

# Molecular Interactions Within the Cell Network, Scale and Complexity







MBI receives funding from the National Science Foundation Division of Mathematical Sciences and is supported by The Ohio State University. MBI adheres to the AA/EOE guidelines.

### **Table of Contents**

Mission and Goals
Directors and Staff8
Postdoctoral Fellows
Committees14
Visitors
Emphasis Year Program
Workshop 1: Network Biology22
Workshop 2: Signal Transduction and Gene Regulatory Networks
Workshop 3: Synthetic Biology40
Workshop 4: Inference in Stochastic Models of Sequence Evolution
Workshop 5: Mathematical and Experimental Approaches to Dynamics
Workshop 6: Transport in a Cell
Workshop 6: Transport in a Cell 62   Current Topic Workshops Workshop for Young Researchers in Mathematical Biology 68   Computational Challenges in Integrative Biological Modeling 74   Mathematical Developments Arising from Biology 80   Biofilms and Infectious Disease 84
Workshop 6: Transport in a Cell
Workshop 6: Transport in a Cell 62   Current Topic Workshops 68   Workshop for Young Researchers in Mathematical Biology 68   Computational Challenges in Integrative Biological Modeling 74   Mathematical Developments Arising from Biology 80   Biofilms and Infectious Disease 84   Summer Undergraduate Education Program 90   Colloquia and Seminars 92
Workshop 6: Transport in a Cell 62   Current Topic Workshops 68   Workshop for Young Researchers in Mathematical Biology 68   Computational Challenges in Integrative Biological Modeling 74   Mathematical Developments Arising from Biology 80   Biofilms and Infectious Disease 84   Summer Undergraduate Education Program 90   Colloquia and Seminars 92   Public Lecture 95
Workshop 6: Transport in a Cell 62   Current Topic Workshops 68   Workshop for Young Researchers in Mathematical Biology 68   Computational Challenges in Integrative Biological Modeling 74   Mathematical Developments Arising from Biology 80   Biofilms and Infectious Disease 84   Summer Undergraduate Education Program 90   Colloquia and Seminars 92   Public Lecture 95   Future Programs 96

#### **Message from the Director: New Programs at MBI**

During the past year MBI successfully re-competed for funds from NSF. MBI was awarded a five-year grant to continue its missions of stimulating research on the interface between the mathematical sciences and the biosciences and of training young researchers in math biology. The two-year renewal process led to a variety of new programs, which I would like to discuss here. MBI activities mostly fall under five categories (scientific programs, postdoctoral fellows, national impact, education, and diversity) and MBI is developing new programs in each of these categories:



#### WORKSHOPS

Several national panels have noted that the quantification of biology has the potential for revolutionizing biology, while simultaneously creating new fields of mathematics and providing new directions to some mature fields of mathematics. It has been argued that the payoff in both directions will be great, but that the time scale for the impact of biology on the mathematical sciences is longer than is the time scale for the impact of mathematics on biology. Two primary goals of MBI are to promote the impact of the mathematical sciences on the biological sciences and the impact of the biological sciences on the mathematical sciences. We term these math->bio and bio->math. During the first years of its existence MBI stressed math->bio programs; in our new grant we proposed to continue these programs and in addition to feature bio->math programs. The first of these new programs was the very successful November 2009 workshop on Mathematical Developments Arising from Biology organized by John Guckenheimer, Bernd Sturmfels, and Reinhard Laubenbacher.

MBI programs can be divided into two groups: emphasis year workshops (this year's theme was Molecular interactions within the cell: Network, scale and complexity) and current topic workshops (which includes the always successful Workshop for Young Researchers in Mathematical Biology). MBI proposed to have approximately six emphasis year workshops and approximately six current topic workshops each year and we always encourage members of our community to propose ideas for these workshops.

#### INSTITUTE PARTNERS AND MENTORING

During the past year the number of MBI Institute Partners doubled from 19 to 38 (two of which are industrial IPs). This change resulted from a concerted effort by MBI to involve IP researchers in Institute programs. Perhaps the most dramatic of these was the effort to promote the offsite mentoring of MBI postdoctoral fellows. Indeed, this year, five MBI post-docs have had a least one off-site mentor. We very much encourage Institute Partner researchers to sign up to be a possible MBI post-doc mentor.

#### EARLY CAREER AWARDS AND LONG-TERM VISITORS

This year was the first year for a new MBI program that makes it possible for young researchers (those who are untenured, but who are in tenure-track positions) to spend between three and ten months in residence at MBI specifically to take advantage of the MBI emphasis year program. This past year, David Romano (Grinnell University) received the first MBI Early Career award. In addition, MBI continued to encourage senior researchers to participate in MBI emphasis year programs by making long-term visits of one-month or more.

#### **EDUCATION PROGRAMS**

MBI is experimenting with several new education programs two of which started last year. Each quarter MBI now sponsors a course in mathematical biology broadly construed. Martin Feinberg (Chemical Engineering and Mathematics, OSU) gave a course on The Mathematics of Chemical Reacting Networks in the winter quarter and Laura Kubatko and Dennis Pearl (Statistics, OSU) gave a course on Statistical Phylogenetics in the spring quarter. We plan to have these courses broadcast by live video streaming.

#### **DIVERSITY AND OUTREACH**

During the past year MBI established a Diversity Committee co-chaired by Carlos Castillo-Chavez and Trachette Jackson. This committee made a number of suggestions the most visible of which is the MBI Visiting Lecturer Program (VLP). This program will support visits by a number of mathematical biologists to campuses with large minority enrollments. In addition, this year MBI restarted its successful public lecture series, but with some new twists: the lectures are held at COSI (the Columbus Science Museum) and have been organized with the cooperation of Metro High, a local STEM high school.

All-in-all 2009-10 has been an active and exciting year for MBI, its staff, and its directors. The driving principle behind most changes at MBI has been the desire to further encourage interaction between MBI and its constituent communities. So I end this message with the open invitation to those who work on the interface between the mathematical and biosciences to participate in, to propose, and to lead MBI programs.

Marty Golubitsky

# **MISSION AND GOALS**

#### MISSION STATEMENT

MBI offers a vigorous program of research and education, and fosters the growth of an international community of researchers in mathematical biology.

The mission of MBI is:

- To foster innovation in the application of mathematical, statistical, and computational methods in the resolution of significant problems in the biosciences;
- To foster the development of new areas in the mathematical sciences motivated by important questions in the biosciences;
- To engage mathematical and biological scientists in these pursuits; and
- To expand the community of scholars in mathematical biosciences through education, training, and support of students and researchers.

To support this mission, MBI programs are designed to reinforce and build upon existing research efforts in the mathematical biosciences, and to inspire and accelerate the expansion of the community and its intellectual growth. These include emphasis year programs, current topic workshops, education programs, and research projects. The administrative and governance structure of the MBI are designed to support the mission of the Institute.



Dan Siegal-Gaskins, Erich Grotewold, and Greg Smith.

#### **INSTITUTE PARTNERS**

MBI welcomes the participation of other academic institutions in the MBI Institute Partner Program. This program uses MBI matching funds to subsidize the travel expenses of IP member researchers to allow their participation in MBI programs.

In addition, each year MBI provides up to \$15K to support conferences in mathematical biology held at IP institutions.

IP representatives are invited to annual meetings to explore research and educational opportunities and provide input for future institute programs. IP members also receive MBI newsletters, proceedings, and annual reports.

#### **Current Institute Partners**

- Arizona State University •
- Batelle
- **Boston University**
- Case Western Reserve University
- **Cornell University**
- Drexel University
- Duke University
- Florida State University •
  - Howard University
- IBM
- Indiana University-Purdue University Indianapolis
- Instituto Gulbenkian de Ciencia
- Iowa State University
- Legacy Good Samaritan Hospital
- Michigan State University
- New Jersey Institute of Technology
- Ohio State University
- Ohio University
- Penn State University
- Princeton University
- University of California, Irvine
- University of Cincinnati
- University of Exeter
- University of Georgia
- University of Houston



Chandan Sen, Chuan Xue, and Avner Friedman.

- University of Iowa
- University of Maryland, Baltimore County
- University of Michigan
- University of Minnesota
- University Notre Dame
- University of Nottingham CMMB
- University of Oxford
- University of Pittsburgh
- University of Southern California
- University of Utah
- University of Washington
- University of Waterloo
- Vanderbilt University
- Virginia Tech

#### **Postdoctoral Fellow Mentoring Program**

Each MBI Postdoctoral Fellow has three mentors: a professional mentor from the MBI Directorate and two research mentors (one from the mathematical sciences and one from the biosciences). In a program begun in 2009 the research mentors may be either at Ohio State University or at one of partner institutions. Indeed, some MBI Postdoctoral Fellows had research mentors at Nottingham, Iowa, UC Irvine, Utah, and Virginia Tech.

The directory of research mentors can be found at the following web page:

http://mbi.osu.edu/postdoctoral/mentoring.html

#### EARLY CAREER AWARD PROGRAM

Early Career Awards enable recipients to be in residence for stays of at least three months during an Emphasis Year Program at the Mathematical Biosciences Institute. The 2011-12 Program is on Stochastics in Biological Systems (see http://mbi.osu.edu/2010/scientific2011.html). Awardees will engage in an integrated program of tutorials and workshops tied to the scientific theme and are expected to interact with local and visiting researchers.

Early Career Awards are aimed at non-tenured scientists who currently have continuing employment and who hold a doctorate in any of the mathematical, statistical and computational sciences, or in any of the biological, medical and related sciences.

#### **MBI PROGRAM IDEAS**

MBI programs are aimed at bringing mathematical scientists and bioscientists together to discuss ways in which the mathematical sciences are being used to solve significant problems in the bio and biomedical sciences or how problems from the biosciences are opening new areas of research for mathematicians, statisticians, and computational scientists.

MBI encourages members from the mathematical sciences or the biosciences community to propose ideas for MBI programs.

MBI programs fall roughly into three categories:

- Semester or yearlong emphasis programs (consisting of a number of related workshops and supported by MBI long-term visitors).
- Current Topic Workshops (typically stand alone meetings of up to one week).
- Education programs.

For details on how to submit an idea for an MBI program, visit:

http://www.mbi.osu.edu/suggestions.html



Larry S. Schlesinger, Judy Day, and Avner Friedman

### DIRECTORS & STAFF



#### MARTY GOLUBITSKY, DIRECTOR

The Director provides the scientific leadership, promotes the institution's mission and goals, and is responsible for the overall management and resource development of the institute. The director reports to the Board of Trustees.



MICHAEL REED, SENIOR SCIENTIFIC ADVISOR The Senior Scientific Advisor designs and implements new

The Senior Scientific Advisor designs and implements new programs that promote the mission of the Institute. The Senior Scientific Advisor will represent the Institute in the Director's absence.



**PROFESSOR HELEN CHAMBERLIN, ASSOCIATE DIRECTOR** Works with the director of diversity issues.



**PROFESSOR YUAN LOU, ASSOCIATE DIRECTOR** Oversees the postdoctoral fellow program and edits the MBI Newsletter.



**PROFESSOR DENNIS PEARL, ASSOCIATE DIRECTOR** Responsible for the education programs, as well as the evaluation process.



**PROFESSOR ANDREJ ROTTER, ASSOCIATE DIRECTOR** Provides leadership for relations between MBI and the Ohio State Medical Center.



**PROFESSOR TONY NANCE, ASSISTANT DIRECTOR** Duties include oversight of the day-to-day operation of the MBI offices and supervision of the institute staff.

Four **Associate Directors** provide scientific advice and support to the director. All Associate Directors are involved in the mentoring program for postdoctoral fellows. Each Associate Director oversees at least one aspect of MBI.



#### NIKKI BETTS, FINANCIAL AND HR MANAGER

Manages all human resources and financial activity in the MBI, including visa, travel, and reimbursement related activities. She also helps with program and reporting activities.



#### STELLA CORNETT, WEB COMMUNICATIONS SPECIALIST

Manages the web site; handles all advertising including web and print; creates and distributes brochures, flyers, newsletters, posters, and annual report booklets; and receives participant abstracts and presentation materials and places them on the web.



#### CARTER SCHOENFELD, SYSTEMS SPECIALIST

Provides support to users of MBI computer and presentation facilities, assists Michael with systems maintenance, and contributes to web programming projects.



#### **REBECCA MARTIN, OFFICE ASSOCIATE**

Provides direct office support for the Director; serves as primary point of contact for people within and external to the MBI; sends letters of invitation to all workshop and tutorial participants.



#### MICHAEL SIROSKEY, SYSTEMS MANAGER

Responsible for technology at MBI, including maintaining and upgrading servers, desktop and laptop machines; handles hardware and software evaluation and procurement decisions; responsible for presentation and telecommunication facilities; provides support on space renovation project; and supervises web activity.



#### MATT THOMPSON, PROGRAM ASSISTANT

Assists in fiscal processing, registration, reimbursements, human resources, and event coordination; responsible for information given to all visitors.



CASEY JACOBS, STUDENT WORKER



**CAITLIN NABER, STUDENT WORKER** 

Our **Student Workers** provide critical logistic and clerical support for MBI events, including materials, advertising, and data management.

# POSTDOCTORAL FELLOWS

#### **COHORT 2007**



**Huseyin Coskun** (Computational and Applied Mathematics, University of Iowa). Huseyin's research area is interdisciplinary: it is a combination of mathematics, biology and engineering. He is principally interested in applied mathematics, partial differential equations, and inverse problems. He developed models for cell movements which incorporate different components of the phenomena, such as mechanics and molecular dynamics that have been studied separately, into a single model. In that sense the models can be considered as 'systems biologic' approach. He also formulated model based inverse problems for parameter and unknown function estimation. Neither this system biologic approach nor the inverse problem formulation have been studied previously, in the area of cell motility.



**Judy Day** (Mathematics, University of Pittsburgh). Judy's research interests are primarily focused on problems that have potential to translate directly to medicine in the care and treatment of the critically ill. In particular, she has worked to form and analyze mathematical models (systems of ordinary differential equations) to explore the non-linear interplay of the various components of inflammation. Inflammation is a complex process not well understood and many potential therapies to control inflammation have failed. Thus, in addition to developing models to understand the inflammatory response, she is also interested in using these models to explore potential therapies to correct immune dysfunction. Consequently, she has been investigating the use of nonlinear model predictive control as one method by which this might be accomplished.



**Rasmus Hovmoller** (Systematic Zoology, Stockholm University, Sweden). Rasmus's current research interest is in phylogenetic studies of emergent infections disease with a focus on Influenza. By creating a genealogy over virus sequences, and mapping them geographically we can trace the events that enables a bird flu virus to infect humans. Influenza viruses have a segmented genome, consisting of 8 separate single-strand RNA fragments coding for 10 proteins.

Reassortment between different strains of Influenza has been thought to cause the large pandemics. The Spanish flu of 1918 is believed to have originated as strain that jumped hosts directly bird to humans, while the Hong Konf flu of 1968 is though to have passed through a genetic reassortment between relatively benign bird flu and human flu viruses in pigs. These assumptions are based on the immunological characteristics of surface proteins: the Hong Kong strain appeared to have one protein from pig flu, and another from seasonal human flu. With new methods and computer implementations, we can examine possible genomic rearrangements in a rigorous phylogenetic context. He will also be working on insect molecular phylogeny, focusing on bluet damselflies, with a group at the Department of Entomology.

#### **COHORT 2008**



**Deena Schmidt** (Applied Mathematics, Cornell University). Deena's interests are in applying probability to problems in population and evolutionary genetics and molecular biology. Her Ph.D. research focused on stochastic models of DNA regulatory sequence evolution in organisms of different population sized. She's currently working on a gene regulatory network model of an experimental system derived from the lambda switch (bacteriophage lambda) and looking for noise-induced oscillations due to a small number of molecules in the system. This is in collaboration with Timothy Newman (Arizona State University) and Vincent Noireaux (University of Minnesota). Thus far at the MBI, she is working on two projects: stochastic models for the evolution of gene expression, and the relationship between stochastic models and their corresponding mean-field approximations which is important in describing various biological systems.



**Dan Siegal-Gaskins** (Physics, University of Chicago). Dan is currently using a combined experimental and mathematical approach to understand the mechanisms that lead to cell fate determination. In particular, he is investigating whether a simple gene regulatory network underlying the development of unicellular leaf hairs (trichomes) in the model system Arabidopsis thaliana has the capacity for bistability, and if that bistability can explain the characteristic trichome pattern. He is also studying the role of global leaf properties in selecting the location for the very first trichome cell differentiation event.



**Chuan Xue** (Applied Mathematics, University of Minnesota Twin Cities). Chuan Xue's research involves multi-scale modeling in bacterial pattern formation and wound healing. She received her Ph.D. in mathematics from the University of Minnesota in Aug. 2008 under the direction of Hans G. Othmer. In her thesis, she focused on unveiling the mechanism of spatial pattern formation in the bacterial colonies found in her collaborator's lab. She developed a hybrid cell-based model which incorporated intracellular signal transduction, cell movement and extracellular signal dynamics. The model yields biologically-based explanations to radial and spiral stream formation in P. mirabilis colonies. To reduce the computational cost due to large number of cells, she lifted the cell-based model to a continuum model by deriving macroscopic chemotaxis equations of cell density using perturbation techniques and moment closure methods. She is also working on mathematical models for ischemic wound healing. The goal is to understand how the supply of oxygen affects the wound healing process and how hyperbaric treatment helps with chronic wound closure in patients with circulation diseases.

#### **COHORT 2009**



**Erik Bloomquist** (Biostatistics, UCLA). Erik is currently working with investigators in the College of Medicine and the College of Veterinary Medicine on the classification of biomarkers for the detection of urological disease. In addition, he is working on a genetic epidemiological study with investigators in the College of Medicine. In the autumn quarter, he taught a graduate course in the College of Public Health at Ohio State. In addition to his work at Ohio State, he is collaborating on an ancient DNA study with an investigator at Penn State, and is writing a review of phylogeography.

# POSTDOCTORAL FELLOWS



Julia Chifman (Mathematics, University of Kentucky). Julia's research primarily focuses on phylogenetic invariants and their performance under the coalescent process (joint work with collaborator and mentor Laura Kubatko, OSU). Although much progress on phylogenetic invariants has previously been made for a single gene, it is well-known that gene trees are not topologically equivalent with the species tree. By deriving phylogenetic invariants within the coalescent framework, her work extends their utility to the case of multi-gene data for which the goal is inference of the species-level phylogeny. She is also interested in the analysis of the mammalian iron metabolism using algebraic tools (joint work with Biosciences mentor Reinhard Laubenbacher, Virginia Tech).



**Shu Dai** (Mathematics, Duke University). Shu is interested in the PDEs arising from mathematical biology and physiology, bifurcation analysis and numerical analysis. Currently he is working on applied dynamical systems and mathematical cardiology.



**Marisa Eisenberg** (Biomedical Engineering, UCLA). Marisa's research focuses on building ODE models of neuroendocrine regulation and developing methods to address identifiability and parameter estimation questions. Her dissertation research centered on building physiology-based feedback control system models of the human hypothalamic-pituitary-thyroid axis, to address several clinical issues in thyroidology. She is extending this work in several new directions involving other hormone regulatory axes, circadian rhythms, and thyroid cancer. Marisa is also working on identifiability and parameter estimation of nonlinear ODE models and exploring identifiability applications in phylogenetics.



**Sam Handelman** (Biological Sciences, Columbia University, New York). Sam works with mentor Dan Janies and collaborator Jesse Kwiek (both of the Medical school) in identifying sequence/evolutionary features of HIV associated with *in utero* mother to child transmission. This project involves analysis of large scale sequence data via phylogenetic and structural methods. Sam also works to extend these methods in other viral and microbial systems. In an overlapping project, Sam works with mentors Joe Verducci (of the Dept. of Statistics) and Dan Janies to develop new techniques in statistics to robustly identify differences in viral sequences that correlate with phenotypic differences. Phenotypes of particular interest include transfer between different host species, compartmentalization in tissues, and antibiotic resistance.



**Harsh Jain** (Applied Mathematics, University of Michigan). Harsh is interested in biochemically-motivated modeling of angiogenesis and cancer growth, and the simulation of novel therapies targeting both cancer cells and tumor vasculature in order to evaluate their potential and aid in the design and execution of clinical trials of these drugs. While at MBI, he will work on a project with researchers at the University of Michigan wherein he will combine multi-scale modeling with model-driven experimentation to develop a comprehensive and predictive computational model that will aid in the characterization of AT-101, a small molecule inhibitor of the anti-apoptotic members of the Bcl-family of proteins, as an anti-angiogenic and anti-cancer agent in the treatment of head and neck cancers. He also plans to collaborate with researchers at OSU and Nottingham



**Suzanne Robertson** (Applied Mathematics, University of Arizona). Suzanne's research interests are in the area of mathematical ecology and epidemiology. She received her PhD in Applied Mathematics from the University of Arizona under the direction of J. M. Cushing. Her thesis work focused on the spatial patterns formed by stage-structured species as a result of density dependent dispersal between life-cycle stages. At the MBI, she is working with Joe Tien (Mathematics, OSU) on modeling the spatial spread of Cholera. She has also started working with Ian Hamilton (EEOB, OSU). They are looking at the role of disease in habitat selection.



**Rebecca Tien** (Ecology and Evolutionary Biology, Cornell University). Rebecca is interested in mathematical and computational ecology with applications to population dynamics, management and conservation of natural resources, particularly as they relate to aquatic ecology. She is currently working with Elizabeth Marschall and Yuan Lou on the biomagnification of PCBs and other heavy metals and their potential effects on food web interactions and population dynamics in Lake Erie.



**Yunjiao Wang** (Mathematics, University of Houston). Yunjiao works in the fields of nonlinear dynamical systems and system biology. Her recent research on nonlinear dynamical systems focuses on studying dynamics of networks, especially on studying general theory of coupled cell systems, functions of motifs, coupled oscillators and transtivity of oscillations. Her current research on system biology focuses on studying heterochronic signaling pathways in C. elegans and nuclear factor \kappaB signaling pathways, especially its interaction with other signaling pathways in the cell and between cells.



**Kun Zhao** (Mathematics, Georgia Institute Of Technology). Kun's research interests are in the area of analysis and applications of nonlinear partial differential equations (PDEs) in various branches of science and engineering. In his Ph.D. thesis, he studied qualitative behavior of solutions to initial-boundary value problems for several systems of nonlinear evolutionary partial differential equations arising from fluid dynamics and civil engineering. At MBI he is currently working on several systems of PDEs arising from mathematical biology, namely, the Keller-Siegel type chemotaxis models and the Cahn-Hilliard-Hele-Shaw equations.

# COMMITTEES

#### **BOARD OF TRUSTEES (BOT)**

The Board reviews the institute management and programs and advises and approves the strategic priorities of the institute. The Board consists of individuals with leadership experience in the public and private sectors, and of recognized scientists in fields related to MBI activities. The Board meets annually to review the institute management and programs and to advise and approve the strategic priorities of the institute.

#### **Current Members**

- Rita R. Colwell, University of Maryland, College Park (12/31/10)
- John Guckenheimer, Cornell University (12/31/11)
- Kirk E. Jordan, IBM T.J. Watson Research Center (12/31/11)
- **Robb Krumlauf**, Stowers Institute for Medical Research, Kansas City, MO (12/31/10)
- Mark Lewis, University of Alberta, Canada (12/31/11)
- Robert M. Miura (Chair, 2009-2010), New Jersey Institute of Technology, Newark, New Jersey (12/31/11)
- Blake Thompson, Battelle, Columbus, OH (12/31/12)
- Michael Waterman, University of Southern California (12/31/12)

#### **Past Members**

- **Barbara Kunz**, Battelle Memorial Institute, Columbus, OH
- **Stephen Ruberg**, National eHealth Collaborative, Senior Research Fellow at Eli Lilly

### SCIENTIFIC ADVISORY COMMITTEE (SAC)

SAC reviews MBI programs and suggests and decides on annual programs and organizers. The Committee consists of internationally recognized mathematical scientists and bioscience researchers from academia and industry. SAC meets annually to review the institute programs, to suggest and decide on new annual programs, and to give advice regarding programmatic goals.



#### **Current Members**

- Linda Allen, Texas Tech University (12/31/11)
- Mark Chaplain, The SIMBIOS Centre, University of Dundee (12/31/10)
- Thomas Daniel, University of Washington (12/31/12)
- Mark Denny, Stanford University (12/31/10)
- **Bard Ermentrout**, University of Pittsburgh (12/31/11)
- Suzanne Lenhart, University of Tennessee (12/31/10)
- Naomi Leonard, Princeton University (12/31/11)
- Mark Lewis (Chair 2009-10), University of Alberta (12/31/09)
- Andre Longtin, University of Ottawa, Canada (12/31/12)
- Paul Magwene, Duke University (12/31/11)
- L. Mahadevan, Harvard University (12/31/12)
- Karl J. Niklas, Cornell University (12/31/10)
- Lior Pachter, University of California, Berkeley (12/31/10)
- Steven Rust, Battelle Memorial Institute, Columbus, OH (12/31/11)
- Stanislav Shvartsman, Princeton University (12/31/10)

#### **Past Members**

- Reka Albert, Pennsylvania State University
- Adam Arkin, University of California, Berkeley
- Herb Bresler, Battelle Memorial Institute, Columbus, OH
- Leah Edelstein-Keshet, University of British Columbia
- Lisa Fauci, Tulane University
- Louis Gross, The University of Tennessee
- Sorin Istrail, Brown University
- Nicholas P. Jewell, University of California, Berkeley
- Kirk Jordan, IBM Computational Biology Center, Yorktown Heights, NY
- Jim Keener, University of Utah
- **Douglas Lauffenburger**, Massachusetts Institute of Technology
- **Gregory Mack**, Battelle Memorial Institute, Columbus OH
- Philip Maini, University of Oxford
- Claudia Neuhauser, University of Minnesota
- Alan Perelson, Los Alamos National Laboratory
- Linda Petzold, University of California, Santa Barbara
- Mike Reed, Duke University
- John Rinzel, Courant Institute of Mathematical Sciences, New York University
- Stephen Ruberg, Eli Lilly and Company, Indianapolis
- James Sneyd, University of Auckland, New Zealand (12/31/10)
- Terrence Speed, University of California, Berkeley
- John Taulbee, Procter & Gamble Company, Cincinnati
- Terry Therneau, Mayo Clinic College of Medicine, Rochester, MN
- Frank Tobin, GlaxoSmithKline
- John Tyson, Virginia Polytechnic Institute and State University
- Steven Vogel, Duke University
- Michael S. Waterman, University of Southern California
- Raimond L. Winslow, The Johns Hopkins University School of Medicine and Whiting School of Engineering



#### LOCAL SCIENTIFIC ADVISORY COMMITTEE (LSAC)

The LSAC consists of members of The Ohio State University community. It helps identify current topics workshops, suggest ideas for future emphasis programs and organizers, and potential mentors for postdoctoral fellows.

