based virtual screens and QSAR modeling. Lower throughput, higher-accuracy techniques are also often used: these include rational drug design or lead optimization via construction of detailed binding models. Additional low throughput methods include quantum chemical analysis of ligands, and relative binding free-energy calculations (via Linear Response or MMPB/SA). CADD is also involved in more exploratory applied research; these include, for example, improving the quality and utility of homology models through molecular dynamics simulation and developing better docking protocols. A brief critique of current methodologies and possible future directions (e.g., QM/MM methods) is presented in the context of the trade-off between cost and accuracy.

S5

From Genome Sequence to Protein Structure and Drug Discovery: computational pipeline with bioinformatics and supercomputing

Yutaka Akiyama
Director, Computational Biology Research Center (CBRC)
National Institute of Advanced Industrial Science and Technology (AIST), Japan

The Computational Biology Research Center (CBRC) is engaged in a variety of bioinformatics research activities covering such themes as automatic gene finding, expression analysis, gene regulatory network estimation, protein structure prediction, protein-protein docking, and virtual screening against chemical compound databases.

We are conducting in-depth research using large-scale computing resources such as, the Magi cluster system (1040 processors, Pentium III 933MHz), the AIST super cluster system (2048 processors, Opteron 2GHz, plus other 1000 processors), and the Blue Protein system (8192 processors, Blue Gene p440-based 700MHz, 22TFLOPS peak).

We have applied our large-scale parallel computing techniques in our genome research including systematic search for human GPCR genes (Suwa *et.al*) and genome-level annotation for micro-organism like *Aspergillus Oryzae* (Asai *et.al*). Those required extensive calculation for genome sequence parsing with large hidden Markov models.

One of our major research highlights is the development of FORTE (Tomii *et.al*), a system for predicting three dimensional protein structures from amino acid sequences by applying profile-profile comparisons (*Bioinformatics*, 20, 4, 2004). This method was devised to detect similarities in three-dimensional structures of compared proteins, even if they have no clear sequence similarities. With this system, CBRC participated in the Sixth Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction (CASP6) in 2004. The CBRC-3D team placed the third position among more than 200 prediction teams in “FR/H (fold recognition)” category. We have been also participating the CAPRI competition for protein-protein docking prediction. CBRC obtained top-level results among 37 teams in CAPRI7. A model selection scheme, similar to the one in CASP, was devised for docking (Hirokawa, *et.al*).

We have also developed CoLBA (Comparative Ligand Binding Analysis) system for efficient virtual screening procedure (Hirokawa, *et.al*). Using this scheme, we can utilize even a theoretical (predicted) protein structure as a docking target, practically.

We believe that bioinformatics techniques are getting matured with the help of supercomputing and soon we can build an automatic computational pipeline from genome sequence to protein structure and protein-protein or protein-compound docking.

S6

Strategies for Improved Vaccine Design

Nikolai Petrovsky

Australian Centre for Adjuvant Research, Flinders Medical Centre

Many challenges remain in the design of modern day vaccines. These include the identification of immunogenic and yet safe vaccine antigens, the problem that pure recombinant antigens are generally far less immunogenic than whole killed vaccines and the time taken for design and validation of new vaccines. This has created a major need to develop new strategies for vaccine development. Computer-aided design has potential to transform vaccine research. Similarly, newer more potent adjuvants are helping increase the immogenicity of subunit vaccines. In light of new vaccine technologies, adjuvants are needed for use with mucosal vaccines, DNA vaccines, cancer and autoimmmunity vaccines. Each of these areas are highly specialised with their own unique needs in respect of suitable adjuvant technology. This paper examines key areas in particular immunomics and adjuvant design technologies that are transforming the vaccine development field. Finally it will also address some of the remaining impediments and barriers to development and registration of new human adjuvants for infectious diseases, allergy, autoimmunity and cancer.

S7

Beyond Epitope Mapping: Making Vaccines and Modifying Therapeutics

Annie De Groot,
Bill Martin, Luisa Marcon, Paul Knopf, Betty Bishop, Julie McMurry, Dan Rivera, Michelle Klutzer, Christine Boren, Claire_Rodriguez-Annon and David Weiner

TB/HIV Research Laboratory, University of Pennsylvania, and EpiVax, Inc.

http:://www.Brown.edu/Research/TB-HIV_Lab; http://www.EpiVax.com

We are in the process of making epitope-driven vaccines for TB, HIV, HPV, EBV, Smallpox, Tularemia, and *Helicobacter pylori* vaccines using, as a point of departure, our immunoinformatics tools EpiMatrix, ClustiMer, Conservatrix, EpiAssembler, and VaccineCAD. We are also using these tools to alter protein therapeutics such as alpha interferon and beta interferon, which are used for the treatment of HCV and multiple sclerosis, respectively.

We believe that this approach to developing vaccines and therapeutics is justified because immune response to a limited set of epitopes derived from proteins, whether "self" or from an infectious pathogen, may be sufficient for competent protection. Clearly, this is the case for some vaccines against pathogens such as Hepatitis B. Immune recognition of every potential epitope derived from the pathogen's genome does not appear to be required for protection from infection and disease, since vaccination with the HbSAg of HBV and not the entire virus, is almost always protective. Similarly, "immunome-derived vaccines" are based on the concept that immune response to a subset of epitopes that interface with host immune system (the immunome), and not the whole organism (represented by the proteome, or genome) can be sufficient for protection. Competent immune responses to cancer are also probably restricted to the neoplasm's "immunome" although the set of antigens that drive successful immune response to cancer cells has proven more difficult to uncover.

We have been using bioinformatics sequence analysis tools, epitope-mapping tools, micro arrays and high-throughput immunology assays to discover the components of the immunomes, which are then used to compose new vaccines. We have successfully generated de novo immune responses to several of our new immunome-derived vaccines in HLA transgenic mice, and we are currently testing them in challenge models. My presentation will review the work of my laboratory and the work of other researchers who are applying immunoinformatics tools to making immunome-derived vaccines.

S8

Automatic construction of specialized databases, exemplified by the VBASE2 database, an integrative database for Immunoglobulin Variable Genes

Werner Müller

Department of Experimental Immunology, GBF German Research Centre for Biotechnology, Braunschweig, Germany

VBASE2 is an integrative database of germ-line V genes from the immunoglobulin loci of human and mouse. It presents V gene sequences extracted from the EMBL nucleotide sequence database and Ensembl together with links to the respective source sequences. Based on the properties of the source sequences, V genes are classified into 3 different classes, namely class 1=genomic and rearranged evidence; class 2=genomic evidence only and class 3=rearranged evidence only.

References to other immunological databases (KABAT, IMGT/LIGM and VBASE) are given to provide all public annotation data for each V gene. The VBASE2 database can be accessed either by the Direct Query interface or by the DNAPLOT Query interface. The Sequences given by the user are aligned with DNAPLOT against the VBASE2 database. Direct Query allows to enter sequence IDs and names (Field 1), choose species, locus, V gene family and class (Field 2) or search for 100% sequences (Field 3). At the DNAPLOT Query, a sequence given by the user is aligned with DNAPLOT against the VBASE2 database. The DNAPLOT program offers V gene nucleotide sequence alignment referring to the IMGT V gene unique numbering.

The VBASE2 database was used to annotate the proximal part of the immunoglobulin heavy chain locus (IgH) of the 129/Sv mouse strain. 117 variable genes (V genes) and pseudo genes have been identified, representing 12 of 15 murine IgH-V gene families. 47 V genes are supposed to be functional. Of that, 27 sequences are described in the germline configuration for the first time. Furthermore, 2 new D segments have been identified. The order of D segments in the 129/Sv strain is DFl16.3, DST4.2, DFl16.1, DSP2.9, DSP2.3, DST4.3, DFl16.2, DSP2.5, DSP2.2, DSP2.11, DSP2.2, DSP2.8, DSP2.7, DST4, DQ52.

The availability of 1.5 Mb genomic sequence of the IgH locus allows the comparison between different mouse strains on the level of the coding sequence as well as on the level of the locus' structure. The 129/Sv mouse has been assigned to the IgHa haplotype. Unexpectedly, the comparison of the D region, J region and constant region genes revealed differences to the BALB/c strain which also has the IgHa haplotye. The comparison of the two IgHa strains to the C57BL/6 strain of the IgHb haplotype will give insight to the evolution of immunoglobulin variable gene segments. The genomic sequences are available from the EMBL database by the accession numbers AJ851868-AJ851885.

S9

Transcriptional and translational variation from an immunomics point of view

Christian Schönbach

Immunoinformatics Team, Advanced Genome Information Technology Group RIKEN Genomic Sciences Center, Yokohama, Kanagawa 230-0045, Japan

The immune system is much more complex than the sum of its parts. Computational and functional genomics approaches have revealed some gene-to-phenotype connections and how certain parts work together on cellular level. However a large number of protein-coding and non-coding transcripts in mouse and human are of unknown functions. Apart from filling the knowledge gaps in the parts list we need to elucidate the mechanisms of interaction and regulation. Alternative splicing, translation initiation and post-translational mechanisms contribute to the complexity of the immune system and have relevance for non-mendelian etiologies of diseases. This presentation will highlight some of the results and potential non-canonical mechanisms of transcriptional and translational variations in context of increased functional complexity.

S10

Pipeline pharmacogenetics: Phase I, Phase II, & Phase III & Surveillance

Allen D. Roses, MD

GlaxoSmithKline

Careful clinical phenotyping is the basis for differential diagnoses as well as clinical responses to treatment. Quite simply, the patient phenotype is made up of the history, physical exam, and laboratory tests. Disease diagnosis, disease prognosis, or response of disease to treatment [surrogate markers] are all fundamentally focused on the patient's clinical status – with effects of treatments set in that framework. The term "biomarkers" is generally used to describe laboratory-based measurements related to clinical phenotypes [diagnostic, prognostic or surrogate biomarkers]. Regulatory agencies are accountable for evaluating the risk/benefit ratio of new medicines, that is, the safety versus the efficacy of a treatment. This is an exercise that is based fundamentally on the most discriminating clinical information.

Chronic Obstructive Pulmonary Disease [COPD] will be used as an example to demonstrate the heterogeneity of clinical phenotypes in evaluating efficacy of medicines. COPD is a syndrome of multiple phenotypes that are defined by a non-sensitive and non-specific "diagnostic" biomarker called FEV1 [forced expiratory volume in 1 second] as a measurement of EFL [expiratory flow limitation]. The standard for registration of a therapy is improvement of patients' EFL impairment as measured by FEV1. However, there are major clinical variables segment patient populations into different pathologies, with potentially different anatomical and physiologic capacities to respond. This heterogeneity is a critical issue, as there can be a wide variety of pathologic state associate with impaired FEV1. Patients can have different disease processes, different severity of the same process, or quite distinct physical capabilities when viewed through other measures of phenotype. COPD provides an excellent example of defining efficacy across a variable syndrome using a response to a non-specific biomarker [FEV1]. Examples of variable phenotypes will be illustrated.

Can diagnostic biomarkers segment the clinical population of patients with impaired expiratory flow into groups that are more likely to respond to a particular intervention? High resolution CT scanning represents a diagnostic method that allows characterization of anatomical heterogeneity: having the capacity to segment a population of COPD patients with destructive anatomical features ["bubble lung"] from those patients with preserved lung parenchyma – <u>even if patients have the same impaired FEV1 measurement.</u>

"Response to treatment" also provides prognostic information for COPD patients - those who respond with benefit to treatment have a better prognosis. Yet we measure therapeutic response for registration purposes solely on the primary measurement of FEV1, despite the existence of other more sensitive and specific measures of diagnostic and physiologic responses. Specific airway conductance [SAC] is another clinical method that can provide additional information regarding quantitative changes in patient's clinical phenotype that are more sensitive than FEV1. This is a specialized technique that is currently being evaluated as a surrogate endpoint for FEV1 in bioequivalence studies in Japan.

Hyperpolarized 3-Helium MRI represents a relatively simple method to measure lung functions both anatomically and physiologically. Images provided by this technique can be used for quantitative measures of anatomical status and physiological function over time. 3-He MRI can be used to examine disease progression as well as response to treatment in real time. Exquisite quantitative measures are possible to both classify patients disease severity as well as to measure effects of treatments.

Scientists are generally familiar with the use of genetic information for diagnosis of a "disease." When the disease is actually a syndrome of multiple phenotypes, it is still possible to apply genetic methods – as long as the different sub-types are clearly defined. When patient groups are carefully segmented it is possible to examine predictors of diagnosis, prognosis, and drug response using genome-based profiles. These profiles might include polymorphisms of complex susceptibility genes or empirical profiles of DNA markers that have not previously been known to be **<u>associated</u>** with differentiated phenotypes. In the context of efficacy pharmacogenetics the most critical element is still the definition of the phenotype to describe the effect of the drug before and after treatment. Therefore, for maximum specificity it is useful to provide the most complete phenotypic measurements that are possible to document. The added advantage of genetic profiling is that a standardized set of population variants can be assessed in each patient rapidly, and the **<u>association</u>** of specific markers with phenotypes, or changes in phenotypes, be measured quantitatively. It therefore becomes possible to define the segment of the COPD population most likely to respond to a treatment modality. In addition, of course, there is the added advantage of being able to characterize patients who experience adverse events using the same genome-wide representative genetic panels.

Recent examples of pipeline pharmacogenetic analyses in Phase 1, IIA, IIB and III will be illustrated. In these cases, the phenotype being investigated is either efficacy or adverse events. The use of genome-wide SNP screening panels and its applicability to multiple ethnic populations will also be demonstrated and discussed.

S11

High throughput biology: from genome diversity to transcriptional regulation

Hiroyuki Aburatani

Genome Science Division, Laboratory for Systems Biology and Medicine,
Research Center for Advanced Science and Technology, University of Tokyo

As genome sequences of various organisms have become available one after another, the high-density oligonucleotide array has become a very powerful analytical platform for functional genomics, besides the regular gene expression and genotyping analysis. A single array with 5 μm feature size, which generates signal intensity data from more than 6 million probes, would provide us with various novel research applications. In this seminar I present several examples for array-based high throughput analyses of human genome functions.

The genotyping array has been already utilized to analyze allelic gene dosage and copy number polymorphisms (CNPs). We have recently developed a novel algorithm called Genome Imbalance map (GIM) to determine the allelic gene dosage using genotyping array, and applied it for molecular karyotyping of cancer cells. Even a tiny homozygous chromosomal deletion has been detected with high accuracy, as well as genomic amplification. Integrated with genotyping information, which are obtained simultaneously in a single assay, a status of uniparental disomy is precisely detected, which has not been accomplished by the conventional BAC array analysis. Moreover, frequency and variation of CNPs will be elucidated using copy number analysis by genotyping array. The genomic tiling array has been applied not only for novel transcript discovery, but also for identification of transcriptional factor binding sites and epigenomic analysis, like DNA methylation detection and characterization of chromatin modification, through ChIP (chromatin immunoprecipitaion) on Chip analysis. The data analysis and integration will be a big challenge for computational biology.

References

Ishikawa S, Komura D, Tsuji S, et al. Allelic dosage analysis with genotyping microarrays. **Biochem Biophys Res Commun.** 333(4):1309-1314, 2005

Ge X, Yamamoto S, Tsutsumi S, et al. Interpreting expression profiles of cancers by genome-wide survey of breadth-of-expression in normal tissues. **Genomics** in press

S12

System-level identification of Mammalian Circadian Clocks.

Hiroki R. Ueda

Team Leader, Laboratory for Systems Biology and Subunit Leader, Functional Genomics Subunit, Center for Developmental Biology, RIKEN, Kobe Japan.
Visiting Professor, Tohoku University.
Visiting Professor, Tokushima University.

Increasing number of genome sequences have been determined in various model organisms from E. coli to H. sapiens. These genome projects are driving a paradigm shift in life science, from the molecular level to the system level, and accelerating innovations based on genome-scale resources and information. Compared to the increasing needs for system-level understanding of biological phenomena using genome-scale research infrastructure, there are few studies fully utilizing genome-scale research infrastructure toward systematic identification of genetic networks underlying complex and dynamic biological phenomena, especially in mammals. Main difficulty in full exploitation of genome-scale research infrastructure is seamless integration of both computational and experimental technologies.

To address these issues, we attempt to combine both computational and experimental technologies to develop system identification strategy toward the system-level understanding of biological phenomena and apply them for identification of the genetic network underlying complex and dynamic biological phenomena. In this symposium, we will introduce the outline of system identification strategy including (i) genome-wide gene expression analysis, (ii) genome-wide promoter analysis using mammalian promoter database, (iii) in vitro real-time monitoring of gene expression (iii) high-throughput screening of regulators, (iv) in vitro phenotype analysis.

The power of this strategy can be demonstrated by its concrete applications for the specific biological systems. We applied this strategy to the system identification of transcriptional networks composed of 16 transcription factors underlying mammalian circadian rhythms, and revealed the complex transcriptional circuits governed by simple design principles, and topologically identified Achilles' heel of mammalian circadian clocks, which was functionally verified using a newly developed in vitro system. In this symposium, we will also report current progresses on system-level identification of mammalian circadian clocks.

References:

Hiroki R. Ueda et al. "A transcription factor response element for gene expression during circadian night" Nature, 418, 534-539, 2002.

Hiroki R. Ueda et al. "Molecular-Timetable Methods for Detection of Body Time and Rhythm Disorders from Single-time-point Genome-wide Expression Profiles" PNAS, 101, 11227-11232, 2004.

Hiroki R. Ueda et al. "System-level Identification of Transcriptional Circuits Underlying Mammalian Circadian Clocks" Nature Genetics, 37, 187-192, 2005.

S13

MC's Comment: The Translational Research Informatics: A bridge toward the Genome Medicine

Jun Nakaya

Associate Professor, Clinical Genome Informatics Center, School of Medicine, Kobe University

In the post genomic era, the bridging between the genome science and the clinical medicine is the fundamental demand. The Translational Research Informatics (TRI) is the informatics that tries to support the translational research, to establish bridging through information, and to facilitate the translational research. The translational research is the transitional bridging research of the basic research and the clinical research. It covers not only the drug development but also the medical device development and the novel medical methodology. From the informatics side of view, the TRI can be located as the interdisciplinary informatics of the bioinformatics and the medical informatics. Essentially it encompasses bridging these two informatics and their integration. Supporting the translational research from the informatics-side needs the following informational terms according to the results compiled from NIH and our analysis:

1. The prediction tools for the patient's future clinical situation
2. Accumulation of the cause-and-effect cases of the translational research
3. Bioinformatics Tools for the discovery, the simulation, and the prediction
4. The clinical informatics tools for the data transition
5. The clinical informatics tools to defend the patient's security without losing informational accuracy
6. Bilateral knowledge translation methodology between the basic researchers and the clinicians
7. Integrated database of the clinical data and the omics data
8. Integrated and standardized data exchanging format
9. The Standardized data sharing technology with keeping privacy

In the translational research field, the technologies derived from these terms must be evaluated by the three criteria. The most important criterion is the safety of the patient, the secondary important criterion is the clinical efficiency, and the thirdly important criterion is the industrialization. From a view of these three criteria, the translational research path is the most critical path in the entire developmental path between the basic research and the clinical practice. This means that the translational research path is the key pathway to optimize the drug development and the other medical developmental processes. According to the results of the NIH and our analysis, the most effective issue to achieve this optimization is to establish the accurate prediction of the future clinical state of the patients during the translational research phase as indicated in term 1. Here the term 2 is fundamental and essential issue to establish the term 1. The role of the translational research informatics is to make the medical research more safe, comfortable, understandable, efficient, and cost-effective through supporting the translational research with the informational technology. The TRI can be a bridge to establish the Genome Medicine and the other future Medicines.

S14

Omics-based systems approach to diseases
-A new paradigm for clinical medicine in the post-genomic era-

Hiroshi Tanaka

Department of Bioinformatics, Tokyo Medical and Dental University

In spite of the rapid increase of the bio-molecular (“omics”) data such as genome, transcriptome, proteome and so forth, the substantial application of omics data to clinical medicine has not been realized yet. This might be ascribed to the fact that there has neither been explored nor established the paradigm or systemic way where the omics data are incorporated into and efficiently utilized in clinical medicine.

We have just started the national projects for comprehensive omics-based disease case database. In constructing the database, we adopted a new paradigm for viewing the disease as a unified system, in which the clinical-pathophysiological-omic hierarchy is involved. The concrete realization of this paradigm is explored to be established as a comprehensive data scheme of clinico-omic database. Our leading concept for developing database is that, “diseases self-organize into a unified system”, we call this paradigm “omics-based systems pathology”. In our approach, first, we collect the molecular, pathological and clinical data to store into two hierarchical data models, which we call the preliminary disease database; one of these two data hierarchies is traditional one which is comprised of clinical symptomatology, pathophysiology, etiology and the other is omics data hierarchy composed of genome, transcriptome, proteome and their related pathway (signaling, gene-regulatory, and metabolic ones).

The relations between these two categories of data are explored by applying the data-mining method and correlation analysis. After establishing the inter-hierarchical relation between the clinical-pathophysiological-omic data, we integrate these data into the unified hierarchical disease model, which we developed for each of generic class of diseases, such as infectious disease, cancer, neurological disease, common metabolic disease, autoimmune disease. With this structured disease case database as a central part, we are developing the user-directed functions such as deductive retrieval which replies various inquiries based on the structured clinico-omic relation, for example, given the clinical syndrome, infer the possible molecular process of disease, or given the omics data, predict the prognosis of disease, etc. In the first term of this project the disease is confined to cancers (hepatic, colon and oral). In constructing this omics-based system pathology database, we aim to establish new clinical medicine integrating clinical and omics data and specialized for individual patient.

S15

Future Vision: A possibility of Genome Information toward the Personalized Medicine and Personalized Healthcare.

Hiroshi Mizushima

Head of Bioinformatics Section. Center for Medical Genomics. National Cancer Center Research Institute.

Human Genome Sequencing has been completed and personalized medicine and medical-care is the next scope of research and development. Many researches have been done collecting medical information and biological or genomic information, and analyzing for association study finding Genome/Proteome - Phenome relationships.
Recently, Nutritigenomics is another field, to find out the relations between nutrition and its effect over genomic background. State of the art research is now active on discovering relationship between these phenotypic information and genomic or biological information.
In this talk, I want to list up some of the examples of these relationships, and possible future prospects using genomic information not only for medical care but also for health care to keep healthy life.
I will also discuss about the security issues and ethical guidelines for handling the personal privacy information such as medical record and genome information, along with necessity of standardizing related information for data exchange.

S16

Role of drug transporter studies in the drug discovery and development; Molecular multiplicity, substrate specificity and drug-drug interaction

Yuichi Sugiyama

Department of Molecular Pharmacokinetics, Graduate School of Pharmaceutical Sciences, The University of Tokyo

Drug transporters are expressed in many tissues, such as the intestine, liver, kidney, and the brain, and play key roles in drug absorption, distribution and excretion. The information on the functional characteristics of drug transporters provides important information to allow improvements in drug delivery or drug design by targeting specific transporter proteins. In this presentation, I will summarize the significant role played by drug transporters in drug disposition, focusing particularly on their potential use during the drug discovery and development process. The use of transporter function offers the possibility of delivering a drug to the target organ, avoiding distribution to other organs (thereby reducing the chance of toxic side-effects), controlling the elimination process, and/or improving oral bioavailability. It is useful to select a lead compound that may or may not interact with transporters, depending on whether such an interaction is desirable. The expression system of transporters is an efficient tool for screening the activity of individual transport processes. The changes in pharmacokinetics due to genetic polymorphisms and drug-drug interactions involving transporters can often have a direct and adverse effect on the therapeutic safety and efficacy of many important drugs. In future, the *in silico* prediction of the affinity of drug candidates for the transporters may become possible. We have also constructed a comprehensive database for membrane transport proteins called "TP-Search" (URL: http://www.TP-Search.jp/). Each record has been extracted from approx. 3500 articles published from 1968 to the present. TP-Search contains information about more than 90 transporters in humans and rodents, including substrate/inhibitor/inducers, tissue distribution, pathophysiology, knockout animals, gender differences and drug-drug interactions. The database also gives a brief description of the experimental methods and all information available in this database is linked to the original references in PubMed.