- Sudha Agarwal, Oral Biology
- Irina Artsimovitch, Microbiology
- John Buford, Physical Therapy
- Ralf Bundschuh, Physics
- Helen Chamberlin, Molecular Genetics
- James Cogdell, Mathematics
- Meg Daly, Evolution, Ecology, and Organismal Biology
- Andrea Doseff, Heart and Lung Research Institute, Molecular Genetics, and Internal Medicine
- Avner Friedman, Mathematics
- Martin Feinberg, Chemical Engineering
- Paul Fuerst, Evolution, Ecology and Organismal Biology
- Erich Grotewold, Plant Biology
- Richard Hart, Biomedical Engineering
- Tim Huang, Center for Integrative Cancer Biology
- Daniel Janies, Biomedical Informatics
- Doug Kniss, Obstetrics and Gynecology
- **Stanley Lemeshow**, School of Public Health, Center for Biostatistics
- **Gustavo Leone**, Molecular Virology, Immunology, and Medical Genetics
- Shili Lin, Statistics
- Yuan Lou, Mathematics
- Thomas Magliery, Chemistry
- Stuart Mangel, Neuroscience
- Elizabeth Marschall, Evolution, Ecology, and Organismal Biology
- Deborah Parris, Molecular Virology
- Dennis Pearl, Statistics
- John Reeve, Microbiology
- Andrej Rotter, Pharmacology
- Wolfgang Sadee, Pharmacology
- Larry S. Schlesinger, Division of Infectious Diseases and Center for Microbial Interface Biology
- Petra Schmalbrock, Radiology
- Chandan Sen, Surgery
- Amanda Simcox, Molecular Genetics
- **Parthasarathy Srinivasan**, Computer Science and Engineering and Biomedical Informatics
- **Don Stredney**, Biomedical Applications, Ohio Supercomputer Center

### COMMITTEES

#### **DIVERSITY PLAN**

The goal of the MBI diversity efforts is to help shape the mathematical biology community in a way that represents the diversity of our society. MBI will work towards this goal on two levels. First, MBI actively seeks diversity among its participants in gender and ethnicity. Second, MBI sponsors activities that promote mathematical biology and its varied opportunities in the academic community.

Specifically, MBI builds and maintains diversity by the following.

- 1. Boards and Advisors: Ensure representation of underrepresented groups among the MBI standing committees.
- 2. Scientific Workshops and Emphasis Programs: Include members of underrepresented groups as members of emphasis year and workshop organizing committees and ensure broad representation among workshop participants.
- Training of Younger Scientists: Ensure broad representation among postdoctoral fellows and build exposure of younger scientists to mathematical biology.
- 4. Awareness Workshops: Periodically host workshops on Opportunities in Mathematical Biology for Underrepresented Groups.

In addition, MBI will pursue the following strategies:

- Participate in meetings of minority scientists, such as the Society for Advancement of Chicanos and Native Americans in Science (SACNAS) and the Historically Black Colleges and Universities Undergraduate Program (HBCU-UP), to provide information about MBI, recruit participants to MBI activities, and inform young scientists about opportunities in mathematical biology.
- 2. Build relations with academic institutions having strong minority enrollments.



- Advertise MBI programs both broadly and to targeted audiences, including meetings of mathematical biology societies and minority-serving science societies.
- 4. Evaluate the implementation of the MBI diversity plan annually.

#### **Diversity Committee**

- Helen Chamberlin, Molecular Genetics, The Ohio State University (ex officio)
- Carlos Castillo-Chavez, Mathematics and Statistics, Arizona State University
- Joan Herbers, Evolution, Ecology, & Organismal Biology, The Ohio State University; President Elect AWIS
- Trachette Jackson, Mathematics, University of Michigan
- Yi Li, Chair, Mathematics, University of Iowa
- Maeve McCarthy, Mathematics & Statistics, Murray State University; Executive Director AWM
- Aziz Yakubu, Chair, Mathematics, Howard University

#### **VISITING LECTURER PROGRAM**

The Mathematical Biosciences Institute developed a Visiting Lecturer Program in 2009. The program sponsors visits of mathematical biologists to institutions that have large numbers of students who are members of groups that are underrepresented in the mathematical sciences community. The purpose is to encourage members of these groups to go to graduate school and to develop careers in the mathematical biosciences. Visiting Lecturers will deliver a lecture on mathematical biology that is accessible to undergraduates and will meet with students and interested faculty. The phrase "underrepresented group" is understood to mean African-Americans, Hispanics, Native Americans and women. It is an important goal of the National Science Foundation to increase the participation of these groups in the sciences, so as to increase the strength of the American scientific workforce.

Department Chairs of a math sciences or a biosciences department can initiate discussions about bringing a Visiting Lecturer to their campus by sending an email to

> Marty Golubitsky, Director Mathematical Biosciences Institute mg@mbi.osu.edu

The web page (http://www.mbi.osu.edu/about/vlprogram.html) contains a list of Visiting Lecturers.

MBI expects to support five visiting lectures each year.





Carlos Castillo-Chavez (Arizona State University)

#### Lecturers

- Janet Best, Ohio State
- Emery Brown, MIT
- Erika Camacho, Arizona State
- Carlos Castillo-Chavez, Arizona State
- Ricardo Cortez, Tulane University
- Isabel Darcy, University of Iowa
- Lisette de Pillis, Harvey Mudd
- Lisa Fauci, Tulane University
- Marty Golubitsky, Ohio State
- Christine Heitsch, Georgia Tech
- Fern Hunt, Howard and NIST
- Trachette Jackson, University of Michigan
- James Keener, University of Utah
- Nancy Kopell, Boston University
- Jonathan Mattingly, Duke University
- Asamoah Nkwanta, Morgan State
- Michael Reed, Duke University
- Miranda I. Teboh-Ewungkem, Lafayette College
- Talitha Washington, University of Evansville
- Abdul-Aziz Yakuba, Howard University

# VISITORS

#### LONG TERM VISITORS 2009-2010

- Fernando Antoneli, Universidade Federal de Sao Paulo
- **Dieter Armbruster**, Department of Mathematics, Arizona State University
- Xinfu Chen, University of Pittsburgh
- Chris Fall, Department of Anatomy and Cell Biology, University of Illinois-Chicago
- Chirove Faraimunashe, University of Botswana
- Lisle Gibbs, EEOB, The Ohio State University
- Yangjin Kim, Department of Mathematics, University of Michigan-Dearborn
- **David Romano**, Department of Mathematics, Grinnell University
- **Richard Schugart**, Mathematics and Computer Sciences, Western Kentucky University
- Martin Wechselberger, School of Mathematics and Statistics, University of Sydney
- **Chang-Hong Wu**, Department of Mathematics, National Taiwan Normal University
- Najat Ziyadi, Cadi Ayyad University, Morocco

#### ANTICIPATED VISITORS 2010-2011

- Abdul Aziz Fall, Gaston Berger University, Senegal
- Kesh Govinder, Mathematical Sciences, University of KwaZulu-Natal
- Maria Leite, University of Oklahoma
- Annie Lindgren, NSF
- Edward Lungu, Mathematics, University of Botswana
- Ian Stewart, Mathematics, University of Warwick, UK
- **David Sumpter**, Mathematics Department, Uppsala University, Sweden
- Barbara Szomolay, Center for Integrated Systems Biology, Imperial College, London
- Satoshi Takahashi, Zayed University
- **Tunde Tajudeen Yusuf**, Federal University of Technology Akure, Nigeria
- Carl Toews, Mathematics, Duquesne University

#### COURSE RELEASE 2009-2010

### Mathematics

- Janet Best
- Ching-Shan Chou
- Avner Friedman
- Chiu-Yen Kao
- Yuan Lou
- Akos Seress
- Joe Tien

#### **Statistics**

- Laura Kubatko
- Shili Lin
- Tao Shi
- Joe Verducci

#### **Biomedical Engineering**

- Samir Ghadiali
- Yi Zhao

#### **Evolution, Ecology, and Organismal Biology**

Ian Hamilton

#### **Electrical and Computer Engineering**

Kevin Passino

#### **Biochemistry**

Mark Foster

#### **Medicinal Chemistry**

• Chenglong Li

### ANTICIPATED COURSE RELEASE 2010-2011

#### **Mathematics**

- Ching-Shan Chou
- Avner Friedman
- Bo Guan
- Chiu-Yen Kao
- Yuan Lou
- Joe Tien

#### **PROGRAM PARTICIPATION**

The chart below shows the total number of participants for each MBI event during the 2009-2010 Emphasis year. **The total number of participants this year was 876.** 



#### **Statistics**

- Jason Hsu
- Laura Kubatko

#### **Electrical and Computer Engineering**

Kevin Passino

### MOLECULAR INTERACTIONS CELL: NETWORK, SCALE, &

#### **ORGANIZING COMMITTEE**

- Reka Albert (Physics & Biology, PSU)
- Eivind Almaas (Biotechnology, NUST)
- Laszlo Barabasi (Physics, Northwestern)
- Jeff Hasty (Bioengineering, UCSD)
- Ian Holmes (Bioengineering, UC Berkeley)
- Anatoly Kolomeisky (Chemistry, Rice University)
- Jane Kondev (Physics, Brandeis University)
- Hao Li (Biochemistry & Biophysics, UCSF)
- Ron Weiss (Electrical Engineering, MIT)

Biological processes can be characterized by different degrees of complexity at microscopic (genes, molecules), mesoscopic (protein-DNA complexes) and macroscopic (cells, organisms) levels. Historically, all biological systems have been studied at different levels. However, an increasing amount of experimental results and theoretical studies suggest that a more comprehensive system approach would tackle better biological problems. It would require a collaboration and intensive exchange between experimental and theoretical researchers from physics, chemistry, biology, mathematics, computer science, and engineering.



### WITHIN THE COMPLEXITY



The proposed activity will answer the following fundamental questions: What are the properties of biological networks? How do they function? How do genes come together to form networks, and how can we use bioinformatics to discover such networks? Can our understanding of the fundamental mathematics inform the design of those bioinformatics methods? How is information transferred in cells? What role can synthetic biology perform in aiding our understanding of real life processes? How can different subjects of biological systems interact together to create effective dynamic systems?

Specific sub-areas of molecular and cellular biology generate their own sets of problems and mathematical challenges, to be addressed by individual workshops throughout the year. For example, how do cells develop, control, and regulate highly-efficient, highly-selective and robust biological transport? What are the algorithms and models that can help elucidate RNA structure and function?

# **2009 2010**

What are the basic pathways of cell-to-cell signaling? How can we design genetic regulatory networks with targeted function for synthetic biology? What are the mathematical principles behind DNA-protein interactions and the coordinated regulation of gene expression? The over-arching theme of the workshops bridges multiple scales, from the molecular to the cellular, in pursuit of the fundamental biological principles guiding the structure, evolution, and maintenance of these networks.

A unifying long-term goal of the proposed activities is to develop a unified approach to study the complexity of biological systems within cells. Such a comprehensive view of biology will require an application and development of new mathematical methods. Current approaches include hidden Markov processes, stochastic dynamics, graph theory, partial differential equations, discrete mathematics and other tools of probabilistic modeling, machine learning and computational analysis. As in the past, it is expected that new frontiers in biology will both benefit from and stimulate the development of novel mathematical techniques.



#### Workshop 1 Network Biology: Understanding Metabolic and Protein Interactions September 14-18, 2009

#### ORGANIZERS

Eivind Almaas, NUST Laszlo Barbasi, Northeastern University

Report written by MBI postdocs Huseyin Coskun, David Romano, and Yunjiao Wang.

#### **OVERALL SUMMARY**

As network approaches have become an important tool to study a wide range of complex systems for which traditional reductionist approaches have enjoyed limited success, maybe the biggest enthusiasm and triumphs have been noted in biology. In particular within the cell, the variety of interactions between genes, proteins and metabolites are well captured by network representations. The dramatic availability of quantitative data from largescale genomic experiments has begged for systemic approaches with the ability to simultaneously integrate information from multiple sources. In response, the advent of systems biology methods has been heavily influenced by network methods. Although recent network analyses have shed light on organizational principles of the proteome as well as the metabolome, there is, however, an increasing need for developing even more sophisticated, integrative approaches as higher quality data is becoming available. These challenges include developing systematic methods for integrating proteomic and metabolic information, thus coupling their mostly separated analyses; incorporating spatial localization of cellular constituents, and developing new tools to include stochastic and time varying measurements.

It is noteworthy that most network oriented workshops and conferences have an interdisciplinary and broad focus, as network approaches flourish in many fields. However, there is a need to bring biologists together with network scientists to discuss sharply defined topics within network biology. The goal of this workshop was to facilitate information exchange between biologists (experimental and theoretical) and network scientists, making them aware of each other's capabilities and methodologies, as well as fostering collaborative interactions. Analysis and modeling of metabolic and protein interaction networks often involve graph theory, optimization, and statistics.

#### DAY 1: MONDAY, SEPTEMBER 14

Graph Concepts and Applications to Protein Interaction Networks Réka Albert (Pennsylvania State University)

Réka opened the workshop with a tutorial on graph theory and its application to protein interaction networks. First, she gave several examples of biological networks, including the food web and gene regulatory networks, and discussed how graphs can be used to represent real networks. She then introduced various graph-theoretic concepts that are useful for studying networks in general --- for example, node degree, in-degree, out-degree, path, and connectivity --- and others that are useful for characterizing large networks, such as the degree distribution and clustering coefficients. She pointed out that if the degrees of a network have a power-law distribution, this indicates a heterogeneous topology. Nodes or subgraphs with higher connectivity may as a consequence be more important for, but there are exceptions.

System-level Analysis of Cellular Metabolism Using Conttraint-based Methods Ali Navid (Lawrence Livermore National Laboratory)

Navid began with modeling yersinia disease and compared the classic kinetic model with a flux balance constraintbased model. He pointed out that the kinetic model requires many parameters and initial conditions, while flux balance, constraint-based metabolic model only requires the statement of metabolic reaction stoichiometry, enumeration of the demands that a metabolic network must meet, and the experimental measurement of certain specific parameters. He then described the method in detail: the flux balance approach is based on stoichiometric matrices and using optimization methods to analyze models. It can be used, for example, to identify redundancy and establish the importance of genes. Their ease of formulation, versatility in use and broad spectrum of application make metabolic flux balance models a potentially significant new method for the analysis of metabolic physiology and the design of optimal bioprocesses.



Réka Albert (Pennsylvania State University).

Understanding Protein Function on a Genome-scale Through the Analysis of Molecular Networks Mark Gerstein (Yale University)

The number of genes is enormous. The first question is how to represent function at a whole-genome scale; in this context, representation in terms of networks provides a universal language. Network connectivity can be used to determine hubs and bottlenecks. Gerstein pointed out that where bottlenecks occur can be of great importance in gene regulatory networks. In miRNA networks, a measure called RE-score (related to the number of genes down-regulated by the miRNA) is used to indicate the degree of importance; RE-score can be used to classify cancers. Gerstein then discussed the dynamics of networks. He described how to use correlation analysis to determine how a network changes across different environments, which pathways are used more often than others, or which are used as biosensors. Finally, Gerstein discussed variation in protein networks: which parts of networks vary most in sequence and which are under selection (positive or negative). He mentioned that positive selection often takes place at the network's periphery and that central points are usually not subject to positive selection. He then expanded this discussion to include miRNA target networks, concluding with the observation that more highly regulated genes are under more constraints in such networks.

Key Issues in Structural Molecular Evolution Eugen Shakhnovich (Harvard University) This talk was about the physical and evolutionary principles that govern the folding of proteins into their unique biologically active structure. The talk was in two parts: a global view of the protein universe by using PDUGs (Protein Domain Universe Graphs), and a microscopic model of Darwinian evolution. Multidimensional PDUGs can be built based on structure, function, and phylogeny. An investigation into the origin of such unusual global properties of PDUGs reveals a Big Bang scenario where the entire protein universe evolved from a small number of original genes via duplication and divergence. The PDUG approach is somewhat limited. Shakhnovich suggested the PDUG be partitioned into organisms by using sequence similarity. By using thresholds, organismal subgraphs can be built. Organismal subgraph analysis tends to lead to more reasonable phylogenies. Overall, the PDUG approach provides the possibility of a robust structure-based construction of phylogenetic trees. He concluded by presenting a microscopic, physics-based model of fold discovery and evolution that allows visualization of the Big Bang process and suggests a quantitative explanation of this process, including an explanation of the exponents of scale-free PDUGs.

#### DAY 2: TUESDAY, SEPTEMBER 15

Synthesis and Conditional Analysis of Signal Transduction Networks Réka Albert (Pennsylvania State University)

The first talk on Tuesday began with a model of drought signaling in plants. Albert pointed out that experimental observations are mainly indirect. She then continued by giving two general algorithms for building networks from indirect evidence: binary transitive reduction with critical edges, and pseudo-vertex collapse. In the second half of the talk, Albert focused on discussing how graph-theoretical analyses can be extended to incorporate negative regulation and synergistic regulation by several components.

#### Dissecting the Functional Importance of Gene Circuit Architecture

Gurol Suel (University of Texas Southwestern Medical Center)

Suel demonstrated why nature chooses one type of gene regulatory circuit instead of others by comparing the capacity of the circuit to that of other alternative synthetic circuits.

First Steps in Modeling Human Metabolism on a Genomic Scale Eytan Ruppin (Tel-Aviv University)

The talk began with a brief primer on constraint-based modeling of metabolism, followed by a description of

several methods developed by Ruppin in his lab. The first method concerned tissue-specific modeling. This method is aimed at finding an existing metabolic model which best fits the observables by using persistent correlations between observed data with numerical data generated by models. One potential application of this method is prediction of metabolic gene regulation. Ruppin then went on to describe his work on inborn error metabolic disorders, generating predictions of metabolic profiles in biofluids for hundreds of these diseases (MSB09). Finally, he described some ongoing projects, such as developing a generic approach for building tissue-specific metabolic models and providing a computational account for metabolic alterations in cancer.

Metabolic Flux: Key Indicator of Cell Physiology and Determinant of Cell and Metabolic Engineering Gregory Stephanopoulos (MIT)

Stephanopoulos summarized methods for high-resolution determination of metabolic fluxes using stable isotopic labeling methods. He then showed how metabolic fluxes can be applied to identify rate-controlling steps in metabolic networks and thus direct the modulation of metabolism at the genetic level in order to amplify fluxes for the overproduction of fuels and chemicals. Fluxes can also be used along with transcriptional and metabolite data from steady state yeast cultures to elucidate the functions of the yeast global regulator Gcn4p. Some general biological principles can also be revealed: rewiring of metabolic flux by transcriptional regulation, and metabolite-enzyme interaction density as a key biosynthetic control determinant. Finally, Stephanopoulos drew the conclusion that fluxes are a critical indicator of the state of cellular metabolism and represent an irreplaceable guide for metabolic engineering.

#### DAY 3: WEDNESDAY, SEPTEMBER 16

Network and State Space Models: Science and Science Fiction Approaches to Cell Fate Predictions John Quackenbush (Dana-Farber Cancer Institute and the Harvard School of Public Health)

John described how the massive amount of data gener-



ated by the technologies spawned by the "Omics revolution" have both lead to a new understanding of biological systems as information management systems (managing information generated at multiple levels of biological organization gene protein organelle) and to a recognition that sequence data alone is not enough to understand these systems; an understanding of the interactions and relationships occurring within these systems is needed. Part of the challenge of dealing with massive amounts of data is not just the sheer quantity of data, but also the often ad hoc and idiosyncratic ways in which this data is represented in practice, and that it can be stored in different formats and on incompatible IT systems. Two projects through which Quackenbush and colleagues have attempted to address this issue are GCOD (GeneChip Oncology Database), which was intended to be a central warehouse for raw gene expression data associated to cancers, and GeneSigDB, which provides this data in a standard form with standard annotations. With access to standardized data, it's possible to leverage this data to learn something about the network structure that produced it. One method to do this in the context of microarray data is the Bayesian network approach (used successfully by Friedman et al (2000)), where potential networks are evaluated according to their probability given the data, and then altered to increase this probability. Quackenbush and his student Amira Djebbari decided to apply this approach to the ALL/AML dataset from Golub et al to deduce an underlying network. However, since the problem of finding an optimal network through an exhaustive search is infeasible for general networks, they decided to seed the search process by using other gene expression and protein-protein interaction (PPI) data ("priors") to identify candidate networks with which to start the search. (The use of priors was suggested by Wolpert and MacReady (1996) in the context of a more general class of problems.) These priors were data regarding gene co-occurrence in gene expression data available in the literature (PubMed), and PPI data from Rual et al (2005), and optimal networks were found for the four conditions corresponding to the inclusion or exclusion of the priors. Notably, the network obtained by using the PPI prior alone was a significant improvement over the one obtained without priors, and in fact contained real interactions that did not appear in the PPI data but did appear in the literature; the one obtained by using both priors included a part of the cell cycle interaction network. Quackenbush finished by going on to describe what he called the "science fiction" suggested by the state space approach, namely the possibility of formulating a biological principle analogous to the path integral formulation of quantum mechanics, in which information about all possible future paths is used to make accurate probabilistic predictions about the future of a given path.

#### Integrative Analysis of Metabolic and Transcriptional Regulatory Networks for Human Pathogens Jason Papin (University of Virginia)

Jason presented two ways in which metabolic network analysis can be used to make progress against the problem of infectious disease, and concentrated on the examples of Leishmania major and Pseudomonas aeruginosa to illustrate these approaches. Both approaches begin with a reconstruction of the metabolic network of the organism, which proceeds in several steps. In the case of L. major, the reconstructed metabolic network was used to identify all possible "knockouts" that would be predicted to result in growth rate of zero; these knockouts are referred to as "lethal" and the corresponding genes as "essential". The collection of reactions associated to these genes were then compared to databases of FDAapproved drugs to identify drugs which target these reactions, particularly, inexpensive drugs, and more importantly, drugs with low toxicity since current treatments for Leishmaniasis can be toxic. This search resulted in several candidate drugs, three of which were selected for testing against L. major. One of these, Antabuse, seemed to be similar in effectiveness to Amphoteracin B., a drug currently used for treatment of Leishmaniasis, while being very well tolerated and inexpensive. In the case of P. aeruginosa, a study by Levesque et al (2005) identified a large set of genes associated to the ability of P. aeruginosa to infect and survive in the rat. Specifically, mutants of P. aeruginosa were produced using a technique called signature-tagged mutagenesis. These were then tested to see which failed to survive in the rat, and the genes that had been altered in this smaller group of mutants were identified. The products associated to these genes are presumably virulence factors -- agents that promote infection -- and in principle, the reconstructed metabolic network could be used to identify metabolic reactions that are essential, not to the organisms as a whole, but to the synthesis of these virulence factors. Papin ended his talk by describing how work comparing reconstructions for P. aeruginosa and P. putida have led to a better understanding of the metabolic networks of both, including the ways in which genes with sequence similarity may be associated to qualitatively different reactions in the two organisms.

### Insights from Large-scale Protein Structure Network on the Evolution of Protein Function

Eivind Almaas (Norwegian University of Science and Technology)

Eivind talked about a network-based approach to the problem of ascribing function to proteins of unknown function. Under the simplifying assumption that the functions performed by a particular protein are determined by the geometric --- as opposed to the chemical --- structure of its active sites, the function associated to a particular active site of a given protein can be inferred from that of an active site of similar geometric structure whose function is known. In this case, a comprehensive network of proteins based on their structural-functional similarity would be useful both clinically, by allowing for the identification of multiple drug candidates with similar function, and also for gaining insight into the evolutionary origins of protein structure and function since it is known that the genes that give rise to active sites are highly conserved. As a first step in the direction of constructing such a network, a structural similarity was constructed for a restricted class of proteins, namely metalloproteins, since the metal atoms contained in a protein are often associated with active sites of the protein. This was done by searching the Protein Data Bank (PDB) for all proteins containing any of a particular list of biologically relevant metals. The resulting network was then used to try to shed light on the case of 2axt, an enzyme that splits oxygen from water and is critical to photosynthesis in plants and other organisms, and whose evolutionary origins cannot be gleaned from sequence data. A more general analysis of the network showed that, although like metals seem to be more often linked to like metals, the network is not segregated by metal. So, proteins containing different metals often share structural similarity. Most promisingly, however, it was found that proteins are most frequently linked with proteins with the same Enzyme Commission (EC) number, a number which is used to classify enzymes according to their activity, providing

further support for the hypothesis of a strong association between structure and function.



Systems Analysis of Cellular Networks Under Uncertainty Joërg Stelling (ETH Zurich)

Joërg spoke about the efficient ways of developing models in systems biology, where complexity and lack of high quality quantitative data often lead to massive data generation as way of addressing the associated high level of uncertainty. Stelling's principal suggestion was to try to exploit all forms of prior knowledge (including current conflicting hypotheses), produce an ensemble of potential models for which very targeted experiments can be designed to discriminate among the models, and also determine those consequences which follow directly from a knowledge of the underlying network alone, independently of any parameter values that may be taken on by particular models of the network. Flux balance analysis, the matrix-based method described earlier in Papin's talk, is an example of this kind of approach, where the network of metabolic reactions gives rise to a space of biologically-feasible values for the vector of reaction rates, a space which can be described geometrically as a polyhedral cone. The remainder of Stelling's talk was devoted to an example illustrating models that incorporate prior knowledge and conflicting hypotheses, namely the target of rapamycin (TOR) pathway in yeast. A detailed ODE model was developed based on the consensus model, but despite the limited data available, the consensus model could not reproduce the time course of the concentration of the Tip41/Tap42 complex. The good models further separated into two groups: those that predicted that the addition of rapamycin would decrease an initially large concentration of a particular complex, and those that predicted a low concentration of the same complex, with or without the addition of rapamycin. A second experiment showed the latter to be the case, pointing to the second group as the best models and suggesting a novel revised organization of the pathway.

#### DAY 4: THURSDAY, SEPTEMBER 17

Functional Insights from Protein-Protein and Genetic Interaction Maps Nevan Krogan (University of California, San Francisco)

Nevan addressed pathways and complexes. He noted how pathways can be considered fundamental units of cell biology, but how their relationship to each other is difficult to define. Comprehensive tagging and purification experiments have generated networks of interactions that represent most stable protein complexes in yeast cells. Krogan described this work and showed how the analysis of pairwise epistatic relationships between genes complements the physical interaction data, and that furthermore, this analysis can be used to classify gene products into parallel and linear pathways.

The Impact of the Solvent Capacity Constraint on Cell Metabolism

Alexei Vasquez (Institute for Advanced Study)

It is experimentally observed that after o.6 hours, E. coli starts acetate metabolism. With access to a mixture of different carbon sources, such as glucose, galactose and acetate, substrate concentration and the hierarchy of substrate utilization by E. Coli change over time. He discussed how crowding coefficients could be estimated from the available data, and showed that there is a striking agreement between model and experiment. He noted that glucose uptake rate and growth rate are limited, and that there is high variation for large values. He described a quasi steady state flux balance model for analyzing the



Ali Navid (Lawrence Livermore National Laboratory).

substrate utilization hierarchy. Predicted versus measured agreement was shown to be not perfect, but good, and through modification of the model, the agreement between experiment and prediction improved. In the last part of the talk, he introduced a kinetic model. With this model he attempted to answer whether there are optimal rates of carbon uptake that guarantee the highest rate of glycolysis. In this model more than one reaction is involved, and solvent capacityconstraints are included in the model. The results show that the fit between predicted glycolysis rate and the measured rate is not good. Vasquez suggested that the model needs some fine tuning, such as by finding good kinetic parameters. He concluded by suggesting further study of how the addition of solvent capacity can improve metabolic modeling.

#### Network Medicine

#### Albert-Laszlo Barabasi (University of Notre Dame)

Barabasi presented two papers that showed the connection between diabetes and obesity. He addressed the question of how we think about human diseases from the perspective of networks. Diseases are the result of the breakdown of a module in the network. It is important to associate networks to diseases. One such construction is called the diseasome, where genes and the diseases they are associated to are connected by edges. He reported that many diseases actually are connected to each other. Cancers are connected to each other through a couple of genes, for example. If only the essential genes taken into consideration, a high correlation with hubs is obtained, but in nonessential genes no correlation is observed. Essential genes have a high correlation with expressing more tissues. On the other hand, he pointed out that essentiality it has more phenotypes. Some people have many diseases and it is important to determine whether the relationships among these diseases are statistically significant. It turns out that they are. He showed that the genes connected at the metabolic level show a higher

level of comorbidity. In a phenotypic disease network, diseases are connected if they are comorbid, i.e. more likely to be observed together. Barabasi introduced a database for comorbidity, which divide the data into different groups. If a person gets one disease, then all the diseases the person can get are in the neighborhood of that disease. People do not get diseases randomly; they get diseases that are within a close range according to this comorbidity network. This is statistically significant.