(References) 1) Mizuno N., Niwa T., Yotsumoto Y. and Sugiyama Y.: *Pharmacol Rev*. 55: 425-461 (2003), 2) Shitara Y., Sato H. and Sugiyama Y.: Evaluation of drug-drug interaction in the hepatobiliary and renal transport of drugs. *Annu Rev Pharmacol Toxicol* 45: 689-723 (2005), 3) Nishizato Y., Ieiri I., Suzuki H., Kimura M., Kawabata K., Hirota T., Takane H., Irie S., Kusuhara H., Urasaki Y., Urae A., Higuchi S., Otsubo K. and Sugiyama Y. :Polymorphisms of OATP-C (SLC21A6) and OAT3 (SLC22A8) genes: consequences for pravastatin pharmacokinetics. *Clin Pharmacol Ther* 73: 554-65 (2003), 4) Sasaki M., Suzuki H., Ito K., Abe T. and Sugiyama Y.: Transcellular transport of organic anions across a double-transfected Madin-Darby canine kidney II cell monolayer expressing both human organic anion-transporting polypeptide (OATP2/SLC21A6) and Multidrug resistance-associated protein 2 (MRP2/ABCC2). *J Biol Chem* 277, 6497-503 (2002)

P6

Structural Analysis and Discrimination of Outer Membrane Proteins

M. Michael Gromiha and Makiko Suwa

Computational Biology Research Center (CBRC), AIST
2-42 Aomi, Koto-ku, Tokyo 135-0064, Japan

Fishing outer membrane proteins (OMPs) from genomic sequences is one of the challenging problems in Bioinformatics. We have systematically analyzed the characteristic features of the 20 amino acid residues in globular and OMPs and we found that the residues, Glu, His, Ile, Cys, Gln, Asn and Ser show a significant difference between the amino acid compositions of globular and outer membrane proteins. The higher occurrence of Ser, Asn and Gln in OMPs than globular proteins might be due to their importance in the formation of β-barrel structures in the membrane, stability of binding pockets and the function of OMPs.

We have devised a statistical method for discriminating OMPs from other globular and membrane proteins. Our approach correctly picked up the OMPs with an accuracy of 89% for the training set of 337 proteins. On the other hand, our method has correctly excluded the globular proteins at an accuracy of 79% in a non-redundant dataset of 674 proteins. Furthermore, the present method is able to correctly exclude α-helical membrane proteins up to an accuracy of 80%. The method based on dipeptide composition also showed similar results.

The membrane spanning segments in OMPs have been predicted using neural network based methods. We have developed a web interface for discriminating OMPs and predicting the membrane spanning segments in OMPs. It is freely available at http://psfs.cbrc.jp/tmbeta-net/.

References:

(1) M.M. Gromiha and M. Suwa (2005) *Bioinformatics* **21**, 961-968.

(2) M.M. Gromiha et al. (2005) *Nucleic Acids Res.* **(in press)**.

P7

Molecular Dynamics Simulations of HIV-1 Proteases – Insight into the Multi-drug Resistance conferred by Non-active Site Mutation L90M -

Hirotaka Ode[1], Sabrou Neya[1], Masayuki Hata[1], Wataru Sugiura[2], Tyuji Hoshino[1,3]
Graduate School of Pharmaceutical Sciences, Chiba University[1], National Institute of Infectious Diseases[2] , JST[3]

Abstract

The infection by human immunodeficiency virus type 1 (HIV-1) essentially requires the enzymatic activity of virally encoded protease, so called HIV-1protease (HIV-1 PR). This protease is an attractive target for AIDS treatment. Hence, various protease inhibiters (PIs) have been designed, and now, seven PIs are clinically available in JAPAN. However, the combination of a short retroviral life cycle and the high mutation rate leads to the selection of drug-resistant HIV-1 variants. Among the patients who failed in treatment with each SQV or NFV, which is one of the PIs, L90M mutation is often seen in HIV-1 PR. L90M is a primary mutation responsible for the resistance to both SQV and NFV. In addition, it appeared to be associated with the other PIs. But, a crystal structure of PR reveals that the amino acid at codon 90 is located at the non-active site, which has no direct interactions with substrates or inhibitors. It is difficult to understand how the non-active site mutation L90M affects the inhibitor binding. In this study, we executed molecular dynamics simulations and investigated the resistant mechanism by L90M mutation. We simulated the HIV-1 PRs complexed not only with the inhibitors (SQV, NFV, LPV) but also with two substrates, using AMBER7 program package. The simulations revealed that non-active site mutation L90M affected the binding with SQV and NFV. L90M caused conformational changes at the active site of PR-SQV or PR-NFV complex, especially on the environment of the residues at 80 loop. In addition, L90M decreased the binding energies between PR and these inhibitors. Whereas, L90M hardly affected the binding with another potent inhibitor LPV and with the substrates of HIV-1 PR. Thus, those conformational changes at the active site might be an index for the fold resistance. These simulations also suggest that non-active site mutations could affect the active site conformations of ligand-binding proteins.

References

1) Johnson V. A., et al. : Top. HIV Med. vol. 13, pp. 51-57. (2005)
2) Ode H., et al. : J. Phys. Chem. B. vol. 109, pp. 565-574. (2005)

P8

GLYMM: a program for modeling a 3D structure of a carbohydrate chain and a lipid membrane: the application to molecular dynamics simulations of glycolipid (GM1) membranes

Kenichi Mori[1], Tsubasa Takaoka[1], Koji Iwamoto[1], Masayuki Hata[1], Saburo Neya[1], Tyuji Hoshino[1, 2]

Graduate School of Pharmaceutical Science, Chiba University [1], PRESTO, JST[2]

We have developed a computer program called GLYMM (GLYcan and Membrane Modeling) that can construct a 3D structure of a carbohydrate chain and a lipid membrane for modeling a lipid microdomain such as a lipid raft. A module for modeling a carbohydrate requires only a sequence that can be obtained from the database GLYCAN in the GenomeNet server at Kyoto University, which enables us to make the 3D structure easily and quickly. A module for modeling a lipid membrane is a GUI-based program, in which only to click and drag your mouse button are needed for making a membrane model, however complex the membrane is.

To apply GLYMM to pharmaceutical research, we selected the aggregation mechanism of Alzheimer β-peptides (Aβ) on glycolipid bilayers. Aβs are known to aggregate on GM1 (glycolipid) including membranes, and the aggregation is enhanced as the concentration of cholesterol in the membrane increases. We made two types of membranes including GM1, POPC, sphingomyelin (SM), and cholesterol (CHL) by GLYMM, and performed molecular dynamics simulations of these membranes. Interestingly, we observed microdomains consisting of the GM1 molecules on the membrane where the ratio of GM1:SM:CHL was 20:40:40, while the GM1 molecules were scattered on the other membrane where the ratio of GM1:POPC was 20:80. Furthermore, we found that the GM1 molecules in the cluster form were located at intervals of 0.7 or 1.4 nm, which was compatible with the intervals of the appearance of the hydrogen bond donors or acceptors of Aβ. Because a GM1 has many hydrogen donors or acceptors, our results suggest that GM1 clusters would become a template for hydrogen bonds for interacting with Aβ, which would be a seed for Aβ aggregation.

REFERENCES

1. Kakio, A, Nishimoto, SI, Yanagisawa, K, Kozutsumi, Y, Matsuzaki, K.: Cholesterol-dependent formation of GM1 ganglioside-bound amyloid beta-protein, an endogenous seed for Alzheimer amyloid, J. Biol. Chem. vol. 276, no. 27, pp. 24985-90 (2002)

P9

Computational Chemistry Platform for Drug Metabolomics

Michihisa KOYAMA[1], Satoshi KAWAGOE[1], Kota KASAHARA[1], Hideyuki TSUBOI[1], Akira ENDOU[2], Momoji KUBO[1, 3], Carlos A. DEL CARPIO[1], Ewa BROCLAWIK[1], Kazumi NISHIJIMA[4, 5], Tetsuya TERASAKI[4, 6], Akira MIYAMOTO[4, 1]

Department of Applied Chemistry, Tohoku University[1], Institute of Fluid Science, Tohoku University[2], PRESTO, Japan Science and Technology Agency[3], New Industrial Creation Hatchery Center[4], Development Division, Mochida Pharmaceutical Co. Ltd.[5], Graduate School of Pharmaceutical Sciences, Tohoku University[6]

To understand metabolic properties of drug in human bodies at the early stage of the design process is important for predicting clinical characteristics of drug or avoiding later stage attrition. Expectations toward *in silico* methods are greatly increasing in the current post-genome era. In order to predict the uninvestigated metabolic properties of newly designed drug molecules and to deal with the chemical reaction involved in the metabolic process, computational chemistry method based on quantum chemistry is necessary. In this study we newly propose a computational chemistry platform for drug metabolomics, which can be defined as a metabolomics study specific to drug metabolism. Computational chemistry platform here consists of various original computational chemistry programs developed in our group. Fig. 1 schematically illustrates the computational chemistry platform for drug metabolomics. Metabolic enzyme that explains main pathway to metabolite is predicted by combining Monte Carlo method and chemical engineering modeling. By adopting interacting structure predicted by Monte Carlo method as an initial structure, hybrid tight-binding quantum chemical molecular dynamics calculations are carried out to predict metabolites of the target drug molecules. Our platform realizes the prediction of metabolic properties of drug molecules in the early stage of drug design process and will change the paradigm in the current drug design process. Details of the platform will be explained together with our future visions.

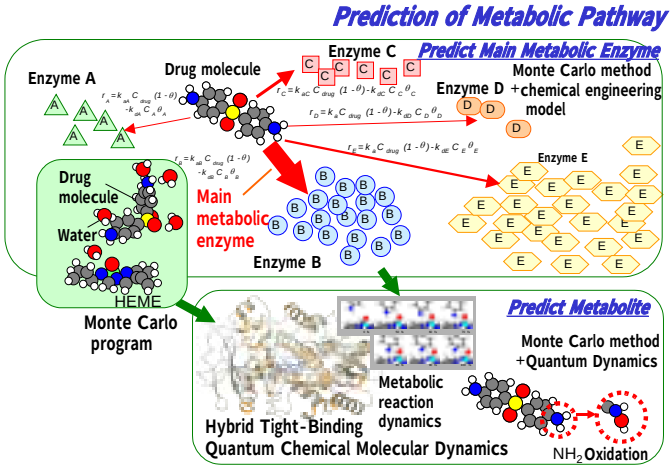


Fig. 1 Schematic of computational chemistry platform for drug metabolomics

P10

Theoretical Study on the Mechanism of Theophylline Metabolism by Compound I of Cytochrome P450

Mohamed Ismael[1], Abdul Rajjak Shaikh[1], Hideyuki Tsuboi[1], Michihisa Koyama[1], Akira Endou[2], Momoji Kubo[1,3], Carlos A. Del Carpio[1], Ewa Broclawik[4], Kazumi Nishijima[4,5], Tetsuya Terasaki[4,6], Akira Miyamoto[1,4]

Department of Applied Chemistry, Tohoku University[1], Institute of Fluid Science, Tohoku University[2], JST-PRESTO[3], New Industry Creation Hatchery Center, Tohoku University[4], Development Division, Mochida Pharmaceutical Co., Ltd.[5], Graduate School of Pharmaceutical Sciences, Tohoku University[6]

Theophylline is a xanthine derivative methylated at N-1, N-3 positions. Theophylline has a prominent place in the treatment of asthma, also it was found that theophylline inhibits the progression of neuroendocrine lung carcinogenesis in hamsters with hyperoxic lung injury[1]. Upon oxidation by hepatic microsomal oxidases, the theophylline is N-demethylated at N-1, and N-3 tertiary amine nitrogen atoms and is also hydroxylated at C-8 position to produce 3-methylxanthine, 1-methylxanthine, and 1,3-dimethyluric acid, respectively. Among these theophylline metabolites, two main enzymes are responsible. The demethylation is known to be catalyzed specifically by the CYP1A2 enzyme and the hydroxylation is catalyzed by the CYP2E1 enzyme. In this paper, we describe the C-8 hydroxylation of theophylline by a compound I of CYP450 from DFT calculations. We present new aspects of this enzymatic reaction, in particular how a C-H bond of the substrate is activated and dissociated from the energetic point of view. The calculated spin density distribution shown in Table 1 demonstrates that the Fe=O moiety of the compound I has two unpaired parallel spin electrons, which are weakly coupled with one unit of spin mainly located on the thiolate ligand and partly located on the porphyrin ring. In the initial stages of the reaction theophylline comes into contact with compound I to form a kind of encounter complex where the face of the C8 atom of theophylline is oriented toward the iron atom and close to the oxygen atom of compound I. This suggests that the hydrogen abstraction mechanism produces 1,3-dimethyluric acid.

Table 1. Calculated Mulliken charges and spin densities for the Fe and O atoms and the CH_3S and porphyrin moieties

	Fe		O		CH_3S		Porphyrin	
	Charge	Spin	Charge	Spin	Charge	Spin	Charge	Spin
Compound I	0.72	0.9	-0.34	0.8	0.00	-0.5	-0.38	-0.2
Theophylline complex	0.76	0.9	-0.43	0.7	0.00	-0.4	-0.33	-0.2
TS	0.68	0.8	-0.44	0.4	0.03	-0.2	-0.28	-0.1
Hydroxylated complex	0.59	0.7	-0.44	0.0	0.04	0.4	-0.66	0.0

REFERENCE

1. Hildegard M. Schuller, B. Porter, A. Riechert, K. Walker, R. Schmoyer, Lung Cancer, 45 (2004), 11-18.

P11

An in Silico discriminant model constructed for HERG potassium channel inhibitors

Motoi Tobita[1, 2] and Tetsuo Nishikawa[1]

1. Reverse proteomics research institute, 2-6-7 Kazusa-kamatari, Kisarazu-si, Chiba 292-0818, Japan,

2. Advanced Research Laboratory, Hitachi, Ltd. 1-280 Higashi-Koigakubo, Kokubunji-shi, Tokyo, 185-8601, Japan

Abstract

HERG attracts attention as a risk factor for arrhythmia, which might trigger *torsade de pointes*. A highly accurate classifier of chemical compounds for inhibition of the HERG potassium channel is constructed using support vector machine. For two test sets, our discriminant models achieved 90% and 95% accuracy, respectively. The classifier is even applied for the prediction of cardio vascular adverse effects to achieve about 70% accuracy. While modest inhibitors are partly characterized by properties linked to global structure of a molecule including hydrophobicity and diameter, strong inhibitors are exclusively characterized by properties linked to sub-structures of a molecule. This work was supported by a grant from NEDO project of the Ministry of Economy, Trade and Industry of Japan.

P14

Evaluation of flexible molecular docking programs for virtual screening

Kenji Onodera [1], Kazuhito Satou [1], Hiroshi Hirota [1]
Riken Genomic Sciences Center [1]

Abstract

Typically two methods, docking performance and scoring performance, are used as measurements to evaluate molecular docking programs. However, it seems that above two measurements are not suffice to evaluate programs' screening performances. Several researches have reported ranking performances for virtual screening with only some ligand-protein complexes. Since it is known that each docking program performs differently due to the nature of active sites, an evaluation based on a limited number of ligand-protein complexes may lead wrong conclusions. In addition, effects from input ligand conformations and from random tables in the docking programs are overlooked and those have not been tested. Thus, we evaluated three docking programs (DOCK, AutoDock, and GOLD) for virtual screening.

To obtain the reliability of the docking programs, more than 100 PDB registered ligand-protein complexes were tested with three input ligand conformations each. Since docking score and docked position in the same ligand-protein complex are different each time due to random number table used in these programs, calculations were repeated 1000 times each for the complex to obtain the ranges of docking scores and positions. Then, the docking scores were compared to docking scores from NCI diversity set screening results to obtain ranking ranges for the ligand-protein complexes in screening results with NCI diversity set. The results showed evaluations from the docked positions and the docking scores did not estimate the screening performance well, and ranges of scores and RMSDs differ among the docking programs when different ligand conformations were tested or the same calculations were repeated. Currently many docking programs are available and many parameters can be modified in docking programs. Thus, this evaluation method can be helpful to evaluate and improve them by finding the best parameter settings according to the test results.

References

1) Kontoyianni M, Sokol GS, McClellan LM: Evaluation of library ranking efficacy in virtual screening. J Comput Chem. 2005, 26(1):11-22.

2) Kellenberger E, Rodrigo J, Muller P, Rognan D: Comparative evaluation of eight docking tools for docking and virtual screening accuracy. Proteins. 2004, 57(2):225-242.

P15

Implementation of a Haplotype Reconstruction Algorithm and a tagSNP Selection Algorithm for Large-scale SNP Information Processing

Sang Jun Kim [1], Sang-Soo Yeo[2], Sung Kwon Kim[3]
Chung-Ang University [1, 2, 3]

The human diversity comes from mutations(esp. SNPs) that appear in human genomes. Association studies between genes and human diversities will take a lot of cost and time, if they are done by a long DNA region. To reduce cost, many researchers have studied for haplotyping using the computational methods, and tagSNP selection of a DNA region of human samples. A lot of researches solve these problems to reduce the computation time and memory consumption, but they have not biological meaning.

In this paper, we propose MarSelHR, which is a Haplotype Reconstruction system and MarSelTS that is a tagSNP selection system. The MarSelHR has three steps for haplotype reconstruction. In the first and second step, MarSelHR infer respectively haplotype pair from unambiguous genotypes and reference set registered haplotype pair. In the last step, we divide the reference set into several blocks using LD-base block partitioning approach, and apply the reconstruction algorithm to all of the ambiguous genotypes with the partitioned reference set.

MarSelTS's algorithm compute LD coefficients between each pair of SNPs, and find potential block boundary candidates, and then optimal blocks determined each potential block by a dynamic programming algorithm based on object function. Finally, it select tagSNP for optimal block by entropy based calculation.

We tested and analyzed MarSelHR's and MarSelTS's performance with Daly data , Patil data, and artificial haplotype dataset that is generated by haplotype data generating programs.

References

1) A. Clark. "Inference of haplotypes from PCR-amplified samples of diploid populations". Molecular Biology and Evolution, 7(2): 111–22, 1990.

2) K. Zhang, "Dynamic Programming Algorithms for Haplotype Block Partitioning : Applications to Human Chromosome 21 Haplotype Data", ACM RECOMB, pp. 332-340, 2003.

Acknowledgement

This work is supported by Korea Science and Engineering Foundation (R01-2003-000-11573-0)

P16

Computational study of conformational preference in the tetrahedral intermediates of the acylation step of ester hydrolysis catalyzed by lipase

Yu TAKANO[1,2] and K. N. HOUK[2]

[1]Institute for Protein Research, Osaka University

[2]Department of Chemistry and Biochemistry, University of California Los Angeles

Enzymes work as catalysts of an enormous variety of biochemical reactions forming living systems. The geometrical and physical complementarities of enzymes to transition states and tetrahedral intermediates are necessary for efficient catalysis. The interior of the active site of enzymes imposes restrictions on the conformational freedom of the transition states and the tetrahedral intermediates.

Lipases are versatile biocatalysts to catalyze hydrolysis of esters and esterification of the carboxylic acids involved in a whole range of bioconversion reactions. Moreover, because of accepting the broad range of substrate while remaining high selectivity, synthetic organic chemists have used these enzymes for regio- and/or stereoselective hydrolyses and esterifications [1]. Three-dimensional X-ray crystallographic structures of lipases from a number of different species show striking similarities to each other. In particular, the bound phosphonate inhibitors exist in only a few of the possible conformations. Conformational analysis has been performed for the tetrahedral intermediates bound to lipases using molecular dynamics (MD) simulations in order to understand how these conformational factors are related to catalysis by enzymes [2]. The conformational preference has been compared to those of the transition states and tetrahedral intermediates in the acylation step of ester hydrolysis computed with density functional theory (DFT) [3].

The distributions of conformations in the tetrahedral intermediates bound to lipases indicate that these enzymes limit the conformation of the tetrahedral intermediates. The conformational preference of the enzyme-bound tetrahedral intermediates in the acylation step is different from those of the phosphonate inhibitors bound to lipases. Comparison to the computed reaction pathways of the acylation of methyl acetate implies that the binding cavities of lipases stabilizes the methoxide dissociation step of transacylation more than the association step and probably harnesses the reaction mechanism in the gas phase for their catalysis rather than in water [2].

REFERENCES

1. R. D. Schmid, R. Verger, Angew. Chem. Int. Ed. vol. 37, pp. 1608–1633 (1998).

2. Y. Takano, K. N. Houk, in preparation.

3. Y. Takano, K. N. Houk, J. Phys. Chem. A vol. 108 pp. 11740-11751 (2004).

P17

Computational Analysis of CYP3A4 Mediated Metabolism of KCNQ2 Potassium Channel Opener

Shaikh Abdul Rajjak[1], Mohamed Ismael[1], Hideyuki Tsuboi[1], Michihisa Koyama[1], Akira Endou[2], Momoji Kubo[1,3], Carlos A. Del Carpio[1], Ewa Broclawik[4], Kazumi Nishijima[4,5], Tetsuya Terasaki[4,6], Akira Miyamoto[1,4]

Department of Applied Chemistry, Tohoku University[1], Institute of Fluid Science, Tohoku University[2], JST-PRESTO[3], NICHe, Tohoku University[4], Mochida Pharma Co., Japan[5], Graduate school of Pharmaceutical Sciences, Tohoku University, Japan[6]

The Cytochrome P450 enzymes play a central role in drug metabolism by catalyzing the biotransformation of a wide variety of xenobiotics. CYP3A4 is a major CYP450 isoform with broad substrate specificity and is estimated to be involved in the metabolism of approximately 50% drugs used in humans. A detailed understanding of its mechanism is vital for predicting biotransformation of pharmaceuticals and other xenobiotics.

DFT calculations were carried out using DMol[3]. Docking studies were carried out using LigandFit module in Cerius2. Molecular dynamics studies were performed using New-Ryudo, software developed in our laboratory.

Complexes of the oxyferryl heme model with cysteine and a ligand were optimized using accurate DFT methods. Our studies indicate that ligand forms weakly bound complex (Fig.1) and next step is the formation of the bond between aromatic ring carbon and oxyferryl oxygen. This step requires activation and involves charge transfer from the ligand to the active site, which indicates that hydroxylation may proceed predominantly through an electrophilic pathway via a cation-like σ complex. This enzymatic mechanism involves proton shuttles mediated by the porphyrine ring through the N-protonated intermediate, which directs the proton either to the oxygen atom to form alcohol or to the ring carbon atom to produce ketone product. In addition, formation of epoxide product was also investigated. Subsequently DFT calculation for isolated porphyrine and ligand structure, docking and MD studies were also carried out in protein environment. This work provides insight into the detailed mechanism and helps to explain and predict the electronic contribution of ligand substituents to reactivity in this enzymatic process.

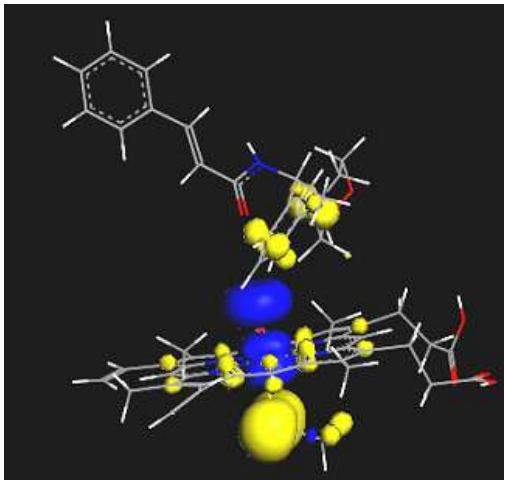


Fig. 1: Spin density distribution on weakly bound complex

REFERENCE

1) Wu YJ, Davis CD, Dworetzky S, Fitzpatrick WC, Harden D, He H, Knox RJ, Newton AE, Philip T, Polson C, Sivarao DV, Sun L, Tertyshnikova S, Weaver D, Yeola S, Zoeckler M, Sinz MW, J Med Chem vol.46, no.18, pp.3778-3781 (2003).

P18

Computational Study on the Enantioselectivity of Lipase Enzyme toward Non-Natural Organic Compounds

Yoshihiro MORI, Yoichiro YAGI, Yoshinobu NAOSHIMA
Faculty of Informatics, Okayama University of Science

Lipase enzymes coupled with their ease of use are now widely used as an environmentally friendly and useful biocatalysts for obtaining enantiomerically pure and biologically important compounds. As part of our continuous works toward utilizing biocatalysts in organic synthesis,[1)] we required a computational modeling of the interaction of two popular lipases, *Pseudomonas cepacia* lipase (PCL) and *Candida antarctica* lipase typeB (CALB) with chiral ligands, in order to explain the enantioselectivity for non-natural organic compounds shown by these two enzymes. This presentation will describe our progress in the molecular dynamics (MD) simulation of lipase enzyme-ligand interactions.[2)]

The PDB X-ray crystallographic structures of PCL (1hqd) and CALB (1lbs) were optimized by an initial preparation with Glide, followed by energy minimization using OPLS-AA force field implemented in MacroModel. Twenty chiral ligands, including aromatic and aliphatic esters, were initially minimized with OPLS-AA. Each structure thus obtained was prepared by employing a combination of molecular mechanics and MD simulated annealing technique and the subsequent Hartree-Fock 6-31G** calculations. Conformation search toward the lipase-chiral ester complexes obtained by flexible docking simulations with the Glide was carried out by Mixed Monte Carlo Multiple Minimum/Large Scale LowMode procedure. Further, we performed a 300-step and 100 ps MD simulation with the timestep of 1.5 fs for conformations possessing lower energy and shorter C-O distance (the distance in each complex between the carbonyl carbon of chiral ester and the oxygen of the active site Ser side chain OH), among the conformations found in Monte Carlo calculations. The MD simulations indicate that the C-O interatomic distance is shorter for the fast reacting (*R*)-enantiomer than for the slow reacting (*S*)-enantiomer. It is therefore envisaged that the difference of the C-O distances between two enantiomers can be related to an expression of the enantioselectivity of lipase enzymes such as PCL and CALB.

REFERENCES

1) Y. Naoshima, T. Kimura, Y. Mori, and M. Kamezawa, *Recent Res. Devel. Org. Bioorg. Chem*., **6**, 1-9 (2004). 2) Y. Naoshima, Y. Mori, and Y. Yagi, The 15 International Conference on Genome Informatics 2004, Poster and Software Demonstrations, P129-1-P129-2.

P19

Extraction of Information on Chemicals-CYP3A4 Interactions from Literature

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

Abstract

CYP3A4 is the most abundant cytochrome P450 isozyme in the human liver that is responsible for the metabolism of almost 50% of known drugs. To understand functional classification of CYP3A4 interactants and their structure-function relationship is important, in order to predict the metabolism of new chemical entities in drug discovery and drug-drug interactions in clinical applications. In this study, we performed information extraction of interaction between chemicals and CYP3A4 from MEDLINE abstracts. An available human language processing systems “GATE” is used as a platform, and two name dictionaries that involve >100,000 compounds and >30,000 enzymes respectively were constructed from MeSH datasets. In addition, 21 verbs appearing frequently in the abstracts were selected from a corpus of two thousand MEDLINE abstracts obtained by the query “CYP3A4”. Based on dictionaries, key verbs and simple syntactic analysis (based on sentence type, verb voice, coordinating conjunction, noun phrase and verb position patterns), a set of rules were constructed to extract information on chemical-CYP3A4 interactions. Validation of the rules was carried out using another corpus of randomly selected two hundred MEDLINE abstracts. More than 85% recall and 97% precision on name identification were achieved. We are now analyzing common structural features of the CYP3A4 interactants extracted.

References

1) Guengerich FP: Role of cytochrome P450 enzymes in drug-drug interactions. Adv. Pharmacol. 43:7-35, 1997.

2) Hamish C, Diana Ma, Kalina B, Valentin T, Cristian U, Marin D, Mike D, Niraj A, Ian R: Developing Language Processing Components with GATE Version 3 (a User Guide) For GATE version 3.0 (January 2005), The University of Sheffield 2001-2005.

3) Blaschke C, Andrade MA, Ouzounis C, Valencia A: Automatic extraction of biological information from scientific text:protein-protein interactions. Proc. Int. Conf. Intell. Syst. Mol. Biol. 60-7, 1999.

4) R. Gaizauskas, K. Humphreys: Using a semantic network for information extraction. Nat. Lang. Eng., 3(2/3):147-169, 1997.

P20

Spectral Analysis of the Daily Prescription Variations at a Pharmacy for Infectious and Non-Infectious Diseases

Katsunori SEGAWA[1], Tatsuya NAKANO[1], Kotoko NAKATA[1], Kazushige Ijuin[2], Noriko Hatanaka[3] and Yuzuru HAYASHI[1]

National Institute of Health Sciences[1], Tanashi Yakuhin Co. Ltd.[2], Kakunoki Pharmacy[3]

The time variation in the number of viruses in a human body as well as in a society of human being can be described by a mathematical model for the changes in the size of populations whose members can be born and die [1-3]. We analyzed the daily variations in the number of formulations prescribed at a pharmacy (Tanashi Yakuhin) from the viewpoints of spectral analysis. The power spectral density of the variations for an influenza anti-viral agent (Tamiful Dry Syrup) is observed to be right downward, whereas that for vasodilator (Norvasc) and antidiabetic agent (Daonil) is parallel. It is concluded that for infectious diseases, the prescription amount at a day is correlated with the prescription amount at another consecutive day, but for non-infectious diseases, they are independent of each other. The stochastic properties of prescriptions will be useful for the study on the prediction of the future prescription amount for the different types of diseases [4].

We also tried to examine computer viruses, which were attached to e-mails for National Institue of Health Sciences, and got an interesting result [5].

References

[1] Feller, W., An Introduction to Probability Theory and Its Applications, John Wiley ‘ Sons, New York, 1968.

[2] Hino, M. Spectral Analysis (Supekutoru Kaiseki), Asakura Shoten, Tokyo, 1982.

[3] Sumi, A., Olsen, L.F., Ohtomo, N., Tanaka, Y., Sawamura, S., Jpn. J. Appl. Phys., 42, 721-733 (2003).

[4] Ijuin, K., Hatanaka, N., Segawa, K., Nakano, T., Nakata, K. and Hayashi, Y.: Spectral Analysis of Prescriptions at a Pharmacy for Infectious and Non-Infectious Diseased. (in preparation)

[5] Segawa, K., Nakano, T., Nakata, K. and Hayashi, Y.: Computer viruses act like viruses. (in preparation)

P21

Computational study on ATP hydrolysis by ABC transporter subunit HisP

Qiang Pei[1], Hideyuki Tsuboi[1], Michihisa Koyama[1], Akira Endou[2], Momoji Kubo[1,3], Carlos A. Del Carpio[1], Ewa Broclawik[4], Kazumi Nishijima[4,5], Tetsuya Terasaki[4,6], Akira Miyamoto[1,4]

[1]Department of Applied Chemistry, Graduate School of Engineering, Tohoku University, [2]Institute of Fluid Science, Tohoku University, [3]JST-PRESTO, [4]NICHe, Tohoku University, [5]Developmen Division, Mochida Pharmaceutical Co. Ltd., [6]Graduate School of Pharmaceutical science, Tohoku University

There is great clinical interest in ATP binding cassette (ABC) transporters not only because of their involvement in multidrug resistance but also because many disease-causing mutations have been identified in proteins of this family in humans. HisP protein is the nucleotide binding domain of ABC transporter complex $HisQMP_2$, the periplasmic histidine permease of S. typhimutium and E. coli. Binding and hydrolysis of ATP or the other nucleotides is considered to power the transport process by transforming conformational changes from the NBDs to the TMDs, which lead to the movement of substrate across the membranes.

DFT calculations were carried out using $DMol^3$. Molecular dynamic simulation were carried out by OPLS-AA force field using MacroModel (Schrödinger, Inc.). QM/MM molecular dynamics studies were performed using Hybrid-Colors, which was developed in our laboratory.

Two mutant models were made based on the initial x-ray crystal structure of HisP protein which was taken from protein data bank (PDB ID: 1b0u [1]). One is a phosphorylated structure, in which Ser41 was bound with a phosphate group in silico; the other one is catalytic activated model, water407 was substituted by a Mg^{2+}, which has been mentioned by Huang and colleagues in 1998.[1] After the energy of three structures were minimized by OPLS-AA force field, we performed a molecular dynamic simulation on three models. ATP binding site was selected from initial structure and optimized using accurate DFT methods. Our studies indicate that ATP molecule and Mg^{2+} and residues forms weakly bound complex (Fig. 1) and next step is the broken of the bond between γ phosphor atom and the oxygen 8. In addition, isolated ATP, ADP and AMP molecules were also investigated, from our calculation, the energy of 12 kJ/mol was released during ATP hydrolysis. QM/MM studies were also carried out using Hybrid-colors.

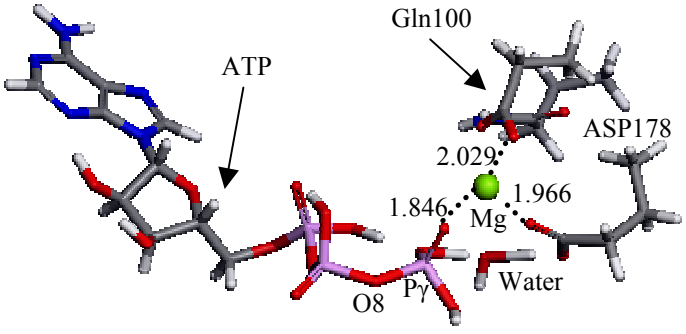


Fig. 1: Optimized structure of ATP binding site, the distances between Mg and ATP, two residues are also shown in the figure.

REFERENCE

1) L. Hung, I. Wang, K Nikaido, P.Liu, G. Ames and S.Kim, *Nature*, 396 (1998) 703-707.

P22

In silico screening for the human $P2Y_{12}$ receptor

Yosuke Nonaka, Takeshi Hiramoto, Kiyoshi Baba and Norihisa Fujita

Laboratory of Pharmcoinformatics, Department of Bioinformatics,
Ritsumeikan University, Kusatsu, Shiga 525-8577

We have searched endogenous ligands acting on the human $P2Y_{12}$ receptors, one of the G-protein coupled receptors (GPCRs) by *in silico* screening against biological compound-database that contained more than 500 animal-metabolites. The 3D model of the $P2Y_{12}$ receptor necessary for the screening was obtained by the homology modeling using the previously constructed 3D model of $P2Y_1$ receptor as a template [1),2)]. The *in silico* screening using "AutoDock" resulted in selection of cyteinylleukotrienes and 5-phosphorybosyl 3-pyrophosphate (PRPP) with high free energy changes, in addition to the known $P2Y_{12}$ ligands such as 2MeSADP and ADP. These candidates were subjected to the intracellular Ca^{2+} assay using the recombinant CHO cells expressing $P2Y_{12}$ receptors. We found that one of cysteinylleukotrienes and PRPP acted on the $P2Y_{12}$ receptor as agonists with the EC_{50} values of 1.3nM and 7.8nM, respectively. Furthermore, we have analyzed the phylogenetic relationship of the P2Y, P2Y-like and cysteinylleukotriene receptors based on the sequence alignment followed by an evolutionary analysis. The analyses showed that the cysteinylleukotriene receptors belonged to the P2Y-related family. Moreover, an orphan GPCR was detected in a P2Y subfamily.

1) Hiramoto. T, Nonaka Y, Inoue K, Yamamoto T, Omatsu-Kanbe M, Matsuura H, Gohda K, and Fujita N. (2004) Identification of Endogenous Surrogate Ligands for Human P2Y Receptors Through an *in silico* Search. *J.Pharmacol.Sci.* 95: 81-93

2) Hiramoto T, Nemoto W, Imai T, Kikuchi T and Fujita N. (2002) Construction of hypothetical three-dimensional structure of $P2Y_1$ receptor based on Fourier transform analysis. *J Protein Chem.* 8: 537-545

P23

Systematic comparisons of consensus scores for computational ligand-docking

Akifumi Oda[1], Keiichi Tsuchida[1], Tadakazu Takakura[1], Noriyuki Yamaotsu[2], Shuichi Hirono[2]
Toyama Chemical Co. Ltd.[1], Kitasato University[2]

A prediction of a binding mode between a target protein and its ligand molecule plays an important role in structure-based drug design (SBDD) trials. FlexX is one of the major ligand-docking programs developed for predicting the three dimensional structure of the protein-ligand complex, and in this program predicted structures are evaluated by FlexX score. Because the ability of single scoring function to discriminate between adequate and inadequate complex structures is sometimes insufficient, the concept of “consensus score”, in which multiple scoring functions are used simultaneously, reported recently[1)]. Although many consensus score strategies are proposed and they are effective for evaluations of complex structures, there are few studies that compare these strategies[2)]. Furthermore, for the selections of complex model candidates, systematic discussions about combinations of scoring functions have not been carried out.

In this study, we compared all possible combinations of nine consensus scoring strategies (number-by-number, number-by-rank, rank-by-number, rank-by-rank, percent-by-number, percent-by-rank, vote-by-number, vote-by-rank, and vote-by-percent) with nine scoring functions (FlexX score, GOLD score, DOCK score, PMF score, ChemScore, DrugScore, PLP, ScreenScore, and X-Score). Thus, 4599 types of consensus scores were compared. The systematic naming of consensus scoring strategies was also proposed. The results indicate that the “vote-by-number” strategy, in which the ligand-docking models are selected by majority vote of scoring functions, is effective for determining the candidates of plausible protein-ligand complex. However, for “vote-by-” strategies the adjustments of parameters were difficult because two different thresholds were needed in these strategies. In consensus scoring strategies required only one threshold, “number-by-number” gave the best results. By incorporating these consensus scores in FlexX program, reasonable complex models seem to be obtained more efficiently than by independently using FlexX score. These strategies may also be useful for other docking programs to improve the scoring scheme, and the SBDD is expected to be more effective by these improvements.

References

1) P. S. Charifson *et al.*, *J. Med. Chem.*, 1999, **42**, 5100-5109.
2) R. Wang *et al.*, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 1422-1426.

P24

A Multilocus Simulation Study to Multiple Testing for Association Test

Woosung Yang [1], Jun Nakaya[1], Norihiro Sakamoto[1]
Kobe University of Clinical Genome Informatics Center[1]

Advances in maker technology have made a dense marker map a reality. The alleles at tightly linked marker loci may themselves be associated and, therefore, the tests conducted at each locus may not be independent. When alleles at different marker loci are associated, the Bonferroni correction may lead to a conservative test, and hence a power loss. Ideally we would like the experimentwise significance level of the test without being overly conservative, in order to increase the power of our test. To archive this goal we propose a modified Monte-Carlo approach to determine significance. We examine the case of tightly linked markers with varying amount of association between them. Using the computer simulations, we compare the power when either the Bonferroni or modified Monte-Carlo procedure is used to determine significance. This proposed procedure guarantees the correct experimentwise level, regardless of dependencies between the individual test and can be applied whenever it is suspected that markers examined have high amounts of association, or as a general approach to ensure appropriate significance levels and optimal power.

References

1. Casella G, Berger R.: Statistical inference. Pacific Grove, CA: Wadsworth and Brooks. (1990)
2. Kaplan NL, Martin ER, Weir BS: Power studies for the transmission/disequilibrium tests with multiple alleles. Am J Hum Genet vol.60, pp.691-702 (1997)
3. Knapp M, Seuchter SA, Bauer MP: The haplotype-relative-risk method for analysis of association in nuclear families. Am J Hum Genet vol.52, pp.1085-1093 (1993)
4. Martin ER. Kaplan NL. Weir BS: Test for linkage and association in nuclear families. Am J Hum Genet vol.61, 439-448 (1997)
5. Risch N, Merikangas K: The future of genetic studies of complex human diseases. Science vol.273, pp1516-1517.
6. Speilman RS, Ewens W: The TDT and other family based tests for linkage disequilibrium and association. Am J Hum Genet vol.59, pp.983-989 (1996)

P25

Kinetic assessment of lenampicillin absorption using Caco-2 cells: in vitro study aiming for rational design of prodrug

Takashi MIZUMA [1], Sayaka SAKAGUCHI [1], Sachie TANAAMI [1] and Masahiro HAYASHI [1]

Department of Drug Absorption and Pharmacokinetics, School of Pharmacy, Tokyo University of Pharmacy and Life Science (TUPLS) [1]

Although lenampicillin has been used as an ampicillin prodrug to improve intestinal absorption, the absorption process of lenampicillin has not been characterized in detail. Therefore, we studied the transport and metabolism of lenampicillin using Caco-2 cells. Lenampicillin and ampicillin were determined by HPLC. Concentration of lenampicillin, which was added to the apical side, decreased with time. On the basal side, lenampicillin appeared and the amount increased over time. Ampicillin also appeared on the basal side, indicating that lenampicillin is in part metabolized to ampicillin during the absorption process. Transport clearance of lenampicillin from the apical to basal sides across the Caco-2 cell monolayer was approximately 150 times higher than that of ampicillin. Similar results were observed in the uptake clearance of lenampicillin from the apical side into Caco-2 cells. However, the transport clearance of lenampicillin from the basal to apical sides was higher than that from the apical to basal sides. In addition, lenampicillin decreased transport of rhodamine123, a P-glycoprotein substrate, from the basal to apical sides. Verapamil and cyclosporine enhanced transport of lenampicillin from the apical to basal sides. These results suggest that the influx (uptake) of lenampicillin by passive diffusion is higher than the efflux by P-glycoprotein, resulting in enhanced intestinal absorption. (Supported in part by the Japan Society for the Promotion of Science (T. M.)).

P26

Database and analysis of binding site residues and ligands in membrane protein-ligand complexes

M. Xavier Suresh, M. Michael Gromiha and M. Suwa

Computational Biology Research Center (CBRC), AIST,

2-42, Aomi, Koto-ku, Tokyo 135-0064, Japan.

Abstract:

Membrane proteins perform various functions such as biological transduction and transport processes by mediating the flows of ions, energy and information across membranes. The structural analysis on the interactions between the membrane proteins and ligands provides deep insights into the mechanism of protein-ligand interactions and their function. An increasing number of experimentally determined structures of proteins would help to derive the structural principles in protein-ligand complexes. In this work we developed a database of membrane protein-ligand interactions from the information available in PDB and other structural databases. The interactions between amino acid residue in membrane protein and the functional group of the ligand have been analyzed through atom contact, hydrophobic behavior, location of the residues, etc. The preliminary results obtained on the analysis will be presented.

P27

GeneT2D: A database for the genetic variation in type II diabetes

Yu LIN[1], Jun NAKAYA[1], Norihiro SAKAMOTO[1]
Kobe University Graduate School of Medicine, 7-5-2 Kusunokicho, Chuo-ku, Kobe, Hyogo 650-0017, JAPAN[1]

OBJECTIVE: Although, as for a common multi-genetic disease, Type II Diabetes (T2D) is more likely triggered by the individual lifestyle along with the plural genetic variation than the genetic disorder such as the type I Diabetes (T1D); comprehensive understanding of the multi-caused genetic background of T2D is essential to establish the pharmacogenomics and the gene based lifestyle modification. Here we present a genetic variation database of T2D (GeneT2D) that includes the comprehensive T2D related gene information. GeneT2D will be used for finding the relationship between the genotype and the phenotype of T2D.

METHOD: The data in GeneT2D are collected from the public databases, such as the Genetic Association Database, the Human Genome Variation Database, and other gene-related databases; meanwhile, we developed a scan tool to update the susceptibility genes from PubMed. The gene's name, the position, the variation information (partly including SNPs), and the gene annotations were collected and confirmed manually. An open source RDBMS named MySQL is used to manage the data. A web based interface was built with the PHP scripting language skill and the Apache HTTP Server technology for the data representation.

RESULT: There are 124 T2D related genes in the database at the present, and a daily scan tool is used to inform the new published susceptibility genes. Each entry includes the gene symbol, the DNA position, the functional effects, the reference, the gene variation, the statistics, the confirmed SNPs, and the redirection information to the other public databases. GeneT2D is organized by MySQL with a web interface.

References

1) M. Alan Permutt, Jonathon Wasson, Nancy Cox: Genetic epidemiology of diabetes. J Clin Invest. Vol.115, No.6, pp.1431-1439 (2005).
2) Joel N. Hirschhorn: Genetic Approaches to Studying Common Diseases and Complex Traits. Pediatr Res. Vol.57, No.5 Pt 2, pp.74R-77R (2005).
3) Laura Dean, Johanna McEntyre: The Genetic Landscape of Diabetes. Published online in NCBI (2004).

P28

Amino acid sequence analysis of *Saccharomyces cerevisiae* proteome in terms of electric charge distribution

Runcong KE [1], Shigeki MITAKU[1]
Department of Applied Physics, Graduate School of Engineering, Nagoya University[1]

Experimental and theoretical studies have indicated the importance of electrostatic effects in individual protein structure and its stability. However, common features of such effects in all proteins from organism proteome have hardly been investigated. In our previous work, we analyzed net charge distribution of all amino acid sequences for genome *Drosophila melanogaster* and showed that the charge distribution was well fitted to Gaussian distribution, suggesting random amino acid sequences in genome level. In this work, we generated random sequences which have the same amino acid composition, number of proteins and the size distribution as *Saccharomyces cerevisiae* proteome. The histogram of charge distribution from random sequences was sharper than one from proteome sequences, indicating the positive correlation of charged residues in proteome sequences. Therefore, we calculated correlation function of charged residues in all amino acid sequences between *S.cerevisiae* proteome and random sequences. The correlation function in *S.cerevisiae* proteome could be well fitted to exponential curve with positive correlation, suggesting that electrostatics interactions within proteins of eukaryote proteomes are repulsive on average. In contract, the distribution of correlation function in random sequences appeared almost constant. The role of charge correlation in the protein folding will be discussed.

REFERENCES

1. C. Nick Pace, Roy W. Alston, and Kevin L. Shaw: Charge-charge interactions influence the denatured state ensemble and contribute to protein stability, *Protein Sci*. vol.9, no.7, pp.1395-1398 (2000).
2. Huan-Xiang Zhou: A Gaussian-chain model for treating residual charge-charge interactions in the unfolded state of proteins, *Proc. Natl. Acad. Sci. U.S.A.* vol.99, no.6, pp.3569-3574 (2002).
3. Runcong Ke and Shigeki Mitaku: Gaussian analysis of net electric charge of proteins in the *Drosophila melanogaster* genome, *Chem-Bio Informatics J*. vol.4, pp.101-109 (2004).
4. Runcong Ke and Shigeki 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* (in press) (2005).