#### Interactome Networks and Human Disease Marc Vidal (Harvard Medical School)

Marc introduced four critical ideas in biology: (i) All life is composed of cells and the cell is the simplest unit exhibiting the characteristics of life; (ii) the gene is the basis for heredity, DNA makes RNA and that makes protein; (iii) evolution by natural selection; and (iv) life is based on chemistry. He added that, in his view, the fifth great idea is systems biology: the idea that biological organization is based on logical and informational processes and structures. Vidal introduced a network perturbation model for diseases. Based on the fact that the protein-protein interactome is involved in most processes, he proposed a hypothesis: Global and local properties of interaction networks relate to biology. He suggested an analogy between the disease network and a high school social network; all pair wise combinations for possible physical interactions needed to be tested, which can be done by using the yeast 2 hybrid method. Artifacts are eliminated and orthogonal assays are used to clear false positives. This forms an empirical framework for binary interactome mapping. Vidal then summarized the conditions under which a biological atlas can be constructed. He mentioned that no more than 1/3 of the human genome is available to beused for interactome network, so the possible interactions are sampled, and experiments are performed on those, with different experimental methods giving different protein interaction networks. For the purpose of developing a biological atlas of functional maps, the trancriptome, phenome, interactome, and localizome should be integrated into network studies. He reported that 1200 disorders are associated with approximately 1700 genes. There is a significant overlap between PPI and disease networks. There is not a one to one correspondence between genes and diseases. A novel technique has been proposed, called edgetic perturbation, to understand disease networks. The Human Gene Mutation Database (HMGD) has more than 50000 alleles; this is an enormous amount of information that could be used for network studies. Half of these mutations are specified as in frame and they cause diseases. Vidal then compared the distinct node removal technique to the edgetic perturbation technique, and mentioned that for a particular gene product they might cause distinct diseases.

Systems Biology of Genetic Interactions in Yeast Metabolism Balazs Papp (Biological Research Center, Hungary)

Papp talked about genetic interactions (epistasis) and non-independence of mutations, and suggested that epistasis is interesting because it provides rich functional information, gives insights into the fitness function, and influences many evolutionary processes. The reason for choosing metabolism as a way to understand epistasis is that it's one of the best characterized cellular subsystems. Analysis is also facilitated by the availability of high quality mathematical tools for genome scale metabolic networks and the availability of experimental techniques. The data presented came from two sources: data collected through metabolic SGA miniarrays and quantitative epistasis data generated by the Boone lab. Papp showed that interactions are not randomly distributed, negative epistasis is enriched among duplicate pairs (~16x), protein-protein interactions are over-represented in positive epistasis, and that distribution across broad functional categories is seen in epistatic genes. He also discussed which network properties correlate with epistasis and the effect of local network topology on epistasis. Papp noted that both positive and negative epistasis are enriched for local short paths, however the local enrichment for positive and negative epistasis are different. Both positive and negative genetic interactions are enriched within traditionally defined metabolic subsystems (e.g citric cycle). Directionally coupled pairs (Gene1->Gene2->Gene3) are enriched by a factor of 35 in positive epistasis and by a factor of 10 in negative epistasis. Papp suggested using flux balance analysis for predicting epistasisis, and noted that a small part of negative epistasis interactions could be predicted with high precision, but that the other predictions are not that impressive. Correct predictions help to interpret empirically observed negative genetic interactions. The model is then used to understand epistasis among duplicate pairs. Many duplicates do not show negative epistasis under a given condition. The FBA model accurately predicts epistasis (p<0.01), but sequence similarity or co-expression do not (p>0.11). Papp noted that even the simple case of epistasis among duplicates is a system level property that cannot be captured by duplicate similarity alone, and suggested that the model helps to interpret epistasis. Automated network refinement was achieved by using genetic algorithms to find the optimum parameters, which included, for example, biomass composition and reaction reversibility.

The Capacity for Multistability in Small Networks Dan Siegal-Gaskins (The Ohio State University)

Siegal-Gaskins spoke about determining whether the network can support multiple steady states. His motivation for studying this problem is that doing this with differential equations is hard since many parameters are not known. One can do parameter-free analysis using graphtheoretical approaches, where edges are labeled as positive or negative. Siegal-Gaskins discussed the Thomas and Kaufman conjectures, and noted that if one assumes mass action kinetics, the system equations take the form of a chemical reaction network (CRN). In this context, one can use three tools to find multistability in the network: (i) a computational method, (ii) graphical technique, and (iii) chemical reaction network toolbox. As an example to illustrate the importance of multistability, he noted that if the product of gene binds monomer or dimer matter, then only one of these is bistable. He then discussed the topological features necessary for bistability and the best parameter-free method for establishing if a network has the capacity for bistability, which considers all subnetworks that consists of a single transcription factor gene X and protein P. He limited the analysis to basal production, degradation, activation, repression, and dimerization only, and checked these with all the methods introduced. For the two-component motifs, there are a few more than (~40000), and the INJ method can rule out bistability in 6.7%, however not all classes of subnetworks are equal. He then compared INJ to CRNT and concluded that CRNT seems to be the best tool for studying this, but is not automatic. He compared only 25 two-component subnetworks and applied both methods. In the trichome case, it has been argued that a positive auto-regulatory relation is not necessary. Siegal-Gaskins also mentioned that direct positive feedback appears to be a requirement for bistability in one-component subnetworks.

#### DAY 5: FRIDAY, SEPTEMBER 18

Inferring and Encoding Modules in Biological Networks Chris Wiggins (Columbia University)

Chris addressed the question of whether topology affects optimal information processing. He introduced and discussed a paper (Guet et al Science (2002)), and presented some measurements from the paper.



The Global Landscape of Genetic Interactions in Yeast Chad Myers (University of Minnesota)

Myers began his talk by introducing reverse engineering biology through perturbation. He described how genes are knocked out to try to understand the network, which makes it important to understand how informative single perturbations are. Yeast studies have shown that there are around 1100 genes essential for life and others that are mostly not effective (~5000 genes). Myers noted that global sampling of interactions (including mutant alleles for essential genes), sensitive and reliable assay (low false negatives and false positives), quantitative single, double mutant phenotypes, and complete pairwise genetic interaction network all together will lead to analysis of global network topology. He also introduced a quantitative definition of genetic interaction. To use double mutant data, Myers set measuring reliable, quantitative genetic interactions from double mutant colonies as a goal. The challenge is that the systematic effects account for most of the variation in the data (such as spatial restrictions). Myers noted that normalization greatly increases the quality of the analysis, and so is important in practice. Single mutant finesse estimates that SGA colony data correlate with other assays and colony size was a good measure for this study. Since the study provides a preliminary view of the global yeast genetic interaction network, it helps to understand the connections in the pathway for different biological functions. Myers then addressed the

correlation between genetic interaction degree and functional characteristics, and noted that genetic interaction hubs are highly multi functional. He found that hubs tend to have multifunctionality and interaction frequency varies across biological processes. For example, chromatin/ transcription-related proteins tend to have lots of connections between each other.

### Computational Biology from an Operations Research Perspective

Allen Holder (Rose-Hulman Institute of Technology)

Holder summarized the method of flux balance analysis (FBA) and discussed ideas about haplotype interference. He provided some historical background for operations research (OR), also referred to as "the science of better," and described its origins in World War II. OR often uses network analysis (at least since 1950s). There is burgeoning interest in OR as a tool for computational biology. In the language of OR, FBA is a linear program with constraints. Holder defined the balance metabolic system that is a 6-tuple system. The objective can be to optimize growth, for example, and also find the parameters that give the optimum objective. Degeneracy (25% to 40%) is a common problem in this kind of analysis. It is sensitive to seemingly meaningless environmental changes such as oxygen rate. Cycling behavior does not explain most of the degeneracy. With regards to FBA, Holder defined the minimal environment by parameter estimation, which provides a unique solution. He noted that among the groups of resources, there are elements that can compensate for each other, and growth rate depends on these uptakes. Growth rate will not change by moving in the space, and the minimal environment is like the natural minimum. Holder gave an economic interpretation that may possibly lead to meaningful mathematical analysis, and discussed an example related to enzyme production. He summarized the Nonsubstitution Theorem and introduced a dynamic extension of the theorem. For the haplotype problem he defined the diversity graph, and introduced the Pure Parsimony problem, which is the problem of finding a minimum diversity subgraph example for an inference problem. Polynomial time subclasses for arbitrarily large graphs have been identified. Genotypes can be ordered to geta lattice-like structure on the collection of genotypes, where each connected component is a Glattice.

#### CONCLUSION

This was a well-attended, lively workshop, which clearly fostered discussion and interaction among the participants during the conference, and promises to yield active collaboration among them in the future.

#### Workshop 2 Dynamics of Signal Transduction and of Gene-Protein Regulatory Networks November 2-6, 2009

#### ORGANIZERS

Reka Albert, Penn State University Andre Levchenko, Johns Hopkins University Alex Mogilner, UC, Davis John Tyson, Virginia Tech and State University

Report written by MBI postdocs Marisa Eisenberg, Deena Schmidt, and Yunjiao Wang.

#### **OVERALL SUMMARY**

The processes by which cells obtain, evaluate, and respond to environmental information are complex, formed by a network of signaling reactions that generate a wide variety of robust, specific responses to different signals and cues. This workshop focused on the dynamics of signal transduction and gene-protein regulatory networks, with the aim of fostering deeper and more fruitful collaborations between theorists and experimentalists.

Experimental studies of biochemical reaction networks have revealed many different dynamic behaviors of interest, such as excitability, oscillations, multiple steady states, and signal propagation. Mathematical models can provide useful insights into the dynamical principles underlying information processing by switches and clocks in living organisms, as well as highlighting unexpected properties, such as hysteresis and critical slowing-down, which can be tested in the laboratory.

Analysis of these signaling networks often utilizes a variety of mathematical tools, including graph theory, discrete mathematics, dynamical systems, signal processing theory, and statistical mechanics. In many situations, new approaches are needed to adapt tools developed for engineering applications (such as control theory) to life science problems. Of crucial importance are algorithms and software to enable modelers to build larger, more complex and realistic models of information processing in cells.

#### DAY 1: MONDAY, NOVEMBER 2

Tutorial: New Biology from Modeling, or—What Can Function Tell Us About Structure? Andre Levchenko (Institute for Computational Medicine, Johns Hopkins University)

Levchenko began with an overview of the model development process and the coupling between model and experiment. He introduced the basic idea— to take known, existing experimental information, develop a model (which he pointed out is in many ways more an art than a science), and then continue through a (repeated) feedback loop of testing the model using experiments, refining the model, and testing again. Levchenko then gave several examples of the basic modeling process in different contexts.

His first example addressed how to infer information about the structure of a system from only steady state measurements. Levchenko introduced the yeast mating process (explored in more detail in Tim Elston's talk later in the workshop) of sending out extensions (called shmoos), merging with another yeast cell, and mixing chromosomes. To merge successfully, the yeast must extend in the appropriate direction, which requires sensing pheromone gradients. Levchenko's group utilized a microfluidic device to explore this system, using crosschannels to create a pheromone gradient. Although they were only able to measure steady state responses to various pheromone levels, they found that two kinases, Fus3 and Kss1, play major roles in regulating this process—Kss1 provides negative feedback regulation, whereas the Fus3 pathway includes positive feedback. However, this positive feedback is dependent on the nonlinearity of the Kss1 pathway and is "not a single, simple positive feedback loop."

Levchenko's second example discussed red blood cell development. There are many possible pathways by which cells may differentiate into red blood cells, making the choice of a model difficult. He proposed that when screening models, we should use biology's tendency towards robust responses to stimuli as a criterion for model selection. Levchenko's group screened approximately 300 potential model topologies, fitting each to data. During this process, they varied the parameters for each model and determined the mean and variance for the resulting distribution of fits. A robust model should have a fairly small variance-i.e. the response should be relatively consistent under changes to the parameter values. Levchenko's group used this criterion to score the models, yielding a subset of robust models which shared a common feature: feedforward loops increasing the efflux of cells from state 2. Based on the biology, this suggested that Fas-FasL apoptosis may be involved in the differentiation process. Levchenko's group tested this hypothesis via experiments, confirming a link between Fas, FasL, and state 2.

In the third example, Levchenko's group used a synthetic biology approach to explore switch-like behavior in quorum sensing in Vibrio fisheri. They replaced the existing autoinducer in the Vibrio fisheri signaling network with their own, and uncovered switchlike bistable behavior and hysteresis effects. They built a differential equation model to explore how the system bistablity changes under various amounts of glucose.

The final example explored the NF- $\kappa$ B pathway, and provided an example of the sort of insight that can be gained when a model breaks (as all must eventually). Levchenko closed his tutorial with the take home message that there are many different ways to approach modeling, whether by starting with a detailed model and trying to simplify it, or if there is little information available, investigating a family of models and exploring which features are important to describe the data. One key concept that can help with this process is the idea of robustness.

### Dynamic Regulation of the PDGF Receptor Signaling Network

Jason Haugh (North Carolina State University)

Jason explored regulation and crosstalk in the plateletderived growth factor (PGDF) signaling network. PGDF is involved in cell cycle progression and cell proliferation, and stimulates fibroblasts in wound healing. Haugh introduced the PGDF signaling cascade, which incorporates multiple levels of signaling and a variety of different pathways. This network can be partitioned into two major interconnected subnetworks: the Ras and PI<sub>3</sub>K pathways. The regulatory dynamics of these two canonical pathways are complex, including positive and negative feedback, as well as significant crosstalk between pathways.

Modeling these pathways necessitates integration of biochemical/biophysical processes across multiple scales

and utilizing data from a variety of experimental techniques. By coupling a simplified model of these pathways with a host of experimental data, Haugh's group built an ordinary differential equation model, and systematically eliminated network topologies to uncover crosstalk relationships between PI<sub>3</sub>K and Ras. They found unidirectional crosstalk from PI<sub>3</sub>K to Ras, with PI<sub>3</sub>K insulated from changes in Ras/Erk (see figure). They found that Erk regulation was shaped by at least two negative feedback loops and that there were multiple saturation points in the model/network.



Haugh used a global search algorithm for parameter estimation, yielding an ensemble of ~10,000 parameter sets. Because he was working with an ensemble of models, it was difficult to make predictions, but it was possible to quantify the effects of the two pathways. Haugh's group has since refined the model further using a similar approach to include desensitization of MEK kinase activity. This success with the PGDF signaling networks demonstrates the usefulness of this approach as a method for systematically quantifying signal transduction networks.

Positive and Negative Molecular Signaling Networks Controlling the Fate of Macrophages Liwu Li (Virginia Polytechnic Institute and State Univ.) and Jianhua Xing (Virginia Tech)

Li and Xing talked about positive and negative molecular signaling networks controlling the fate of macrophages. Li focused on the biological problem and Xing focused on how to design and use a model of immune regulation networks. Macrophages contain negative and positive



regulatory loops that finely control the expression of proand anti-inflammatory genes. Li described inflammation as a double edge sword: it helps fight infection and heal wounds, but excessive inflammation can lead to serious problems such as sepsis and heart disease. We'd like to understand the underlying molecular mechanisms responsible for activation and differentiation of macrophages and T helper cells. Receptor proteins, known as IRAKs (interleukin receptor-associated kinases), recognize a diverse array of microbial cues and relay the signals through a serious of intracellular signaling components. Each IRAK member has unique regulation and role in mediating innate immune response. Li's lab has shown that both feed forward and feedback controls exist in macrophages that regulate NFkB-mediated expression of pro-inflammatory genes, as well as nuclear-receptor (NR) mediated expression of anti-inflammatory genes. They also found that the cross-inhibition of NFkB pathway and NR pathway potentially gives rise to two bistable anti- and pro-inflammatory states. In particular, they demonstrated that IRAK-1 is a critical signaling component following diverse microbial stimulants, and that it is involved in regulation of several key transcription factors that are known to be involved in the activation and differentiation of macrophages and T helper cells. Knocking-out of several key molecular players can skew the bistable state to the direction of antiinflammatory flavor, and serve as viable targets for the development of anti-inflammatory therapies.

A Systems Biology Analysis of Yeast Chemotrophic Growth Tim Elston (University of North Carolina)

Tim gave the next talk on a systems biology analysis of yeast chemotrophic growth: dynamics of signal transduction and of gene-protein regulatory networks. Saccharomyces cerevisiae (yeast) have the ability to propagate as haploids. The two mating types, *a*- and  $\alpha$ -cells secrete type-specific pheromones in order to attract a mate and form an  $a/\alpha$  diploid. Stimulation with pheromones leads to a well-defined series of events required for mating, including MAPK phosphorylation, new gene transcription, and morphological changes. In particular, a-cells undergo chemotrophic growth, meaning that they elongate in the direction of increasing pheromone concentration. Elston's lab studies cell differentiation and gradient sensing in yeast. His talk focused on computational and experimental investigations designed to reveal the signaling events that lead to chemotrophic growth in yeast. After giving background information about information processing and developmental decisions in yeast, he discussed the role of Bar1 in chemotrophic growth, and then described models for polarization and gradient sensing. Bar1 is known to degrade pheromones, but it also seems to sharpen the pheromone gradient so that *a*-cells avoid each other. This allows yeast cells to dynamically modulate their environment to achieve better mating efficiency. Lastly, by merging the two individual models for polarization and gradient sensing, Elston finds that a Turing instability can account for each phenomenon, and that cell shape seems to influence polarization.

Noise, Robustness and Memory in Bacterial Chemotaxis Victor Sourjik (University of Heidelberg)

Bacteria such as E. coli navigate in chemical gradients by performing temporal comparisons of ligand concentrations. In the absence of a gradient, cells perform a random walk (the so-called run and tumble) which allows them to efficiently explore their environment. In presence of a gradient, the random walk becomes biased: an increase in attractant concentration rapidly suppresses tumbles and results in longer runs in the desired direction. This initial response is counteracted on a longer time scale by an adaptation system that regulates pathway activity through chemoreceptor methylation. Bacterial chemotaxis is affected by several types of noise, from stochastic ligand binding to Brownian motion to stochastic protein expression, and much of the pathway evolution appears to have been driven by the selection for robust signal processing under these conditions. Sourjik used experiments, bioinformatics tools and computer modeling to look at the effects of the most prominent type of noise. He showed that such gene expression noise is compensated both by the robust pathway topology and by the chromosomal organization of chemotaxis genes. Chemotaxis response results from a balance between excitation and adaptation, and response depends on the adaptation rate. He concluded that heterogeneity of adaptation rates in an E. coli population can be beneficial. The pathway seems to use noise in the expression of adaptation enzymes to broaden the range of environmental gradients that a chemotactic population as a whole can follow.



Eric Shea-Brown (University of Washington)

#### **Poster Session**

The first day concluded with a poster session and reception. Posters presented included:

- Sandip Kar: Exploring the roles of noise in the eukaryotic cell cycle
- Zhanghan Wu: Amplification and Detection of Single-Molecule Conformational Fluctuation through a Protein Interaction Network with Bimodal Distributions
- Baris Hancioglu: The Dynamics of Exit from Mitosis in Budding Yeast
- Debashis Barik: Bistability by Multiple Phosphorylation of Regulatory Proteins
- James Lu: Mapping dynamics to mechanisms: qualitative inverse problems in biology
- Yongfeng Li: A mathematical study of Goldbeter-Koshland model for open signaling cascade with forward activation
- Assieh Saadatpour-Moghaddam: Attractor analysis of

asynchronous Boolean models of signal transduction networks

- Ruisheng Wang: Boolean modeling of microarray data reveals novel modes of heterotrimeric G protein action
- Colin Campbell: A Generalized Model of Immune-Surveillance of Tissue Specific Tumors
- Dirk van Zwieten: Discrete event modeling: Glycolysis and energy balance
- Jon Young: A Simplified Ras-Raf-Mek-Erk Pathway with a Single GFR
- Jianhua Xing: Systems biology and nonequilibrium statistical physics

#### DAY 2: TUESDAY, NOVEMBER 3

Resistance to Diet-Induced Obesity James Liao (UCLA)

James talked about resistance to diet-induced obesity in terms of synthetic metabolism and ensemble modeling (EM). He presented experimental results from various studies on obesity, and then he discussed modeling strategies for synthetic biology and genetic engineering, including how to construct a mathematical model to aid in analysis/design. He focused on the glucose-fatty acid cycle and noted that eating too much food leads to an excess of sugar in the body which then leads to triglyceride accumulation in the liver. One way to treat obesity is to find a way to increase fatty acid metabolism in the presence of glucose. Liao noted that that plants and bacteria digest fats differently from mammals, and this is due to a set of enzymes called the glyoxylate shunt that are missing in mammals. Liao's team cloned genes from E. coli that would express the shunt and then introduced them into liver cells in mice. They found that the glyoxylate shunt cut the energy-generating pathway of the cell in half, allowing the cell to digest fatty acid much faster than normal. It also created an additional pathway for converting fatty acid into CO2. Mice expressing the glyoxylate shunt that were fed a high-fat diet remained skinny, compared with control mice fed the same diet. These mice have increased whole-body fat oxidation, decreased cholesterol, triglycerides, and adipose tissue. Lastly, Liao talked about the role of ensemble modeling to fine tune the design of the desired phenotype in the experiment. The idea is to build an ensemble of models that reach the same steady state but span different kinetics. Perturbation (overexpression, knock-outs, etc.) shifts both the transient and steady states depending on the kinetic parameters used. One then compares the perturbation to the data and retains those models that match the range of experimental data.

#### An Integrated Approach for Drug Development and Customized Therapy

Zoltan Oltvai (Univ. of Pittsburgh School of Medicine)

Integration of advances in genome sequencing and analysis, network biology, structural biology and computational chemistry may have the potential to revolutionize drug discovery and may allow customization of drug therapy. Oltvai described the role of informatics and computational approaches in medicine and pathology, drug discovery in bacterial metabolic networks, and what he called "3P" medicine (personalized, predictive, and preventative). Pathology as a field is somewhat behind compared to other medical fields, characterized by a cluttered, paperbased and physical slide-based system. It needs to transition into digital pathology where all slides are scanned into digital images. The challenge today is that one diagnosis does not fit all. Doctors need to identify disease subtypes systematically using a standardized method, in part using data that is incomprehensible for the pathologist. Next generation sequencing technology allows for routine sequencing of clinical isolates. Despite increasing the amount of data and the ease of generating data, they are under utilized in diagnosis: disease trajectories are not quantitatively predicted and treatment modalities are not calculated. Oltvai showed an example of three morphologically challenging kidney tumors and emphasized the need to develop new ways of visualizing data. His group developed software in C that takes information from tumor slides and yields diagnostic decision support to aid in tumor classification. In particular, it gives a likelihood score for the different tumor categories. It also does a comparative analysis based on SNP array data.

#### An Endogenous Gene Expression Level-to-Rate Converter Provides a Fitness Advantage Leor Weinberger (UCSD)

Historically, signal transduction circuits have been known to differentiate between signals by amplifying inputs to different levels. Weinberger introduced a level-to-rate converter circuit, i.e. novel transcriptional circuitry that dynamically converts greater input levels into faster rates without increasing the final equilibrium level. He noted the trade-off between rate and level: getting faster rates at the expense of achieving higher levels. Weinberger described how this trade off occurs in the human herpesvirus (cytomegalovirus). The CMV transcriptional master



circuit includes the Major Immediate Early locus (MIE), and its promoter which drives two alternative-splice variants: IE1 and IE2. IE2 is essential for viral replication, but is highly cytotoxic to the cell. It also transactivates many CMV genes and disarms the cellular defenses. Weinberger's group used time-lapse microscopy to study the infection of live cells in real time. Combining experimental approaches with mathematical modeling, they showed that level-to-rate conversion results from a highly self-cooperative transcriptional auto-regulatory loop encoded by IE2. They noted that cutting out IE2 auto-regulation eliminates level-to-rate conversion and severely impairs viral replication. Hence, level-to-rate conversion yields a significant fitness advantage for the virus. Even minimal IE2 feedback circuits, lacking all other viral elements, maintain this fitness advantage via level-to-rate conversion. Lastly, Weinberger talked about a new approach for antiviral therapy to reach superspreader groups, specifically for HIV/AIDS. He introduced the idea of therapeutic interfering particles (TIPs) and then discussed a multiscale model of therapeutic TIP therapy for HIV/AIDS. They found that TIP therapy overcomes mutational escape of the virus, drastically outperforms even the most optimistic HIV vaccine, and is robust to behavioral disinhibition.

### Structural Sources of Robustness in Biochemical Reaction Networks

Guy Shinar (Weizmann Institute of Science)

An important property of many biological systems is robustness, or the capacity for sustained and precise function even in the presence of environmental disruption. We'd like to know if there are common network features incorporated by different biochemical modules to ensure the robustness required. Shinar introduced the concept of absolute concentration-robustness (ACR): A biochemical system is said to exhibit ACR relative to an active molecular species if the concentration of that species is identical in every positive steady-state the system might admit. He presented one main theorem to determine when a system exhibits this property. The function of an ACR-possessing system can be protected even against large changes in the overall supply of the system's components, such as those arising from cell-to-cell variability or from variations in the same cell over time. This integration of mathematics and chemistry allows the identification of subtle structural characteristics that will impart ACR to any mass action network possessing them. For example, these core network features provide a common source for the strong concentration robustness observed experimentally in two very different systems: E. coli EnvZ/ OmpR osmoregulation and IDHKP/IDH glyoxylate-bypasscontrol system. The same structural foundation supports a large variety of biochemical networks for which strong concentration robustness is essential.

#### Afternoon Short Talks: Oleg Igoshin, Christian Darabos, Jaewook Joo

Multiple Positive Feedbacks Lead to Bistability and Low-Pass Filtering in Gene Regulatory Network Module Controlling Hematopoiesis Oleg Igoshin (Rice University)

Oleg looked at how multiple positive feedback loops lead to bistability and how low-pass filtering in a gene regulatory network module controls hematopoiesis. He proposed a method to quantitatively characterize the regulatory output of distant enhancers using a biophysical approach. In particular, he used experimental results from transcriptional reporter libraries to determine free energies of protein-protein and protein-DNA interactions. He then applied his method to a network module that plays an important role in cell fate specification of hematopoietic stem cells. This network consists of three transcriptional regulators that positively regulate each other's transcription by acting at distant enhancers; this network is called the Scl-Gata2-Fli1 triad. Igoshin's model showed that this triad is inherently bistable with irreversible transitions in the presence of physiologically relevant signals such as Notch, Bmp4 and Gata1. He used the model to predict the sensitivity of the network to mutations and to indicate a possible role of the triad architecture in filtering transient stimuli. This work yields a new method for estimating binding free energies directly from transcriptional reporter expression data.

Are Cells Really Operating at the Edge of Chaos? A Case Study of Two Real-life Regulatory Networks Christian Darabos (University of Lausanne in Switzerland)

In the title of his talk, he asked the question: Are cells really operating at the edge of chaos? To answer this question, he presented two cases of genetic regulatory networks found in the literature (mouse embryonic stem cells and the yeast cell-cycle) that have been modeled using random Boolean networks. Darabos began by citing Kauffmann's hypothesis that living organism cells operate in a region bordering order and chaos, the so-called critical regime. The goal was to identify under what conditions this hypothesis holds. This property gives living organisms both the stability to resist transcriptional errors and external disturbances, and the flexibility to evolve. Darabos then introduced random Boolean networks (RBNs) and their properties. They compared the behavior of the two systems under the original RBN update function and a novel additive function that takes into account the activation and repression of a gene. They believe this new update function is more biologically realistic. Lastly, Darabos emphasized their numerical results which showed that there exist values in both update functions that allow the models to operate in the critical region, and that these values are comparable in the two GRNs.