P29

Folding dynamics of 10 residues protein, chignolin

Atsushi SUENAGA, Tetsu NARUMI, Noriyuki FUTATSUGI, Ryoko YANAI, Yousuke OHNO, Noriaki OKIMOTO, Makoto TAIJI
Computational and Experimental System Biology Group, RIKEN Genomic Sciences Center

Characterization of the folding dynamics and free energy landscape of proteins are important steps in understanding the protein folding mechanism. To understand those of the protein element, such as α-helix and β-hairpin, provides us information in early event of protein folding process. Computer simulation is very effective method to elucidate the folding dynamics and conformational landscape of the protein element at the atomic resolution. Here we performed folding, unfolding and refolding molecular dynamics simulations of 10 residues protein, chignolin, which folds into β-hairpin structure (1). We analyzed the wide range of conformational space obtained from these simulations. The chignolin folded β-hairpin structure with going down the funnel-like energy hill in which the protein reduces conformational entropy in our simulations. In the folding process, a specific pathway did not exist, rather multiple pathways converged to dense conformations. These results indicate that the “folding funnel theory” (2) is appropriate for interpretation of the folding mechanism of the chignolin.

REFERENCES

1) Shinya Honda, Kazuhiko Yamasaki, Yoshito Sawada, Hisayuki Morii: 10 residue folded peptide designed by segment statistics, Structure, vol. 12, pp. 1507-1518 (2004).
2) Peter E. Leopold, Mauricio Montal, José N. Onuchic: Protein folding funnels: a kinetic approach to the sequence-structure relationship, Proc. Natl. Acad. Sci. USA, vol. 89, pp. 8721-8725 (1992).

P30

Massively parallel calculation of binding free energy for biomolecules

Masakatsu Ito[1], Yoshiaki Tanida[1], Hideaki Fujitani[1],
Michael R. Shirts[2], Guha Jayachandran[2], Christopher D. Snow[2], Vijay S. Pande[2]
Fujitsu Laboratories Ltd.[1], Chemistry Department, Stanford University[2]

Although drug-discovery process would be more efficient with accurate estimation of binding free energies, such estimation with less than 1.0 kcal/mol error still remains a challenge to computer simulations. We improve this accuracy by applying massively parallel processing system, BioServer to use Bennett acceptance ratio (BAR) method[1]. The BAR method gives statistically optimal estimation of free energies based on a set of molecular dynamics (MD) trajectories. BioServer parallelizes the generation of this set by running GROMACS MD program on multiple processors with varying initial velocities and coupling parameters. Typically the simulation for FKBP-ligand complex is accelerated by 400(~ 33 × 12) processors.

We examined the accuracy of our estimation using eight FKBP ligands complexed with their receptor. Generalized Amber force field parameters were assigned to the ligands using AM1-BCC[2], and AMBER99 parameters to FKBP receptor. Figure 1 shows the comparison of our obtained values with experimental ones. The calculated free energy differences are well correlated to experimental inhibition constants. The root mean square deviation is 0.36 kcal/mol.[3]

This result implies that our system can improve the efficiency of screening process of inhibitor candidates using the estimation of absolute free energies.

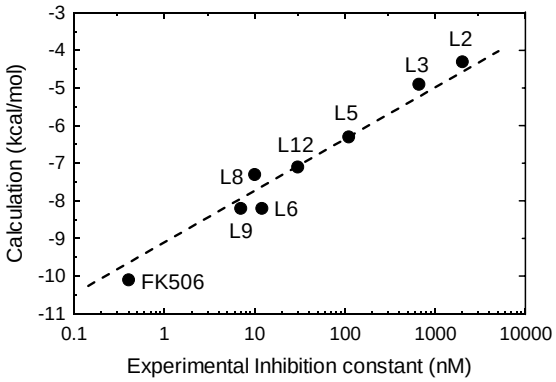


Fig. 1: Comparison of calculated binding energy with experimental values

REFERENCES

[1] M. Shirts, and V.S.Pande, J.Chem.Phys.,**122**,144107(2005); M. Shirts, E.Bair, G.Hooker, and V.S.Pande, Phys.Rev.Lett.,**91**,140601(2003).

[2] J.Wang et al. , J.Comput.Chem.,**25**,1157 (2004).

[3] H.Fujitani, Y.Tanida, M.Ito, M.R.Shirts, G.Jayachandran, C.D. Snow, V.S.Pande , ACS National Meeting March 13 (2005).

P31

Construction of the Web-Accessible comprehensive transporter database for drug discovery and development, “TP-Search” (Part II)

Takafumi Ochiai [1], Naoki Ozawa[12],Yoshimasa Hama [1],Katsuhiko Taki[3], Toshiyuki Ueda[4], Noriko Sugawara[4], Yukiko Abe[4], Hitoshi Matsui[4], Kiyomi Ito[5], Kazuya Maeda[6], Hiroyuki Kusuhara[6] and Yuichi Sugiyama[6]

Advanced Research Institute for Science and Engineering, Waseda University[1], Wyeth K.K. [2], Nihon Visual Science.Inc[3], CAC Corporation[4], Hoshi University[5], Graduate School of Pharmaceutical Sciences, the University of Tokyo[56]

Recently, researchers involved in drug discovery and development have increasingly paid more attention to the transporter as one of the determinant factors influencing drug disposition and pharmacokinetics. Over the last several years, so much information about transporters has been published that there is now a need for a publicly accessible database with comprehensive well-organized information about transporters for drug discovery and development purposes. To assist this, we have constructed a transporter database “TP-Search” on the WEB. (URL:http://www.TP-Search.jp/)

In our previous work, this database covered a huge amount of information about 80 kinds of transporters in humans, rats and mice, and had been prepared from data extracted from about 2600 research articles, and provided several properties of transporters, such as substrate/inhibitors, tissue distribution, regulation of expression level, knockout animals, gender differences and drug-drug interactions. Now, the database covers about 90 kinds of transporters, and has been prepared from data extracted from about 3300 research articles, and provides additional property of pathophysiology. The database is updated every three or four months. Users can search for information not only on transporter name but also on compound name and organ name. And all information is directly linked to the original references in PubMed, which ensures that the users can confirm its validity and obtain more details. TP-search will be updated periodically and add properties of transporters that researchers involved in drug development have great interest, to keep pace with the current advances in transporter research.

References

1) Naoki, O, et al. “Transporter Database, TP-Search: A Web-Accessible Comprehensive Database for Research in Pharmacokinetics of Drugs”, Pharmaceutical Research, Vol. 21, No. 11, November 2004, 2133-2134

P32

Particle simulation approach for subcellular spatio-temporal events

Ryuzo Azuma [1], Tetsuji Kitagawa[2], Tomoyuki Yamamoto[3],
Takeshi Arikuma[2], Hiroshi Kobayashi[4],Akihiko Konagaya[1,2]
Genomic Sciences Center, RIKEN [1], Mathematics and Computing Sciences, Tokyo Institute of Technology[2], School of knowledge Sci. JAIST [3],
Graduate Sch. of pharmaceutical Sciences, Chiba Univ.[3]

Abstract

Sub-cellular dynamics of molecules is now an important topic for biologists thanks to the advances in molecular imaging and microscopy. The single particle tracking (SPT) and single fluorescent molecule video imaging (SFVI) technologies have contributed to gaining understanding of how these molecules move in space as individuals. Unfortunately, however, it seems that only a limited knowledge is obtained about how we systematically infer the relation between the molecular and cellular level phenomena. We have been focusing on this problem and tried to bridge the gap between the molecular-level and cellular-level properties by making use of a particle simulation technology we developed recently. This technology has following characteristics suitable for a bottom-up level expression of molecular interactions in space. (1) It expresses a single molecule as a particle, (2) it explicitly assigns spatial coordinates to particles that move in a 3D space, (3) these particles interact with each other by stochastic processes, and (4) any geometry of membrane structures can be generated so that they constrain movements of some molecules. A typical snapshot obtained by a simulation based on this technology is shown in Fig.1. The line plot is a trajectory from a single particle movement, which resides initially in the cytoplasm and is translocated to the nucleus afterward. Thus our simulation method enables not only highly flexible and sophisticated model constructions but also discuss how molecular-level dynamics affect cellular-level properties. This feature will be demonstrated further by showing visualized data from simulations with a highly detailed spatial expression of a cell.

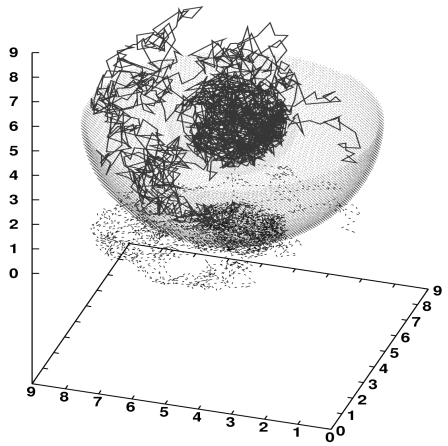


Fig. 1: A trajectory of single protein molecule translocated from the cytoplasm to the nucleus through a nuclear pore.

P33

Spontaneous Docking MD Simulations of Aspartate Racemase - Roles of Amino Acid Residues in Docking Process -

Okimasa Okada[1], Tomohiro Seko[1], and Masafumi Yohda [2]
Fuji Xerox Co., Ltd. [1], Tokyo University of Agriculture and Technology [2]

The molecular mechanism of the aspartate racemase from the hyperthermophilic archaeum, *Pyrococcus horikoshii* OT3, has been investigated based on its crystal structure[1)]. Although two sulfur atoms in the active site are thought to act as acid and base at the racemization process, the distance (0.96 nm) between their sulfur atoms is longer than the estimated preferable distance of 0.8 nm for functional cooperation. Molecular dynamics (MD) simulations in a wide temperature range have shown that hinge-like molecular motion of the whole racemase with conformational changes of the side-chains of the active cysteine residues was assigned to the narrowing of the distance between the two sulfur atoms. In this study, we will present the results of MD simulation on the roles of other amino acid residues and also docking of substrate.

We observed stochastic conformational changes of the side-chain of the Tyr160 locating at the entrance of the active site near the enzymatically active temperature. This conformational change induces large displacement of the phenol group to cover or uncover the active site as if it is functioning as a gate. We succeeded in the docking of the L-aspartic acid into the racemase by MD simulations without adding any artificial interactions. The L-aspartic acid, which was initially placed about 2 nm away from the sulfur atoms in the active site, spontaneously entered into the racemase with the assistance of amino acid residues locating around the cleft. The observed docking process enabled us to propose the molecular mechanism and roles of the amino acid residues by comparing with the results of the mutation analyses. The roles of water molecules will also be discussed at the presentation.

REFERENCE

1) L. Liu, K. Iwata, A. Kita, Y. Kawarabayashi, M. Yohda and K. Miki, *J. Mol. Biol.* 319 (2002) 479-489.

P34

Transplantation of the PEACH5.8 from UNIX to Windows System, and Practice in the Simulation of TNA

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

PEACH Ver.5.8[1], a program for molecular dynamics simulation of biochemical molecules, was transplanted from UNIX to Windows computer[2,3]. Visual Fortran was utilized as the compiler.

TNA, (L)-α-threofuranosyl oligonucleotides, would be considered a possible ancestor of RNA. The TNA strands can also pair up with complementary strands of RNA and DNA. MD simulations of TNA model structure were performed at several different temperatures. The structures of 1N1O and 1JGR in Protein Data Bank[4] were used as the initial structure of the simulations. The 1N1O molecule is B-form DNA duplex containing (L)-α-threofuranosyl nucleotides, and 1JGR is just B-form DNA which have the same base sequences. MD simulations were carried out under the periodic boundary condition by using the parallel computing with MPICH Ver1.2.5[5]. The number of solvent water molecules for the simulation was 10831 molecules. The simulations were performed at the atmospheric pressure for 500ps with a 0.4ps coupling parameter.

The results of MD simulations indicated that the 1N1O structure is a little bit unstable, compared with 1JGR structure with the data of the root mean square deviations (RMSD) of the structure.

References

1) http://staff.aist.go.jp/y-komeiji/peach/peach.html
2) J.Comput.Chem.Jpn, Vol.2, No.4, pp149-154(2003)
3) http://www3.nibh.go.jp/~nakata/peach/peachw.html
4) http://www.rcsb.org/pdb/
5) http://www-unix.mcs.anl.gov/~ashton/mpich.nt/References

P35

Semi-automatic interpretation of experimental results based on conceptual network

Asako Koike [1,2] and Toshihisa Takagi [1]

Dept. of Computational Biology, Graduate School of Frontier Science, The University of Tokyo, Kiban-3A1(CB01) 5-1-5, Kashiwanoha Kashiwa, Chiba, 277-8561, Japan [1], Central Research Laboratory, Hitachi Ltd. 1-280 Higashi-Koigakubo Kokubunji, Tokyo, 185-8601, Japan [2]

With the exploitation of high-throughput technologies such as DNA/Protein array, RNAi, and differential display, the demand of automatic/semi-automatic interpretation of experimental results based on the amount of published text is increased. We have developed a system called BioTermNet to cluster genes/concepts and calculate the explicit and implicit relationships between multiple concepts based on information extraction and information retrieval techniques to efficiently interpret high-throughput experimental results. The relationships between multiple concepts are mainly presented using three methods: association matrix, its hierarchical tree, and conceptual network. An overview of the relationships between multiple genes/concepts can be obtained by using the association matrix and its hierarchical tree, while the detailed and/or causal relationships can be presented by the conceptual network. The association matrices are calculated using the vectors whose components are the co-occurred concepts using Lnu term weighting and protein interactions extracted by syntactic analysis, individually. The conceptual network is calculated by the hybrid method of Lnu term weighting and the protein-interactions and functions using syntactic analysis (PRIME data developed our laboratory). The adequacy of the definition of gene relatedness is verified using physically interacting gene pairs and genes with the same gene ontology annotation. When this system is used to interpret microarray data of cancer cells, it creates a plausible distance matrix, its hierarchical tree, and an understandable conceptual network connecting similarly expressed genes and other concepts. Further, this system presents reasonable commonly related concepts such as diseases, phenotypes, drugs, and functions of similarly expressed genes with/without mediating protein/gene interactions. This system is also useful for knowledge discoveries such as to find potentially new target diseases for drugs. The BioTermNet system is available at http://btn.ontology.ims.u-tokyo.ac.jp:8086/.

P36

Development of Effective Computer-Aided Drug Design Strategy

Noriaki Okimoto, Atsushi Suenaga, Noriyuki Futatsugi, Ryoko Yanai, Tetsu Narumi, Yosuke Ohno, Makoto Taiji（Times New Roman/Times12pt）
Computational and Experimental Systems Biology Group, RIKEN Genomic Sciences Center

Lead discovery is one of the most important components in rational drug design. The structure-based virtual screening is very effective for lead discovery[1,2]. To predict ligand–receptor complex structures and estimate their binding affinity, many docking programs, such as DOCK, AutoDock, and FlexX, have been widely used. However, the ability on estimation of binding free energy for the docking programs is not accurate because most of them do not consider the flexibility of proteins and the solvation effects. Therefore we used a docking approach of flexible ligand – flexible receptor in an aqueous environment based on molecular dynamics simulation, and try to develop more cost-effective and more accurate computer-aided drug design strategy. In this study we evaluated the ligand binding affinities on several different proteins and research more efficient method to estimate the binding free energy between ligand and receptor.

References

1. Kuhn, B., Gerber, P., Schulz-Gasch, T. & Stahl, M. Validation and Use of the MM-PBSA Approach for Drug Discovery. *J Med Chem* **48**, 4040-8 (2005).
2. Wang, J., Kang, X., Kuntz, I.D. & Kollman, P.A. Hierarchical database screenings for HIV-1 reverse transcriptase using a pharmacophore model, rigid docking, solvation docking, and MM-PB/SA. *J Med Chem* **48**, 2432-44 (2005).

P37

Novel Type of Secondary Structure Breaker in Soluble Proteins

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

Breaking signal of secondary structure put strong limitation on the tertiary structure of protein. Pro and Gly are well-known secondary structure breakers and enhance the occurrence of disordered structure. In addition to these residues, amino acids with amphiphilic side chains having both a polar group and a flexible hydrocarbon (Arg, Lys, His, Glu, Gln) are more likely to be present in disorder region.[1] Previously, we found that these amphiphilic residues commonly exist at the terminal region of transmembrane helices.[2-4] The amino acid composition of disordered region in soluble proteins together with those at terminal region in membrane proteins strongly suggest that clusters of amphiphilic residues generally correlate with breaking points for secondary structures and amphiphilic clusters are thought to be a new type of secondary structure breaker.

Herein, we examined the distribution of amphiphilic clusters together with Pro and Gly in amino acid sequence of soluble protein to determine the environment that drives a segment to form a loop. In addition to Pro and Gly clusters, clusters of amphiphilic residues were found to be secondary structure breakers depending on specific conditions, including the average hydrophobicity, the helical periodicity, the density of Ser and Thr and presence of Trp and Tyr clusters. We carried out principal component analysis of environmental factors to identify candidate breakers in secondary structure breaking regions. Predicted breakers were located in the loop segments with accuracy of 90%, covering 63% of all loop segments.

References

1. Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C. and Obradovic, Z. Intrinsically disordered protein. J Mol Graph Model. **19**, 26-59 (2001).

2. Mitaku, S., Hirokawa, T., and Tsuji T. Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane-water interfaces. Bioinformatics. **18**, 608-616 (2002).

3. Hirokawa, T., Boon-Chieng, S. and Mitaku S**.** SOSUI: Classification and secondary structure prediction system for membrane proteins. Bioinformatics. **14**, 378-379 (1998).

4. Mitaku, S. and Tsuji T. Features of transmembrane helices useful for membrane protein prediction. CBIJ. **4**, 110-120 (2004).

P38

A solvent site-dipole field mediating DNA-protein binding

Nobuyuki Hamasaki[1], Daisuke Mitomo[1], Junich Higo[1], Akihiko Yamagishi[1], Hiroh Miyagawa [2]
School of Life Science, Tokyo University of Pharmacy and Life Science[1]
Taisyo Pharmaceutical co., Ltd[2]

It has been believed that binding of bio-molecules, realized in enzyme-substrate or DNA-protein complexes, is controlled by a random diffusion of the bio-molecules. This implies that the binding occurs by chance only when the binding sites of the bio-molecules encounter to each other with appropriate relative orientations. On the other hand, a computer-simulation work proposed another binding mechanism that the bio-molecules can approach to each other through a field, i.e., solvent site-dipole field (1-3). In this mechanism, water-molecule orientations coherently order in space between the bio-molecules even when the two bio-molecules are apart by about 20 Å to each other. Physico-chemically saying, this ordering makes the electrostatic interactions longer range than expected from a continuum theory. However, no experimental support on the existence of the solvent site-dipole field has been provided yet, because highly time-dependent measurement at an atomic resolution is required to detect the field. If two bio-molecules are known to bind experimentally and also computationally, analysis on the simulation data may provide a key to understand the binding mechanism. For this purpose, we chose a DNA-peptide complex (DNA-protein HMG-I(Y) fragment complex (4)) for the simulation. We observed that the two bio-molecules, which were initially put apart by about 20 Å to each other in explicit solvent, bound during the simulation. The analysis showed that the solvent site-dipole field mediate the binding. We will report details of the analysis.

References

1. Higo J., Nakasako M.: J. Comp. Chem., **23**, 1333 (2002).
2. Higo J., Sasai M., Shirai H., Nakamura H., Kugimiya T.: Proc. Natl. Acad. Sci. USA, **98**, 5961 (2001).
3. Yokomizo T., Yagihara S., Higo J.: Chem. Phys. Lett. **374**, 453 (2003).
4. Huth R.H., Bewly, C.A.: Nature Struct. Boil., **4** , 657 (1997).

P39

Development of a Three-Dimensional-Structure Database of Natural Metabolites (3DMET) and Problems about Conversion from 2D to 3D-Structures.

Miki Maeda and Hisataka Numa
Genome Research Department, National Institute of Agrobiological Sciences

We have been developing a three-dimensional-structure database of natural metabolites (3DMET[1]) and the first release of the database is accessible since July 1st, 2005. The database provides 3D-structures of metabolites and several molecular properties based on the 3D-structures. The structures are searchable by their properties. For example, the molecular weights of compounds can be specified between 100 and 200. The database includes 1123 compounds with 3D-structures in this release.

The 3D-structures were deduced from the 2D-strctures of the COMPOUND database (a part of the LIGAND database[2]) of KEGG[3]. The 2D mol files of COMPOUND were converted to 3D SYBYL mol2 files by using SYBYL[4]/CONCORD[5] and MOE[6]. Every structure was minimized by the MM calculation module of each program. Respective results by SYBYL and MOE may not reflect their initial chiralities. Therefore, the chiralities of the two structures were compared by aRSchirality of MOE. The compound data was added to the 3DMET database if the following two conditions were satisfied: (1) The chiralities of the two structures obtained from the same compound were identical and (2) the data of the COMPOUND database included the PATHWAY, REACTION, or ENZYME line. In this version, only the structures whose chiralities were strictly matched each other were only collected. However, some identical structures were excluded due to different descriptions in the data derived from the order of hydrogen added by the programs. These structures will be recovered in the future release.

References

1) 3DMET: A Three-Dimensional-Structure Database of Natural Metabolites. http://www.3dmet.dna.affrc.go.jp/
2) Goto, S., Nishioka, T., and Kanehisa, M., LIGAND: Chemical Database for Enzyme Reactions. Bioinformatics, vol. 14, pp591-599 (1998)
3) KEGG: Kyoto Encyclopedia of Genes and Genomes. http://www.genome.jp/
4) SYBYL, Tripos Inc. St. Louis, MO, USA.
5) CONCORD™, Optive Research, Inc., Austin, TX, USA (Now, Tripos Inc.).
6) MOE (Molecular Operating Environment), Chemical Computing Group, Montreal, Canada.

P40

How effective for fold recognition is a potential of mean force that includes relative orientations between contacting residues in proteins?

Sanzo Miyazawa [1], Robert L. Jernigan[2]
miyazawa@smlab.sci.gunma-u.ac.jp, Faculty of Technology, Gunma University, Japan[1]
L. H. Baker Center for Bioinformatics and Biological Statistics, Iowa State University, USA[2]

"Statistical potentials", which are estimated from statistical preferences observed in protein crystal structures, are widely used to recognize and fold protein structures as well as to assess interactions between proteins. A pairwise isotropic contact potential between the 20 types of amino acids is a typical statistical potential. In order to extend the capability of the contact potential, we estimate the statistical distribution of relative orientations between contacting residues from a database of protein structures and evaluate the potential of mean force for relative orientations between contacting residue. A total contact energy for contacting residue pair is evaluated as a sum of the isotropic contact energy and the present orientational potential. Polar angles and Euler angles are used to specify two degrees of directional freedom and three degrees of rotational freedom for the orientation of one residue relative to another in contacting residues, respectively. A local coordinate system affixed to each residue based only on main chain atoms is defined for fold recognition. The 4435 protein domains defined in SCOP-1.61 were used with sampling weights determined on the basis of a sequence identity matrix between them; the effective number of contacting residue pairs used is equal to 1467302. The number of contacting residue pairs in the database will severely limit the resolution of the statistical distribution of relative orientations, if it is estimated by dividing space into cells and counting samples observed in each cell. To overcome such problems and to evaluate the fully-anisotropic distributions of relative orientations as a function of polar and Euler angles, we choose a method[1] in which the observed distribution is represented as a sum of δ functions each of which represents the observed orientation of a contacting residue, and is evaluated as a series expansion of spherical harmonics functions. The sample size limits the frequencies of modes whose expansion coefficients can be reliably estimated. High frequency modes are statistically less reliable than low frequency modes. Each expansion coefficient is separately corrected for the sample size according to suggestions from a Bayesian statistical analysis. As a result, many expansion terms can be utilized to evaluate orientational distributions. Also, unlike other orientational potentials, the uniform distribution is used for a reference distribution in evaluating a potential of mean force for each type of contacting residue pair from its orientational distribution, so that residue-residue orientations can be fully utilized to recognize protein structures. The zero energy level of the orientational potential, which is formulated as a logarithm of the probability density, is defined such that the expected value of orientational energy for the native folds is equal to zero for each type of contacting residue pair. It is shown by using decoy sets ("Decoys'R'Us") that the discrimination power of the orientational potential in fold recognition increases by taking account of both the polar and the Euler angle dependences, and becomes comparable to that of a simple contact potential. In the result, the total energy potential taken as a simple sum of contact, orientation, and backbone (ϕ, ψ) potentials identifies native structures better than any other method including a CHARMM potential. In addition, the results strongly indicate that all these energy terms complement each other and are needed to recognize native structures in a wide range of decoys from near native to denatured structures.