Stochastic and Heterogeneous Dynamical Response of NFxB upon Lipopolysaccharide Insult to Live Macrophages Jaewook Joo (Penn State)

Jaewook talked about the stochastic and heterogeneous dynamical response of NF-кВ upon lipopolysaccharide insult to live macrophages. The kinetics and key controlling components of the Toll-Like Receptor 4(TLR4)-mediated innate immune response to infectious stimuli are poorly understood. Using computational modeling and live cell imaging, Joo showed how different lipopolysaccharide (LPS) dosage levels elicit different immune responses in individual immune cells. Their study focused on the LPSinduced nucleo-cytoplasmic translocation dynamics of NF-kB, one of endpoint proteins in the TLR4 signaling pathways. An integrative approach of computational modeling and time-lapse fluorescence microscopy was used to investigate the kinetic mechanisms of NF-KB translocation dynamics in single cells. They developed a stochastic model of the NF-kB signaling pathway that includes multiple negative and positive feedback loops, which predicted that the NF-kB nucleo-cytoplasmic translocation oscillates in LPS-stimulated individual cells. Another prediction was that the extrinsic noise-originated cell-to-cell variability, modeled as the different kinetic conditions of the individual cells prior to the LPS stimulation, diversifies the shuttling patterns of NF-kB. Joo showed that both model predictions were experimen-

tally validated. He noted that although the biological functionality of NF- $\kappa$ B oscillatory shuttling remains to be proven, this systems biology study illustrated the highly stochastic and heterogeneous nature of the immune response in single cells.

#### DAY 3: WEDNESDAY, NOVEMBER 4

Discrete Dynamic Modeling of Signal Transduction Networks: Survival Signaling in T-LGL Leukemia Réka Albert (Pennsylvania State University)

Albert constructed models of signaling networks for cytotoxic T-cells in T-LGL leukemia using a Boolean framework where nodes are described by discrete states, and edges indicate causal relationships. Inhibitors were represented by NOT gates, conditional activation by AND gates, and independent activation by OR gates (with the more general OR gate used as the default, unless additional information was available). To implement time, they used asynchronous updating in which the state of each node is updated in random order, thereby also adding an element of variability to the model behavior. Albert's group validated their model against experimental data, and also provided several new predictions—they found 2 steady states, where approximately 70% of cells die, and 30% survive (the choice of steady state was determined by the timing and initial conditions). She also identified minimal initial conditions and several key mediators needed to produce a T-LGL-like state. If the state of any of these key mediators is reversed, the T-LGL-like state is lost. These predicted mediators capture the major known mediators of T-LGL, as well as some novel predictions. Their simulations also suggested that T-bet needed to be active together with NF-KB to reproduce the low IL-2 production phenotype in leukemic T-LGL, which they have since confirmed experimentally. Albert concluded her talk by noting that the network-based discrete dynamic model has predictive value and allows for the discovery of new strategies and hypotheses. In this framework, the topology of the interaction network plays a determining role in the system behavior, and their methodology can be further refined by incorporating known timing information into the update order.

Random Ideas About Biology Jehoshua (Shuki) Bruck (Caltech)

Jehoshua took us on a tour through three major concepts: random, biology, and ideas. Bruck argued that mathematical biology needs to develop the 'language' of biology—the mathematical machinery to think about biological systems properly and design the tools we need. Bruck next delved into a little bit of history on literacy, and Alexander Luria's 1930's study of literacy in the Soviet Union. As people transitioned to literacy, their ability to abstract and generalize increased—Bruck asserted that we have not (yet) reached mathematical literacy with respect to biology. To address this issue, he contends we must understand the process that led to literacy in other things (e.g. in engineering, etc.). Using George Boole as an example, he discussed how reasoning with, say, relay circuits requires the 'syntax' of Boolean algebra.

Similarly, Bruck suggests that if biological, stochastic networks are our topic of interest, we need to find new abstractions with which to reason about these networks. Bruck explored one potential framework, using networks with edges controlled by Bernoulli random variables—essentially a set of probabilistic relay circuits. Using this framework, they developed algorithms for synthesizing probabilistic relay circuits. Bruck's group plans to develop the mathematical syntax related to biology by focusing on synthesis.

Learning Signaling Pathway Structure from Single Cell Data Karen Sachs (Stanford University)

Karen discussed using Bayesian network models to go from phospho-molecular profiling to signaling pathways. Her group uses flow cytometry to measure data protein data at a single cell level, yielding thousands of data points. By adopting this approach, they are able to test multiple conditions with details for the entire population (i.e. a histogram of signal for each condition and each protein). To model this data, her group utilized Bayesian network models, where network nodes represent the level/ activity of a given protein, and edges represent influence/ dependency between proteins. Using this framework, they determine a conditional probability distribution be-


German Enciso (UC, Irvine)

tween proteins, where each node has a probability of a state given its parents. However, the Bayesian framework also necessitates that these networks must be acyclic—a problematic assumption as biology has many known cycles. Dynamic Bayesian networks offer a possible solution, by letting each node represent a given species at a particular time, allowing cycles to be represented this way. By perturbing different species within the network, Sachs' group was able to build an accurate model of the T-cell signaling network, as well as demonstrate the usefulness of the dynamic Bayesian network modeling framework.

Phenomenological Models of Regulatory Networks Ilya Nemenman (Emory University)

Ilya utilized methods from statistical physics and model selection to simplify network models, reducing the number of degrees of freedom. Nemenman introduced a stochastic model of kinetic proofreading, and showed that although there are a number of different probabilistic behaviors and ways that molecules can be assembled, once proofreading is included, the resulting behavior depends only on the Peclet number, with low Peclet number yielding an exponential distribution and high Peclet number yielding a deterministic case. Although the microparameters play a role in determining the system behavior, they do so only by affecting this single major macroparameter.

This suggests a classic adage, often useful when modeling biological systems—to build the simplest model possible (but not simpler). Utilizing the Bayesian Information Criterion (BIC), his group developed a parsimonious model of heat evasion in C. elegans, revealing a potential connection between thermotaxis in C. elegans and chemotaxis in E. coli.

Patterns of Oscillation in Coupled Systems Marty Golubitsky (MBI, Ohio State University)

Using a very general model framework, Marty Golubitsky

and Ian Stewart have been able to draw conclusions about the dynamics of networks, based only on their topology/architecture. Golubitsky began by introducing the general form of network system to be studied. There can be multiple types of nodes and arrows (couplings), where the type of arrow or node indicates a different mathematical form. For this talk, he focused on regular networks, where the nodes and arrows are all of the same kind. Differential equations associated with these networks have synchronous dynamics since the diagonal obtained by setting all nodes equal is flow invariant. The two-node network  $x_{i}=f(x_{i},x_{i}), x_{i}=f(x_{i},x_{i})$  provides an example. This network yields two types of periodic solutions: in-phase oscillation, and half-period out of phase oscillation. Golubitsky continued by examining Hopf bifurcation in this network. He noted that the Jacobian is

$$J(\lambda) = \begin{bmatrix} \alpha(\lambda) & \beta(\lambda) \\ \beta(\lambda) & \alpha(\lambda) \end{bmatrix}$$

where alpha is the linearized internal dynamics and beta is the linearized coupling. Symmetry implied that such Jacobians have invariant subspaces each of which can exhibit a Hopf bifurcation and these types of Hopf bifurcation lead to the periodic solutions mentioned above. Note that a system of differential equations has symmetry if  $F(\gamma x, \lambda) = \gamma F(x, \lambda)$ .

Golubitsky further examined this framework with several other examples (unidirectional and bidirectional rings, etc.), demonstrating the various spatial and spatiotemporal symmetries that are encoded in each. Next, he introduced an application of this methodology to quadruped gaits (e.g. walking, trotting, etc.). We can characterize each gait by its symmetries—by determining the phase shifts for each gait, we can examine what neuronal networks would produce these outputs. Golubitsky showed that four nodes in the network will not suffice—with just four nodes, one cannot distinguish between the gaits 'trot' and 'pace'. They found that the simplest network has  $(Z/4Z) \times (Z/2Z)$  symmetry, with six irreducible representations: pronk (a type of leaping gait), pace, trot, bound, walk, and a new gait—jump. This network can be matched up with known muscle groups (flexor and extensor), and these basic six gaits are seen in many animals, from gerbils to horses. As an interesting addition, Golubitsky also noted that the bipedal walk cannot be obtained from any standard quadruped gait by breaking symmetry between the fore and hind legs. Golubitsky also explored several other perspectives and applications of this general method, looking at a chain with back coupling (which he shows can have robust synchrony even without symmetry), quotient networks, and symmetry in the vestibular system.



#### DAY 4: THURSDAY, NOVEMBER 5

Modeling Cell Polarity and Motility: Signaling to Actin Leah Edelstein-Keshet (University of British Columbia)

This talk represented Leah's joint works with Adriana Dawes, Alexandra Jilkine, Stan Maree, Yoichiro Mori, Veronica Grieneisen, and Ben Vanderlei. Remodeling of the actin cytoskeleton is recognized to be an important process underlying eukaryotic cell motility. However, regulation of the spatio-temporal dynamics of actin is essential in order for the cell to orient and move correctly in response to chemoattractive stimuli. In this talk, Keshet surveyed efforts in her group over the last years to understand this process. In the beginning of the talk, she made a short introduction to cytoskeleton cell motility and continued to describe how motility is probably regulated. Then she raised the question: how does a cell organize, coordinate itself to sustain robust motility? She showed that a module of switch-like proteins can set up robust cell polarity, leading to increased actin nucleation at a cell "front" and increased contraction at the opposite pole ("rear"). Keshet and her lab members set up a one-dimensional model by ignoring both nucleus and thickness. Signal comes from one end of the simplified cell (segment). Then they continued to extend the model to two-dimensional space. 2-D cell motility simulations and analytic treatment of reduced versions of the mathematical model lead to insights about the underlying mechanism.

Enzyme Substrate Competition in a Developing Embryo Stanislav Shvartsman (Princeton University)

Shvartsman first introduced phosphorylation cascade phenomena. It was known that the same kinase could have multiple substrates. That is, the same kinase can have many targeting enzymes. Then he raised the questions:

- (I) Is substrate competition significant in vivo?
- (II) If yes, is it beneficial or detrimental?

By both experiments and mathematical model, Shvartsman showed that MAPK phosphorylation is affected by levels of its substrates, the effect does not depend on transcriptional response to Bcd, direct interaction with enzyme is essential for substrate-dependent control of MAPK phosphorylation, substrate levels control MAPK activity. In the end, Shvartsman drew the conclusion that substrate competition is significant and is beneficial for integrating signal patterns.

Stochastic Models of Cell Cycle Regulation in Eukaryotes John J. Tyson (Virginia Polytechnic Institute & State University, Virginia Bioinformatics Institute)

Tyson began with a short introduction to the cell cycle and introduced a deterministic model developed in 2004. The model accurately accounts for the average phenotypic properties of wild-type cells and 150+ mutant strains. However, the deterministic model cannot account for the considerable variability in cell cycle properties that have been observed among genetically identical cells. This variability is due in large part to small numbers of molecules in yeast cells: 100's - 1000's of molecules of each specific protein and only 10's of molecules of each specific mRNA species per yeast cell. How can the cell cycle function reliably in the face of the large intrinsic molecular fluctuations implied by such numbers? Tyson and his collaborators addressed this question by constructing a realistic model of Cdk regulation in budding yeast that is suitable for exact stochastic simulation by Gillespie's algorithm. The results of this model compare favorably to the extensive statistical properties of budding yeast cell cycle progression collected recently in Fred Cross's laboratory.

### Afternoon Short Talks: Richard Yamada, Shinya Kuroda, Attila Csikasz-Nagy, and Dieter Armbruster

Molecular Noise Enhances Oscillations in the SCN Network Richard Yamada (University of Michigan)

Richard began with an introduction to the SCN network: the SCN network is composed of autonomous oscillators. Each cell is governed by a complex biochemical network involving the core circadian components. In this talk, he focused on addressing the dynamics of coupled oscillators (coupled neuron cells) with noise. By simulating SCN network with 100 individual coupled cells, he showed that transcriptional noise enhances oscillations rather than dampening them.

#### Temporal Coding of Akt Signaling Networks Shinya Kuroda (University of Tokyo)

In cellular signal transduction, information in external stimulus is coded as temporal patterns of signalling activities; however, temporal coding mechanism has been poorly investigated. In this talk, based on a model of the epidermal growth factor (EGF)-dependent Akt pathway in PC12 cells, he showed that the Akt pathway exhibits low-pass filter characteristics, and that this characteristic of the Akt pathway can explain the decoupling effect of peak amplitudes between receptor and downstream phosphorylation. Because low-pass filter characteristic is an intrinsic feature of biochemical reactions, this finding raises a caution in interpreting biological data without temporal information.

Role of Protein Removal in Signaling Attila Csikasz-Nagy (University of Trento)

In many models, it was assumed that total amount of certain protein is fixed. However, this total amount could change as each protein has a degradation rate and production rate, though these rates are relatively small. Csikasz-Nagy proposed a biochemical network unit model to study the effect of such an assumption. Then, based on simulations on several models, he concluded that ignoring one of the rates may lead to (i) adaptation, once and only once, (ii) adaptation for multiple input increases, (iii) switch like response.

Dynamic Simulations of Single-Molecule Enzyme Networks Dieter Armbruster (Arizona State University)

Armbruster presented first a novel method of simulating stochastic biochemical networks using discrete events simulation techniques. Then he applied the technique to the glucose phosphorylation steps of the Embden-Meyerhof-Parnas pathway in E. coli. He showed that a deterministic version of the discrete event simulation reproduces the behavior of an analogous deterministic differential equation model. The stochastic version of the same model predicts that catastrophic bottlenecks in the system are more likely than one would expect from deterministic theory.

#### DAY 5: FRIDAY, NOVEMBER 6

Ratchets, Phase-locking and Cell Cycle Control Frederick Cross (The Rockefeller University)

Frederick addressed the question of why things happen in one cycle. Cross demonstrated a nice and simple model to explain this phenomena. The model gave meaningful predictions which were verified in experiments. Moreover, the model is small and simple enough to have an explicit solution and only contains three parameters, which could be obtained in one experiment.

Elucidating the Architecture of the CDK1-APC Oscillator Joseph Pomerening (Indiana University)

In the final talk, Pomerening focused on addressing the dynamical behavior of the cyclin-dependent kinase 1 (CDK1) - anaphase-promoting complex (APC) oscillator. While protein synthesis, proteolysis, and phosphorylation-dephosphorylation events drive this system in general, it remains unclear how these inputs together confer the overall output of this oscillator. The Wee1-Cdc25-CDK1 module hints at the possible regulation by players that are involved in other aspects of mitotic control, and recent evidence has confirmed relationships between these regulators. Might the involvement of other M-phase kinases in the activation of CDK1 serve to tune the output of this kinase as a function of cyclin stimulus? To answer this question, Pomerening and his collaborators have initiated a systematic analysis of the activities of M-phase kinases in relation to the pattern of cyclin stimulus and CDK1 activity in Xenopus egg extracts. Their overall goal is to map and dissect experimentally the connections of the embryonic M-phase activation network, and to gather and apply these quantitative details towards the refinement of their mathematical model of the CDK1-APC oscillatory system.

### Workshop 3 Synthetic Biology January 25-29, 2010

#### ORGANIZERS

Jeff Hasty, UC, San Diego Ron Weiss, Princeton University

Report written by MBI postdocs Erik Bloomquist, Marisa Eisenberg, and Dan Siegal-Gaskins.

#### **OVERALL SUMMARY**

Synthetic biology is concerned with the design of genetic networks that perform desired functions in single cell and in multi-cellular environments. With the advent of a wide range of genomic and molecular engineering tools has come the ability to engineer cells—to incorporate new genetic circuits or re-wire existing ones to tune cellular processes for new goals. Such synthetic circuits can be used to gain insight into the molecular components of gene regulation and have applications in a variety of areas including tissue engineering, biomaterials, and biosensing.

Building genetic circuits requires a host of components to regulate transcription, translation, phosphorylation, and numerous other tasks. These components are then assembled to form engineered molecular networks, which involves using tools from nonlinear dynamics, statistical physics, and control theory, as well as traditional molecular biology techniques. This workshop discussed foundational technologies for synthetic biology, such as developing a library of gene circuit component parts, and identifying, designing, and constructing modules in molecular networks. The workshop presented recent work designing and constructing gene/protein networks and circuits, with applications in engineering metabolic networks, cancer detection and therapy, biosensing, and energy production.

#### DAY 1: MONDAY, JANUARY 25

Chemically Modified Viral Capsids as Platforms for Drug Delivery and Solar Energy Collection Matthew Francis (University of California, Berkeley)

Francis began by discussing the utility of proteins as a component for materials. Advantages include that proteins are rigid 3D structures, with multiple attachment sites for different objects, can self-assemble (under the appropriate circumstances) and change shape dynamically. While proteins aren't currently standard, he believes that as we become more interested in lowering toxicity and energy costs, they will become more and more important. Thus, a major goal of his lab is to build customized protein components, ranging from chromophores to radiolabels, polymers, and catalysts. The first example protein component he explored was viral capsids. Capsid proteins have excellent self-assembly abilities, and Francis has developed methods for loading these self assembling hollow capsids with chemical cargo, and targeting the loaded capsids to a variety of tasks and locations, including PET and MRI imaging which specifically images tumors, or for delivery of chemotherapeutic agents directly to tumor cells.

Next, Francis discussed some of his group's ongoing research in chromophore development for solar energy collection. He modeled his chromophore set-up after bacterial photosynthetic reaction centers, which collect solar energy much more efficiently than our current systems. Using rod-shaped tobacco mosaic virions to space the chromophores properly for energy transfer, he has developed an effective energy collector which is resilient even under 90% photobleaching.

Lastly, Francis explored the uses of metallothioneins, protein strands that form pockets that can bind metal ions. Metallothioneins form the body's first line of defense against accidental ingestion of mercury and other metals. Metallothioneins have been engineered into plants to address mercury spill areas, and Francis' group have used them to develop a polymer gel which undergoes a shape change in the presence of heavy metals which are hazardous to humans (e.g., Hg, Cd, etc.). Using this engi-



Erik Winfree (Caltech)

neered metallothionein gel, they are developing systems for binding and removing heavy metals from water systems, to make for easier water testing in the field.

Plant synthetic biology for human and environmental use June Medford (Colorado State University)

June presented her work exploring the systems biology possibilities in plants. Medford focused on a few main topics: detector plants (phytodetectors), toggle switches, and renewable energy. She pointed out a currently unmet need to monitor human environments for pollutants, explosives, pathogens, and terrorist agents. Typical detection methods (e.g., lab sample analysis) can be expensive in both time and money. Plants, on the other hand, are present in most human environments and have inbuilt detection systems for many compounds. To build a plant based detector, Medford needed to engineer an input system that can sense compounds of interest to humans, and a readout system which produces changes in the plants that humans can recognize. Building the genetic toggle switches needed for detector plants requires a bistable system with the capacity for memory, and requires a rapid response. Medford has developed plants which can function as multiple-use biosensors (they can 'reset' and be used again to detect a new signal) and indicate the presence of a compound (in this case, TNT) by changing color from green to white. Eventually, she suggests that perhaps we can replace whole-body scanners with a detector garden that you simply walk through and changes to the plant colors are observed. The toggle switch concept is also useful for developing plant applications in renewable energy, utilizing controllable switches in plants and algae.

Limits of Modularity of Cellular Signaling Systems Alex Ninfa (University of Michigan Medical School)

Alex explored the question: What are the factors that af-

fect the modularity of signal transduction systems? In order to design cellular modules and circuits, understanding why modularity can break down is important. Ninfa explored several systems which could yield a wide range of responses, to the point that he argued the system could no longer truly be viewed as a module. He demonstrated bicyclic systems which can display hyperbolic sensitivity or great ultrasensitivity, depending on the circuitry and parameters. Ninfa also explored how retroactivity, where downstream clients can affect the system, can alter the system dynamics, affecting the system bandwidth and reducing the amplitude of responses to time varying input signals. These retroactive effects can be influenced by a variety of factors, such as catalytic rates of cycle enzymes (cycle flux) and frequency of input stimulation. Ninfa concluded by stating that if one has cycles of reversible covalent modification, then every downstream client matters as they form a network which can affect the reaction of interest. Downstream targets of a system matter not just because they compete with each other, but they also can send a signal back upward and slow the response of the system as a whole, having large effects on the dynamics of the system.

A completely in vitro protein engineering pipeline Homme Hellinga (Duke University)

Describing the characterization of protein variants as "central" to many fields, Dr. Hellinga stressed the need for the predictable and rapid prototyping of new proteins, using procedures amenable to large-scale studies. To this end, Dr. Hellinga developed a protein production and analysis pipeline that does not require cloning, instead using oligonucleotides and miniaturized biophysical and functional assays for in vitro construction and characterization. Dr. Hellinga concluded his talk by detailing how synthetic gene sequences can be designed to optimize protein expression in E. coli, by tuning AT/GC composition, secondary structure, and codon usage.

#### DAY 2: TUESDAY, JANUARY 6

Stochastic Gene Expression, Redundancy, and Cell-fate Decisions in Development Scott Rifkin (University of California, San Diego)

Scott explored how stochasticity affects developmental cell fate decisions. The effect of an allele, mutation, or protein variant on phenotype often depends on the genetic or environmental context. But even if both of these are fixed, the effect of an allele can still be indeterminate. For example, genetically identical bacteria show different expression levels. Rifkin's talk explores how stochasticity plays a key role in this indeterminate behavior. His group studies C. elegans, which has an intricate and exact developmental plan, down to the cell (each C. elegans has



Homme Hellinga (Duke University)

959 cells). From one cell, these worms have a pattern of cell division which is almost invariant, so that each cell can be traced back through its progenitors to that very first cell. Rifkin's talk focused on the E-cell lineage of intestinal cells. This system is quite isolated and modular, forming 20 intestinal cells. Using fluorescent imaging data, they were able to determine transcript levels (gene expression levels) for the primary components of the network controlling the intestinal cell fate decision (see figure). In skn-1 mutated cells, some fraction of the embryo cells will still develop into intestinal cells, however, the number of intestinal cells depends on the mutation and the temperature. At 25°C, Rifkin found not only partial penetrance of the lack of intestinal cells, but also partial penetrance of gene expression effects. Med-1,2 and End-3 were almost completely missing, however End-1 and Elt-2 were extremely variable, ranging from absent to normal levels, with End-2 appearing to follow a bimodal distribution in the cell population. Rifkin's group found that End-1 needed to reach a threshold level within a certain time window in order to activate Elt-2. If these conditions are met, then Elt-2 was turned on for that cell, whereas otherwise that cell would remain with Elt-2 turned off. The stochastic variation in End-1, Rifkin discovered, was due to chromatin remodeling affecting stochasticity of gene expression. When the chromatin is open, many transcripts can be made, but if it is closed, all transcription stops, leading to burst-y transcription effects. Skn-1 recruits a

chromatin remodeler CBP-1 which opens the End-1 DNA to make it consistently available for transcription—with Skn-1 disabled, End-1 experiences only transient bursts of transcription. They also found that End-3 may be more responsible than End-1 for Elt-2 activation, so that End-1 and End-3 provide complementary semi-redundant pathways for Elt-2 activation.

#### Synthetic Biology: From Modules to Systems Ron Weiss (MIT)

Next, Ron explored modularity in systems and synthetic biology. He began by examining the layers we use in designing synthetic biology systems—from modules, to devices, to physical objects. Weiss' goal was to take proven engineering principles, and use and apply them in interesting and challenging aspects of biology. One important issue, Weiss pointed out, is incomplete information—as synthetic biologists, one must try to engineer cells before we fully understand them. The main topics Weiss covered included basic modules (digital, analog, multicellular), local rules/global behavior (Turing systems), larger scale system design (such as pathogen sensing and destruction, programmed tissue engineering, and artificial homeostasis).

When it comes to basic modules, he went over some of the components implemented in his lab: digital switches built using noisy components, analog pulse generators, and spiral ring pattern generators. With this in mind, Weiss continued on to discuss Turing patterns, exploring whether his group could design a Turing pattern formation system. His group engineered the cellular machinery and circuits based on Turing systems, and found that the bacterial systems did form speckle patterns, although they may not have the characteristic length scale typically seen in Turing patterns. Based on their data, Weiss suggested that the original Turing model may not be quite accurate, and that it may not be feasible to ignore issues like stochasticity in gene expression.

Next, Weiss explored developing artificial tissue homeostasis for beta cells. His goal is to maintain a population level of beta cells using autoregulated differentiation of embryonic stem cells that counterbalances the autoimmune attacks involved in Type 1 diabetes. They designed a 22-component system architecture, ran simulations, and found that the stem cell population collapsed, followed by the beta cells. The main problem was that the feedback effects were too slow. Instead they added a fast toggle switch and a messy oscillator as a symmetry breaking mechanism, so that only a small fraction of the stem cell population can differentiate at any given time. These simulations yield non-collapsing cell populations, although they are still working on keeping the beta cell population steady rather than oscillating. By using noise effects to their advantage, to keep only a small portion of the stem cell population differentiating, they are making steps towards maintaining an artificially induced beta cell homeostasis.

#### Enabling Technologies for Synthetic Biology Jingdong Tian (Duke University)

Jingdong discussed the development of foundational technology for high throughput gene and genome synthesis, precise gene design and expression control, efficient metabolic pathway engineering, and accurate automated protein engineering. A major problem in gene and genome synthesis is cost. To synthesize a small 5Mbp genome, one needs 250,000 oligos of 40-mers, 60,000 liters of solvents, and costs around 100 million dollars. To improve this Tian's group has been developing methods for chip-based oligo synthesis, which improves costs significantly. His group has also been combining this idea with microfluidics to improve his method. Additionally, they are working on inkjet based synthesis, which allows them to print the chemicals onto chips for the synthesis process. Using their new technology allows for extremely precise synthesis patterns of oligos on a chip. Tian's goal is to develop a high throughput integrated microfluidic gene synthesizer, which could reduce the costs of synthesizing a 5Mbp genome to less that 50 thousand dollars.

Next, Tian addressed systematic gene design for achieving precise protein expression control. Tian's group examined protein expression from 1400 versions of the same gene. He showed that different choices of codon for the same protein yield significant variation in how strongly the proteins are expressed. Utilizing this idea, his group has developed high throughput codon optimization to improve protein expression precision. They have used this technology in several applications, such as HIV vaccine development and increasing insulin production in human cells.

Tian also discussed bioplastic synthesis in the context of optimizing metabolic pathways. The goal of this project is to develop biological plastics which are biodegradable and also useful for medical applications. The group constructed a bioplastic synthesis pathway in an E. coli strain and is exploring questions of how to shape the plastic, and what properties bioplastic has as a material. Lastly, Dr. Tian discussed his work on a molecular engineering protocol for accurate automated protein engineering. Using this protocol, they have designed a clamp based on DNA restriction nucleases by lowering the specificity until it no longer has restriction activity.

### Engineering RNA-Based Chemical Interfaces for Synthetic Biology

Yohei Yokobayashi (University of California, Davis)

Yohei discussed how to build chemical interfaces for synthetic biology circuits. There are many examples of interesting circuits, such as oscillators, logic gates, pulse generators, and others. To make use of these circuits, Yokobayashi points out that we need an interface which can take in inputs (both physical, such as light and temperature, and chemical, such as hormones, drugs, etc.) and return outputs. A chemical interface needs to take molecules, recognize them specifically and change gene expression. There are many examples of this in nature, but to design and build novel transcription factors can be very challenging. Yokobayashi's group has started working with RNA instead of protein for this purpose as it's more 'engineerable', has intuitive secondary structures, and there exist prediction algorithms and experimental methods to confirm these predictions. RNA is also very versatile and can perform a wide variety of roles in the cell, including molecular recognition and regulating gene expression. Using RNA aptamers and ribozymes, they have built chemical interfaces in mammalian cells, which do not require engineered transcription factors, are compact and compatible with a variety of promoters and vectors, and are modular. They have also used riboswitches to engineer chemical interfaces in bacteria. They use a genetic selection algorithm to design the riboswitches and have developed riboswitches that act as on and off switches and logic gates for bacterial processes. This process should be applicable to any switches or circuits with simple on/off output, making it useful to build synthetic biology libraries, in vivo metabolite sensors, chemical engineering, and many other applications.