1. Onizuka, K., Noguchi, T., Akiyama, Y., Matsuda, H.: *Intelligent Systems*, **17**, 48-54, (2002).

P41

MOLWORKS+G: Integrated Workbench for the Chimo-Bio Molecular Design by Grid Computing

Fumikazu KONISHI [1], Toru YAGI[2], Akihiko KONAGAYA[1]
RIKEN GSC, Advanced Genome Information Technology Research Group [1],
BestSystems Inc.[2]

In this paper, we present MolWorks+G an integrated workbench for a molecular design based on Grid computing. In *silico* drug design takes a computational challenge of molecular dynamics (MD) simulations for protein science brought about by the interactions among chemical compounds, genes, and proteins. The target simulations required mass computations will appear to demonstrate significant result in the post-genomic era. In this direction, a peta FLOPS computer will be available for time-consuming and expensive applications such as MD simulation, and a virtual screening for target compound validation will play a key role in drag design for novel drug discovery. In order to perform a virtual screening, we need to prepare a parameter for target chemical compound before. Those compounds had been collected in database. A chemical compound database such as the Cambridge Structural Database1 (CDC) registers 325,000 entries of the crystal structure information of the organic compounds and the organo-metallic compounds, and it will use for a docking simulation by using MD approaches. However there is a realistic issue about the preparation for the force field parameters which are calculated from those each compounds by quantum chemical calculation programs such as Gaussian2 or Q-Chem3. To calculate a force field for each target compound for drug discovery needs a validation process of several candidates' structures by Chemist and/or Pharmacologist.

REFERENCES

1. The Cambridge Structural Database: a quarter of a million crystal structures and rising F. H. Allen, Acta Crystallogr., B58, 380-388, 2002
2. Gaussian - http://www.gaussian.com/
3. Q-Chem - http://www.q-chem.com/
4. Tajima, S., Nagashima, U., Hosoya, H., J. Comput. Chem. Jpn., 2002, 1, 103
5. MolWorks – http://www.molworks.com
6. GAMESS - http://www.msg.ameslab.gov/GAMESS
7. Stewart, J.J.P., J. Comput. Chem., 1989, 10, 209, 221.
8. Pople, J. A., Beveridge, D. L., "Approximate Molecular Orbital Theory", McGraw-Hill, New York, 1970.

P42

A mutation in a cuticle collagen causes hypersensitivity to bisphenol A in *Caenorhabditis elegans*

Masahito WATANABE [1], Nanako MITANI[1], Naoaki ISHII[2], Keizaburo MIKI[3]
Japanese Institute of Pearl Science.[1]
Department of Molecular Life Science, Tokai University School of Medicine[2]
BIOS Inc.[3]

To assess toxicity of various environmental chemical pollutants, we have used the nematode, *Caenorhabditis elegans*. In this study, we isolated a novel mutant, *bis-1(nx3)*, that is hypersensitive to bisphenol A (BPA), one of endocrine disrupting chemicals (EDC). The mutation (glycine substitution) was located on a cuticle collagen gene, *col-121*. Other collagen mutants so far studied, *dpy-2(e8*), *dpy-7(e88*) and *dpy-10(e128*) also showed sensitivity to BPA, indicating that collagen mutations cause hypersensitivity to BPA. These collagen mutants may provide a starting point for a useful tool of EDC assay.

REFERENCE

1) Masahito Watanabe, Nanako Mitani, Naoaki Ishii, Keizaburo Miki: A mutation in cuticle collagen causes hypersensitivity to the endocrine disrupting chemical, bisphenol A, in *Caenorhabditis elegans*, Mutat Res. Vol.570, 71-80 (2005).

P43

*Ki*Bank, major new developments and status in 2005

Junwei Zhang[1], Masahiro Aizawa[1], Shinji Amari[1], Yoshio Iwasawa[2], Tatsuya Nakano[3], Kotoko Nakata[3]

Collaborative Research Center of Frontier Simulation Software for Industrial Science, Institute of Industrial Science, University of Tokyo [1], AdvanceSoft Corporation [2], Division of Safety Information on Drug, Food and Chemicals, National Institute of Health Sciences [3]

*Ki*Bank (http://kibank.iis.u-tokyo.ac.jp), launched on Oct 1st, 2003, is a database developed for computer-aided drug design. It contains protein and chemical structures as well as experimental binding affinities (K_i values) [1, 2]. *Ki*Bank has kept growing since last year. The entries of proteins and chemicals were expanded and some new application functions were added. Presently, over 250 target proteins (including 50 with structures) covering nuclear and membrane receptors, enzymes, transporters, ion channels and ion pumps, nearly 8000 chemicals (including 4000 with structures) and 11600 K_i values are available. With the data update, download buttons for the affinity data and the chemical and protein structural data were added. For each chemical with K_i values, we also provide a page displaying binding affinity data between the chemical and the corresponding proteins. The data in *Ki*Bank have been being widely used in structure based drug design (SBDD), quantitative structure-activity relationship (QSAR) research and some other computational approaches analyzing the protein-ligand interactions [3].

Reference

1. Masahiro Aizawa, Kenji Onodera, Junwei Zhang, Shinji Amari, Yoshio Iwasawa, Tatsuya Nakano and Kotoko Nakata: KiBank: A Database for Computer-Aided Drug Design Based on Protein-Chemical Interaction Analysis, Yakugaku Zasshi. vol.124(9), pp.613-619 (2004).

2. Junwei Zhang, Masahiro Aizawa, Shinji Amari, Yoshio Iwasawa, Tatsuya Nakano and Kotoko Nakata: Development of KiBank, a database supporting structure-based drug design, Comput Biol Chem. vol.28 (5-6), pp.401-407 (2004).

3. Kaori Fukuzawa, Kazuo Kitaura, Masami Uebayasi, Kotoko Nakata, Tsuguchika Kaminuma and Tatsuya Nakano: *Ab initio* Quantum Mechanical Study of the Binding Energies of Human Estrogen Receptor α with Its Ligands: An Application of Fragment Molecular Orbital Method, J Comput Chem. vol.26, pp.1-10 (2005).

P45

Software system for the prediction of mitochondria localization on the basis of physicochemical profiles in amino terminal segments

Toshiyuki TSUJI[1,2], Shigeki. MITAKU[1]
Department of Applied Physics Graduate School of Engineering Nagoya University[1],
21st Century COE program “Frontiers of Computational Science” Nagoya University [2]

Most proteins in mitochondria are synthesized in cytoplasm by genomic information and translocated to mitochondria from the cytoplasmic side. It is considered that translocation machineries recognize some signals in amino acid sequences at the amino terminus. We first prepared two dataset of amino acid sequence for the mitochondrial and cytoplasmic proteins. Then, several physicochemical properties of amino terminal segments were compared between the two types of proteins.
The results indicated that the amino terminal segments of mitochondrial proteins statistically have many positive charges, moderate hydrophobicity, and preferred aromatic residues. About 85% of positive data could be predicted by these parameters.
Similar analysis for the carboxyl terminal segments will be the future work for the complete prediction of mitochondria localization of proteins.

REFFERNCES

1. Roise, D., and Schatz, G., (1988) Mitochondrial presequences *J Biol Chem.* **263**, 4509-4511
2. Kovermann, P., Truscott, K.N., Guiard, B., Rehling, P., Sepuri, N.B., Muller, H., Jensen, R.E., Wagner, R., and Pfanner, N. (2002) Tim22, the essential core of the mitochondrial protein insertion complex, forms a voltage-activated and signal-gated channel *Mol Cell.* **9**, 363-373
3. Abe, Y., Shodai, T., Muto, T., Mihara, K., Torii, H., Nishikawa, S., Endo, T., and Kohda D (2000) Structural basis of presequence recognition by the mitochondrial protein import receptor Tom20 Cell 100, 551-560

P46

Comparative Analyses for Selecting Effective siRNA Target Sequences

Shigeru TAKASAKI [1], Akihiko KONAGAYA[1]
Advanced Genome Information Technology Group, RIKEN Genomic Sciences Center [1]

Abstract

Because the short interfering RNA (siRNA) widely used for studying gene functions in mammalian cells varies markedly in its gene-silencing efficacy, several siRNA design rules/guidelines have been reported recently. Analyzing the reported siRNA design guidelines from qualitative and quantitative points of views, we found that they were not always effective selection rules for many other mammalian genes. Though some rules from the guidelines are suitable for extracting effective sequences for specific genes, they might sometimes be unsuitable for selecting sequences for other genes. Since the gene-silencing efficacy depends very much on the target sequence positions selected from the target gene, we examined 361 effective siRNA sequences from 227 different mammalian cDNAs in the literature. As a result, we got eleven preferred and eight unpreferred nucleotides different from the ones used in the previous guidelines. These sequence-dependent nucleotides could be used as a more general guideline for selecting new siRNA sequences in target genes. We proposed a measure (score) for selecting effective siRNA candidates based on the positional features of specific significant nucleotides and demonstrated the effectiveness of the proposed scores for the recently reported genes. We also discussed the A/T content efficacy and optimal GC content in siRNA sequences.

References

1) Takasaki, S, Konagaya, A: COMPARATIVE ANALYSES FOR SELECTIING EFFECTIVE SIRNA TARGET SEQUENCES, Proceedings of KDDMBS pp.12-28, 2005.

2) Reynolds, A, Leake, D, Boese, Q, Scaringe, S, Marshall, W.S et al.: Rational siRNA design for RNA interference. *Nat. Biotech.* **22**, 326-330, 2004.

3) Ui-Tei, K, Naito, Y, Takahashi, F, Haraguchi, T, Ohki-Hamazaki H et al.: Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res.* **32**, 936-948, 2004.

4) Amarzguioui, M. & Prydz, H.: An algorithm for selection of functional siRNA sequences. *Biochem. Biophys. Res. Commun*. **316**, 1050-1058, 2004.

5) Takasaki, S, Kotani, S, Konagaya, A: An Effective Method for Selecting siRNA Target Sequences in Mammalian Cells. *Cell Cycle* **3**, 790-795, 2004.

P47

Development of a Workbench for Selective Nuclear Receptor Modulator

Naomi Komiyama [1], Tatsuya Nakano [2], Kaori Fukuzawa [3], Junpei Komiyama [4], Masumi Yukawa [5], Kotoko Nakata [6], Tsuguchika Kaminuma [7]
Chem-Bio Informatics Society [1], The National Institute of Health Sciences [2], Mizuho Research Institute [3], The University of Tokyo[4], Tokyo Medical and Dental University [5], IIS, University of Tokyo / Advance Soft Co [6], Hiroshima University [7]

Nuclear receptors are known as important therapeutic targets for many diseases including endocrine and metabolic disorders and cancer. Selective Nuclear Receptor Modulators (SNuRMs) are drug candidate chemicals for these targets. Although there already exist many SNuRM drugs in the market, many of them can not escape from side effect problems. Thus the research on SNuRM is one of the hot spots in drug discovery research. We are interested in this topic from two reasons. One is to use them as sample systems to test large scale molecular computing methods such as the FMO (fragment molecular orbital) method. Second it is to use them as materials for tutorial and training of computational chemistry particularly for rational drug design. Our eventual goal is to provide essential methods, materials, and sample numerical data to start modeling and analyzing ligand-binding domains of nuclear receptors and to carry docking simulation of nuclear receptors with ligands. We called this facility as the Workbench for Selective Nuclear Receptor Modulator study.

Among 48 superfamily members of nuclear receptors, about half of them are identified some ligands. There already exists more than 80 X-ray crystallography data deposited in PDB, some of which are only receptors and some are with ligands. In addition there are large amount of K_i values. We worked on these data to produce data sets from which one can easily start 3D modeling and energy calculations. The Workbench has sample data files for nuclear receptors, their ligands, their DNA response elements (DNA sequence patters of binding site of the nuclear receptors), 3D graphical models for nuclear receptors and their complex with ligands, data analyzing tools, and model viewers. The workbench has been developed in parallel with carrying a large scale molecular docking study FXR.

References

1) T. Kaminuma, *CBI Journal*, **3**, 130-156 (2003)
2) Fukuzawa, K., et al., *J. Comp. Chem.*, **26**, 1-10 (2005)

P48

In silico analysis of biosynthesis rate limiting factor in cell-free system

Naoya FUJITA [1], Motoaki SEKI[2], Kazuki SHINOZAKI[2], Kengo KINOSHITA[1],
Tatsuya SAWASAKI[3], Yaeta ENDO[3], Kenta NAKAI[1],
Inst. of Med. Sci., Univ. of Tokyo[1], GSC, RIKEN[2], CSTRC, Ehime Univ.[3]

For the functional analysis of proteins, it is fundamental to get proteins themselves. Wheat germ cell-free system, which is an *in vitro* biosynthesis system using wheat germ extracts, is a helpful method for the purpose. It was technically improved to yield sufficient amounts, but depending on the proteins it still results in low production. Using informatics, we searched the reason for these low biosynthesis rates, and especially focused on potential interactions and entanglements between amino acid sequences on polysome. 425 protein kinases from *Arabidopsis thaliana* were used as dataset.

As possible interactions, we assumed disorders and coiled coils and predicted them using DISOPRED2 and COILS. Disordered regions are flexible and dynamic parts of a protein. Coiled coil structures are constructed between α-helices, and usually dimers and trimers are made. The result is that those two interactive regions are suggested as rate limiting factors, and two tendencies are observed.

If the N-terminus is a highly disordered region, the yield will be low. However, proteins that have few disordered region in the N-terminus are distributed in wider range of biosynthesis rate. Therefore, in the low disorder case, we checked the coiled coil structure. The difference of yield distribution between proteins that contain coiled coils and other proteins is significant and leads to another tendency. For proteins with low disorder, if a coiled coil structure exists, the yield will not be high. Some additional experiments are currently being carried out in order to improve our results.

In summary, we showed that the sequence of the N-terminus region of protein kinases of *A. thaliana* plays a significant role in determining their biosynthesis yields. A high disorder or presence of coiled coils in this region will result in a low yield. Whether these tendencies also apply to other protein classes and are valid in different organisms remains to be determined.

References

1) Tatsuya Sawasaki, Tomio Ogasawara, Ryo Morishita, and Yaeta Endo: A cell-free protein synthesis system for high-throughput proteomics, Proc. Natl. Acad. Sci. USA 99, pp.14652-14657 (2002)

P49

Cis-Regulatory Machinery Unit (CiRMU) Model: A portable system for studying gene transcriptional processes

Masumi Yukawa [1], Yoshitomo Tanaka [1], Hiroshi Tanaka [1], Kotoko Nakata [2], Naomi Komiyama [3], Tsuguchika Kaminuma [4]
Tokyo Medical and Dental University [1], IIS, University of Tokyo / Advance Soft Co [2], Chem-Bioinformatics Society [3], Hiroshima University [4]

Descriptions and analyses of gene expression processes are now emerging as one of the most highlighted areas of bioinformatics. In addition to the existing transcription factor databases new bioinformatics tools for searching target genes of transcription factors, analyzing promoter structures, identifying cis-regulatory elements, identifying DNA response elements, and elucidating and making databases of gene regulatory networks have been developed. Some of these systems are of large scale and open for public, while some are quite small and used only by limited number of researchers. We are developing a small scale computer program called CiRMU（Cis-Regulatory Machinery Unit）. The system is designed to assist for extracting and editing data and knowledge relevant for making logical frame work models of gene expression processes.

The design consideration of this system is as follows: (1) it has interface to search public websites for transcription factors, their ligands, their cofactors, and their target genes in addition to its own small database storing these data, (2) it has graphical user interfaces that enable researchers to trace and analyze pathways and networks that consist of cell signal transductions that trigger gene expression, protein synthesis, and metabolic processes by the resultant proteins, (3) it has storage and retrieval capabilities for original papers, (4) it is easy to be implement and be update, (4) when several units of this system is combined they can produce causal multiple loops (networks) of gene expressions.

Gene expression processes are intrinsically complicated and dynamic. They depend on tissue and cell types that are basically related to the cell lineage, environment of the cells, and also timing. Therefore we are not aiming to design general purpose system but focusing our initial effort on special type of transcription factors, namely nuclear receptors, that are deeply related to various metabolic syndromes.

P50

A Knowledge Environment for Adipocyte and Adipokine Research

Tsuguchika Kaminuma [1], Masumi Yukawa [3], Naomi Komiyama [2],
Yoshitomo Tanaka [3], Hiroshi Tanaka [3]
Hiroshima University [1], Chem-Bio Informatics Society [2],
Tokyo Medical and Dental University [3]

One of the emerging needs in Bioinformatics in the post genome sequence era is how to handle vast amount of data and knowledge. Although many of the current knowledge offering systems claim user friendliness, it is not so easy for biomedical researchers to touch on the most recent information and to access the proper knowledge sources without any help of information specialists. What will be need in the near future is some agent either of human, computer, or their combination that facilitate easy access environment to the rich biomedical knowledge resources.

In order to develop such an agent type system we made a preliminary study choosing a specific biomedical research area, namely adipocyte and adipokine. Adipocytes are cells in adipose tissues and adipokines are the molecules produced in such cells. In the past decade these cells and molecules are identified as the key players in metabolic syndromes particularly at their onset stages. Adiponectine, leptin, resistin, TNF-alpha, IL-6 are some of these molecules. Currently mutually reciprocal roles of adiponectin and TNF-alpha are attracting special interests of clinical researchers in these fields. Structurally they belong to the same superfamily yet function quite oppositely. Adiponectin give good effects to obesity and diabetes while TNF-alphas work badly.

We made detailed survey of large number of public databases from genome databases to disease-oriented databases to find data and knowledge on Adiponectin and TNF-alpha relevant for resolving this problem. These surveys include gene structure, promoter structure, cis-regulatory elements, DNA binding factors that trigger gene expressions of these two molecules, ligands to these biding factors, target genes of these binding factors, protein isoforms, proteins that interact to these molecules, pathways and networks that include these molecules as their nodes.

It was found that the relevant knowledge is tremendous, and the part of these difficulties.

P51

Protein-protein interaction sites prediction and protein-protein docking by the methods of MD, grid scoring and the pair-wise interaction potential in CAPRI Rounds 3-5

Genki Terashi[1], Mayuko Takeda-Shitaka[1], Daisuke Takaya[1], Katsuichiro Komatsu[1], Kazuhiko Kanou[1], Mitsuo Iwadate[1], Hideaki Umeyama[1]
School of Pharmaceutical Sciences, Kitasato University[1]

CAPRI (the Critical Assessment of PRedicted Interaction) is a communitywide experiment to assess the capacity of protein-docking methods to predict protein-protein interactions. We developed a protein-protein docking program that was applied to the all targets of CAPRI. In CAPRI Rounds 1-2, we assumed that, because there are many ionic charges that weaken electrostatic interaction forces in living cells, the hydrophobic interaction force might be important entropically. As a result of Rounds 1-2, the predictions for binding sites and geometric centers were acceptable, but those of the binding axes were poor because only the largest benzene cluster was used for generating the initial docking structures. These were generated by fitting of benzene clusters formed on the surface of receptor and ligand. In CAPRI Rounds 3-5, the grid scoring sum on the protein-protein interaction surface and the pairwise potential of the amino acid residues, which were indicated as coming easily into the protein-protein interaction regions, were used as the calculation methods along with the smaller benzene clusters that participated in benzene cluster fitting.
Good predicted models were obtained for Targets 11 and 12. When the modeled receptor proteins were superimposed on the experimental structures, the smallest Ligand-rmsd values corresponding to the rmsd between the model and experimental structures were 6.2 and 7.3 Å, respectively.

1. Terashi G, Takeda-Shitaka M, Takaya D, Komatsu K, Kanou K, Iwadate M, Umeyama H. Searching for protein-protein interaction sites and protein-protein docking by the methods of molecular dynamics, grid scoring and the pair-wise interaction potential of amino acid residues. Proteins 2005;60(2); in press.
2. Komatsu K, Kurihara Y, Iwadate M, Takeda-Shitaka M, Umeyama H. Evaluation of the third solvent clusters fitting procedure for the prediction of protein–protein interactions based on the results at the CAPRI blind docking study. Proteins 2003;52:15-18.
3. Yoneda S, Umeyama H. Free energy perturbation calculations on multiple mutation bases. J Chem Phys 1992;97:6730-6736.

P52

Protein Structure Prediction in the 6th round of Critical Assessment of Techniques for Protein Structure Prediction (CASP6) using CHIMERA and FAMS

Mayuko TAKEDA-SHITAKA[1], Genki TERASHI[1], Daisuke TAKAYA[1], Kazuhiko KANOU[1], Mitsuo IWADATE[1], Hideaki UMEYAMA[1]
School of Pharmaceutical Sciences, Kitasato University[1]

Homology modeling is the method that constructs 3D model structures of unknown proteins based on known homologous 3D structures. Homology modeling has become a key component in post-genomic era because of huge amounts of protein sequences and the growing number of known structures. The Critical Assessment of Techniques for Protein Structure Prediction (CASP) is the blind contest of protein structure prediction. In CASP6, the *CHIMERA-group* predicted full-atom models of all targets using SKE-CHIMERA. SKE-CHIMERA is a web user interface system for protein structure prediction that allows human intervention at necessary stages, using enormous information from our own data and from publicly available data. Using SKE-CHIMERA, we iterated manual step (template selection and alignment by in-house program CHIMERA) and automatic step (three-dimensional model building by in-house program FAMS). The official CASP6 assessment showed that *CHIMERA-group* was one of the most successful predictors in homology modeling, especially for FR/H (Fold Recognition/Homologous). Here, we introduce the method of *CHIMERA-group*, and discuss its successes and failures in CASP6.

References

1. Takeda-Shitaka M, Terashi G, Takaya D, Kanou K, Iwadate M, Umeyama H. Protein Structure Prediction in CASP6 using CHIMERA and FAMS. Proteins; submitted.
2. Takeda-Shitaka M, Takaya D, Chiba C, Tanaka H, Umeyama H. Protein Structure Prediction in Structure Based Drug Design. Curr Med Chem 2004;11:551-558.
3. Yoneda T, Komooka H, Umeyama H. A computer modeling study of the interaction between tissue factor pathway inhibitor and blood coagulation factor Xa. J Protein Chem 1997;16:597-605.
4. Takeda-Shitaka M, Umeyama H. Elucidation of the cause for reduced activity of abnormal human plasmin containing an Ala55-Thr mutation: importance of highly conserved Ala55 in serine proteases. FEBS Lett 1998;425:448-452.
5. Ogata K, Umeyama H. An automatic homology modeling method consisting of database searches and simulated annealing. J Mol Graphics Mod 2000;18:258-272.