#### Phenotypes in the Design Space of Biochemical Systems Michael Savageau (University of California, Davis)

Savageau gave the final talk for the day, exploring how molecular phenotypes are derived from their underlying gene networks. He presented two fundamental unsolved problems: first, how to go from genotype and environmental information to a model of that organism, and second, how to determine the phenotypic repertoire of a given model. One can sample a wide variety of parameters and get some range of phenotypes, but one would never know if they missed some important combinations

of parameters—the combinatorial explosion is much too huge to check the entire parameter space. His primary goals are to develop a method for constructing the design space and define qualitatively distinct phenotypes, analyze and compare phenotype fitness, and to measure tolerance to changes that would take an organism from one phenotype to another.

Savageau discussed the characteristics and uses of the design space, which dimensionally compresses the parameter space, lumping parameters together into parameter combinations which control the phenotypes. He introduced several examples of this idea, such as the lytic vs. lysogenic decision for phage lambda. He found ten valid phenotype regions for the Lambda phage model, some corresponding to a lysogenic phenotype and others to a lytic phenotype, along with a hysteresis phenotype region which suppresses noise. He also examined the sensitivity and robustness of the system and the global tolerances for the lysogenic phenotype, i.e. the parameter fold changes allowable to remain within the lysogenic phenotype, and determined which parameters are phenotypically sensitive (i.e. required for maintaining a lysogenic phenotype). Using this method allows one to get a bound on the possible number of phenotypes, and explore the types of dynamical behaviors and parameter tolerances for each phenotype.

#### DAY 3: WEDNESDAY, JANUARY 27

Toward Logic Calculations in Mammalian Cells Kobi Benenson, (Harvard University's FAS Center for Systems Biology)

Kobi motivated his talk by presenting a hypothetical "biological computer" that could operate as a molecular "doctor" capable of diagnosing and treating disease at the single cell level. The design challenge lies in reprogramming signaling and transcriptional networks. Dr. Benenson described a systematic framework for building networks to receive prescribed cellular inputs, perform a specified computation, and produce a desired outcome. The building blocks of these networks include transcription factors, mRNAs, and microRNAs. To demonstrate his technique, Dr. Benenson constructed a molecular "logic engine" in mammalian cells using three small RNAs, and also presented preliminary results showing a molecular sensor operating autonomously in Drosophila embryo lysate.

Biology is Technology: The Promise, Peril, and New Business of Engineering Life Rob Carlson (Biodesic)

Dr. Carlson began by describing some of the recent drivers of biological technology development, including economic growth, the desire for material and energy efficiencies, carbon load reduction, curiosity, and a growing need around the globe for access to food, water, and energy. Comparing biotechnology development to airplane design, he discussed the role of scale and regulation in the biotech industry. He further outlined how the global capacity for synthetic biology applications, combined with dropping margins, will very quickly lead to technological revolutions in fields as disparate as pharmaceuticals, agriculture, and energy.

Biology by Design: Reduction and Synthesis of Cellular Networks

Lingchong You (Duke University)

Dr. You described attempts to match network architecture with system properties such as homeostatis or oscillations. His work focused on the cell cycle restriction point, and using high-throughput protein quantification at the single cell level he was able to show that a bistable circuit involving the proteins Myc, Rb, and E2F underlies cell cycle entry. The importance of the simple Myc-Rb-E2F system can not be overstated, as virtually all human cancers have deregulation in this network.

A Synthetic Biology Approach Towards Improving Microbial Synthesis Kristala Jones Prather (MIT)

Dr. Jones Prather's research goals are to improve chemical production by microbes with metabolic engineering, to reconstitute natural pathways in unnatural hosts, and to construct new metabolic pathways. As an example of a successful application of metabolic engineering, Dr. Jones Prather's group constructed a synthetic pathway for the production of glucaric acid, deemed a "top-value added chemical" from biomass, using glucose in Escherichia coli. The glucaric acid production efficiency was improved with the creation of synthetic scaffolds to colocalize essential enzymes in the pathway.



Kristala Jones Prather (MIT)

Modeling Synthetic Biology as Systems Biology Andy Ellington (University of Texas at Austin)

Ellington discussed the relationship between synthetic biology and systems biology. He argued that modularity is likely to be largely unrealistic, although modular approaches in synthetic biology will still be useful. He cautioned that evolutionary machines have their own 'agenda,' and may not always do what we engineer them for. Additionally, he argued that there is no such thing as a truly replaceable part—biological parts will behave differently under different genetic backgrounds. He also showed several examples of how one can take advantage of where synthetic biology and gene circuit engineering goes awry to learn more about the system and improve gene circuit engineering methods, such as making ribozymes and in bacterial edge detection (where bacterial colonies reproduce the edges of an image or shape using pigment proteins). Ellington discussed how building models of the systems one works with can help one adjust for this nonmodularity-rather than building modularity de novo, one should examine the system itself more carefully. Rather than a zero-one Boolean framework, he suggests we approach synthetic biology using an amorphous computing model. For example, in bacterial edge detection, although the same cells may not always produce the pigment protein in two copies of the same image, but the edge will nonetheless be detected in both cases.

Reverse and Forward Engineering of Retroviruses David Schaffer (University of California, Berkeley)

Dr. Schaffer first described two major clinical problems that make it difficult to find a cure to HIV: 1) rapid viral evolution and 2) the presence in the body of long-lived latent reservoirs of infected T cells. The complex HIV system encompasses numerous fundamental biological processes, including signal transduction and epigenetics, and has proven to be a medically important model system for investigating evolution of mammalian gene regulation. Dr. Schaffer's group conducted rigorous experimental and computational analyses to elucidate mechanisms that govern gene expression dynamics in a lentiviral model of HIV-1. It is believed that an understanding of these mechanisms can enable the development of new therapies. Dr. Schaffer further utilized directed evolution of viruses to change target receptor binding specificity and efficiency, and applied protein engineering and library selection tools to engineer retroviral vectors with greatly enhanced safety.

#### DAY 4: THURSDAY, JANUARY 28

Information in Protein Sequences: Combinatorial and Statistical Design Tom Magliery (The Ohio State University)

Tom described his lab's effort to explore amino acid substitutions and their affect on the stability, dynamics, and solubility of proteins as well as their interactions with other proteins. Although these questions may seem straightforward, they remain unanswered to this day, even with technologies such as Rosetta. Two possible reasons for a lack of answers have been the time-consuming nature of current biochemical analysis as well as the vastness of possible protein sequences. To actually tackle these issues, Dr. Magliery uses high-throughput/combinatorial technologies and novel statistical techniques. Most of the techniques are wet lab techniques directed from the theoretical point of view.

Control of Noise and Measures of Modularity in Synthetic Networks

Kyung Hyuk Kim (University of Washington)

Kim focused on the control gene circuit noise levels and the development of what he deems, "stochastic control analysis." In particular, Dr. Kim described the aspects of A) concentration fluctuation control, B) flux concentration control, and C) time-scale separation measure for multi-time-scale dynamical processes. During the first part of his talk, he elaborated on these three points as well as the development of metabolic control systems. In the second part of his talk, he elaborated on the question "when a functioning gene-circuit drives downstream



Lingchong You (Duke University)

components, how many of them can be connected without affecting the functioning circuit?" The concept is similar to that of fan-out in electrical engineering.

Systematic Construction of Nucleic Acid Circuits for Cellfree, Enzyme-free Environments Erik Winfree (Cal Tech)

Winfree first compared biology to a programmable chemistry in spatial and temporal dimensions. If this is the case, Dr. Winfree asks: "what makes up the biological design space for engineering purposes?" In other words, what is a programming language for chemistry and biology? Synthetic biology provides one such example that starts with an organism and changes one circuit. Another example is the bottom-up synthetic biology approach, similar to taking apart a radio and putting it back together again. Nanotechnology provides another such programming language. In this particular talk, he focused on strand-displacement technology under the third category. In particular, how to use biology to mimic electrical engineering logic gates, i.e. transistors.

Formal Approaches to the Design, Analysis, and Control of Synthetic Gene Networks Calin Belta (Boston University)

Calin described formal approaches to the development

and control of synthetic gene networks. In particular, by making the analogy between toggle switches, oscillators, counters, concentration range detectors, and logical gates in electrical systems and the respective gene circuits with similar behavior, he explored how electrical engineering techniques can be used to fine-tune gene circuits such as toggle switches. In his efforts, Dr. Belta has created several software tools to explore gene networks, including RoVerGeNe and LTLcon, available at his website: http://hyness.bu.edu/index.php? page=software.

Gene Circuit Noise Mapping: A Tool for Characterization and Discovery

Michael Simpson (Ridge National Laboratory)

Michael described Gene Circuit Noise Mapping, i.e. how to use noise to characterize a particular gene circuit. This characterization is multidimensional, allowing us to visualize different aspects. In the first dimension of noise magnitude and correlation, we see aspects of autocorrelation; the second dimension provides information on gene regulatory structures and kinetics; the third dimension gives information on very subtle interactions. As an example, he used noise mapping across the human genome in Jurkat cells to reveal an intrinsic chromatin burst rate.

Recoding Viral RNA Genomes through Chemical Synthesis: Novel Genetics and Practical Applications

Eckward Wimmer (Stony Brook University School of Medicine)

Eckward described some practical benefits from the recoding of viral genomes. In particular, he explored how these modifications of viral genomes will allow for the design of novel vaccine candidates, the study of "hidden material" in the human genome, and the elimination of harmful recombination in viruses. As an example, Dr. Wimmer discussed how the modification of the poliovirus genome can also be used for novel strategies in polio vaccine development. In particular, viral recoding allows one to avoid harmful recombination that can lead polio from a live attenuated virus. Arguably most importantly, in the next decade, viral recoding will be somewhat affordable, that it will be become the dominant method of molecular

#### DAY 5: FRIDAY, JANUARY 29

Orthogonality in Biochemical Networks Sven Panke (ETH Zurich)

Panke described the application of topological ideas to enzymatic networks, with applications to large-scale biotech companies and the pharmaceutical industry. In his lab, Dr. Panke uses metagenomic and high-throughput technologies for the development of sugar networks of oligosaccharides. These data-rich technologies give instant feedback on the network dynamics allowing for fine-scale tuning. Moreover, these large-scale efforts give information on higher-level interaction in the networks. Nevertheless, difficulties include the insulation from alternative noise structures as well as the optimization of the dynamics. Overcoming these challenges, however, will provide worthwhile avenues to dynamical orthogonality.

From Artemsinin to Biofuels: The Role of Synthetic Biology in Industrial Biotechnology at Amyris Timothy Gardner (Boston University)

Timothy discussed the role of synthetic biology at Amyris, an integrated renewable products company that specializes in advanced renewable fuels and chemicals. One example of the successful use of synthetic biology is a strain of yeast engineering by Amyris for use in the fermentation of sugarcane. The company is currently working on developing a yeast strain that can produce artemisinin, a compound used to treat multi-drug resistant strains of malaria. Artemisinin production is an excellent application of synthetic biology, as the current supply of the drug is not sufficient to match demand and chemical synthesis is too expensive. Amyris is also using engineered organisms to develop a diesel fuel that can operate at extremely low temperatures and that has a higher cetane number, comparable energy density, and 90% lower greenhouse gas emissions than traditional diesel fuel.

Modeling the Dynamics of a Synthetic Gene Oscillator Matthew Bennett, (Rice University)

Bennett first summarized the various mathematical modeling strategies that one could adopt to simulate genetic networks. Ultimately, any modeling strategy should be validated by experimental analyses. This can be easily done using tractable model systems such as genetic oscillators, naturally-occurring dynamical circuits that can be used as benchmarks for testing mathematical models of both synthetic and native gene networks. Dr. Bennett created a fast, robust and tunable synthetic gene oscillator in E. coli, and in doing so revealed fundamental flaws in the ways in which many dynamical systems and cellular signaling pathways are studied computationally. Discrepancies between the experimentally observed dynamics and the mathematically predicted behavior led to new insights into the importance of fast reactions and dynamical delay in gene regulatory networks.

Genetic Clocks from Engineered Oscillations Jeff Hasty (University of California, San Diego)

Dr. Hasty pointed out that although there are a large number of oscillators existing in nature--which makes them particularly good targets for model deduction--many of these systems, such as circadian networks, are extremely complex. It is simpler to design and build less complex networks and systematically increase complexity in order to mimic and understand natural systems. Dr. Hasty was able to build a synchronized quorum of genetic clocks that, when placed in a large microfluidic device, exhibited traveling waves of fluorescence. In the experimental setup, the initial bursts of fluorescence are cell densitydependent: a region with a high density of cells sets the pace of the oscillations, medium density regions are excitable, and low density regions don't transmit waves at all. Dr. Hasty also showed that oscillators of this kind can be entrained. The results of this synthetic biology work have led to new biological discoveries, including the insight that the exact form of degradation terms in both the mathematical model and real system is more important than was previously believed.

### Workshop 4 Inference in Stochastic Models of Sequence Evolution

ORGANIZERS Ian Holmes, UC Berkeley Gerton Lunter, University of Oxford

Report written by MBI postdocs Eric Bloomquist, Harsh Jain, and Sam Handelman.

#### **OVERALL SUMMARY**

The increasingly high-throughput nature of biological research has created a need for quantitative modeling. This is particularly true for genetic and evolutionary studies that use DNA sequence data as their raw material. This has been recognized for a long time, and particularly since the completion of the human genome sequence, the field of stochastic modeling of biological sequence data has flourished. Nevertheless, the changes that sweep through the biological sciences are so dramatic and fast-paced, that there is a danger that researchers working on the interface of mathematical modeling and biology, who often have a background in more classical and slower-paced fields as mathematics, statistics, or physics may not remain fully up-to-date with the current questions on the biological coal-face.

#### DAY 1: MONDAY, FEBRUARY 22

Parallel Mutations and Partial Sweeps Graham Coop (UC, Davis)

Dr. Coop discussed aspects of natural selection and its signatures in the human genome. With the recent large scale sequencing efforts, such as the Human Genomic Diversity Panel, the International MapHap Project, and the upcoming 1,000 Genomes project, evolutionary biologists can now search for, in a scale not capable before, signatures of natural selection and associated phenotypes. Some examples include genetic markers for skin pigmentation, height, and predisposition to disease. These steps of observation, however, are only the first steps in



Carolin Kosiol (University of Veterinary Medicine, Vienna)

this field; now theoretical models of genome wide evolution can not only be formulated, but tested on actual data. Dr. Coop is most interested in the relationship between migration and geography and what they can tell us about genomic selection. As examples, using the HGDP, he showed that selective sweeps spanning all of Eurasia are quite old, but those of more recent origin do not fully spread across Eurasian. Both of these are indicative of partial sweeps. To explain mechanisms of parallel sweeps and muation, Dr. Coop employed a model of dispersal (similar to crystallization models) that resulted in the prediction that in species with little neutral structure, parallel sweeps and adaptation may be relatively common.

A Coalescent Process Markov in Time and Space Thomas Mailund (University of Aarhus)

With the completed or near-complete genomic sequences of humans, chimps, gorillas, and many other species, investigators have begun to analyze genomes for signatures of selection, and the speciation history. Due to the size of the genomes, however, and the present of persistent recombination, accurate statistical inference is typically computationally infeasible. To handle, this Dr. Mailund proposes to avoid the use of full ancestral recombination graphs, and instead to employ hidden Markov models to take care of recombination. The complexity of the Ancestral recombination graphs scales in terms of sequence length, but with 3 billion base pairs, full estimation is near impossible. He proposes to use a hidden Markov model approach to capture the flexibility of the ancestral recombination graph, but to simplify it enough so that estimation becomes feasible. The details of the model are quite simple to write down, but this is does not make the model less appealing. On the contrary, due to its flexibility, if something can be modeled using the coalescent at two nucleotides, the model of Dr. Mailund is applicable. The model will likely see much use in the upcoming years.

Coalescent HMMs for Inference of Population Genetics Processes in Ancestral Species Mikkel H. Schierup (University of Aarhus)

In the previous talk, Dr. Mailund described the coalescent hidden Markov model for genomic sequences. Absent from his talk, however, was any data or inference from genomic sequences. Dr. Schierup provides this missing facet. In particular, he describes the adaptation of the coalescent hidden Markov model to incomplete lineage sorting problems, i.e. species tree-gene tree reconciliation, how well the coalescent hidden Markov model approximates the true coalescent process, and how to correct these biases on real data. Dr. Schierup then applies these adaptations to a Human-Chimp-Gorilla dataset. He also applies these results to an Orangutan divergence time dataset, estimating a time of 12 million years ago for this split. Note that due relatively distant divergence, incomplete lineage sorting on the Orangutan divergence dataset is small with less than 1% throughout the genome. At the end of his talk, he brings up the important point that all coalescent based methods, including his coalescent hidden Markov model, can tell us little about speciation history in humans from 0.5 million years ago to 6 million years ago since all convergence likely occurs before this. Nevertheless, inclusion of Neanderthal genomes provides a detour around this problem, and will be quite useful in the upcoming decade. Moreover, with the improvements in ancient DNA sequencing, epitomized by the recent Nature publication of a whole human genomic sequence, will make collecting Neanderthal sequences much easier.

#### DAY 2: TUESDAY, FEBRUARY 23

Efficient Calculation of Various Summary Statistics for Endpoint Conditioned Continuous-time Markov Chains Asger Hobolth (University of Aarhus)

The sequence substitution process can be thought of as a missing data problem. In the simplest case, we know the endpoints of a sequence problem; essentially we know where the process started and where the processes ended; we do not know what happened between these two time points. Luckily a continuous time Markov chain allows to model such a process, but to also integrate out these missing data. Moreover, maximization of the CTMC using the EM algorithm is relatively straightforward, if we can calculate the expected number of substitutions given two end-points. To calculate these numbers, three methods can be employed: Eigen value decomposition, uniformization, and matrix exponentiation. Dr. Hobolth discusses the advantages of the second strategy and shows that it provides easy accesses to several summary statistics of interest, such as the covariance.

#### Molecular Fitness Land and Seascapes Michael Lassig (University of Cologne)

We have a good understanding of adaptation at the organismic level, where it is primarily a non-equilibrium dynamical phenomenon. Dr Lassig is interested in the detection of these dynamics at the molecular levels. He argues that a comprehensive theory of molecular evolution must include stochastic forces of mutations and genetic drift, as well as evolutionary histories. It should further distinguish between compensatory evolutions from adaption which is viewed as a non-equilibrium phenomenon: the genomic response to time-dependent selection. His approach thus distinguishes between the static concept of fitness landscapes from the dynamics of fitness landscapes themselves, referred to as seascapes. He demonstrated his methods by providing examples where adaptive evolution could and could not be inferred, despite the presence of positive selection. In the case of the evolution of transcription factor binding sites and the processing of pre-miRNA to mature mi-RNA in plants there was evidence for equilibrium evolution at moderate selection coefficients. This could be quantified by fitness landscapes that depend on appropriate biophysical free energy phenotypes. However, deviations from equilibrium were observed when genomic comparisons in drosophila were carried out providing evidence for fitness seascapes driving adaptive evolution. This lead to a theorem that states that fitness flux increases with genetic drift as a nearly universal evolutionary principle. Dr.





Abdoulaye Diallo (University of Quebec, Montreal)

Lassig concluded by observing that the newly emerging field of non-equilibrium thermodynamics has a wide variety of tools relating to fluctuation theorems that can be exploited for genomic and evolutionary inferences.

Empirical Codon Models for Comparative Re-sequencing Data

Carolin Kosiol (University of Veterinary Medicine, Vienna)

Differences between sequences can be modeled at many different spatial levels. The three most common include models of nucleotide substitution, model of amino acid substitution, and models of codon substitution. This talk focuses on the model for codon substitutions, specifically empirical codon models. Due to the large size of a codon rate matrix, 61x61, accurate models for substitutions typically do not have enough parameters. The alternative approach is to take a nonparameteric model and use the empirical divergences to estimate rates. These models are completely flexible since they allow for double and triple nucleotide codon changes; nevertheless, on real data, these types of changes do not appear to occur frequently. From this, and several examples on genomic datasets, Dr. Kosiol suggests that the re-introduction of parameterized models to empirical models may be advantageous. In particular, he is looking to estimate heterogeneity of selection pressures between different amino acid pairs. After this, Kosiol presented some preliminary

work, as well as some roadblocks such as limited data and recombination.

Parameter Estimation in Models for Sequence Alignment Ana Arribas-Gill (Unversidad Carlos III de Madrid, Spain)

Alignment for homology underlies almost all research in molecular phylogenetics. As raw output, these methods simply produce strings of characters that need to be aligned before any methods can be developed. Early alignment methods, such as clustal typically were built upon heuristics until the early 1990's when statistical models for sequence alignment came into play. Dr. Arribas-Gill discusses some of these methods and how they fit into the hierarchy of statistical methods < hidden Markov model methods < score methods. In particular, she discusses her work on the pair hidden Markov model for sequence alignment. She also discusses some consistency properties of this model and some simulation work. Arribas-Gill's work is quite significant since it presents some of the first efforts to show consistency of sequence alignment.

A Bayesian Method for Genome-wide Identification of Small Insertions and Deletions Kees Albers (Cambridge University)

Dr. Albers covers the challenges confronted in identifying small insertions and deletions between individuals as part of the "thousand genomes project." In Dr. Albers' approach, when a sequence is found in a given genome read with no exact homology in a reference genome, between mutually adjacent regions and a reference sequence can be interpreted as indels. To ensure that sequencing errors are not mis-interpreted as indels, multiple reads must support the presence of the indel via a local realignment. These methods are implemented in the Dindel package, which encapsulates them in a maximum likelihood framework. This framework produces results that are more robust than comparable methods.

Inference of Insertions and Deletions for Ancestral Genome Reconstruction and Phylogenetics Analyses Abdoulaye Diallo (University of Quebec at Montreal) Dr. Diallo presents an approach to reconstruct ancestral sequences using methods that account for insertion and deletion of genetic characters, as well as substitutions between characters. This is an exceptionally difficult and important outstanding problem in phylogenomics, given the limitations of currently available approaches. Dr. Diallo's approach employs Hidden Markov Models (HMM), which rapidly and robustly identify the most likely pattern of insertions and deletions to explain the observed data. These HMM provide both a reconstruction of each ancestral sequence and a confidence for each insertion or deletion at each ancestor; he presents a complimentary approach to aid in visualization of these complex data using a directed acyclic graph (instead of a tree). Diallo's introduces two new heuristics in his HMM method: restricting the available states to only those that are consistent with the phylogenetic tree, and disallowing a transition directly between "conservation" and "extension of insertion" states. With these heuristic adjustments, his method efficiently and robustly reconstructs common ancestors with relatively large numbers of taxa, producing a roughly 100-fold improvement in speed compared to exact, non-heuristic methods, with comparable results. Dr. Diallo's method is available as a web service at http:// ancestors.bioinfo.ugam.ca



#### DAY 3: WEDNESDAY, FEBRUARY 24

Models for Gene Gain, Gene Loss, and Gene Movement Matthew Hahn (Indiana University)

Dr. Hahn proposes novel modifications to existing models of gene gain and gene loss over the course of evolution. He combines overall studies of gene gain and loss rates (which show rapid gene gain/loss in the primate linneage, particularly near the division of humans and chimpanzees) with studies on particular trees that support the result, indicating that the apparent rapidity in the primate linneage is not a short-branch-length effect; however, it may be a predictable result from small

population sizes. When Dr. Hahn relaxes the assumption that gene gain and gene loss rates are equal and uniform across all gene families, he finds that those gene families showing high rates of gain and loss are also those which show high rates of sequence substitution (In Mammals, GOIDs corresponding to immune defense, neuronal development and intercellular transport.) He finds that, in general, these gene gain and loss rates are greater than proportional to the gene copy number of the corresponding family. Dr. Hahn speculates that this may arise from a combinatorial effect, if recombination events are a primary means to initiate gene duplication. However, Hahn also emphasizes that his overall results are reproducible under a variety of sensible model assumptions, and are robust against errors of the type commonly introduced by gene tree methods. Finally, he introduces a qualityof-assembly parameter, which enables his models to be robust to low-quality genomic sequences, as errors in a genome could otherwise appear as gene gain or loss.

#### Many-core Algorithms for Statistical Phylogenetics Marc Suchard (UCLA)

Oftentimes in molecular phylogenetics, investigators are left with a quandary between realistic models and computational speed. On one hand, investigators want models to accurately capture evolutionary processes; on the other hand, investigators want answers in a reasonable amount of time. For most small lab groups, this means at most, answers in a month or two. Dr. Suchard starts off by discussing a problem where investigators want to use a codon model to infer the ancestral history of carnivores. Using standard software, the group predicted this process would take over a year to a complete, far too long for reasonable inference. The group knew they wanted to use codon models, so they asked Dr. Suchard to speed up the process. To do this, he employed fine scale parallelization of the Felsenstein peeling algorithm on graphics processing cards. When examining Felsenstein's peeling algorithm, the process essentially breaks down a matrix times a vector done numerous times, numerical linear algebra. Knowing that graphics process cards can parallelize these steps relatively easy, he integrated Felsenstein's peeling algorithm using CUDA and OpenCL frameworks. This has essentially two punch lines. One, it is fast, almost 200 times faster than using a single processor. Two, it is inexpensive, since these cards typically cost only \$500, substantially less than huge supercomputing clusters.

DNA Transposons: Germline Invaders with a Lasting Impact on Genome Evolution

Cedric Feschotte (The University of Texas, Arlington)

The spread and survival of DNA transposons and their contribution to the emergence of lineage-specific functions is an interesting and challenging question in the field of

genome biology. In his talk, Dr Feschotte demonstrated the use of bioinformatic and phylogenetic tools such as sequence divergence, nested insertions, and comparative genomics in order to decipher the evolutionary history of DNA transposons in mammals. His results pointed to most elements being very old. However, he identified a set of DNA transposon families which he called Space Invaders or SPIN, whose consensus sequences were approximately 96% identical in the genomes of rats, bushbabies, the little brown bat, opossum, and a South American lizard. In contrast, SPIN elements were undetectable in other species. He argued that such a distribution coupled with an overall lack of selective constraint acting on these elements, was incompatible with vertical inheritance, but strongly indicative of horizontal transfer. He further shed light on the mechanism of transfer, which was most likely through the transmission of trypanosomes by a group of blood sucking insects called triatomines. Dr. Feschotte concluded his presentation by observing that these horizontal transfers of DNA transposons have contributed significantly to raw genomic variations in multiple mammalian and tetrapod species.

Inference of Population Structure Daniel Falush (Science Foundation of Ireland)

Dr. Falush first began by talking about the conditional likelihood framework of Li and Stephens. Essentially this method takes a hidden Markov model approach to find the ancestry of a particular DNA locus. The Markov model has two features, one a jump in ancestry and two a probability of mutation for a particular SNP. Using this particular model, Dr. Falush then began to discuss the application of this model to data from the Human Genomic Diversity Panel. As expected, when applying the Li and Stephens model to the HGDP, most individuals most closely mapped to individuals from the same population. He then applied Structurama to the data from Europe. Structurama is a no-admixture model that will classify individuals in distinct populations. Unexpectedly, some populations were split into subpopulations, while other groups from entire countries such as Russia and Spain fellow into a single population. In a total sense, however, the models due suggest strong population structures within the European continent, even though geographic



Anna Kedzierska (Polytechnical University of Catalonia)

distances are within reasonable lengths.