P53

Performance of multi-stage docking simulation with a fast Poisson-Boltzmann solver and re-scoring methods

Hiroaki FUKUNISHI [1], Jiro SHIMADA[1], Daisuke TOKUSHIMA[1], Takumi WASHIO[1], Hiroshi KURAMOCHI[2]
NEC Corporation [1], NIPPON KAYAKU CO., LTD.[2]

In order to find novel drug candidates in compound library by *in silico* screening, we performed "Multi-stage docking simulation" in which molecular model is changed, stage by stage, from the simple/empirical one to the more realistic/rigorous. Specifically, our approach comprises the following four stages:

(1) Docking simulation by FlexX to find plausible docked poses to the active site.
(2) Re-scoring of the compounds and poses by our newly developed score.
(3) Refining the poses and estimating binding energy on the basis of molecular mechanical model incorporated with solvation effect through a fast Poisson-Boltzmann solver (MM + PBSA).
(4) Estimating more accurate binding energy by full-atom MD and by QM/MM simulations, though this final stage is beyond the scope of this report.

We applied our approach to DNA Gyrase, which plays an important role in replication and transcription of DNA. Since DNA Gyrase is an essential enzyme in prokaryotes and there is no highly homologous enzyme in mammalian, it is considered to be an ideal target for antibiotics. According to *in Silico* screening result published, Gyrase is a difficult target for FlexX, compared with other targets[1]. Our result indicated that MM+PBSA combined with our newly developed rescoring method was superior to the FlexX and other scoring as implemented in SYBYL. Figure 1 shows binding structures docked by FlexX and subsequently optimized by MM+PBSA.

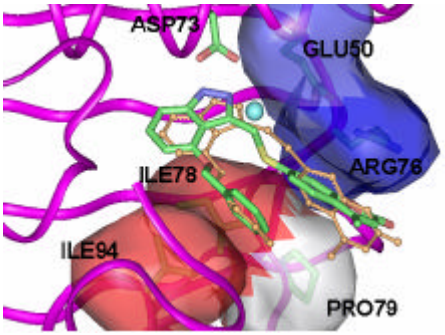


Fig.1 Binding structures docked by FlexX (orange) and optimized by MM+PBSA (by atom); fig by InsightII.

References

1) M.Stahl, M.Rarey, *J.Med.Chem*., vol. 44, 1035 (2001).

*Current address (T.W.): JST and Tokyo University

P54

FAMS Complex: A fully automated homology modeling system for protein complex structures

Mayuko TAKEDA-SHITAKA[1], Genki TERASHI[1], Daisuke TAKAYA[1], Kazuhiko KANOU[1], Mitsuo IWADATE[1], Katsuichiro KOMATSU[2], Kinji FUCHIKAMI[2] and Hideaki UMEYAMA[1]

School of Pharmaceutical Sciences, Kitasato University [1], In-Silico Sciences, Inc.[2]

Formation of protein-protein complexes contributes many biological functions. Therefore, three-dimensional structures of protein complexes are essential for deeper understanding of protein functions and the mechanisms of diseases at molecular level. In comparison with individual proteins, however, complex structures are difficult to solve experimentally because of the technical limitations. Thus a method that can predict protein complex structures is highly valuable. In this study, we developed new software FAMS Complex; a fully automated homology modeling system for protein complex structures consisting of two or more molecules. It requires only sequences and alignments of target protein as input and constructs all components simultaneously and automatically. FAMS Complex will become the essential tool for structure-based drug design such as *in silico* screening to accelerate drug discovery before the experimental structure is solved. Moreover, in this post-genomic era that the huge amounts of protein sequence information are available, the major goal is to determine protein-protein interaction networks on a genomic scale. FAMS Complex will also contribute to this goal, because its procedure is fully automated and is suited for large-scale genome wide modeling.

REFERANCES

1) Takeda-Shitaka, M.; Takaya, D.; Chiba, C.; Tanaka, H.; Umeyama, H. Curr. Med. Chem., 2004, 11, 551.
2) Takeda-Shitaka M., Nojima H., Takaya D., Kanou K., Iwadate M., and Umeyama H. Chem. Pharm. Bull., 2004, 52, 643.
3) Ogata, K.; Umeyama, H. J. Mol. Graphics Mod., 2000, 18, 258.

P55

Fully automated protein structure predictions at CASP6 using homology-modeling server SKE-FAMSD

K. Kanou[1], M. Iwadate[1], G. Terashi[1], D. Takaya[1], M. Takeda-Shitaka[1] and H. Umeyama[1]

[1] - Department of Biomolecular Desig School of Pharmaceutical Sciences, Kitasato University

To construct high-quality predictions of protein structures, fully-automatic homology-modeling server "SKE-FAMSD" calculate some alignments for amino acid sequences using 7 kinds of methods, BLAST [1], PSI-BLAST, PSF-BLAST, RPS-BLAST, IMPALA, FASTA and Pfam. PSF-BLAST is PSI-BLAST whose profile sequence group of PSSM construction process is revised, and the selection criterion is E-value<=0.001 from template PDB sequence on PSI-BLAST search. For selecting the "best alignment" in all alignments calculated by 7 kinds of methods, the score-function that was constructed by model length, homology% and degree of secondary structure agreement between PSI-PRED and STRIDE was defined:

$$score = f(k_i, Hom, Len, SS)$$

Len is residue length of model protein. Hom indicate homology % value, the ratio between the number of match residues and Len. SS is so called Q3 value, degree of secondary structure agreement between PSI-PRED and STRIDE. k_i are coefficients. The subscript number "i" indicate kind of alignment method, 0 is PSI-BLAST, 1 is BLAST, 2 is RPS-BLAST, 3 is Family-BLAST, 4 is IMPALA, 5 is FASTA, 7 is Pfam.

The next step is the construction of structures using best scored alignment by homology-modeling program FAMS [2]. In the gap-regions modeling, the fragment selection procedure of FAMS was modified to select fragment of same protein family. Therefore selection criteria were RMSD of fitting, and degree of SCOP [3] ID agreement between template PDB and fragment.

Finally, for refining constructed models by FAMS, both of Energy Minimization and Molecular Dynamics(MD) are applied. The condition for MD calculation (temperature, time, position constrain, torsion angle constrain etc..) was optimized using high resolution X-ray structures.

We had participated CASP6 experiment (http://predictioncenter.llnl.gov/casp6/Casp6.html) in fully-automatic section, and all target protein structures were predicted using "SKE-FAMSD" server.

References

[1]. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs *Nucleic Acids Res/Nucleic Acids Res* **25**, 3389-3402.

[2]. Ogata, K. and Umeyama, H. (2000). An automatic homology modeling method consisting of database searches and simulated annealing *J Mol Graph Model/J Mol Graph Model* **18**, 258-272, 305-256.

[3]. Lo Conte, L., Brenner, S.E., Hubbard, T.J., Chothia, C. and Murzin, A.G. (2002). SCOP database in 2002: refinements accommodate structural genomics *Nucleic Acids Res/Nucleic Acids Res* **30**, 264-267.

P56

Development and application of a molecular dynamics simulation system : prestoX

Satoru Kubota [1], Yoshifumi Fukunishi [2], Ikuo Fukuda [1], Masaru Horie [1], Katumi 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: prestoX. The aim of this system is general simulation for protein-ligand complexes and membrane proteins. The latest version, prestoX version 3 will be released on this December.

We developed a new *in silico* screening method, multiple targets screening (MTS) method, based on a multi-receptor vs. multi-ligand docking affinity matrix, and examined the robustness against the change of the scoring system. According to this method, compounds in a database were docked to multiple proteins. The compounds, which likely bind to the target protein among these proteins, are selected as the members of the candidate-hit compound group. We prepared the two new scores, one is the ΔG score, which is designed to reproduce the protein-ligand binding free energy, and another is the hit-optimized score, which is designed to maximize the hit ratio of *in silico* screening. Using the Sievgene docking score, ΔG score and the hit-optimized score, the MTS method is more robust than the multiple active-site correction scoring method. Also we provided methods for classifications of proteins and ligands and a method to design a focused library.

In addition, the component methods of prestoX were designed for parallel computers and GRID computer systems.

This work was supported by grants from NEDO and METI.

P57

Transcriptome according to a thermodynamic view

Tomokazu KONISHI [1]
Bioresource Sciences, Akita Prefectural University[1]

While decoding genomic information is routine work for a cell, it remains challenging for researchers. Indeed, although genome sequences have been determined for many organisms and the transcriptome has been measured under various conditions, we are still unable to predict a transcriptome from genome sequences, and the genome cannot be interpreted in terms of a transcriptome. Without an appropriate “grammar” for the genome, and due to the difficulty of dealing with the sheer volume of the genome information, an alternative approach is necessary. This poster presents a model that describes the embodiment of quantitative information in a genome from a thermodynamic perspective. The model is derived from a physical interpretation of biochemical mechanisms, and describes the concentration of each transcript according to the three forms of energies representing the functions of proteinous factors in a cell.

1. Tomokazu Konishi: Redeveloping Microarray Data Treatment Methods for Transcriptome Study, Pharmaceutical Discovery, Sep 15, 45-50 (2004)
2. Tomokazu Konishi: Three-parameter lognormal distribution ubiquitously found in cDNA microarray data and its application to parametric data treatment, BMC Bioinformatics, vol. 5, pp. 5 (2004)
3. Tomokazu Konishi, Masanori Yoshida, Kenya Shibahara: SuperNORM: A Computer Program for the Parametric Normalization of Microarray Data, Genome Informatics, vol. 14, pp. 292-293 (2003).
4. Tomokazu Konishi: Parametric treatment of cDNA microarray data, Genome Informatics, vol. 13, pp. 280-281 (2002).

P58

Automated genotyping of human CYP2C19 SNPs by a novel SNP typing system and development of a database of it

Yukari KUREBAYASHI [1,2], Naoko MATSUMOTO[1,2], Fumiko KAKIHARA[1,2]
Mitsuharu HIRAI[3], Setsuo HASEGAWA[1], Masafumi YOHDA[2]
Sekino Clinical Pharmacology Clinic [1],
Tokyo University of Agriculture and Technology [2] , ARKRAY, Inc. [3]

Proton pump inhibitors (PPIs) are widely used for therapy of peptic ulcer disease. They are metabolized by CYP2C19 in the liver. The frequency of 2C19 poor metabolizers is relatively high (approximately 20%) in Japanese population, and the genetic variations result in differences of their kinetics and pharmacological actions, e.g. clinical differences in *Helicobacter pylori* eradication rate. In this study, we developed a novel automatic SNPs-typing system and applied it for the genotyping CYP2C19.

The developed system is based on analysis of a melting curve of probe DNA bound to the target SNP site using a fluorescence quenching probe. The system enables automated and multiplex SNP genotyping from sample preparation. The results of SNP typing are consistent with the results obtained by allele specific primer PCR method.

We have also developed software for automated typing from the output data, melting curves, and built a data base of CYP2C19 genotypes. This full automation analyzing system can be translated to the clinical setting for order-made medicine using PPIs for peptic ulcers, which are the most common diseases and need optimized therapy in Japan.

References

1) Frieder Langenbucher: Handling of computational in vitro/in vivo correlation problems by Microsoft Excel: Ⅰ. Principles and some general algorithms, European Journal of Pharmaceutics and Biophamaceutics, Jan;53(1):1-7. (2002)
2) Furuta T, Shirai N, Sugimoto M, Ohashi K, Ishizaki T.: Pharmacogenomics of proton pump inhibitors. Pharmacogenomics. Mar;5(2):181-202.(2004)

P59

Gene expression profiles of the effect of cryoprotectants on the freeze-thaw damage in yeast

Yuko Momose [1], Masakazu Yamaoka[2]
The National Institute of Advanced Industrial Science and Technology [1],

Among the methods of culture collection storage, freezing is very convenient and low cost. But some cultures get serious damage and cannot survive after freeze-thaw so that we need suitable cryoprotectant for such organisms. In order to understand the effects of cryoprotectant, we analyzed the gene expression profile of yeast before and after freeze-thawing with and without that cryoprotectant.

The yeast Saccharomyces cerevisiae was frozen at -20 ºC (cooling rate, 1 ºC/min) for 1 week with or without cryoprotectant and thawed at 30 ºC for 5min. After freeze-thaw, we checked viability using flow cytometry. We analyzed the gene expression profile of yeast before and after freeze-thawing with and without those cryoprotectants.

We had already studied the genome-wide responses of yeast cells to various kinds of stresses, we compare that of freeze-thaw with or without cryoprotectant to the other stress, which were both physical and chemical, we analyzed with the responsive and stress specific genes we selected, and the hierarchical clustering analysis of gene expression profiles was carried out.

P60

Analysis of drug-resistant mutation patterns of HIV-1 protease under highly active antiretroviral therapy using Bayesian network modeling

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

INTRODUCTION: The emergence of antiviral drug resistance has been the largest obstacle in the treatment of AIDS for the last decade. It has been reported that the multiple mutations are usually required for the development of resistance for these anti-HIV drugs, because some mutations need to act as accessory mutations to increase the level of resistance or to compensate for losses in viral fitness[1]. Characterizing the evolutionary patterns of within-patient HIV is crucial to the challenge of anti-HIV drug resistance acquirement. In this study, we employed the Bayesian network approach for characterizing the patterns of drug resistant mutations and applied it to a large data set of HIV-1 protease gene (273 sequences) serially collected from a single patient under HAART (highly active antiretroviral therapy) over four years.

METHODS: Bayesian network is a graphical model which represents probabilistic dependency. We implemented a program to construct such a network using both selected mutation sites and drug combination sets (DCs) as discrete variables. Prior to the analysis, 19 mutation sites which had information entropy greater than 0.2 bits were chosen. The selected mutation sites were 10, 11, 12, 14, 15, 20, 22, 30, 43, 46, 55, 62, 73, 74, 77, 84, 88 and 90. Six different DCs were administrated for the patient, and each protease sequence was associated with one of the DCs.

RESULTS and DISCUSSION: The constructed network of 273 HIV-1 protease genes showed that there were different mutation patterns for different DCs. The dependencies between the mutations were clarified to a certain extent by this network. For instance, mutations at site 88 were inferred to have strong interaction with the ones at site 30 on the network, consisting with the finding that these mutations compensate for losses in viral fitness each other. We suggest that Bayesian network approach may be useful for detecting and characterizing drug-resistant mutation patterns, with the potential of contributing to evaluation of anti-HIV therapy.

References

1) Wu TD et al. *J Virol.* 2003 Apr;77(8):4836-47

P61

Comparison of the validation methods with linear and non-linear classifier

Chihoko YOSHIMURA[1], Taizo HANAI[2], Masahiro OKAMOTO[2]

Drug Discovery Lab., Hanno Research Center, Taiho Pharmaceutical Co., LTD[1], Laboratory for Bioinformatics, Graduate School of Systems Life Sciences, Kyushu University[2]

Abstract

In microarray data analysis, we can use thousands of gene's expression values but the number of samples is relatively limited. Therefore, to build classifier for the purpose of outcome prediction or cancer subtype detection from microarray data tends to result in "overfitting". So it is important to estimate the classification error for the general data, and it is called "validation method" for classifier to know this error. In this study, we tested some validation methods with linear and non-linear classifier. The dataset used in this study was divided into training dataset and test dataset. And we compared the estimated error, calculated from training data based on validation method and true error, derived from test dataset. As the dataset, we used the lymphoma patient's gene expression data[1] and built the classifier to predict 5 years survival. This dataset consisted of 210 samples. 140 samples are used for building classifier and model estimation, and the rest 70 samples were used for verification of the true error.

To build the classifier, we used linear Support Vector Machine (SVM) and Multiple Linear Discriminant and Fisher's Linear Discriminant as linear classifier, and SVM with polynomial and gauss kernel function as non-linear classifier. With these classifier, as the validation methods, resubstitution, 10-fold Cross Validation, Leave One Out Cross Validation, ξ-α estimate, and .632 bootstrap method were tested.

Here we report that the linear SVM with .632 bootstrap had the smallest difference between estimated error and true error.

Reffernce

[1] The dataset is available at http://llmmp.nih.gov/DLBCL/

P62

Computational Proteomic Differential Analysis: *i-OPAL*

Mitsuhiro Kanazawa[1,3], Hisae Anyoji[1], Hsiao-kun Tu[1], Takao Kawakami[1,2], Toshihide Nishimura[1,2], Atsushi Ogiwara[1,2]

[1]Medical ProteoScope Co., Ltd., [2]Clinical Proteome Center, Tokyo Medical University, [3]Graduate School of Pure and Applied Sciences, University of Tsukuba

In recent years, analytical technology of proteomics has been significantly improved, and it is focused as one of methods to discover biomarker or target molecule that characterize certain diseases, symptoms, responses of treatments, adverse events by certain drugs and so on. Several experimental methods for proteome differential display have been developed to discover biomarkers based on the labeling methods, but the analytical efficiency of those methods is viewed with suspicion because of sampling bias and artificial errors.

We therefore developed computational proteomic differential analysis system *i-OPAL* (*i*nternal standard guided *O*ptimal *P*rofile *Al*ignment) to discover biomarkers based on LC-MS measurements. This analysis requires no artificial labels to compare the profiles derived from different conditions. It not only overcomes sampling bias and artificial errors, but also realizes the extensive and unbiased quantitative analysis.

i-OPAL's basic approach consists of 2 steps: one is *Intra-group* alignment (*Step 1*) and another is *Inter-group* alignment (*Step2*). *Step 1* bundles peptide signals into sheaves using dynamic programming based alignment sample by sample in a group, and creates consensus signal map of its group. Here, the volume of each bundled signal indicates peptide expression levels by its group. *Step 2* aligns those consensus signal maps of each group, and compares bundled signal volumes as quantitative difference of related location between signal maps. The bundled signal that has significant quantitative difference among groups is biomarker or target molecule itself.

We believe this approach becomes a huge breakthrough of biomarkers discovery and the elucidation of disease/drug mechanism.

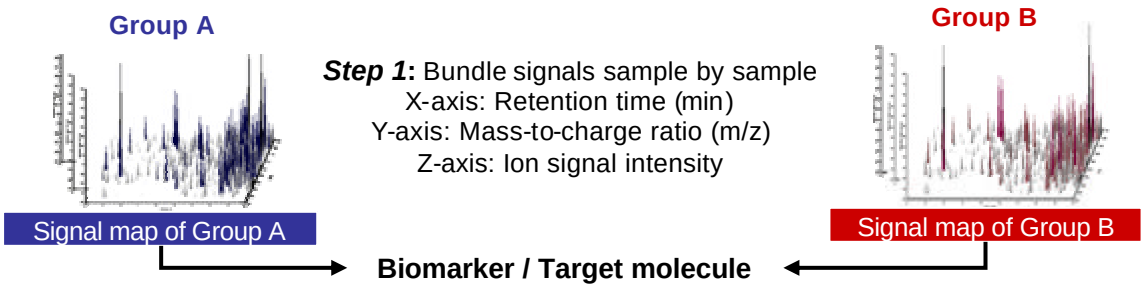


Figure: *i-OPAL* basic analysis steps

References

1) V. Pravdova, B. Walczak, and D.L.Massart, A comparison of two algorithms for warping of analytical signals, *Anal. Chim. Acta*: 77 - 92, 2002

P63

NMR-based Metabolomics—2/3
2) How to analyze NMR-spectra of mixture samples by chemometric method; in examples of vegetable metabolome

Masako Fujiwara [1], Itiro Ando[2], and Kazunori Arifuku [1]
[1]JEOL DATUM LTD,
[2]Envioronmental Research Center LTD

We measured ^{1}H NMR spectra of over 175 kinds of green, oolong, black and other tea infusions using JEOL ECA-500spectrometer. The chemometric analysis combined with PCA and SIMCA was used to discern the category of the tea spectra by 'Aile2 for metabolome' software. This approach for latent-information extraction can be applied to the biological samples as a powerful tool for whole-system diagnostics and metabolic function.

References

1. Gweawlle Le Gall, Lan J.Colquhoun, and Marianne Deferenez: Metabolite Profiling Using ^{1}H NMR Spectroscopy for Quality Assessment of Green Tea, *Camellia sinesnsis* (L.), J. Agri. Food Chem. 52, pp 692-700,(2004)
2. Henrik Antti, Elaine Holms and Jeremy Nicholson: Multivariate Solutions to Metabonomic Profiling and Functional Genomics. Homepage of Chemometrics, Editorial October 2002
3. Kazunori Arifuku, Keiko Hirakawa, Kaoru Koike, Kyoko Uekusa, Masako Fujiwara, and Youkichi Ohno: Integrated NMR-Spectroscopic and Chemometirc Analysis of Mixtures; Two Hundred Kinds of Tea, 46th Experimental NMR Conference, poster 287, 2005
4. Nicholas V. Reo: NMR-based Metabolomics, Drug and Chemical Toxicology, vol25, No.4, pp375-382, 2002

P64

Structural analysis by molecular dynamics simulations: Protein-protein interactions of Ras-Raf and Ras-RalGDS complexes

Noriyuki Futatsugi [1], Mikako Shirouzu [2], Atsushi Suenaga [1], Noriaki Okimoto [1], Tetsu Narumi [1], Toshikazu Ebisuzaki [3], Shigeyuki Yokoyama [2], Akihiko Konagaya [1], Makoto Taiji [1]

Computational and Experimental Systems Biology Group, RIKEN Genomic Sciences Center [1], Protein Research Group, RIKEN Genomic Sciences Center [2], Ebisuzaki Computational Astrophysics Laboratory, RIKEN [3]

In this study, we investigated protein-protein interactions of Ras protein and its targets[1] (Raf1 RBD, RalGDS RBD) by using molecular dynamics (MD). Comparing simulations with experimental results, we could know the detailed atomic-level origin of the affinity change in mutants, and confirm the prediction ability of the simulations which is important for large-scale *in silico* screening. For the Ras-Raf1 and Ras-RalGDS complexes, we have performed simulations both for wild type and mutant Ras (E37G/I36A). The mutations of these residues are known to affect loss of the binding affinities. In the MD simulations with using very fast special-purpose computers, the MDGRAPE-2, we found obvious changes in the binding region between the wild type and the mutant Ras complexes. In the mutant Ras complex, some distances in the binding region became longer and hydrogen bonds were broken. In addition, we performed to calculate free energies by the MM-PBSA analysis and free energies in the mutant Ras complexes were larger than those of the wild type Ras complexes. Furthermore, we found that the binding patterns of the Ras-Raf1 and the Ras-RalGDS are different in motion of explicit water molecules and these water molecules are involved in the binding mechanism. In brief, it guesses that the mutation of amino acids decreases binding affinities vastly. These results indicate that the mutations cause local structural changes in the binding region of Ras and this local change makes affinity lower.

References

1) M. Shirouzu et al. *Biochemistry*, **38** 5103-5110 (1999)

P65

NMR-based Metabolomics—3/3
3) High-field NMR-spectroscopy for metabolomic analysis; An application to urine samples from model rats

Tadashi Nemoto [1], Taeko Kataoka [1], Itiro Ando [2], Katsuo Asakura [3], Kazunori Arifuku [4], Kenji Kanazawa [1], Masako Fujiwara [4]

National Institute of Advanced Industrial Science and Technology (AIST) [1], Environmental Research Center LTD[2], JEOL LTD [3], JEOL DATUM LTD [4]

NMR spectroscopy is known as powerful tool for structural proteomics. We tried high-field NMR spectroscopy to metabolomic analysis for discriminating healthy or unhealthy and other physiological conditions.

Rat urine samples were measured as 1D ^{1}H NMR spectrum by using a JEOL ECA-800 spectrometer (800.03MHz). Without any spectral assignment, spectrum were converted into numeric datasets and then subjected to pattern recognition analysis by "Alice2 for Metabolome". It resulted clear classification reflecting the disease status and/or circadian rhythm of the rats.