#### DAY 4: THURSDAY, FEBRUARY 25

#### Model Selection for Mixture Models

Anna Magdalena Kedzierska (Polytechnical University of Catalonia)

The central theme of Ms. Kedzierska's talk was to demonstrate how about how math and biology are co-dependent and how each nourishes the other. Her work focused on looking at equivariant evolutionary models such as the Kimura, the Jukes-Cantor and the strand symmetric models. In an innovative application of algebraic geometry, she started out by looking at the group of all permutations on the set of nucleotides. She then proceeded to identify the subgroups that leave the transition matrices of the various models unchanged. For instance, the dihedral group corresponded to the Kimura 2-parameter model. Further, regarding the entries of the transition matrices as model parameters that depend on the model chosen, a polynomial mapping from this space of parameters to the space of joint probabilities was defined. This is an improvement over current homogenous methods as these assume that the rates at each edge of a phylogenetic tree are constant. Here, this assumption was relaxed to allow for unequal rates among lineages. Kedzierska proceeded to show how it was possible to characterize the above linear spaces as corresponding to the space of all mixture models on a tree of a certain size. This had a useful application in model selection. The irredundant set of linear equations describing this linear space could be used for assessing goodness of fit to data. Linearity implied that MLE methods used for this fitting would provide global estimates as opposed to local estimates. Further, she also presented an application of this method to model identifiability by obtaining a bound for the number of mixtures for which model parameters are identifiable. For the future, she plans to develop a webbased tool enabling access to the generating equations of such spaces of mixtures for a given model. Also under development is an interface for performing model fitting on real-life data.

Estimation of Ancestral Population Sizes and Divergence Times from Complete Human Genome Sequences Adam Siepel (Cornell University)

Dr. Siepel explained that when looking at recently evolved organisms, the concept of phylogeny no longer applies. Instead, the branches of the phylogeny need to become tubes, since these organisms now evolve in a population. Because we think of it as a population now, an interesting question is to infer times of speciation and ancestral population sizes. Most naturally, Dr. Siepel suggests that we focus on ancestral sizes and speciation times of humans, especially know that we have several publicly available genomes. To do this, he makes use of MCMCcoal written by Rannala and Yang to infer ancestral population sizes into the past. After analysis, their methods well agree with previous estimates provided by other methods, especially a large bottleneck with the out of Africa hypothesis. Afterwards, he discussed some issues with migration and how this affects the resulting analysis.

### Accelerated and Biased Nucleotide Evolution in the Human Genome

Katherine Pollard (Gladstone Institutes, UC, San Francisco)

Dr. Pollard's work involves the exploitation of the Chimp genome project and the similarity between the Chimp and Human genome to understanding the differences underlie the speciation and subsequent lineage-specific evolution. In other words, she would like to answer the question "what makes us human?" By using tools from comparative genomics, her group was able to identify over 500 genomic elements, called HARs that are highly conserved in vertebrates but show evidence of significantly accelerated substitution rates in humans. These are mostly in non-coding DNA, often near genes associated with transcription and DNA binding. She further demonstrated that combining experimentation with bioinformatics was crucial to understanding the contribution of HARs to human-specific development, biology, and health. In the second part to her talk, Dr. Pollard described how using a likelihood ratio test implemented in a specially developed software tool called phyloP, her lab was able to identify protein-coding sequences with an accelerated rate of base substitutions along the human lineage. By observing that these have a tendency in general to contain clusters of AT to GC or weak-to-strong biased substitutions, she concluded that a recombination-associated process, such as biased gene conversion (BGC), is driving fixation of GC alleles in the human genome. Thus, elements such as the HARs and proteins that appear in positive selection screening can be explained by both positive selection and exclusively neutral forces.



Bayesian Gene-tree Reconstruction and Learning in Phylogenomics Matt Rasmussen (MIT)

With the increase in sequencing, evolutionary biologists now have a wealth of data to infer evolutionary histories from molecular sequences. This wealth of data, however, takes much longer to analyze making the wealth of data both a curse and a blessing. Dr. Rasmussen is interested in eliminating the curse but keeping the blessing, especially in problems involving gene loss and duplication with applications to gene tree-species tree reconciliation. In particular, Rasmussen has implemented a novel reconciliation method that fixes the species tree to help improve prediction of particular gene tree methods. One novel feature of SPIDR, his method, is that it incorporates a novel birth-death process for gene loss. Moreover, this method is relatively easy to compute, making SPIDR computationally efficient. To test his method, he showed some simulation studies that compared SPIDR to make classical methods such as MrBayes, PHYML and a competing group in Sweden. Unexpectedly, SPIDR showed significant gains over competing methods with no increase in computational time. In fact, in several instances, SPIDR showed equal computational complexity as PHYML, mak-

ing the method useful for phylogenomic pipeline applications. Since SPIDR attempts to use maximum posterior graphs, searching becomes easier. As such, Rasmussen found a way around computation of the Felsenstein likelihood, making SPIDR attractive for many genomic applications.



Kees Albers (Cambridge University)

#### DAY 5: FRIDAY, FEBRUARY 26

Independence and Dependence of the Nucleotide Positions in the Evolution of Ribosomal RNA Structure Robin Gutell (The University of Texas at Austin)

Dr. Gutell describes a family of analyses on multiple sequence alignments (MSA) of structured RNA molecules from different organisms. His methods use covariation (non-independence between positions/columns) to identify common structural features, going beyond established results that depend primarily on canonical basepairing. Gutell emphasizes the importance of biological diversity, and of sensitivity to lineage-specific effects, in studies of RNA structure especially in the ribosome. When applied in this way, these methods reconstruct adjacent, anti-parallel watson-crick base pairing purely from the sequence information. Based on this result, he suggests that it may be possible to identify other, fundamental structural features using the same methods. Following a broad review of the history of structure determination in the ribosome (for which a nobel prize was recently awarded), he suggests that, while the majority of the tertiary contacts found in the ribosomal structure are not reflected in covariation of the sequences, the potential of the covariation approach is better appreciated as a means towards identifying the underlying rules that govern the formation of three-dimensional structure in RNA molecules, an area of opportunity for sophisticated mathematical approaches. Finally, Dr. Gutell presents a phylogenetic-history-informed method to identify covarying positions in nucleotide sequences.

Models for Amino Acid Substitution and Gene Duplication With Roots in Molecular and Evolutionary Processes David A. Liberles (University of Wyoming)

Dr. Liberles introduced novel methods to jointly model evolution at the organismal and protein-biochemical levels. In his models, apparent conflicts between the effect of thermodynamics on amino acid substitutions and the effect of population sizes on evolutionary rates are jointly modeled and reconciled. The joint model incorporates Grantham Matrices (for thermodynamic differences between amino acids) and Kimura et al. equations to determine the likelihood that an amino acid substitution will be fixed as a function of evolutionary distances and population sizes. The joint model is then optimized to best-predict sequence substitutions in 7 pairs of relatively-closelyrelated organisms.

The optimized model produces a number of unexpected results: selective pressure appears mainly to be exerted through relatively conservative amino acid substitutions and the data is not well supported by thermodynamic data alone (contrary to assertions in the literature). In a rousing discussion, Liberles speculated that organisms with small populations may be more tolerant to more extreme amino acid substitutions because they are under more intense selection, or that population bottlenecks may produce more extreme substitutions arising from linkage/hitch hiking. In the second component of his talk, Liberles presented models for the retention of duplicated genes. First, he reviews the biological/evolutionary mechanisms which may confer an advantage to retained gene duplicates: gain of function in one duplicate, partial loss of function in both duplicates, increased "dosage" of the gene, "robustness" meaning tolerance against subsequent loss of the gene (which is controversial), or some combination or derivative of these four effects. Using an additional family of models which account for differences between whole genome and local duplications, as well as possible effects of protein function or fold, he shows that observed retention of duplicated genes is consistent with those models built on the assumption that duplicated genes are retained when one of the two duplicates adopts a new function.



Extracting Useful Information from Protein Sequences Stephen Chamberlin (Foundation for Applied Molecular Evolution, Florida)

Dr. Chamberlin's talk focused on the divide which exists between computational and experimental biology and his ideas on how to bridge this, namely by understanding molecular biology information within a historical context using available computational tools, and communicating useful and scientifically relevant facts back to biologists and chemists who would ideally like to make better use of genomic data, which in its current form is highly disorganized. By considering the evolutionary histories of aromatase in pigs and alcohol dehydrogenase 1 in humans, he showed the importance of using varied tools such as ancestral sequences, molecular dating, Ka/Ks ratios, crystallography, fossils, geological data and homoplasy in order to build a well supported hypothesis. He further demonstrated how it was possible to build the complexities of first order models in order to account for differences between expected versus predicted behaviors. Thus, by combining information from different disciplines such as planetary biology, paleobiology, computational or molecular biology, biochemistry and synthetic chemistry, a complex biological question could be solved to the satisfaction of a broad scientific audience by providing answers in their own language and standards of proof. Chamberlin concluded his presentation by making the point that we can learn a lot more about systems biology by considering some of the relevant histories and studying how adaptation has forced change in the primary and related systems.

### Workshop 5 Mathematical and Experimental Approaches to Dynamics of Protein-DNA

#### ORGANIZERS

Jané Kondev, Brandeis University Hao Li, University of California, San Francisco

Report written by MBI postdocs Julia Chifman, Suzanne Robertson, and Rebecca Tien.

#### DAY 1: MONDAY, MARCH 8

Everything Interesting in Biology is Done by Molecular Motors Zev Bryant (Stanford University)

Bryant and coworkers have studied the structural dynamics of DNA protein complexes. In particular they are interested in DNA gyrase; a molecular motor that winds up DNA to generate the supercoil structure. DNA gyrase is useful to bacteria in which this protein resides in the following ways:

- It is more compact than relaxed DNA.
- It removes positive supercoils.
- The negative supercoiling structure that gyrase helps create destabilizes the helix, thus facilitating local denaturation for processes such as transcription.

Gyrase uses a chiral DNA wrap to control the directionality of topoisomerization so that it is always introducing negative supercoils. Every time gyrase attaches it causes that section of DNA to wrap around the enzyme twice, leading to the negative supercoil structure. This was determined by studying the motor activity of gyrase using a rotor bead-tracking assay.

Through this experiment, Bryant et al. were able to determine how gyrase functions in coiling the DNA strand. On the basis of this data a mechanochemical model for processive gyrase activity model was created. This model has gyrase attach to DNA and then very rapidly the DNA wraps around this enzyme twice, replicating the patterns observed through experimentation.



Wilma Olson (The State University of New Jersey)

The enzyme pauses at every 2nd rotation. This pause acts as a sort of reset before the next step. Next question is to look at the role of contraction (after the 2nd rotation) simultaneous with the process of rotation. At low ATP, one can see a pause after one rotation. However, the major effect of lowering ATP is not to generate this pause, but to lengthen the lifetime of the achiral state (i.e. the pause at the 2-rotation mark). That implies that ATP can bind not just at the chiral wrap state (beginning of rotation 1), but also at the achiral contracted state. From these results a revised model was created where exit from the chiral wrapped state is accelerated by ATP binding. Velocity, processivity and appearance of an intermediate pause state should all depend on ATP and they do.

DNA Looping Probabilities and Semi-Classical Path Integrals John H. Maddocks (Swiss Federal Institute of Technology)

There were two goals to this research: 1) to assume a known sequence-dependent parameterized coarse grain model of DNA and show how to compute looping probabilities; 2) Use a multi-scale approach to extract coarse-grain rigid base-pair, or rigid base, model approaches. To meet these goals, the researchers started with rigid interbase pair deformations (shift, slide, rise, tilt, roll, twist). The configuration variable for an N junction oligomer is then 6N dimensional.

A special feature of DNA is that it is a 1D object. His goal was to determine the pdf of the final configuration at a given place and orientation. The sort of biological ques-

tion that gets answered using this method: if you bind one end of a loop, what is the probability that you form the other end of the group? Given an initial configuration, what is the probability of getting a final configuration of a specified form? The probability depends very heavily on the actual sequence of DNA. Have to integrate over all the paths satisfying both initial and final conditions. How to integrate over all these paths depends heavily on the level of coarse grading that is used. Using path integral methods, he obtained a sequence-dependent formula for the probability of loop formation, which also included the case of minicircle cyclization.

#### Single-molecule studies of DNA replication

Antoine M. van Oijen (Centre for Synthetic Biology, Groningen)

The focus of the research is on DNA replication and single-molecule experiments on the replisome. A replisome is a complex molecular machine that carries out replication of DNA. It is made up of a number of subcomponents (such as helicase, gyrase, polymerase...) that have a specific function in the replication process of DNA. The experimental approach was as follows: single molecule fluorescence was used to obtain observations of polymerases at the DNA replication fork in a prokaryotic cell. DNA molecules are stretched so that their elastic properties can be used to obtain information on the proteins that are involved in unwinding the double helix in order to copy its genetic information. He worked first with two different polymerases each with different rates of DNA synthesis. Mixing both polymerases together resulted in abrupt changes in the synthesis rate. The experiment started with only the slow unlabeled polymerase and then flowed in the second, faster polymerase. A transition of synthesis rates was observed. Specifically, the fast one swaps out with the slower one (i.e. after some time, faster one takes over the process of synthesis). Another benefit of stretching out the individual DNA molecules to measure the formation and release of DNA loops. The next phase is to look at the same mechanisms of replication in eukaryotic replication. Unfortunately, the mechanisms become much more complex. Researchers are getting closer to understanding mechanisms, however they are nowhere near being able to reconstitute the process in vitro. Van Oijen is currently doing some work with frog eggs to try to mimic the different phases of replication. The measurements obtained in the experiment of both prokaryotic and eukaryotic cells allow the researcher to more deeply understand the role that each of the subcomponents of the replisome play in the unwinding and synthesis of DNA.

DNA Mechanics and Gene Expression Wilma K. Olson (Rutgers, SUNJ, Piscataway) The main goal of this research is to study the looping properties of DNA as a way of understanding gene expression. These loops are created by the adherence of non-specific binding proteins which help to stabilize the complex structure of DNA. The researchers look at the specific roles that a number of these proteins take on the configurational properties of fragments of DNA. On one experiment, the researchers looked at the binding of the Lac repressor assembly to the 01 and 03 operators. This assembly was thought to impede the binding of RNA polymerase at the intervening promoter site, thereby inhibiting the production of gene products. In another experiment on the non-specificity of HU binding to DNA the researchers measured the level of gene expression of Lac as a function of how far apart the operators are. Could take e. coli cells and disable capability of HU to be formed (architectural protein). HU is tied in to Lac expression.

In general, DNA is modeled at the level of base pair steps, using elastic potentials that take an amount of the variability in both the intrinsic structure and the elastic moduli of individual base-pair steps (Hook's Law is used to model this). Theese base pair models takes into account sequence effects such as the deformability of CA-TG vs. AC-GT steps, as well as the coupling of conformational variables such as Roll and Slide. (Roll and Slide are 2 of the 6 possible conformational variables which also include: tilt, shift, rise, twist.) This is how you incorporate sequence independent formation in the data to understand the structure of loops in the DNA molecule. The base-pair level model reduces to the conventional treatment of DNA as an ideal naturally straight, inextensible elastic rod.

Olson estimated the likelihood of DNA cyclization from the number of simulated configurations of a linear molecule with the terminal ends positioned in such a way as to ensure chain ligation. To model this, one starts with a straight, inextensible, isotropic bending, independent twisting elastic rod with only a few base pairs. One can then look at whether the DNA ends come together and form a closed loop or form an oligomer. One well known early result of this approach was that one could abstract a twisting constant from this type of data.

A computational approach can be used to understand the exact mechanism by which HU affects the conformation of DNA. Using this computational approach, the researchers looked back to the role of specific proteins and how they relate to the shape of DNA, the histonelike heat-unstable HU protein introduces localized sharp bending, twisting, and shearing in its DNA binding partner. The HU-bound DNA fragments are able to adopt a variety of bent forms. The binding of HU transforms the smooth, defined stiff polymer into what appears as DNA. Want to know why loops are not formed when HU is not around. Looked at what the repression levels were in the

presence or absence of looping and how this relates to the expression of Lac. The bimodal pattern of looping mediated by these two proteins reflects the tendency of DNA to adopt different types of loops depending on how many base pairs existed in the particular DNA molecule looked at.

In summary, the goal was to show how using a simplified computational approach to understand DNA structure provides new insights into the binding of proteins to DNA that are not possible through the base-pair model approach that has previously been used.

#### DAY 2: TUESDAY, MARCH 9

DNA Architecture and Transcriptional Regulation: The Physics of Genome Management Hernan Garcia (California Institute of Technology)

This presentation focused on developing models of transcriptional regulation and testing them experimentally. Garsia's main point was the analogy between genetic circuits and electronic circuits: Can we actually predict inputoutput function of a genetic circuit? Both the sequence and the physical architecture of DNA determine patterns of gene expression; thus, can we predict patterns from sequences? Statistical mechanics gives input-output function of any regulatory architecture and different inputs allow modification of the input-output function. He presented his team experimental efforts aimed at dissecting repression by DNA looping and the sequence-dependent flexibility associated with the mechanical code of the DNA. Overall, Garcia's presentation emphasized cellular decision-making, explored regulatory architecture and DNA mechanics in living cells.

Modeling and predicting DNA-binding specificity Gary Stormo (Washington University Medical School)

Stormo's presentation focused on two parts: modeling the specificity of DNA-binding proteins and predictive recognition models. Given specificities for many proteins of a specific class one can also predict the binding specificities of novel proteins, allowing for the design of new proteins with unique specificities. Gary Stormo is interested in intrinsic specificity of DNA-binding proteins that is how does protein distinguish different DNA sequences. His focus is on data based approach and ideally he would like to have a Kd data for many, preferably all, possible binding site sequences.

Many different kinds of data are available for modeling the specificity of a DNA-binding protein, and the quality of the model depends on both the type of data used and the algorithms for estimating binding energies. Gary discussed his team approaches for modeling from several different types of data. They analyze how the affinity of certain proteins for DNA changes depending on the sequence of the DNA, and conversely for changes in the key residues of the protein. They have developed methods that allow them to obtain accurate, quantitative data in large amounts, which can be used to develop models for the interactions.

Micromechanical study of protein-DNA interactions and chromosome structure John Marko (Northwestern University)

Marko's team is focused on trying to understand how

protein-DNA interactions work and how huge DNA molecules are folded up into active protein-DNA complexes called chromosomes. Experiments on bacterial chromosome folding proteins where presented. One of the lessons that his team did draw from these experiments is that the kinetics of protein-DNA interactions are very different form the way we like to teach our students about ligands binding to receptors. Future directions for the team include combining fluorescence microscopy and force microscopy in experiments on DNA-protein complexes and whole chromosomes, and in-vivo studies of coupling of chromosome dynamics to gene expression.

First-principles calculation of DNA looping in tethered particle experiments Philip Nelson (University of Pennsylvania)

Nelson began his presentation by displaying stimulating questions: Of what use is a theorist, anyway? - "Well, somebody's got to do the theory." Ok, smart guy -- of what use is theory? - "Well...Theory is often needed to



Eran Segal (Weizmann Institute of Science) extract information from data."

The presentation concentrated on three parts:

- Sometimes theory suggests a new kind of measurement that tests a model more stringently, or distinguishes two different models more completely than previous measurements.
- 2. Sometimes our model is not obviously connected with what we can actually measure experimentally, but theory makes a connection.
- 3. Sometimes the model that interests us involves the behavior of actors that we can only see indirectly in our data; theory may be needed to separate them from each other, and from noise.

The main point of Philip's presentation was that theory is needed every day and that it is our microscope and it helps us to see the invisible.

#### DAY 3: WEDNESDAY, MARCH 10

"Pfam-wide" determination and inference of transcription factor DNA sequence specificities Tim Hughes (University of Toronto)

Understanding gene regulation and genome function involves the understanding of DNA sequences, protein-DNA/RNA interactions, protein-protein interactions, and epigenetics. There is a lot of missing information in each of these areas. Hughes's talk focused on improving our understanding of the protein-nucleic acid "interactome" by looking at transcription factors (TFs) and DNA-binding domains. This insight may ultimately help answer questions such as how eukaryotic cells know where to find promoters. Dr. Hughes's lab uses DNA-binding data from Protein Binding Microarrays to measure sequence preferences of TFs and develop methods for inference of TF sequence specificity. He has found that overall sequence identity in the DNA-binding domain is a very good predictor of sequence preferences, so similar sequences result in similar binding activity (at least 65% sequence similarity has been shown to be a good indicator of similar binding activity). These results are being used by Dr. Hughes to create a "pfam-wide" analysis of DNA-binding domain sequence preferences for a wide variety of eukaryotes. He has constructed a database, called the Catalog of Inferred Sequence Binding Preferences, in order to store known and inferred sequence preferences.

Transcriptional Lego: Predictable Control of Gene Expression by Manipulating Promoter Building Blocks Eran Segal (Weizmann Institute of Science)

Dr. Segal's talk works towards the goal of one day being able to "read" the genome by looking at DNA sequences and being able to understand how regulatory sequences encode for transcriptional behavior. The organization of nucleosomes and the factors controlling where they bind in vivo are critical pieces of this puzzle.

Dr. Segal first addressed the question of how transcription factors (TFs) find their targets in genomes, as the same sequence may appear multiple times in the genome. Intrinsic nucleosome organization may assist in directing TFs to their appropriate sites in the genome, helping them decide if a sequence is a functional binding site or a happenstance occurrence. There is less nucleosome occupancy in functional sites, which leaves them more accessible to TFs. Dr. Segal shows that Poly(dA:dT) sequences (sequences of multiple A's or T's) which are abundant in eukaryotes, disfavor nucleosome formation. They act as boundaries for nucleosome formation creating a depletion of nucleosomes in their proximity. He presented experimental evidence showing that the closer a Poly(dA:dT) sequence is to a binding site, the greater the degree of nucleosome depletion.

Dr. Segal discussed whether nucleosome organization determines gene expression kinetics, or whether the presence of a boundary matters for expression. He has found that nucleosome disfavoring sequences causally affect transcriptional output. The longer and more perfect the Poly(dA:dT) sequence, the stronger the transcriptional response. He also found a decrease in transcription with distance from the boundary. Nucleotide disfavoring sequences can be used for predictable tuning of transcription levels at very fine resolution, by playing with perfection of sequences and distance from boundary. This may be a genetic mechanism by which genomes can fine tune expression during evolution.

Dr. Segal also presented a human example, where functional p53 binding sites have been shown to have high intrinsic nucleosome occupancy. He proposed p53 may be directed to the functional sites by factors involving higher order chromatin structure. High intrinsic nucleosome oc-

cupancy regions may resist chromatin compaction, rendering them more accessible.

Mechanisms of ATP Dependent Chromatin Remodeling Enzymes Geeta Narliker (University of California, San Francisco)

Chromatin structure affects the regulation of gene expression by providing a way to compartmentalize the genome into active and inactive states. Compacted chromatin restricts access to DNA while open areas are accessible. Chromatin structure is not fixed, but a dynamic process of conformational changes. Chromatin remodeling motors help set up either open (SWI/SNF motor class) or closed (ACF motor class) chromatin states, but the mechanisms behind this process are not yet well understood. By understanding how ACF motors operate, one can hypothesize why there are different outcomes between the ACF and SWI/SNF motor classes.

A key feature of the closed state is the regular spacing of nucleosomes along the genome. In this talk, Narliker addressed the question of how ACF motors may work to rearrange nucleosome structure by equalizing the length of the DNA on either side of nucleosomes to achieve this defined spacing. She suggested that ACF functions as a dimeric motor in which two ATPases take turns in moving the nucleosome. Furthermore, the two ATPases face each other, allowing the nucleosome to move back and forth rapidly in order to attain regular spacing.

While Dr. Narliker provided evidence that ACF operates as a dimeric motor, all evidence suggests that SWI/SNF motor classes function with a monomeric ATPase. This difference may explain the different biological roles of the two motor classes.

A Simple Biophysical Model of Nucleosome Occupancy and Energetics

Alexandre V. Morozov (Rutgers University, BioMaPS Institute for Quantitative Biology)

The presence of nucleosomes, 147 basepair-long segments of DNA wrapped around a histone octamer, influence the binding of transcription factors (TFs) to DNA through the physical occlusion of binding sites. In this talk, Morozov



Geeta Narlikar (University of Californa San Francisco)

investigated what factors determine the position of nucleosomes on DNA by studying the sequence specificity of histone-DNA interactions. He inferred nucleosome energies and occupancies from high throughput data and nucleosome positioning experiments (occupancy profiles) taking steric exclusion between nearby nucleosomes into account, and he fitted nucleosome energies to DNA sequence features. Morozov found that the best model for predicting in vitro nucleosome occupancies for yeast and E. coli genomic DNA is an N=2 position independent model, and the location of nucleosomes is largely determined by the dinucleotide content of the underlying DNA sequence rather than periodic patterns or longer sequence motifs.

#### DAY 4: THURSDAY, MARCH 11

Nucleosome positioning and chromosome structure from archaebacteria to man Jonathan Widom (Northwestern University)

Genomes encode an additional layer of genetic information that is superimposed on top of other codes of regulatory information that have been previously understood. This information controls the 3D organization of the chromosomal DNA molecules in space. Specifically, genomes care where nucleosomes are located on the DNA. Information to bias where the nucleosomes are located is actually written into the genomic sequence. In the last few years, researchers have become increasingly good at predicting where nucleosomes will be and also what the consequences are for their particular location within the

#### DNA.

Why do genomes care? For one, nucleosome wrapping competes with protein biniding. Second, cells carefully control the amount of nucleosomes well beyond their saturation point. As a consequence, different portions of the DNA are competing with themselves for this limited amount of nucleosomes. Genomic sequences control the outcome of that self-competition. Consequence of this is that two sites of the DNA with the same piece of code might be different in their affinity for nucleosome binding. The site that has a lower affinity is then free for a protein to bind at that site. Here, researchers are interested in deciphering the nucleosome positioning code to determine what sections of DNA the nucleosomes "like". They are interested in detecting non-random patterns that lead to binding. For instance, nucleosomes are more likely to bind at the DNA helical turn where there is a T followed by an A in the genetic code. When the DNA are out-of-phase, a different pattern of sequence-preference is found. This type of information can be used to predict location of nucleosomes. The model that the researchers were able to construct from this information turns out to be strongly predictive of nucleosome occupancy in vivo.

The second part of the talk looked at improving the current maps that are available. This is important for several reasons:

- 1. We want to better understand competition with transcription factors (DNA binding proteins).
- 2. The structure of the chromatin fiber depends on the location of the nucleosomes.
- 3. The quality of the model depends on an accurate understanding of nucleosome location.

A chemical approach to nucleosome mapping can be done. Through this process roughly 30,000 nucleosomes have currently been mapped with a very fine scale approaching zero base pairs. Interestingly, when comparing the position of nucleosomes across a diverse range of organisms, it is found that the positioning code of nucleosomes may be conserved from archaebacteria to man.

#### Structuring a prokaryotic chromosome

Paul Wiggins (Whitehead Institute for Biomedical Research)

According to Wiggins, the human genome is really no more than a parts list. His work currently focuses on both explaining the physical structure of chromosomes in the cell and how the structure of chromatin (the complex of DNA and protein that makes up chromosomes) affects gene expression. The stochasticity of chromosome organization was investigated by fluorescently labeling genetic loci in live E.coli cells. In spite of the common as-

sumption that the chromosome is well-modeled by an unstructured polymer, measurements of the locus distributions reveal that the E.coli chromosome is precisely organized into a nucleoid filament with a linear order. Loci in the body of the nucleoid show a precision of positioning within the cell of better than 10% of the cell length. The precision of inter-locus distance of genomically proximate loci was better than 4% of the cell length. The measured dependence of the precision of inter-locus distance on genomic distance singles out intra-nucleoid interactions as the mechanism responsible for chromosome organization. From the magnitude of the variance, we infer the existence of an as-yet uncharacterized higher-order DNA organization in bacteria. We demonstrate that both the stochastic and average structure of the nucleoid is captured by a fluctuating elastic filament model.