REFERENCES

1. E. Holmes, A. W. Nicholls, J. C. Lindon, S. Ramos,2 M. Spraul, P. Neidig, S. C. Connor, J. Connelly, S. J. P. Damment, J. Haselden, J. K. Nicholson, Development of a model for classifcation of toxin-inducedlesions using 1H NMR spectroscopy of urine combined with pattern recognition. *NMR Biomed.*, 11:235–244 (1998)
2. J. K. Nicholson, J. Connelly, J. C. Lindon, E. Holmes, Metabonomics: a platform for studying drug toxicity and gene function, *Nat. Rev. Drug Discov*., 1:153-61(2002)
3. T. M. D. Ebbels, E. Holmes, J. C. Lindon, J. K. Nicholson, Evaluation of metabolic variation in normal rat strains from a statistical analysis of ^{1}H NMR spectra of urine. *J. Pharm. Biomed. Anal.*, 36: 823-833(2004)
4. Kazunori Arifuku, Keiko Hirakawa, Kaoru Koike, Kyoko Uekusa, Masako Fujiwara Youkichi Ohno: Development of Analytical software for NMR-based Metabolomics; "Alice2 for Metabolome" The Annual Meetings of the NMR Society of Japan, proceeding, 2P092, (2004)

P66

NMR-based Metabolomics—1/3
1) Newly developed software; “Alice2 for Metabolome”; Integrated NMR spectroscopic and chemomerirc analysis for mixture

Kazunori Arifuku[1], Itiro Ando[2], Masako Fujiwara [1]
[1] JEOL DATUM LTD,
[2] Environmental Research Center LTD

For the analysis of mixture samples by NMR spectroscopy, large number of spectral processing is needed. Each spectrum is bucket-integrated and the datasets are submitted to chemometric solutions. We have developed integrated software packages of ‘Alice2 for Metabolome’(JEOL). The essential modules of Principal Component Analysis (PCA) and Soft Independent Modeling of Class Analogy (SIMCA) are implemented in the software and well-suited and fitted to NMR spectral processing and analyzing for metabolome on the single interface.

References

1. Kazunori Arifuku, Keiko Hirakawa, Kaoru Koike, Kyoko Uekusa, Masako Fujiwara Youkichi Ohno: Development of Analytical software for NMR-based Metabolomics; “ Alice2 for Metabolome” The Annual Meetings of the NMR Society of Japan, proceeding, 2P092, (2004)
2. Kazunori Arifuku, Keiko Hirakawa, Kaoru Koike, Kyoko Uekusa, Masako Fujiwara Youkichi Ohno: Integrated NMR-Spectroscopic and Chemometirc Analysis of Mixtures ;Two Hundred Kinds of Tea, 46th Experimental NMR Conference, poster 287 proceeding, (2005)
3. Svante Wold, Michael Sjostrom,; SIMCA: A Method for Analyzing Chemical Data in Terms of Similarity and Analogy, pp 243-282
4. Svante Wold; Pattern Recognition by Means of Disjoint Principal Components Models, Pattern Recognition, Pergamon Press Vol 8. pp. 127-139 (1976)

P67

Controlling dissociation and association of amyloid protofibrils using high pressure

Yuji O. Kamatari [1,2,3], Shigeyuki Yokoyama[2,3,4], Hideki Tachibana[5], Kazuyuki Akasaka[2,6]

Gifu University[1], RIKEN Harima Institute[2], Riken Genome Science Center[3], The University of Tokyo[4], Kobe University[5], Kinki University[6]

A number of diseases such as Alzheimer's disease, Parkinson's disease, prion disease, and autosomal dominant hereditary amyloidosis from lysozyme variants, are associated with the formation of amyloid fibrils. Understanding of mechanism of amyloid formation and controlling it are a central issue in amyloidosis. In this study, the dissociation and reassociation processes of amyloid protofibrils initiated by pressure-jump have been monitored with real-time ^{1}H NMR spectroscopy using intrinsically denatured disulfide-deficient variant of hen lysozyme. Upon pressure-jump up to 2 kbar, the matured protofibrils grown in several months become fully dissociated into monomers within a few days. Upon pressure-jump down to 30 bar, the dissociated monomers immediately start reassociating. The association and dissociation cycle can be repeated reproducibly by alternating pressure, establishing a notion that the protofibril formation is simply a slow kinetic process toward thermodynamic equilibrium. The outstanding simplicity and effectiveness of pressure in controlling the protofibril formation opens a new route for investigating mechanisms of amyloid fibril-forming reactions. The noted variation in the pressure-induced dissociation rate with the progress of the association reaction suggests multiple mechanisms for the elongation of the protofibril. This reversible system of amyloid protofibril formation offers a particularly simple model system for thermodynamic and kinetic studies of protofibril formation as well as for screening drugs for amyloidosis.

Reference

Kamatari YO, Yokoyama S, Tachibana H, Akasaka K. 2005, Pressure-jump NMR Study of Dissociation and Association of Amyloid Protofibrils. J Mol Biol. 349, 916-921.

P68

The evolutionary mechanism of protein-protein interaction network

Takeshi Hase[1],Soichi Ogishima[1] , So Nakagawa[1],Hiroshi Tanaka[1]
Department of Bioinformatics, Tokyo Medical and Dental University[1]

There are a lot of researches about topological properties of protein-protein interaction network (PIN). Distribution of PIN shows a scale free property [1], as well as a high clustering. It means a network have a lot of triangles. [2].
To uncover the evolutionary mechanism of PIN, concept of duplication is used to construct the evolutionary model. This model can reproduces a scale free network, but is not appropriate in forming high clustering [3]. Vazquez improves duplication model using self-interaction as a driving forces of creating triangles and reproduces a high clustering very well. However, there are no evidences that all triangles are made by duplication of self-interaction proteins [4]. Then, we clarify the rate of contribution of self-interaction for high clustering and find that their contribution rate is only 30%. It suggests the existence of another mechanism for high clustering.
We find another mechanism that creates triangles by connecting new nodes and 2^{nd}-neighbors. In this study, we show the results of our mechanism and the biological feature supports our hypothesis.

References

1. H. Jeong, S.P. Mason, A.-L. Barabasi, Z.N. Oltvai: Lethality and centrality in protein networks, Nature, vol.411, pp.41-42 (2000)
2. A. Wagner: The Yeast Protein Interaction Network Evolves Rapidly and Contains Few Redundant Duplicate Genes, Molecular Biology and Evolution, vol.18, pp.1283-1292 (2001)
3. R. Pastor-Satorras, E. Smith, R.V. Sole: Evolving protein interaction networks through gene duplication, Journal of Theoretical Biology, vol.222, pp.199-210 (2003)
4. A. Vazquez, A. Flammini, A. Maritan, A. Vespignani: Modeling of protein interaction networks, ComPlexUs, vol.1, pp.38-44 (2003)

P69

Is there a code for protein-DNA interaction? Yes, there exists the *analog* code for protein-DNA interaction.

Tomoki YOSHIDA [1], Misako AIDA [1], Akinori SARAI [2]
Center for Quantum Life Sciences (*QuLiS*), and Department of Chemistry, Graduate School of Science, Hiroshima University [1], Department of Bioscience and Bioinformatics, Kyushu Institute of Technology (KIT) [2]

Based on experimental structural data, it was suggested that there was "no code for recognition" in protein-DNA interaction.[1)] It seemed impossible to predict a recognition sequence for protein from the base-amino acid interaction level; no clear one-to-one correspondence is observed in the base-amino acid interaction (redundancy) or no unique conformation is observed for any base-amino acid pair (conformational flexibility). Here, we propose the "**analog recognition code**" for protein-DNA interaction. We take into account both the redundancy and the flexibility in constructing the code.[2)] The analog recognition code is a set of the calculated interaction free energy ($\Delta\Delta G$) landscapes between an amino acid side chain and 32 independent kinds of three stacking base pairs. The $\Delta\Delta G$ values are calculated around the coordinate positions in the major groove; the region having low $\Delta\Delta G$ value represents the preferable recognition position for each side chain.[3), 4)] By means of the analog recognition code we can predict the recognition position for each side chain along any length of DNA in the major groove. There may be no code having a clear one-to-one correspondence, which may be called a "**deterministic**" recognition code, but there exists the "**analog**" recognition code for protein-DNA interaction.

REFERENCES

1) Matthews B W, "Protein-DNA interaction. No code for recognition," *Nature*, **335**, 294-295 (1988).

2) Fabio Pichierri, Misako Aida, M. Michael Gromiha and Akinori Sarai, "Free-energy maps of base-amino acid interactions for DNA-protein recognition," *Jornal of the American Chemical Society*, **121**, 6152-6157 (1999).

3) Tomoki Yoshida, Tomohiro Nishimura, Misako Aida, Fabio Pichierri, Michael M. Gromiha and Akinori Sarai, "Evaluation of free energy landscape for base-amino acid interactions using ab initio force field and extensive sampling," *Biopolymers*, **61**, 84-95 (2002).

4) Misako Aida, Giorgina Corongiu and Enrico Clementi, "Ab initio force field for simulations of proteins and nucleic acids," *International Journal of Quantum Chemistry*, **42**, 1353-1381 (1992).

P70

Investigation of Structural and Dynamic Properties of Importin-β Based on a Molecular Dynamics Study

Yoshinori Hirano [1], Noriaki Okimoto[1,2], Atsushi Suenaga[2],
Naoko Imamoto[3], Toshikazu Ebisuzaki[1]
Computational Astrophysics Laboratory, RIKEN[1],
Bioinformatics Group, Genomic Sciences Center, RIKEN[2],
Cellular Dynamics Laboratory, RIKEN [3]

We performed a series of molecular dynamics (MD) simulations of importin-βs to investigate their structural and dynamic properties. The starting structures in our simulations were two importin-βs, one with the importin-β binding (IBB) domain of importin-α and one without it. Both structures were constructed based on the X-ray crystallographic structure of the full-length importin-β complexed with the IBB domain[1)]. All MD simulations were carried out using the modified Amber 6.0[2)] for MDM. In the simulation results, the superhelical structures of the importin-βs showed a far greater degree of loosening in aqueous solution than the crystal structure. It indicated that the structure of importin-β with IBB domain was tightly packed under crystallization. The dynamics of the importin-β was produced by the motions of the N-terminal domain, C-terminal domain, and connecting region. The free importin-β exhibited a stronger dynamic property than the IBB-bound importin-β, due to the absence of the crystal packing interaction (the IBB domain binding to importin-β). In the IBB-bound importin-β, the C-terminal domain and connecting region formed numerous hydrogen bonds with the IBB domain, and appeared to augment the increase in its binding affinity. The N-terminal domain, on the other hand, showed no interactions with the IBB domain. The dynamic properties of the importin-β observed in our simulations provided an excellent interpretation of the structural change of the importin-β which could be induced upon release of IBB domain.

References

1) G. Cingolani, C. Petosa, K. Weis, and C. W. Müller, Nature, vol. 399, pp.221-229 (1999).
2) D. A. Case, D. A. Pearlman, J. W. Caldwell, T. E. Cheatham III, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, V. Tsui, R. J. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, U. C. Singh, P. K. Weiner, and P. A. Kollman: (1999) AMBER 6. San Francisco, University of California

P71

Matlab/GNU Octave as a Platform for Analyses of Metabolic Networks

Jun Ohta[1]
Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences[1]; Okayama 700-8558, Japan; E-mail: jo25@md.okayama-u.ac.jp

Matlab and GNU Octave are softwares suitable for manipulation of matrices. Both can also manipulate strings and be easily customized using script files and functions, indicating their potential as a platform for a variety of analyses in the area of bioinformatics. I am interested in structure-function relationship in metabolic networks, and have been developing a system for analyses of metabolic networks using Matlab and GNU Octave[1,2,3,4]. Atom is the smallest node in metabolic networks. Atom-level connectivity in metabolic networks is classified into inter-metabolite connectivity through enzymatic reactions and intra-metabolite connectivity through chemical bonds. For detailed analyses of metabolic networks, all the atom-level connectivities are expressed in a matrix, connectivity matrix, in my approach. Then, a variety of analyses are performed on Matlab/GNU Octave using script files and functions including those written by each researcher. This approach has been named 'connectivity matrix method'. Connectivity matrix (CM) can describe both inter- and intra-metabolite atom-level connectivities including information of compartmentation, responsible reactions, stoichiometry, and valency of chemical bonds. Each row of CM is a combination of 3 row vectors, [va_1, va_2, vc], where va_1, va_2 and vc indicate 2 different atoms and one connectivity, respectively. Calculation of the paths connecting 2 specific atoms, discovery of hidden atom-level cycles and calculation of net balance of the reaction sequence constituting such a path with the aid of balance matrix can be performed using m-files for Matlab/GNU Octave. Information of atom-mapping can be written on mol files for visualization of those paths obtained. Information of CM method and CMs for individual reactions are available on the web, www.metabo-info.org. Procedures of CM method on Matlab/GNU Octave will be presented.

References

1) Programs and Abstracts, First IECA Conference on Systems Biology of E. coli, 112 (2003)
2) Molecular & Cellular Proteomics 2(9), 830 (2003)
3) Proceedings for 5th International Conference on Systems Biology, 141 (2004)
4) Course Book, Program and Abstracts, First FEBS Advanced Course on Systems Biology: From Molecules & Modeling To Cells, 140 (2005)

P72

A Grid-based Information Integration System for Drug Discovery by using Freely Accessible Compound Databases

Gen Kawamura[1, 2], Masato Kitajima[1,3], Takahiro Kosaka[1], Susumu Date[1],
Shinji Shimojo[4], Hideo Matsuda[1]
Graduate School of Information Science and Technology, Osaka University. [1],
Aztec-System Co., Ltd. [2], Fujitsu Kyushu System Engineering Limited. [3],
Cybermedia Center, Osaka University. [4]
E-mail: matsuda@ist.osaka-u.ac.jp

In this study, we present an information integration system for drug discovery by using available compound databases and an employing grid technology. Drug discovery involves a number of stages including disease modeling, target protein identification, lead compound identification, and clinical trials. To cope with such a discovery process, we utilize grid technology for integrating various databases[1] and develop metadata for describing relationships between heterogeneous data that belong to different domains (such as, protein-compound interactions) [2]. In particular, to discovery the candidate subset of ligands, we have prepared a library of 1820174 compounds, each with chemical structure, using catalogs of compounds from freely accessible databases[3,4]. The compounds have been annotated with attributes such as MACCS-KEY fingerprint, molecular weight, calculated LogP and added information.

The effectiveness of our system is demonstrated by evaluating results of several example queries.

References

1) I. Foster, C. Kesselman, J.M. Nick, and S. Tuecke, “The Physiology of the Grid. An Open Grid Services Architecture for Distributed Systems Integration.” *Open Grid Service Infrastructure WG, Global Grid Forum*, 2002, http://www.ggf.org/
2) G. Schneider, and S.-S. So, “Modeling Structure-Activity Relationships,” G. Schneider and S.-S. So (ed.), *Adaptive Systems in Drug Design*, Landes Bioscience, Georgetown TX, 2002.
3) M. von Grotthuss, J. Pas, and L. Rychlewski, “Ligand-info, searching for similar small compounds using index profiles.” *Bioinformatics* 2003, 19, 1041-1042.
4) PubChem: http://pubchem.ncbi.nlm.nih.gov/

P73

Conformational analysis of an antimicrobial peptide, Pediocin PA-1

Hitoshi Goto[1], Tatsunori Matsui[1], Toshiyuki Kamakura[1], Naofumi Nakayama[1],
Tatsuya Tominaga[2], Hideo Nakajima[2]
[1] Toyohashi University of Technology, [2] Saitama Industrial Technology Center North Institute

Pediocin PA-1 is a broad-spectrum lactic acid bacteria bacteriocin that shows a particularly strong activity against *Listeria monocytogenes*, a foodborne pathogen of special concern among the food industries. This antimicrobial peptide is the most extensively studied class IIa bacteriocin (pediocin family), and it has been sufficiently well characterized to be used as a food biopreservative. This work focuses on the elucidation of structures and biological activation mechanism of pediocin PA-1 and the derivatives, which are subject to a single-amino acid substitution of various residues. In this conference, our original technique of conformational analysis for over 10,000 unique conformers found by using the exhaustive conformation search method CONFLEX5/MMFF94s will be proposed. In comparison with the structures obtained from some homology modeling tools, it is suggested that CONFLEX5/MMFF94s can produce many of more stable conformations of pediocin and the related peptides (Fig. 1-2). On the other hand, comparison between the conformational analysis and the experimental bioactivity shows that some characteristic residues are closely related to antimicrobial activity.

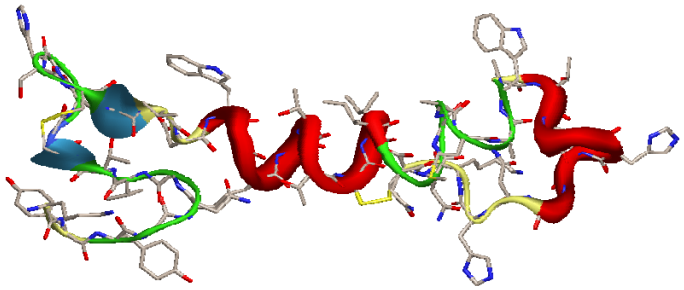


Fig. 1 Homology Model based on SakP[N24C+44C] NMR structure:
MMFF94s ΔE=161.31 kcal/mol, ΔG=153.23 kcal/mol

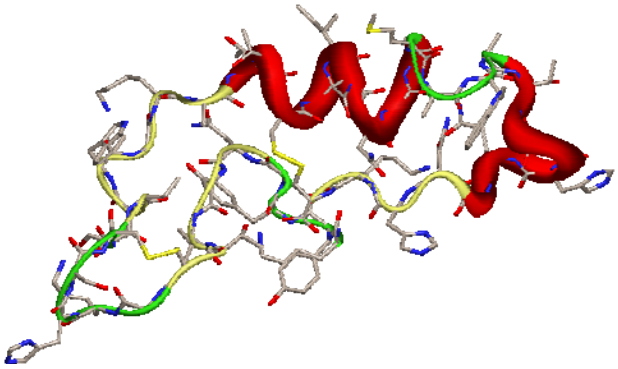


Fig. 2 Most Stable Conformer in CONFLEX/MMFF94s Search: MMFF94s ΔE=0.00 kcal/mol, ΔG=0.76 kcal/mol

This work is performed in the Rational Evolutionary Design of Advanced Biomolecules (REDS) Project, Saitama Prefecture Collaboration of Regional Entities for the Advancement of Technological Excellence supported by JST.

P74

Normal mode analysis of the complex of HIV-1 protease with inhibitor vibrational behavior

Toshiyuki KAMAKURA[1], Hitoshi GOTO[1]
Toyohashi University of Technology[1]

In order to elucidate the predominant structures and biological function of protein-ligand complex, it is important to understand both the kinetical and thermodynamical behaviors in biological environment. Especially, temperature dependency on their biological activity reveals that the thermal fluctuation plays an important role on the dynamic interaction between protein and ligand. In this study, we propose a new conformational analysis approach to understanding protein-ligand complex functions. Practical mixing of normal modes based on thermodynamic analysis for the protein-ligand complex gives an important knowledge for finding many plausible bio-activated conformations. This vibrational simulation will be suggested for the difference of dynamical behaviors between HIV-1 protease and the complex with inhibitor in water. We measured displacements of C_α atoms on 28-60 residues, generally called “flap-region”, each residue from equilibrium structure (Fig 1). In only protease system (**PR**), Gly49 fluctuates with neighbor residues. In protease-ligand system (**PLG**) and protease-ligand in waters (**PLH**), Gly49 does not fluctuate. In these all systems, Ser37’ fluctuates, but smaller than Gly49. Asp25 and Asp25’, which are active site, are motionless in any systems. In the previous MD simulation[1)], the mutation protease, which Gly49 were mutated to Trp, had favorable binding free energy. From these comparing, our approach may be favor for searching the inhibitors.

This work was supported by grants (No. 17300094) from the Ministry of Education, Science, Sports and Culture of Japan.

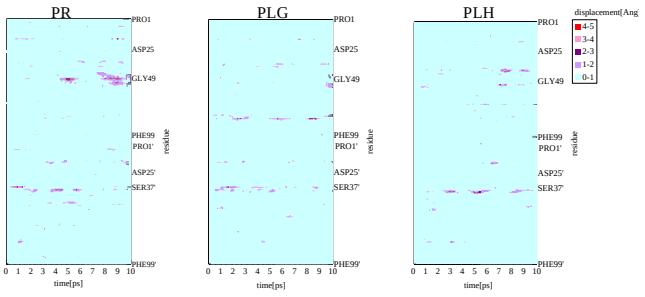


Fig 1 Displacements of Cα of HIV protease and -ligand complexes.
Vertical axis is residue number, and horizontal axis is simulation time. The distances of C_α atoms from equilibrium structure are colored.

References

1) Wei Wang, Peter A. Kollman: Free Energy Calculations on Dimer Stability of the HIV Protease using Molecular Dynamics and a Continuum Solvent Model, Journal of Molecular Biology, vol.303, pp.567-582 (2000).

P75

Genome Functional Analysis based on Systematic Map of Protein Interactions

Carlos A. Del Carpio[1], Pei Qiang[1], Eichiro Ichiishi[2], Michihisa Koyama[1], Momoji Kubo[1,3], Akira Miyamoto[1,2]

Department of Applied Chemistry, Tohoku University[1], New Industry Creation Hatchery Center, Tohoku University[2], JST-PRESTO[3]

With the constant accumulation of genomic data around the world, the need for genome-wide functional analysis is of paramount importance in order to elucidate the intertwined relations among genes, proteins and metabolites. In the present work we present a new methodology for docking unbound protein molecules**.** The process is oriented to serve as a tool for elucidation of the gamut of protein interactions encoded in the genome. The methodology is characterized by its simplicity and easiness of embedment in any rigid body docking process based on point complementarity. It is oriented to allow limited free but not unrealistic interpenetration of the side chains of protein surface amino acid residues. The methodology assists in deletion of atomic-scale details on the surface of the interacting monomers, leading to the extraction of the most characteristic flattened shape for the molecule as well as the definition of a soft layer of atoms to allow smooth interpenetration of the interacting molecules during the docking process. Although the methodology does not perform structural or conformational rearrangements in the interacting monomers, results output by the algorithm are in fair agreement with the relative position of the monomer in experimentally reported complexes. The algorithm performs especially well in cases where the complexity of the protein surfaces is high, that is in hetero dimmer complex prediction. It is oriented to play the role of a fast screening engine for proteins known to interact but for which no information other than that of the structures at the isolated state is available. The importance of the methodology will increase in structural-function studies of thousand of proteins derived from large scale genome sequencing projects being executed all around the globe. Fig.1 shows a complex inferred automatically from the isolated sub-units. The methodology has been embedded in the general system for protein-protein interaction developed by Del Carpio et al[1,2] in recent years.

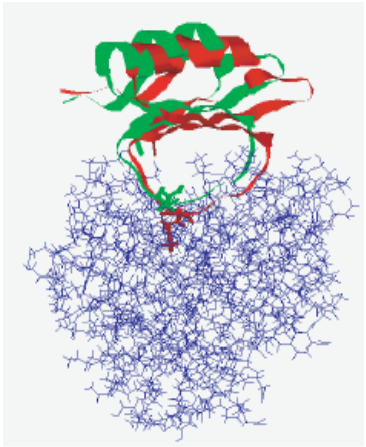


Fig.1. MIAX predicted complex (red) Vs. Crystal structure (blue)

References

1. Del Carpio CA., Peissker T, Yoshimori A, Ichiishi E. *Genome Informatics14:238-249* (2003).
2. Del Carpio CA, Ichiishi E, Yoshimori A, Yoshikawa T; PROTEINS: Structure, Function and Genetics (2002). 48(4):696-732

P76

GenoSearch: High Performance Inter-Genomic Sequence Retrieval System using Suffix Array Technique

Aki HASEGAWA, Akihiko KONAGAYA
Informatics Infrastructure Team, Genome Core Technology Facilities
RIKEN Genomic Sciences Center

Bio-sequence analysis often requires the location and frequency of the occurrences of a given sequence. Several tools have been developed including SSAHA [1], PatternHunter [2], BLAT [3] as well as FASTA [4] and BLAST family [5][6]. However, higher search performance is strongly required to deal with multiple genome sequences and non-redundant protein databases which become tens of giga bytes in total. GenoSearch solves this issue by providing pre-built indices based on suffix arrays [7]. The GenoSearch can extract the location and frequency information of a given sequence from hundreds of genome sequences in sub seconds to several seconds according to the complexity of a query pattern. The query pattern is divided into simpler sub patterns prior to the search. Then the sub query is performed for each sub pattern in planned order. Final result is built up with results of sub query and cached results of frequently searched sub pattern. In this way, the GenoSearch speeds up sequence pattern retrieval. The indices for bio-sequences are updated regularly according to the renewal of the bio-sequence file. The GenoSearch is also available as the Web Service at the GSC site (http://www.gsc.riken.jp/Informatics/).

References

[1] Z. Ning, A. J. Cox and J. C. Mullikin, *Genome Res.*, **11**, 1725–1729 (2001).
[2] M. Li, B. Ma, D. Kisman and J. Tromp, *J. Bioinf. Comput. Biol.*, **2**, 417–439 (2004).
[3] W. J. Kent, *Genome Res.*, **12**, 656–664 (2002).
[4] W. R. Pearson and D. J. Lipman, *Proc. Natl. Acad. Sci.*, **85**, 2444–2448 (1988).
[5] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, *Nucleic Acids Res.*, **25**, 3389–3402 (1997).
[6] Z. Zhang, S. Schwartz, L. Wagner and W. Miller, *J. Comput. Biol.*, **7**, 203–214 (2000).
[7] U. Manber and G. Myers, *SIAM J. Comput.*, **22**, 935–948 (1993).

P77

Evaluation of polymorphism for an organic molecular crystal

Shigeaki Obata [1], Hitoshi Goto[1], Naofumi Nakayama[1]
Toyohashi University of Technology [1]

Crystal structure prediction for organic molecules can provide immense valuable scientific knowledge concerning on crystal growth, polymorphism, rate of dissolution, electronic and optical properties, and etc. Although many computational applications for challenging these problems have been already developed by some research groups, approach to the practical solution seems to be difficult to overcome so far. As a part of our long-standing project toward the goal of predicting the crystal packing of molecular solids, we will introduce our recent work on description of conformational behaviors for an organic molecule, hydroxymalonic acid, packed into the known solid state. As the first step in our approach, the nine stable conformers of hydroxymalonic acid were generated by our original conformation search program CONFLEX, and virtually-constructed crystal structures based on those geometries were subjected to minimization by our original crystal packing application KESSHOU. Lattice energy of each crystal structure was evaluated as the total energy of all intermolecular interactions in MMFF94s potential functions. As the result of the computation, it was found that the most stable crystal structure that has the lowest lattice energy and was constructed from the 5th stable conformer in vacuum has been the best agreement with X-ray crystal structure (Figure 1).

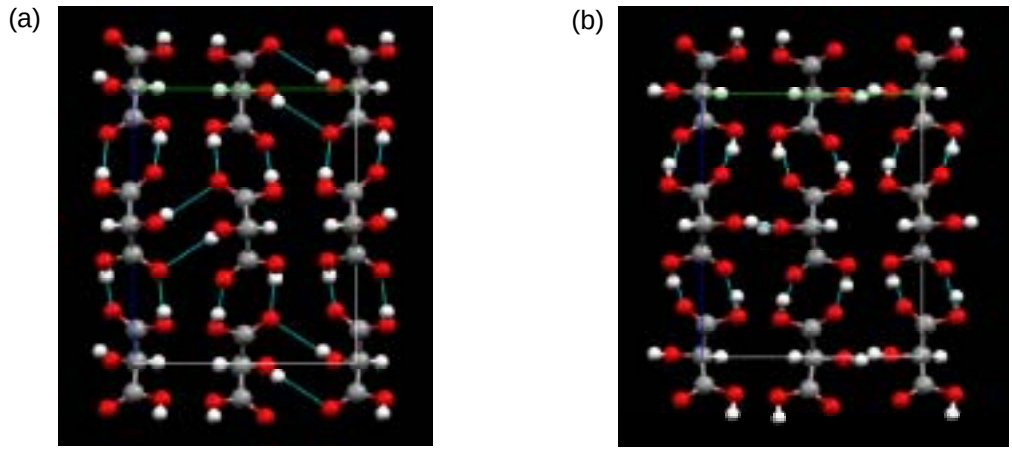


Figure 1 Comparison of Crystal Structures: (*a*) X-ray structure in the Cambridge Structural Database (Refcode: HMALAC10, *a*=4.494, *b*=8.819, *c*=10.882 Å), (*b*) the most stable crystal structure optimized by KESSHOU (*a*=4.557, *b*=8.969, *c*=10.842 Å.).

This work was supported by grants (No. 17300094) from the Ministry of Education, Science, Sports and Culture of Japan.

P78

Workflow for Open Genome Grid Service

Ryo UMETSU[1], Shingo OHKI[1], Akihiko KONAGAYA[1,2]
Advanced Genome Information Technology Research Group, GSC, RIKEN[1],
Bioinformatics Laboratory, Tokyo Institute of Technology[2]

We propose a workflow platform for genome informatics incorporated with the Bioinformatics Web Services constructed on Open Bioinformatics Grid (OBIGrid). Web Services or Grid Services have a great potential to provide a powerful bioinformatics environment based on the service oriented architecture (SOA). The grid services enable for the bioinformaticians to access the public databases and to use the free bioinformatics application tools without considering computation resources nor operation costs. The grid services also provide full options and the same application programming interfaces as provided by the corresponding UNIX commands on a local computer.

However, care must be taken when dealing with genome wide data, especially for scalability and session management. It often happens that a single father task produces thousands of daughter tasks in a bioinformatics workflow. In addition, computational results may become giga bytes order in total. We solved the above issues by providing asynchronous Web Service facilities, a thread pool incorporated with a task scheduler and a global file system on J2EE. The feasibility of our system has been demonstrated by the GLIMMER2-BLASTALL-SELECTIONFILTER workflow for micro bacterial genome sequences.

P79

Drug Interaction Ontology (DIO)
- OWL formalizations of interaction causal logic -

Sumi Yoshikawa[1)2)], Kentaro Watanabe[2)], Kazumi Matsumura[1)2)], Akihiko Konagaya[1)2)]
[1)]Genomics Science Center (GSC), RIKEN (The Institute of Physical and Chemical Research)
[2)] Graduate School of Information Science and Technology, Tokyo Institute of Technology

We previously reported the development of Drug Interaction Ontology (DIO) [1) 2) 3)] designed for inference system of possible drug-drug interactions, based on drug-biomolecule and biomolecule-biomolecule interactions. Ontology was employed for encoding literature information associated with drug behaviors in a body system; metabolism, transport, target reactions, gene expressions etc. Taking account of spatiotemporal nature of interaction events, DIO represents interaction as triadic relationship model; comprising of three name spaces assigned for input, object, and output. For example, in case of drug metabolism reaction, a substrate drug, enzyme and/or cofactors, and metabolized product. Besides triadic relationship model, other features are: 1) it deals with both molecular level direct interactions and molecular pathway level indirect interactions, 2) it deals with reaction effect in terms of positive/negative (i.e. facilitation, inhibition) with orientation, and 3) it provides semantics of “effect” differentiating biochemical interpretations and pharmacological / clinical ones.

DIO was originally implemented as relational database scheme. Here we introduce its representation using an ontology language OWL, edited by Protégé (http://protege.stanford.edu/). The drug interaction causal logic is under implementation by Prolog and we plan to discuss its prototype.

References

1) Yoshikawa S, Konagaya A. DIO: Drug Interaction Ontology – Application to Inferences in Possible Drug-drug Interactions. 2003 In: Proceedings of The 2003 International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences (METMBS '03), June 23 - 26, 2003. Las Vegas, Nevada, USA. 231-236.

2) Yoshikawa S, Sato S, Konagaya A. Application of Drug Interaction Ontology (DIO) for Possible Drug-drug Interactions. In: Chem-Bio Informatics Society Year 2003 Proceedings: 320.

3) Yoshikawa S, Satou K, Konagaya A. Drug Interaction Ontology (DIO) for Inferences of Possible Drug-drug Interactions. In: MEDINFO 2004, M. Fieschi et al. (Eds) :IOS Press;2004: 454-458.

P80

Transition state of a SH3 domain detected with principle component analysis and a charge-neutralized all-atom protein model

Daisuke Mitomo[1] and Junichi Higo[1]

[1]School of Life Science, Tokyo University of Pharmacy and Life Science

To explore nature of transition state of src SH3 domain, we did ten molecular dynamics simulations with using a charge-neutralized all-atom (CNAA) model, currently-proposed, where all of the atoms in the protein and water molecules were explicitly treated, but their atomic partial charges were set to zero. We could assess an effect of side-chain atoms using this model more efficiently than other models. Each simulation was done for 90 ns at 300 K, starting from the native structure, and the protein unfolded in each run. The integrated trajectories (10 × 90 = 900 ns) were analyzed by a principal component analysis, and showed a clear free-energy barrier between folded- and unfolded-state conformational clusters. The transition state (i.e., a conformational ensemble at the barrier) was characterized by conformational deformations of a segment, to which high F values had been experimentally assigned. Contrary, there was no significant free-energy barrier in a two-dimensional profile with using R_g (radius of gyration) and Q value (native-contact reproduction ratio) as the coordinate axes. Analysis suggested that the transition state from the CNAA model has different nature from one found in course-grained chain models without side-chains, and that loosening of atom packing in the segment might be an important determinant for the transition state.

References

1. Jackson SE, Fersht AR. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. Biochemistry 1991;30:10428-10435.
2. Plaxco KW, Simons KT, Ruczinski I, Baker D. Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. Biochemistry 2000;39:11177-11183.
3. Maxwell KL *et al*. Protein folding: Defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. Protein Sci 2005; 14: 602-616.

P81

CBDB: Cell Behavior Database
- Toward understanding high order level biological phenomena -

Michiaki IWAZUME[1], Hiroyuki KANEKO[2], Marina DAN[3]
Brain Science Institute, RIKEN [1], Keio University[2], Hierarchical Biology Lab.[3]

A large amount of molecular level information on biological phenomena, such as embryogenesis, cell proliferation, cell differentiation, signal transduction, transcriptional regulation, cell response, etc. is rapidly increasing because of the progress of genome science in recent years. In order to understand high-order biological phenomena, it is necessary to bridge the gap between the amount of information in the molecular level and the amount of information in the cell functional level. However, the biological knowledge in the cell functional level has not been systematized well.

We have been seeking a method to systematically describe and understand cellular behaviors in morphogenesis. Various morphogenetic processes were manually selected and described according to the following format: “cell(s)=(subject)” “does something=(output matter)” “under a certain condition=(input matter)”. CBDB is a biological database on cell behaviors in the format with images and movies which indicate

To sophisticate and make use of the CBDB, we propose an AI-based approach for more multiple searching of instances and knowledge discovery from the CBDB. As the first step, we attempted to construct the ontology for understanding morphogenetic phenomena. The ontology will also facilitate clearer definition of terms and concepts, and knowledge sharing among communities of various fields.

In this study, We introduce the CBDB and the prototype of ontology-based information retrieval system.

REFERENCES

1. M. Ashberner et al., “Gene Ontology: tool for the unification of biology. Nature Genetics”, Vol.5, 25-29, Nature America Inc., (2000)

2. M.IWAZUME et al. "Ontogenesis Ontology －Toward To Systematically Describing and Understanding Cellular Behaviors in Morphogenesis", The 15th Annual Conference of Japanese Society for Artificial Intelligence, 2B2-6, (2001).

3. R. Stevens, C. A. Goble and S. Bechhofer, “Ontology -based Knowledge Representation for Bioinformatics”, Journal Briefings in bioinformatics, (2000).

P82

Drug discovery based on the structural dynamics of prion

Kazuo Kuwata

Division of Prion Research, Center for Emerging Infectious Diseases, Gifu University, Gifu 501-1194

Several different compounds and antibodies have been reported to inhibit production of abnormal prion *in vitro*, however the mechanism of effect for most of those substances is unclear and there is still no effective drug *in vivo*. Because prion diseases occur as a result of conformational changes in the cellular prion protein, PrP^{C}, it should be feasible to design drugs that interrupt the critical conversion process. We previously observed a partially denatured form of the prion protein using the high-pressure NMR and the CPMG relaxation dispersion methods, and hypothesized that slow exchange dynamics on the micro- to milliseconds time scale might be critical for the subsequent pathogenic conversion. Thus we incorporated the dynamical information in residue level into *in silico* screening and established a novel strategy termed "Dynamics Based Drug Design (DBDD)". According to this DBDD strategy, we could find several low-molecular weight anti-prion substances, by a combination of *in silico* and *in vitro* drug screening. These substances may stabilize the normal conformation of PrP^{C}, with the result of reducing the population of the intermediate conformer, PrP*, thereby inhibiting its conversion to the pathogenic form.

References

Kazuo Kuwata, Yuji O. Kamatari, Kazuyuki Akasaka, Thomas L. James. Slow conformational dynamics in the hamster prion protein. Biochemistry, 43, 4439-4446, 2004

Kazuo Kuwata, Tomoharu Matsumoto, Hong Cheng, Kuniaki Nagayama, Thomas L. James, Heinrich Roder, NMR-detected hydrogen exchange and molecular dynamics simulations provide structural insight into fibril formation of prion protein fragment 106-126. Proc. Natl. Acad. Sci. 100，14790-5，2003

Kazuo Kuwata, Hua Li, Hiroaki Yamada, Guiseppe Legname, Stanley B. Prusiner, Kazuyuki Akasaka and Thomas L. James: Metastable conformers of the hamster prion protein characterized by high pressure NMR. Biochemistry 41, 12277-12283 (2002)

P83

Parametric Normalization for GeneChip and Tiling Array Data

Tomokazu KONISHI [1],
Midori KATAKAMI[2], Kenya SHIBAHARA[2],
Masahiro HAYASHI[3] and Takushi NAKAJIMA[3]
Bioresource Sciences, Akita Prefectural University[1],
INTENSITY Co. Ltd. [2], Skylight Biotech Inc.[3]

Normalization is the first process of data analysis. Since data obtained by GeneChip technology have variations that have been provoked by hybridization and scanning, in prior to higher level of data analyses, such variations should be reduced or canceled; such cancellation of the technical variations is the data normalization. Consequently, appropriate method of calculation as well as proper standard is required for this process. The validity and accuracy of whole data analyses depend on those of the normalization.

In this poster, we introduce a parametric method for normalizing GeneChip data. Parametric methods are a branch of linear normalization methods, and are characterized by their objectivity. Since they use statistical data distribution as the standard, high accuracy can be expected in the analyses if a proper distribution model would be available. We have introduced an appropriate model, the three parameter lognormal distribution model, for to normalize microarray data obtained by Pat Brown-type chips. This poster shows applications for the model to GeneChip data, which base on an alternative experimental platform. Despite the differences of the platform, the model fit well to the data. Using the model, reproducibility and stability of measured ratios were significantly improved. Although the noise levels are different between experiments, we could expect ca. 80-90 % of genes to be found within a reliable range of intensity.

1. Tomokazu Konishi: Redeveloping Microarray Data Treatment Methods for Transcriptome Study, Pharmaceutical Discovery, Sep 15, 45-50 (2004)
2. Tomokazu Konishi: Three-parameter lognormal distribution ubiquitously found in cDNA microarray data and its application to parametric data treatment, BMC Bioinformatics, vol. 5, pp. 5 (2004)
3. Tomokazu Konishi, Masanori Yoshida, Kenya Shibahara: SuperNORM: A Computer Program for the Parametric Normalization of Microarray Data, Genome Informatics, vol. 14, pp. 292-293 (2003).
4. Tomokazu Konishi: Parametric treatment of cDNA microarray data, Genome Informatics, vol. 13, pp. 280-281 (2002).

P84

Consensus scores for efficient selection of computational protein-ligand docking models

Tadakazu Takakura[1], Akifumi Oda[1], Keiichi Tsuchida[1], Noriyuki Yamaotsu[2], Shuichi Hirono[2]

Toyama Chemical Co. Ltd.[1], Kitasato University[2]

Here we compare the performance of nine consensus scoring strategies (number-by-number, number-by-rank, rank-by-number, rank-by-rank, percent-by-number, percent-by-rank, vote-by-number, vote-by-rank, and vote-by-percent), in which multiple scoring functions are used simultaneously to evaluate candidate structures for a protein–ligand complex[1], in combination with nine scoring functions (FlexX score, GOLD score, DOCK score, PMF score, ChemScore, DrugScore, PLP, ScreenScore, and X-Score). In this study, the trade-off between accuracy and efficiency was investigated (the study using conditions under which reasonable models could be obtained for all test systems without exception is reported by A.Oda *et al.* in this symposium).

Our results demonstrate that choosing the most appropriate type of consensus score is essential for model selection in computational docking; the rank-by-number, rank-by-rank, percent-by-number, and percent-by-rank strategies are effective selection methods, and these results differs from the result we obtained for model selection using conditions under which reasonable models could be successfully obtained for all test set without exception (in which the number-by-number strategy was more appropriate). Thus, using the appropriate consensus scoring strategy according to the demands of the situation is essential. For rank-by-number strategy, ChemScore, DrugScore, and PLP were included in many of the consensus scores which can efficiently select the docking poses. It has been reported previously that these scoring functions work well when they are used independently.[2] Our results suggest that they are appropriate not only for independent scoring, but also for consensus scoring. By incorporating these consensus scores into the FlexX program, reasonable complex models can be obtained more efficiently than those selected by independent FlexX scores. These strategies might also improve the scoring of other docking programs and more effective structure-based drug design should result from these improvements.

References

1) P. S. Charifson *et al., J. Med. Chem.*, 1999, **42**, 5100-5109.

2) R. Wang *et al., J. Med. Chem.* 2003, **46**, 2287–2303.

P85

Molecular dynamics simulation of prion using a coarse-grained model

Hironori K. Nakamura[1], Mitsunori Takano[2], Kazuo Kuwata[1]
Division of Prion Research, Center for Emerging Infectious Diseases, Gifu University[1],
Department of Physics, School of Science and Engineering, Waseda University[2]

Abstract

There are fatal, untreatable neurodenerative diseases known as prion diseases such as bovine spongiform encephalopathy (BSE) for cow, scrapie for sheep, Creutzfeldt Jakob disease (CJD) for human etc. It has been considered that prion diseases are caused by the conformational change of prion protein from a normal form (PrP^{C}) to an abnormal form (PrP^{Sc}). The mechanism of the conformational change has been unclear.

To shed light on the conformational transition, we conducted molecular dynamics simulations of a prion protein using a simple chain model. In this study, we used the Janus-type Go model we proposed recently, which is an extended model from an off-lattice Go model where the PDB structure can be set as the native conformation [1]. The Janus-type Go model uses information of two conformations in contrast that only one native conformation is used in the traditional Go model, which is popular in the protein folding study. A putative PrP^{Sc} conformation of PrP27-30 modeled in an amyloid form [2] and a NMR structure of cellar form were used in this study.

Using the Janus-type Go model, we could successfully conduct a simulation from PrP^{C} to PrP^{Sc}. In the trajectory we found a new stable conformation different from both PrP^{C} and PrP^{Sc}. Features of the stable intermediate conformation indicate some possible scenarios for directed elongation of an amyloid fibril and mechanism of a trimer formation. For drug discovery, the newly found intermediate conformation may be used as a target for a docking of ligand, which may control the conformational change from PrP^{C} to PrP^{Sc}.

References

1) Cecilia Clementi, Hugh Nymeyer, Jose N. Onuchic: Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* vol. 298, pp.937-953 , (2000).

2) Cedric Govaerts, Holger Wille, Stanley B. Prusiner, Fred E. Cohen: Evidence for assembly of prions with left-handed β-helices into trimers., *Proc. Natl. Acad. Sci. USA.* vol. 101, pp. 8342-8347, (2004)

P86

Molecular Interaction Between Estrogen Receptor and Their Ligands

Kaori FUKUZAWA [1,2], Yuji MOCHIZUKI [2,3], Shigenori TANAKA [2,4], Kazuo KITAURA[2,5] and Tatsuya NAKANO[2,6]

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

We have theoretically examined binding energies of typical ligands to the human estrogen receptor ligand-binding domain (hER LBD) and analyzed their interaction.[1] Our method, *ab initio* fragment molecular orbital (FMO) method,[2] has been applied for the calculation of ER-ligand complex, by dividing molecules into small fragments. The ligand binding energies and inter-fragment interaction energies (IFIE) were calculated at Hartree-Fock (HF) and MP2 levels,[3] and the orbital interactions such as the charge-transfer from lone-pair to π^* between the crucial fragments were also analyzed based on a configuration analysis (CAFI) [4,5]. The calculated relative binding energies were very well correlated with the experimental values. We found that the values of charge transfer between ER and ligands are related to that of ER-ligand binding energies.

On the molecular interaction between ER and ligand, there are strong electrostatic interactions between ligand and surrounding some charged/polarized residues, and weak van der Waals interactions between ligand and surrounding many hydrophobic residues. Both interactions are equally contribute to the total binding energies. The strongest interaction occurs between Glu353 and ligand, and the main component is charge transfer from the lone pair orbital of carbonyl oxygen in the Glu353 to the σ^*(OH) orbital of phenol group in the ligand. Our approach is the powerful tool to understand detailed molecular interaction in quantum mechanical level.

REFERENCES

1. K. Fukuzawa et.al., J. Comp. Chem. 26, 1-10 (2005).
2. T. Nakano et.al., Chem. Phys. Lett. 351, 475-480 (2002).
3. Y. Mochizuki et.al., Chem. Phys. Lett., 396, 473-479 (2004).
4. Y. Mochizuki, Chem Phys. Lett., 410, 165–171 (2005).
5. Y. Mochizuki et.al., Chem Phys. Lett., 410, 247–253 (2005).
6. ABINIT-MP and BioStation Viewer are freely available from: http://www.fsis.iis.u-tokyo.ac.jp/result/software/index.html

P87

Computer-aided identification of cytokine mimetics derived from plant extracts

Edgardo E. Tulin,[1] Zenaida T. Ecleo,[1] Igor Kurochkin,[2]
Yasushi Okazaki[3] and Christian Schönbach[2]

[1]PhilRootcrops, Leyte State University, Visca, Baybay, Leyte, Philippines; [2]Immunoinformatics Team, RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, Japan and [3]Research Center for Genomic Medicine, Saitama Medical School, 1397-1 Yamane, Hidaka, Saitama, Japan.

Plant extracts offer a promising source of molecules that mimic the function of cytokines in cellular regulation. We aim to identify cytokine-mimetic molecules from five Philippine indigenous food crop plant extracts to isolate active components which are essential in molecular functional characterization and classification of mimetic molecules and targets. Biologically active extracts from yam roots and leaves, taro roots, “lagundi” leaves, sweet potato leaves and “buyo” leaves were purified and active protein components were isolated and subjected to mass spectrometry. Computational analyses of the mass spectrometry data revealed homologs of a tuber storage protein in yam. The protein has been reported as an anti-hypertensive agent and an inhibitory molecule for angiotensin converting enzyme activity. The homologs appear to be growth stimulatory proteins capable of inducing growth of bone marrow cells and spleen cells *in vitro*. This finding necessitates further characterization on biochemical and computational level to validate the cytokine-mimetic effects.

Author Index