Birds eye view at protein-DNA binding and DNA packing Leonid Mirny (MIT)

Mirny studies DNA-protein interactions from the information-theoretical perspective. He also looks at nucleosome and loop mediated cooperativity and DNA folding. In this talk Mirny addressed the problem of how much information is required for a protein to find and recognize a DNA binding site, and whether we can determine from known binding sites how specific recognition of the binding sites should be.

For bacteria, 20-23 bits are sufficient and 20-25 bit of information are provided. Thus bacteria are very specific and we do not expect any spurious binding sites on DNA. For multicellular eukaryotes, however, 27-33 bits of information are required, but only 12 bits are available. This is a huge discrepancy, which would lead to about 30,000 "false hits" assuming 5% of the DNA is accessible. This suggests that while binding is necessary and sufficient for gene expression in bacteria, binding alone is not sufficient for gene expression in multicellular eukaryotes. This begs the question of how cells handle spurious sites. That is, if all sites are equally accessible, how do cells know which sites they should bind to?

Dr. Mirny next addressed models of cooperative binding of transcription factors. He argues that scaffold, loop, and target mediated cooperativity models are all mathematically identical, fitting under a universal framework. The third part of his talk looked at how DNA is packed into a cell, using a Hi-C technique of chromosomal conformational capture and new generation sequencing. Mirny shows that two commonly used models of chromatin folding don't fit the data, and suggests a fractal ("crumpled") globule model in which the chromatin is unknotted and densely packed. The fractal globule is hierarchically organized and separated into chromatin sectors, or territories. The unknotted chromatin is easy to open and close, and to fold and collapse.

### Workshop 6 Transport in a Cell April 12-15, 2010

#### ORGANIZERS

Michael Diehl, Rice University Anatoly Kolomeisky, Rice University

Report written by MBI postdocs Shu Dai, Chuan Xue, and Kun Zhao.

#### DAY 1: MONDAY, APRIL 12

How Kinesin Walks on a Microtubule: A View of the Story So Far Michael Fisher (University of Maryland)

The talk gave a summary of recent findings on how kinesin moves in biological systems. Kinesin is one family of motor proteins, which are enzymes that convert chemical energy into mechanical work. Kinesin walks along microtubule, and they typically step from the minus-end towards the plus-end. Each step forward consumes one ATP. The speed of the motion depends on the load on the motor, with a stalling load corresponding to zero speed. It is argued that the chemical reaction rates depend on the load. Non-published results show 5-10% of the time they take a back step. The exact mechanism of the backward walking is still under debate. The talk concluded with the interesting observation that physicists think that all biological motor proteins have a unified mechanism, whereas biologists think all biological motors are different from each other.

Free Energy Landscape of Knotted/Slipknotted Proteins Joanna Sulkowska (University of California, San Diego)

Joanna Sulkowska started with a brief introduction on the mathematical representation of different knots that are used to describe protein knots. The first knotted protein was discovered in 1994 and untill now it has been shown that about 1% of proteins in the protein database are knotted. Knotted proteins are mainly enzymes and themophylic proteins, which may have a relation with the



Michael Fisher (University of Maryland)

fact that they have a slower degradation rate. She finished with stochastic model that utilized computer simulations to show possible ways of how different knotted proteins are folded.

Spindle Movement and Checkpoint Control During Mitosis in Yeast

John Cooper (Washington University School of Medicine)

Dynein is a major motor protein that involved in the control of the spindle position. Dyneins move microtubule minus ends, and they pull the spindle into the neck between mother and bud in yeast cells. Dynein's function requires another protein, dynactin, and other related molecules to function succesfully. Loss of dynein function and failure to properly position the spindle activates a cell-cycle checkpoint and delays the progression of the cell cycle.

FIONA Looks at Individual Molecular Motors Walk and Run Paul Selvin (University of Illinois)

In this talk, Paul Selvin introduced a technique of Fluorescence Imaging with One Nanometer Accuracy (FIONA), which allows one to observe the movement of a single protein molecule with high spatial resolution. Using this technique, his group found that all protein motors tested so far walk in the so-called hand-over-hand fashion in vitro. This technique also leads to interesting in vivo observations such as passing of cargo from motor to motor or shared cargo by two motors.

The Conserved L5 Loop Establishes Eg5's Pre-powerstroke Conformation Sarah Rice (Northwestern University)

Different kinesin motors can have different regulatory mechanisms. Eg5 is a member of the Kinesin-5 subclass of kinesins. L5 is a structurally conserved loop near the ATP binding site of this motor protein molecule. L5 is a known to be a drug inhibitor binding site, but the role of it was unclear. By site-specifically attaching a magnetic marker to ADP, L5, and other groups in the molecule Sarah Rice and her colleagues found that L5 shows a conformational change that enables Eg5 to bind to microtubules in a prepowerstroke state.

#### DAY 2: TUESDAY, APRIL 13

Dynamic Instability of Organelle Transport Vladimir Gelfand (Northwestern University)

The first part of the talk is about motor-induced organelle transport along microtubules and actins. It was found that kinesins and dyneins require each other in vivo, and they are coupled probably mechanically. The second part of the talk described the active movement of microtubules during the organelle transport. Microtubule movement could originate from the presence of motors on the surface of vesicles and the motors that are bound simultaneously to several microtubules in a bundle.

Toward a Unified Walking Model for Cytoplasmic Dynein Arne Gennerich (Albert Einstein College of Medicine)

Cytoplasmic dynein is a two-headed motor protein that moves towards the minus-end of microtubules in eukaryotic cells. Optical trapping experiments show that cytoplasmic dynein produces an average force of 7 pN, and mainly takes 8nm steps but can take as large steps as 24 nm as well. His group also found that dynein exhibits an asymmetric response to ATP under forward and backward forces, and ATP binding to the first ATP binding site of on dynein head causes rapid head dissociation under the forward load, while ADP binding to the third site decreases dynein's MT-binding affinity under backward load.

Mechanics of Actomyosin Interaction and the Role of Substrate Stiffness on Actin Network Dynamics Sean Sun (Johns Hopkins University)

In this talk, Sean Sun introduced a model of how adhesion complexes function in cells. It predicts that these cell-surface interactions provide a viscous drag that increases with the elastic modulus of the surface. He also derived a force-velocity relation of myosin II, which implies that myosin generates greater force when the adhesion complexes slide slowly. Then, using a simple cytoskeleton model, it was shown that an external force applied to the cytoskeleton causes actin filaments to aggregate and orient parallel to the direction of the force application.

Functional Mechanical Deformations in Nature's Machines Charles L. Brooks III (University of Michigan)

In this talk Dr. Brooks presented findings from the studies, conducted by his research group, on the functionally important mechanical processes in the ribosome and an AAA+ helicase. The research was conducted by examining a number of coarse-graining methods to decompose the functional components of motion in these systems using a combination of elastic deformation theory, Go-type structure centric folding models and Brownian motion. It is found that functional motions associated with tRNA translocation in the ribosome are largely attributed to the lowest energy for deformation eigenvector directions from elastic theory. It was also discovered that the de-



Alan Hunt (University of Michigan)

composition of the elements of translocation of the AAA+ helicase along a single stranded DNA, substrate coupling and asymmetry in substrate binding lead to a simple picture of the essential features needed for functioning of this molecular motor.

Road Signs for Kinesin Transport Kristen Verhey (University of Michigan Medical School)

In this talk Dr. Verhey reported recent progress on understanding of transport in the cytosol of cells. The research conducted was based on the microtubule-based trafficking in mammalian cells. The mean squared displacement of probe particles as a function of time was determined to range from 1 millisecond to 10 seconds in solutions by tracking single molecule of Kinesin-1 in live COS cell. It was

discovered that Kinesin-1 walks preferentially on stable microtubules and is selective for stable microtubules, whereas Kinesin-2 and Kinesin-3 are not selective and can move on more dynamic microtubules too. Therefore, at high frequencies tested solutions exhibit significant elasticity, whereas at low frequencies, they are purely viscous. This viscoelasticity property of protein is attributed to a dense network of weakly-bound chains of protein molecules with characteristic lifetime of 10-100 ms.



Sarah Rice (Northwestern University) Collective Behaviors of Multiple Interacting Kinesin Motors: Can Kinesin Number Regulate Intracellular Transport? Michael Diehl (Rice University)

It is generally believed that multiple motors can transport cargos for much larger distances than single motors, can produce much larger forces, and can advance cargos in fractional displacement sizes. These properties are often described as signatures of cooperative transport behaviors. Whereas, in this talk, Dr. Diehl argued that this conventional wisdom is not rigorously true in some systems. Based on the so-called "macromolecular systems bioengineering" approach and using a conventional kinesindynein model test system, Dr. Diehl and his collaborators investigated some potential influences of collective effects on motor regulation. Their experiment discovered that, in reality, multi-motors system often exhibit cooperative transport behavior. This indeed contradicts with the conventional wisdom that predicts non-cooperative transport behavior in a multi-motors system that only happens in an ideal world.

#### DAY 3: WEDNESDAY, APRIL 14

Motoring Along a Nucleic Acid Strand: Template-dictated Polymerization of Macromolecules of Life Debashish Chowdhury (Indian Institute of Technology)

In this talk Dr. Chowdhury reported recent progress made by his research group on the understanding of the mechanics and dynamics of polymerases and ribosomes. A new stochastic differential equation model has been developed, and it was used to analyze data from X-ray crystallography, cryo-electron microscopy, single-molecule imaging and manipulation s. Based on the analysis of the mechanics and biochemical properties of ribosome, the model was derived by coupling deterministic ordinary differential equations with the Langevin equations that take into account the stochastic noise effect. The model incorporates mechano-chemistry of individual machines and steric interactions and it also accounts for both singlemolecule properties and collective behavior. By calculating the average flux and average density profiles together with the fluctuations, Dr. Chowdhury and his collaborators demonstrated that the model fits into experimental data with a high accuracy.

DNA Chasing DNA: Physics of Homology Recognition and the Secret of Perfect Match Alexei Kornyshev (Imperial College London)

In this talk, Dr. Kornyshev reported evidence that intact, double-stranded DNA have the ability to recognize sequence between one another, from a distance, without the aid of proteins. This recognition indicates the tantalizing possibility that this is an inherent property of DNA that may assist in the pre-alignment of two parental alleles of the same gene, increasing the accuracy and efficiency of homologous recombination between genes — the process responsible for DNA repair, evolution, and genetic diversity. Possible explanation of the phenomenon is given on the basis of the theory of electrostatics and interaction between rigid helical materials. These new findings may thus shed light on ways the nature uses to avoid recombination errors, which underpin a set of genetically determined diseases, as well as contributing to aging.

Modeling Adenosine Triphosphate (ATP) Hydrolysis in Myosin

Alexander Nemukhin (Moscow State University)

In this talk, Dr. Nemukhin reported recent progress made on understanding of the mechanistic properties of adenosine triphosphate (ATP) hydrolysis and conformational transitions in myosin, a prototypical molecular motor.



A combined quantum mechanical - molecular mechanical (QM/MM) approach is introduced, and it is applied to construct the reaction coordinate energy profiles for the ATP hydrolysis in the active site of the protein and to specify reaction products and possible intermediates. This approach achieves a near ab initio representation of the entire system. The modeled transformation, with coordinates derived from the heavy atoms of the crystal structure, is viewed as a part of the overall hydrolysis reaction occurring in the closed enzyme pocket after ATP is bound tightly to myosin and before conformational changes preceding release of inorganic phosphate. Molecular dynamics simulations coupled with the QM/MM studies have been carried out to assist the analysis of the events occurring immediately after the intrinsic chemical reaction. This theoretical method provides a new insight on mechanisms of this fundamental process.

Spindle Pole Mechanics: Architecture, Compliance, and Force Distribution Alan J. Hunt (University of Michigan)

In this talk, Dr. Hunt reported recent progress made by his research group on the understanding of spindle mechanics and roles of specific spindle proteins. In their work, they assembled spindle poles in a cell-free mitotic extract and applied optical trapping to directly assess the mechanics of microtubules emanating from these poles. It is observed that microtubules exhibit bidirectional movements reaching speeds of order nm/s, and attenuated in the presence of AMP-PNP. This is indicative of the molecular motor activity. Microtubule linkages to spindle poles are remarkably compliant and are apparently mediated by only a handful of crosslinkers. The large compliance of microtubule linkage to spindle poles provides a robust, non-brittle mechanical architecture capable of accommodating microtubule movements associated with poleward microtubule flux without coordinated rearrangement of crosslinks. Furthermore, compliance helps to distribute strain amongst microtubules focused at the pole, thereby integrating even antagonistic spindle forces such as those resulting from poleward and anti-poleward chromosome movements without disrupting the overall spindle architecture.

Pushmi-pullyu: The Peculiar Dynamics of Microtubule-based Motors in Chlamydomonas William Guilford (University of Virginia)

In this talk, Dr. Guilford reported recent findings about the dynamics of microtubule-based motors in Chlamydomonas. Their earlier work found that oppositely directed molecular motors are reciprocally coordinated rather than operating in a tug-of-war fashion in which many motors can engage simultaneously to move the extracellular cargo. Dr. Guilford and his research group recently discovered that the velocities of transport are discrete and quantized. These motor proteins routinely travel at much higher velocities than those expected or allowed by our current understanding of processive molecular motors. Furthermore, the force generated on extracellular cargo is not proportional to the contact area with the membrane. These results suggest that motor dynamics and mechanics are unique to this intracellular environment.

Multiple Motors Cooperating in vivo and in vitro: What have we Learned?

Jed Macosko (Wake Forest University)

In this talk, Dr. Macosko reported some recent research results on motor cooperativity. The research was conducted using a so-called minimal load-sharing model that has potential to qualitatively explain all multi-motor results to-date. It is discovered that the number of kinesins matters a great deal when they jointly pull a significant load, and kinesin and dynein cooperativity is important in fast vesicle transport in an intracellular environment of high viscous drag. These new results correct previous assumptions about the motor cooperativity that increasing kinesin could not result in higher microtubule gliding speeds. This is because motors that are attached to their rails, like kinesins, could not achieve higher speeds, based on experiments performed in the absence of any appreciable load.



#### DAY 4: THURSDAY, APRIL 15

1D versus 2D Models for Microtubule Self-assembly: New Tests at the Nanoscale David Odde (University of Minnesota)

Microtubules (MTs) are stabilized by lateral and longitudinal bonds between heterodimeric tubulin segments, and so self-assembly of these biopolymers can potentially be treated as a 2D problem along the axial and circumferential directions. In 1970's several groups investigated the microtubule dynamics instability phenomena using simplified 1D model. In 2002, VanBuren et al. have proposed a new 2D model to account for some structural effects. Recent experimental data of nanoscale fluorescence microscopy and optical tweezers data demonstrate the limitations of the 1D model and confirmed the 2D model predictions. The conclusion is that the kinetics of MTs assembly has been consistently underestimated in the literature by an order-of-magnitude. These findings point out to the model where infrequent and weak alterations of the tubulin-tubulin bonds and tubulin mechanical properties must be taken into account.

Viscoelasticity in Homogeneous Protein Solutions Peter G. Vekilov (University of Houston) To understand the transport in the cytosol of living cells, the transport properties have been investigated in protein solutions stable with respect to any, solid or liquid, phase separation. He discussed the mean squared displacement of probe particles in the time range of 1 millisecond to 10 seconds in solutions of a model protein. The tested solutions exhibit significant elasticity at high frequencies, while at low frequencies, they are purely viscous. The conclusion is that the found intrinsic viscoelasticity of protein solutions should be considered in biochemical kinetics models. These finding might change significantly current views on biological transport.

Illuminating the Way Kinesin-1 Walks Using FRET between the Motor Domains

Erwin Peterman (Vrije Universiteit Amsterdam)

Each 8 nm step of the motor protein kinesin requires the hydrolysis of one ATP and it takes about 10 ms at cellular ATP concentrations. Key aspects of kinesin's walking mechanism are still not fully understood. A novel assay based on single-molecule confocal fluorescence microscopy was used to characterize kinesin's stepping mechanism in vitro. A key advantage of the approach over conventional wide-field methods is that the time resolution is far better, less than 0.1 ms. With this approach, an intermediate state may be identified in the stepping process that lasts 2-3 ms at saturating ATP concentration. In this intermediate state one motor domain is bound to the microtubule and the other is rotated and substantially less than 8 nm away.

Single-molecule Science with A Nanopore: Inspiration from Nature

Liviu Movileanu (Syracuse University)

In this talk, Dr. Movileanu illustrated the nanopore probe techniques in revealing features of nucleic acids and proteins. Such approach proves to be quite powerful, because single molecules and biopolymers are examined at very high spatial and temporal resolutions. Mechanistic understandings of the forces that drive protein translocation through a nanopore have been discussed. These measurements facilitate the detection and exploration of the conformational fluctuations of single molecules and the energetic requirements for their transition from one state to another.

Channel-Facilitated Molecular Transport Across Cellular Membranes Anatoly Kolomeisky (Rice University)

Using discrete stochastic site-binding models, Dr. Kolomeisky investigated transport of molecules across membrane channels. It was shown that the interaction potential between molecules and the channel has a strong effect on translocation dynamics. For small extracellular concentrations, attractive binding sites in the pore accelerates the particle current, however, for large concentrations, repulsive binding sites yield the most optimal transport. The effect of the asymmetry of the interaction potential on the channel transport has also been discussed.

### WORKSHOP FOR YOUNG RESEARCHERS

### Workshop for Young Researchers in Mathematical Biology August 24-26, 2009

#### ORGANIZERS MBI Postdocs

Report written by MBI postdocs Huseyin Coskun and Judy Day.

#### **OVERALL SUMMARY**

The 2009 Workshop for Young Researchers in Mathematical Biology (WYRMB) had over 50 people in attendance, including several international participants. This year, there were over 100 applications received, and the MBI postdoctoral fellows (organizers) decided to open up the workshop to more individuals than those whom the MBI could fund. Thus, after ranking the applications and inviting the allotted number of the very highest ones for which we could provide funding, we additionally invited the next tier of top applicants to participate if they could fund all of their own expenses. This resulted in approximately 10 more participants that were able to attend and take an active part in this year's WYRMB by presenting their research in poster format.

Also new this year, the postdocs decided to have short talks (20 minutes) from among the workshop participants rather than from only the MBI postdocs as was done in the past. These short talk spots were given to the top nine applicants. The workshop participants (tenure-track faculty, postdoctoral researchers, advanced graduate students, and even a few undergraduate students) represented institutions from around the world. The short talk presenters and all others presented a poster at one of the two poster sessions that were held. Prior to each poster session, each presenter gave an overview of their poster material in a 1-2 minutes presentation. Talks and posters covered a wide range of topics such as intracellular dynamics, phylogenetics, epidemiology, optimal control applied to infectious disease control, neuroscience, and ecology.

#### DAY 1: MONDAY, AUGUST 24

Tumor Immune Interaction, Surgical Treatment and Cancer Recurrence in a Mathematical Model of Melanoma Yang Kuang (Arizona State University)

Dr. Kuang presented a recent mathematical model of skin cancer using reaction diffusion equations that describe the evolution of cancer cells, healthy cells, blood vessels, immune cells, necrotic tissue, basement membrane, angiogenic factor and oxygen. The model's aim was to help cancer surgeons decide how much tissue should be cut to avoid metastasis. They found that small metastatic lesions distal to the primary tumor mass can be held to a minimal size via the immune interaction with the larger primary tumor. Numerical experiments further suggest that metastatic disease is optimally suppressed by immune activation when the primary tumor is moderately, rather than minimally, metastatic.

#### Correlative Cellular and Tissue Models of Chemotaxis Zhian Wang (Vanderbilt University)

Dr. Wang started with an introduction of chemotactic movement and individual-based and continuum models of taxis-driven population dynamics. Recent literature shows that individual-based stochastic models (using velocity-jump processes) can be reduced to Keller-Segel models formally in the parabolic limit. He showed that the derivation can be proved rigorously in 1-D, but for higher dimensions it is still an open problem.

Multiscale Modeling in Biology Mark Alber (University of Notre Dame)

Multiscale modeling approaches (typical in systems biology) tend to mix continuous, discrete, deterministic, and probabilistic submodels. First, a computational model of M. xanthus swarming was introduced. Swarming, a collective motion of many thousands of cells, produces colonies that rapidly spread over surfaces. The mechanism and molecular dynamics of forward and reversal motion were discussed and individual and social motilities and the synergies between them were addressed. Simulations and comparisons between the model findings and experiments were presented and outlined. One of the conclusions was that reversals of gliding direction are essential for swarming and that the reversal period predicted to maximize the swarming rate is the same as the period observed in experiments. Another outcome is that order extends over long distances. Cell alignment provides conditions for this to happen and replace chemotaxis. It is stated that swarming strategies used by bacteria might provide swarming algorithms for robotic systems used in controlling traffic flow.

Second, thrombus development was addressed. To prevent the loss of blood following a break in blood vessels, components in blood and the vessel wall interact rapidly to form a thrombus (clot) to limit hemorrhage. Biological models and real mechanisms for platelet dynamics and clot formation were discussed before the quantitative modeling approaches were presented. Molecular mechanisms were also explained. The multiscale model of thrombus formation consists of components such as viscous, incompressible blood plasma; coagulation pathway; quiescent and activated platelets; blood cells; activating chemicals; fibrinogen; the vessel walls and their interactions. The cellular and subcellular Potts models were introduced and simulations for the multi scale model for

thrombus development were presented, with the model outcomes justified by comparison to the experimental data. For example, simulation results demonstrated the development of an inhomogeneous internal structure of the thrombus which was confirmed by the preliminary experimental data. Also, the dependence of the thrombus size on the blood flow rate in simulations is close to the one observed experimentally.

Nonlinear Fluctuation Theorems in Biology Johan Paulsson (Harvard Medical School)

Dr. Paulsson began by summarizing some of the difficulties present in modeling biological systems, particularly those that arise as a result of the rarity of many species and reactions in the cell, nonlinearities, and poor system characterization. He then showed, using tools from functional analysis, information theory and statistical physics, that seemingly mild constraints such as short time delays or finite numbers of molecules can impose limits on noise suppression that are independent of the particulars of the feedback system. The general results were applied to the real-world example of bacterial plasmid replication, where large fluctuations would promote extinction if not for the numerous mechanisms that have evolved to sup-



Participants in the Young Researchers Workshop.

### WORKSHOP FOR YOUNG RESEARCHERS



Johan Paulsson (Harvard Medical School)

press noise.

#### Afternoon Short Talks: Peter Kim, Baris Hancioglu, Joanna Pressley, and Janine Haugh

Emergent Group Dynamics Produce a Robust Primary T Cell Response Peter Kim (University of Utah)

Peter compared and questioned the known models for the T cell proliferation program. Current basic biological theories were reviewed and a summary of experimental data was presented. He explained that it is widely accepted that effector T cells commit to autonomous developmental programs. However, his research focused on the role that adaptive regulatory cells and intrinsic developmental programs play in governing the dynamics of the primary response. Several models were presented and compared, highlighting the importance of feedback control mechanisms in providing a robust developmental program over a wide range of initial stimulatory levels and conditions.

The Dynamics of Exit from Mitosis Baris Hancioglu (Virginia Tech)

Baris introduced the biological background for cell division, which provided the necessary information to dis-

cuss a mathematical model of mitosis used to understand the mechanism for exiting from mitosis. Experimental results by two independent research groups regarding the mechanisms for mitotic exit were compared to each other and the model was able to shed light on the apparently contradictory results from the two labs. The model suggested that the differences could be justified by including more molecular factors into the analysis of the results from either lab.

How Co-colonization Affects Competition of Hospital-acquired and Community-acquired MRSA Joanna Pressley (Vanderbilt University)

Hospital-acquired Methicillin-resistant Staphylococcus aureus (HA-MRSA) is historically an infection seen in immuno-compromised hospital patients and Communityacquired Methicillin-resistant Staphylococcus aureus (CA-MRSA) is a new strain detected in the general community that can infect healthy individuals. Joanna discussed a model that assumes that patients can be co-colonized with multiple strains for MRSA. Their results indicate that, contrary to previous results derived from the assumption that co-colonization is impossible, competitive exclusion is rarely seen with the co-colonization model and that more often it is the case that both strains become endemic in the hospital. In addition to exploring the possibility of competitive exclusion, they also investigated two types of intervention strategies: the efficacy of decolonization and hand-washing compliance.

Modeling Cartilage Regeneration in the Extracellular Environment of a Cell-Seeded Hydrogel Janine Haugh (North Carolina State University)

Janine first explained that as a natural biomaterial, cartilage is important for load support, energy distribution, and lubrication of joints, but can become damaged due to injury or osteoarthritis. Damaged cartilage does not usually repair and regenerate to its pre-injury state. Thus, much research has gone into looking at using nutrientrich hydrogels seeded with cartilage cells as biomaterials for tissue regeneration and repair. Her research was centered on the development of a mathematical model for cartilage regeneration that focused on the local environment of a single cell seeded in a hydrogel scaffold. After the short talks were finished, participants presenting posters at the evening's poster session gave a preview of their work. This was followed by a reception in the MBI foyer where the poster session took place.

#### DAY 2: TUESDAY, AUGUST 25

The second day was organized so that the three panelists serving on the "Math-Bio Jobs in Industry" panel first gave talks describing their specific work environments and the types of research being done at their respective companies.

#### Math Biology Experience - A View from the IT Industry Kirk Jordan (IBM)

Kirk described why health care and the life sciences were important to an IT company, since such things as knowledge management, visualization and data analysis, security and privacy, modeling and simulation, data mining, and data management are all areas where data needs to be transformed into knowledge. The big themes in systems biology such as automated analysis, modeling, simulation, and integration of computational biology and experimental biology are creating needs that the IT industry, along with mathematicians and computational scientists, have the potential to solve, with the goal of using advanced computing systems and computational math to provide solutions of clinical relevance. He described the various



areas of IBM where health care and life science initiatives are the focus and used several examples from his own research and that of his colleagues to show how high-end computing is enabling researchers to solve problems that otherwise would be impossible to solve. These projects have also led to the development of better numerical algorithms. In addition, even the computing abilities have been challenged in the case of certain problems, driving the innovations for high performance computing. He also challenged the audience to think about the ways in which to utilized high performance computing, perhaps rethinking the problem at hand. He challenged them to think about what computing resources (no matter how big) would be needed for a certain problem, perhaps even reconsidering problems previously thought to be "too big." He concluded by presenting some possible opportunities regarding multi-scale biological problems for high performance computing and mathematics to join forces.

Young Researchers in Math Biology Rukmini Kumar (Entelos, Inc.)

Rukmini described the mission of Entelos, which is a company that is focused on developing predictive products and services that help health care and life sciences organizations improve human health. Essentially, they are enabling pharmaceutical companies to understand and predict human response to therapeutic interventions. This is accomplished through the development of large-scale mathematical computer models that simulate human physiology to predict human response and improve decision making throughout the drug discovery "pipeline." She briefly described a model for diabetes as an example and explained the use of virtual patients to determine who a particular drug will help. Such an approach allows the pharmaceutical companies to experiment with their ideas at lower cost and involvement. She emphasized the importance of good communication and presentation skills in her workplace and described the various tasks that she is involved in on any particular day, such as conducting literature surveys or talking with biologists and meeting with academic experts. She also described the types of qualities that her company looks for when hiring, such as strong, proven technical skills (i.e. publication/presentation record), communication skills (ability to convey ideas as well as listen effectively), and ease of interaction, leadership, and enthusiasm.

A Career in Bioscience (Industry and Academia) and a Bit About Mathematics

Herb Bresler (Battelle Memorial Institute)

Herb provided information about Battelle from the perspective of a Math-biologist with an academic background, now working in industry. Battelle Memorial Institute is a non-profit research organization founded in Columbus Ohio in 1929. They employ over 20,000 people worldwide and perform research in life sciences, energy technology, and national security and defense. Battelle provides companies with service contract research, and re-invests the profit into science and technological development.

In the afternoon, a discussion panel featuring the three industry speakers was led by MBI postdoc Judy Day. Four main questions were presented to the speakers in

### WORKSHOP FOR YOUNG RESEARCHERS

succession and audience participation and questions were spread throughout the discussion. The panelists discussed the type of characteristics and qualities that a company looks for when considering hiring a person with a math-biology background as well as how one makes themselves an appealing industry candidate. Much of the discussion focused on the type of information that should be contained in a resume, noting the differences between a resume and an academic CV. Also discussed was the feasibility of someone to transition from industry to academia. Participants found the discussion panel very helpful and informative and a lively discussion continued throughout the allotted time for the panel.

### Afternoon Short Talks: Simone Linz, Berton Earnshaw, and Choonseok Park

Calculating Tree Distances: A Divide-and-conquer Approach Simone Linz (University of California, Davis)

Simone described the necessity of calculating a distance between two phylogenetic trees in evolutionary studies that aim at comparing trees. He described a particular way to calculate such a distance called the rooted subtree prune and regraft (rSPR) distance, but that it is often computationally difficult to calculate because it is an NP-hard problem. Algorithms for finding the exact calculations are not easily available and not always practical. He presented a method for calculating this distance by breaking up the problem in to several small, more tractable subproblems. The resulting algorithm provides a way to calculate the rSPR distance exactly.

A Diffusion-activation Model of CaMKII Translocation Waves in Dendrites Berton Earnshaw (University of Utah)

Berton presented the research his team has performed on Ca2+-calmodulin-dependent protein kinase II (CaMKII), and its role in the expression of synaptic plasticity. Their model, which includes the diffusion, activation, and local translocation of CaMKII in the spines of dendritic nerve cells, reproduces the wave-like spread of CaMKII that is observed experimentally when dendrites are locally stimulated with glutamate or glycine.



Rukmini Kumar, Kirk Jordan, and Herb Bresler

The Intermittency of Synchronized Activity in Basal Ganglia Choongseok Park (IUPUI)

Choongseok gave a talk on the role oscillatory synchronous activity of basal ganglia, compact clusters of neurons in the forebrain associated with several functions such as cognition and motor control. Dysfunction of the basal ganglia is associated with disorders of motion control such as Parkinsons syndrome. Their models describe the dynamics in-between asynchronous and synchronous episodes and show that in the disease Parkinsonism, the state of the basal ganglia circuits are relatively close to the non-diseased, healthy state. The closeness of the two states may be explained by the fact that synchronized oscillations are efficiently produced for movement generation. Immediately following this last short talk, the remaining poster participants gave 1-2 minute previews of their work. This was followed by a second (smaller) reception and poster session in the MBI foyer.

#### DAY 3: WEDNESDAY, AUGUST 26

Socially-induced Reproductive Synchrony in a Sea Bird Colony Shandelle Henson, (Andrews University)

Dr. Henson presented data that showed how the ovulation cycles of glaucous-winged gulls Larus glaucescens
can become synchronized much in the same way that menstrual cycles of women can exhibit spontaneous synchronization. It was observed that the level of synchrony declined with decreasing colony density, suggesting that social stimulation plays a role in the synchronization. Dr. Henson then discussed a discrete-time mathematical model based on the hypothesis that social interactions are in fact responsible for synchronized surges in preovulatory luteinizing hormone, although the exact nature of the social interaction was not discussed.

Model Formation and Optimal Intervention Strategies for Cholera

Rachael Neilan (University of Tennessee)

- in collaboration with Suzanne Lenhart

Rachael explained that for cholera, a diarrheal illness caused by the bacterium Vibrio cholerae, there is still no effective outbreak control strategy. She discussed the formation a mathematical (SIR) model for the spread of cholera that included control functions representing the effects of various types of treatments: oral rehydration therapy, antibiotic treatment, vaccination, and sanitation. Optimal control theory, parameter sensitivity analysis, and numerical simulations were among the tools used to determine the proper intervention strategies during a cholera outbreak that minimized both death (due to disease) and treatment costs. The results highlighted important factors involved with striking an optimal balance among the intervention methods and offered suggestions that could benefit the efforts of policy makers who need to design strategies for controlling an outbreak in a way that is cost-effective.

Optimal Control in Epidemic Models of Rabies in Raccoons Suzanne Lenhart (University of Tennessee)

Suzanne began by giving background information on rabies in raccoons (costs, treatment, death rate, etc.) and then introduced the basics of optimal control theory (including Pontryagin's Maximum Principle). Lenhart discussed frameworks to investigate controls for vaccine distribution as it impacts the spread of rabies among raccoons. The control gives amount and location of the food packets containing vaccine. The goal is to minimize the number of infected raccoons and the cost of distributing the packets. She then presented two epidemic models for rabies outbreaks in raccoons: (1) a system of differential equations with a birth pulse (SEIRV model, where V stands for vaccine) and (2) a discrete temporal/spatial model. In the first model, more vaccine is required when the birth pulse is encountered, and a second round of vaccine is often required. Also, distribution of the vaccine depends on the number of days until the birth pulse. In the second model, optimal bait distribution depends on the initial location of the disease outbreak and the distribution of raccoons throughout the grid. The method can be readily extended to evaluate optimal vaccination distribution strategies with other spatially heterogeneous interactions, such as larger spatial grids and different movement assumptions (including density dependence).

Finally, a panel discussion about "Applying for Academic Jobs" was led by MBI postdoc Huseyin Coskun, and the following individuals served as panelists: Mark Alber, Shandelle Henson, Suzanne Lenhart, Johan Paulsson, Chiu-Yen Kao (The Ohio State University), and Tony Nance (MBI Assistant Director). Many pertinent topics were discussed:

- Finding the most suitable job and employer;
- Credentials;
- Preparation for the applications and application materials;
- Interviews;
- Early steps in the career; and
- Personal relations and contacts.

As expected, the workshop participants were actively engaged in the discussion and the panelists did a wonderful job of addressing the various issues and questions.

#### CONCLUSION

Even though this year's workshop was slightly shorter than last years, it proved to be highly beneficial and enjoyable to everyone in attendance. Several people mentioned that they enjoyed talking with other young researcher's in diverse fields and learning about topics outside of their personal research program. Also, some mentioned that they found people with whom they were going to collaborate. It was clear, especially from observing the poster sessions and lengthier coffee breaks, that the workshop facilitated the establishment of connections between young researchers and the opportunity to learn from the experience of those more established. The facilities and staff at the MBI played an important role in the success of the workshop, and the excellent logistical support was noted by many of the participants. Three quarters of the second day was focused on math-biology jobs in industry, and many noted that the time devoted to this was beneficial. The format used for this day proved to work very well compared to past WYRMBs that only held a panel on this topic. We received much positive feedback from the participants who said they appreciated knowing more about industry options. The overall workshop received very positive feedback from participants who also gave helpful suggestions that will be passed on for next year's Workshop for Young Researchers in Mathematical Biology.

### Computational Challenges in Integrative Biological Modeling October 5-8, 2009

#### ORGANIZERS

Suncica Canic, University of Houston Lisa Fauci, Tulane University Mette Olufsen, North Carolina State University

Report written by MBI postdocs Judy Day, Rasmus Hovmoller, and Chuan Xue.

#### MONDAY: OCTOBER 5, 2009

An integrated Study of Lamprey Swimming or: Going from Neurons to Vortices Avis Cohen (University of Maryland, College Park)

Dr. Cohen provided an overview of the past five years of the research from an interdisciplinary group that is studying lamprey swimming. The goal of the project is to come to an understanding of what all goes into one complete behavior: swimming. This collaborative effort to understand lamprey swimming integrates observations, physiological behavior (such as the sensory systems and the musculature of the animal, as well as the neuronal network), mathematical modeling (e.g., computational fluid dynamics), and robotics.

Fluid-mediated Coupling Between Filiform Hairs in the Cricket Cercal System Tomas Gedeon (Montana State University)

Dr. Gedeon began by mentioning four problems for which the cricket cercal system is a good model: neural encoding of sensory stimuli, evolutionary optimality vs. developmental constraints, biologically inspired sensor design, and fluid-structure interactions. He focused on fluid-structure interactions in his talk, and first described the physiology of the cerci, which serve as sensory organs for the animal. The cerci are covered with tiny hairs that move in response to certain stimuli. The goal of the research is to create a fluid-structure model to answer several questions regarding how these hairs work. To this end, each hair is modeled as a linear oscillator. He compared the current approach they are taking with previous approaches and then described the assumptions and the modeling construction. The model can look at hair to hair viscous interaction, showing that the cercus boundary layer incorporated into the model causes significant decrease in velocity experience by hairs. These results differ from previous studies that do not include the cercus boundary layer.

Two Problems in Vascular Biology: The Platelet Near Wall Excess and a Spatial-temporal Model of Blood Coagulation and Thrombus Formation Under Flow Aaron Fogelson (University of Utah)

To familiarize the audience with the physiology of the problem, Dr. Fogelson gave an overview of the blood clotting mechanism: platelets respond to an injury, become activated and activate other platelets and form a buildup of platelets, which eventually form a blood clot. The enzyme thrombin causes a fibrin gel to be produced which, along with the platelets, form a clot and many molecules, called tissue factors, regulate the reactions involved in thrombin production. Several questions remain open regarding this system and are being explored via a mathematical model which incorporates inactivated and activated platelets, various chemicals, and the fluid through which the platelets and chemicals move. More recently, an extension of the model includes two dimensional space and many of the known interactions of the clotting system. The model addresses the issue of scale between platelet density and the size of individual platelets. Although the model measures density of platelets, the size of the platelets is taken into account in four ways in the model. The model includes terms/equations that add realism to the model by incorporating mechanisms that define a platelet binding region and that limit the amount of platelets entering an already densely filled area. He also talked about the effect that red blood cells have on platelet adherence to the vessel wall, which has been observed but not well characterized. Dr. Fogelson and his students are using immersed boundary methods to look into this topic. Simulations and statistical analysis shows that the platelets accumulate near the vessel walls. The model here uses 'drift-diffusion' stochastic differential equation and diffusion coefficients are extracted. It showed an inverse correlation between the location of red blood cells and that of platelets; where the red blood cells are mostly present (core of the vessel lumen) less platelets are found, whereas the areas where the red blood cells are scarce (vessel wall) the platelets are more likely to be found.

The Challenge of Multiple Scales in the Biological Sciences: Applications in Cerebro-vascular Perfusion Tim David (University of Canterbury, New Zealand)

As implied by the title, Dr. David focused on explaining the multiple scale feature of the vascular network. The diameter of blood vessels ranged from 25 mm (ascending aorta) to 20 micrometers (capillary). Vessels with different sizes have different mechanical properties. The challenge is to integrate vessels of all different scales into one model. He introduced a mathematical model of the blood perfusion in the cerebro-vasculature, which describes the complicated structure of the vasculature by randomly generated vascular trees with different layers representing different scales of arterial vessels. Pressure difference was prescribed to the root and tips of the tree. Numerical results of the blood flow in the tree were presented.



Strong Inference for Systems Biology Daniel Beard (Medical College of Wisconsin)

Dr. Beard emphasized that combining biological research with mathematical computation is important to elucidate some of the long-standing questions in cardiology. The talk invited intense discussion of the centrality of hypothesis-based research. Although different people have different opinions on specific aspects, many of them agreed that mathematical training should be part of medical school training. An Integrated Approach to Tumor Invasion: Bridging Scales and Models

Alexander Anderson (H. Lee Moffitt Cancer Center & Research Institute in Florida)

Dr. Anderson started with a general description of the multiscale property of cancer growth, and followed with their recent modeling efforts in cancer modeling and computation. Their model incorporated biological details at different spacial scales, including subcellular, cellular, and extracellular levels, with a special focus on the role of tumor microenvironment. Applications of these models to prostate cancer were also discussed as the last part of the talk.

#### TUESDAY: OCTOBER 6, 2009

Functional Imaging in Cardiology: Current and Emerging Applications

Stephen H. Little (Cornell University)

Dr. Little presented a review of cutting edge imaging methods in cardiology. With recent developments in imaging, the beating human heart can be rendered in three dimensions. Imaging methods include echo imaging by ultrasound and MRI. By detecting the Doppler shift by 3D echo methods, the flow of blood through the valves and chambers of the heart can be accurately measured, and the movement of the entire heart in the chest can be described.

Dynamic Analysis of 3D-echocardiographies: Deformable Models of Mitral Valve Robert Azencott (University of Houston)

Dr. Azencott demonstrated advances in software for generating patient specific models of the mitral heart valve. By using recent 3D imaging methods, the movement of the individual parts of the mitral heart valve can be tracked and tagged for analysis. His team has developed the open source software package SLICER for analysis and visualization of 3D data, where the tagged images of the heart can be used to generate NURBS models of the mitral valve apparatus and speckle tracking for modeling the dynamic heart movements of individual patients.

Adaptive Numerical Methods for Simulating Biological Fluid Dynamics and Electrophysiology Boyce Griffith (NYU School of Medicine)

Dr. Griffith described new developments in adaptive numerical methods for simulating fluid dynamics in cardiology. The flow of blood through the mitral valve can be modeled using an immersed boundary method that captures the fluid-structure interactions of an elastic structure immersed in an incompressible fluid. However, us-



Carson Chow (NIH-NIDDK)

ing high resolution grids on the entire spatial domain is impractical, and new methods developed by Dr. Griffith reduced complexity by identifying the appropriate level of detail for simulating flow and deformation in different regions of the structure.

Rank-deficient and Ill-conditioned Nonlinear Least Squares Problems in Cardiovascular Modeling Tim Kelley (North Carolina State University)

Dr. Kelley described a problem involving parameter identification in a compartmental model of cardio-vascular system that produced a nearly rank-deficient nonlinear least squares problem. The use of a usual Matlab stiff solver (ode15s) to solve the nonlinear least squares problem involves many problems with the optimization routine. He described the iterative process that is normally used to deal with this problem: the Levenberg-Marquardt (L-M) method and truncated singular value decomposition. The ideal is that the L-M converges to a minimizer or at least the gradient is zero; however, this method only works if the residual value is small. There is a clear gap in the singular values, and the Jacobian and residuals can be computed accurately. Such exact arithmetic does not always occur in problems as in the case of the cardiac model. Thus, he presented another method based upon subset selection. Essentially, this method turns a rankdeficient matrix into one with full rank so that 'normal' algorithms work well on the problem and each iteration does not have to deal with singular values. He described this approach for the particular problem that they were considering and first went over a toy problem to illustrate the method. For the full cardiac problem, they were not able to calculate the Jacobian as accurately (as in the toy model) but instead used a finite difference Jacobian. Nevertheless, the subset method results were superior over the more traditional results.

Parameter Identification Problems from Systems Biology Philipp Kuegler (Industrial Mathematics Institute, University of Linz, RICAM, Austrian Academy of Sciences)

Dr. Kuegler first discussed several types of biochemical reaction networks and used the gene regulatory network of halobacterium NRC-1 to illustrate a specific example of a complex biochemical reaction network. Many different models exist for such networks such as graph models, Boolean network models, deterministic ODE systems, stochastic models, and spatial models. The focus for this talk was on deterministic ODE systems to discuss the process by which to carry out parameter identification. One of the usual ways of determining parameters is to take time series data and map that information to parameter values, which is labeled as Type I parameter identification. Such inverse problems are typically ill-posed and regularization methods are required to avoid large amplifications of error in the data. He presented an example involving a three step metabolic pathway which was described with an ODE model of 20 variables with 35 parameters. The problem was ill-posed due to large error amplifications in the data and the resulting least squares problem diverged. The introduction of some regularization methods improved results and the software package, SOSLib, is now available, which extends the SMBL ODE Solver Library with some of the regularization methods.

However, some problems still have too little data to carry out parameter identification of Type I, and it would be advantageous to map bifurcation and heuristic model behavior to parameter values instead. Many biological systems may show qualitative behavior such as multiple steady states, oscillations, deterministic chaos, etc. Solutions that reveal key components/mechanisms and allow for simple interpretation might be of interest. How does one go from the observed/desired qualitative behavior to parameter values? Kuegler highlighted another example involving cellular apoptosis (programmed cell death). Singling out one of the parameters, it was shown via a bifurcation diagram that steady state behavior depended on the value of that parameter. Suppose that a system operating with a certain set of parameters triggers apoptosis in an excessive manner such that the critical threshold of the parameter increases. He summarized the process by which the parameters can be altered to get this behavior via the solution of an inverse problem of type II.

### Deconstructing Integrative Models: The Case for Model Reduction

George C. Verghese (Research Laboratory of Electronics, MIT)

Dr. Verghese highlighted large-scale structured dynamic models, heterogeneous, mechanistic, and integrative models (not discretizations of PDE). Typically, physiology models are large and grow with increasing knowledge about the system. However, in these models there is much uncertainty in parameters and there are no clearly defined inputs and outputs. The idea is to reduce these structured "gray box" type models so that the result can be interpreted and useful to understanding the larger model. In general, model reduction should examine the integral parts of the model and reflect important features of the system, rather than rely purely on the reduction algorithm used. He discussed several large network problems covering cardiac physiology to world-wide energy networks, showing that the output of the data from such large systems can be fit with models having much fewer parameters compared to a model that would describe the system fully. He described the value of larger models, noting that such models summarize the knowledge about a given system, allow simulation and exploration, as well as suggestions for experiments and hypothesis generation. However, they can fall short due to the difficulty in tuning the parameter values to match available measurements. Furthermore, practitioners cannot typically utilize the full model.

So, how does one go from a large model to something smaller that represents the system behavior well? In other words, for very large systems with many components, is there a systematic process that reveals the significance of a portion of the system? Much can be learned from clinicians themselves who regularly reduce massive amounts of information and data into something that can be reasonably used and applied in a practical situation. Dr. Verghese gave an overview of typically used model reduction methods which do not generally take into account the gray-box structure and showed an example of a recent literature article where the model reduction results are not clearly interpretable and translatable. He then discussed some traditional reduction methods that do preserve interpretability. These include various averaging methods (refinements of mass action kinetics). He introduced a method of "participation factor analysis" on the network where a participation matrix is formed. Once structure insight is obtained, these can be used in the model reduction, with less relevant dynamics being approximated in various ways. He presented an example of glycolytic oscillations, asking what small subset of variables from a 20 dimensional model can capture the oscillations. Normal model reduction techniques required copious amounts of work and only reduced it to a 6th order system. After applying the participation factor analysis on the model, four significant variables were singled out which captured the oscillation, although the amplitude of the oscillations was slightly off.



Giovanna Guidoboni (University of Houston)

#### WEDNESDAY: OCTOBER 7, 2009

Bayesian Approaches for Parameter Estimation and Model Evaluation of Dynamical Systems Carson Chow (NIH – NIDDK)

Dr. Chow began by giving an overview of Bayesian Inference (BI), going over Bayes' theorem and how to apply it to a parameter estimation problem, using a simple example to illustrate. BI tells you the uncertainty in the parameters, given the data, with more data giving better accuracy. This idea can be then applied to model comparison. For instance, given a data set, what is the probability that one model will be better than another, competing model. The computational expense to calculate the integrals containing the appropriate probabilities is typically high or was high in terms of past computational power. This is one of the reasons why Bayesian techniques have only recently been of wide spread use. A model with very high likelihood probability strongly overlaps the prior, leading to one of the main complaints regarding

the Bayesian framework: the necessity of the prior. He made the case that modeling in general is not independent of priors and that BI gives a systematic way of keeping assumptions disclosed and methods consistent. One method of computing BI is the Markov Chain Monte Carlo (MCMC) method for calculating the posterior. There was much discussion about some of the possible difficulties in actually getting to a posterior and that the prior definitely has an important role and contribution to the result. Afterward, he presented the physiology of embryonic stem cells transitioning to extra-placental cells and back again. A drift-diffusion-growth equation was used to model the system and many different kinds of models were developed to capture different ways a mechanism could happen. Bayesian inference was applied to find the best model that fit the given, available data. Dr. Chow went over an additional problem having to do with insulin's affect on glucose and free fatty acids, where there is much individual variation between people. Here again, they tested many different types of models, using BI to choose the best one.

Results and Challenges in Using Mathematical Models of Physiology to Improve the Differential Diagnosis Process Jonathan Rubin (University of Pittsburgh)

In the medical sciences, monitoring of patients produces much quantitative physiological measurements and evaluation of such data is mostly done heuristically. Although there has been a great increase in the amount of data that can be acquired, there has not yet been a great improvement in treatment. The system generating the observation is complex and highly nonlinear. The goal of this research effort is to exploit physiological models as a way to give patient data a usable, clinical interpretation. The particular physiological system considered is a heart/circulation model, since the ICU data and the nature of the ICU suggested this would be a relevant place to begin. The circulation model is composed of five ordinary differential equations containing 19 parameters. Bayesian inference is used to come up with a probability distribution on the model parameter space, which is individualized to each patient. The idea is to make likelihood inferences based on the given data to make choices about treatment; much like a clinician would do regarding a critically ill patient. A

'proof of concept' simulation was demonstrated, where it was shown that "meaningful information can theoretically be extracted in an underdetermined situation" and that a "possible direct correspondence between multimodality in a posteriori distribution and [a] differential diagnosis" can exist. Dr. Rubin then addressed several additional factors that need to be considered, such as optimal sampling from distributions. The use of Markov Chain Monte Carlo methods can present some difficulties since there are several hand-tuning types of things that



Stephen Little (Weill Medical College of Cornell University)

have to be done; however, if the problem is computationally tractable, then a population MCMC with additional features can be used. Dr. Rubin presented an alternative method that he and co-authors have developed: Ratio Approximation Markov Chain Monte Carlo (RAMC), where knowledge of the target density is exploited to obtain a "tuning-free, non-parametric algorithm for generating samples." The conclusion at this stage of the venture is that the RAMC method either matches the results of the optimal Metropolis-Hastings method or does better.

#### Reconsiderations on Parameter Estimations in Physiology Johnny T. Ottesen (Roskilde University in Denmark)

When modeling physiological systems, it is generally the case that several of the parameters are not generally observable. Dr. Ottesen presented several approaches to circumvent these issues, including variants of sensitivity analysis and sequential modeling using heart rate, depression and diabetes in humans as examples.

High Resolution Measurements and Scaling of Cardiovascular Mechanics in Mice Craig Hartley (Baylor College of Medicine)

Dr. Hartley presented research on how to scale mouse models for vascular disease to humans. In most mammals, many parameters such as blood pressure, heart rate and body dimensions scale by a simple formula Y = A \*body weightB, where a and b are constants. Dr. Hartley's lab has invented and built equipment to measure heart rates and blood flow in the mouse vascular system, and by careful modeling they have demonstrated how measurements taken in knockout mice can be applied to the corresponding values for a human.

Mathematical models of arterial blood flow Giovanna Guidoboni (University of Houston)

Arteries, the blood vessels that lead from the heart out into the tissues, are usually modeled as a viscous liquid flowing through a deformable structure. In this system, modeling the fluid-structure interaction including wave propagation and shear stress in the vein walls is a complex problem. Novel methods such as the kinematicallycoupled scheme were presented by Guidoboni and compared to traditional models.

Some Recent Numerical Methods in Electro-Cardiology Alessandro Veneziani (Emory University)

Dr. Veneziani elaborated on new developments in numerical methods in electro-cardiology. The traditional bidomain model of the electrical properties of the heart, a continuum model developed in the 1970s, is computationally expensive. However, recent developments by Veneziani and others have improved the numerical effectiveness of the model.

#### THURSDAY: OCTOBER 8, 2009

Calibration and Validation of Cerebral Blood Flow Models Pierre Gremaud (North Carolina State University)

Dr. Gremaud discussed numerical difficulties in modeling cerebral blood flow raised by complicated vascular structure, viscoelastic property of the vessel, irregular boundary of the flow, as well as the three-dimensional features of the flow. The talk examined calibration and validation in hemodynamic models of the Circle of Willis with different boundary conditions.

Model-based Estimation in Biomedicine Thomas Heldt (Massachusetts Institute of Technology)



Tomas Gedeon (Montana State University)

Dr. Heldt highlighted the challenge of using clinical data being collected in hospitals to guide and improve medical treatment. Thomas showed their success in using computational models of organ systems to derive physiologically meaningful information from data collected in intensive care, peri-operative care and emergency care environments.

Model Design: Evaluating the Parameter Estimation Problem in Cardio-respiratory System Modeling Jerry J. Batzel (University of Graz, Austria)

Dr. Batzel presented a respiratory control system model responding to blood gases and altering ventilation, as well as a cardiovascular control system model incorporating baroreflex control loops that alter resistance, unstressed volume, and heart rate. He then discussed how his group used the model to estimate parameter and suggest experimental designs.

Biomechanics of Epithelial Cancer Initiation. A Tale of Two Models

Katarzyna Rejniak (Moffitt Cancer Center)

Dr. Rejniak presented an integrated approach of modeling the development of epithelial tumors. Model development combined laboratory experiments, image processing, biomechanics, and computational fluid dynamics.

### Mathematical Developments Arising from Biology November 8-10, 2009

#### ORGANIZERS

John Guckenheimer, Cornell University Reinhard Laubenbacher, Virginia Tech Bernd Sturmfels, University of California

Report written by MBI postdocs Julia Chifman, Shu Dai, and David Romano

#### **OVERALL SUMMARY**

This workshop focused on significant theorems, theories and algorithms in mathematics that have been or are being inspired by problems in biology. Topics were chosen from dynamical systems, combinatorics, partial differential equations, probability, statistics, topology, algebraic geometry, and others. The primary goal was to bring new and interesting mathematical questions to the attention of the entire mathematical sciences community. A strong effort was made to attract mathematical scientists who have had little previous contact with the biosciences.

#### SUNDAY: NOVEMBER 8, 2009

Deep PDE Challenges in Biology Mike Reed (Duke University)

The main focus of Reed's presentation was on the history of partial differential equations arising form biology and the impact of biology on numerical analysis and PDE's. The presentation surveyed the works of many young researchers and the challenges scientists face today. Reed emphasized that whenever we look at the biological problem at hand, we discover that the mathematical machinery does not yet exist to solve the problem, and that our challenge is to develop that machinery. His overall message was that in order to find new PDE questions we must look at biology; we need to understand dynamical systems that are active and have life-like properties, systems that are far from equilibrium. Meanders and RNA Folding Christine Heitsch (Georgia Tech)

Heitsch's presentation centered on applications of combinatorics as motivated by molecular biology. She is interested in understanding how folded RNA sequences form structure. Heitsch makes a connection between meanders and RNA folding and proves that meanders are connected under appropriately defined local move transformations. It was pointed out that meanders suggest new approaches to the enumeration question, and also relate to the challenging biomathematical problem of comparing different possible folds for an RNA sequence.

Databases for the Global Dynamics of Multiparameter Nonlinear Systems Konstantin Mischaikow (Rutgers)

Mischaikow began by describing three biologically motivated problems: interpreting computations, model selection and data-driven modeling. He pointed out basic issues such as when specific parameters are not known or not directly computable, they are estimated. Thus, to identify model parameters, one needs to be able to match dynamics produced by the model against that which is observed experimentally. Proposed solution: build a database that catalogs robust global dynamics over a wide range of parameter values. The rest of the talk focused on the issues that need to be resolved: decomposition of the dynamics, describing the dynamics, and identifying dynamics at different parameter values.

On The Role of Positive Circuits in Gene Networks Christophe Soule (CNRS and IHES)

Gene networks can be represented by interaction graphs, which are finite oriented graphs with signs on the edges. Soule began his presentation by pointing out that we have this graph and we have these signs but we know nothing about the parameters; the main issue is what can we infer under the dynamics of the gene networks knowing only this qualitative information and nothing about actual values of the parameters. Some people tried to address this question via simulation, while others through the general properties of the gene networks. Robin Thomas conjec-



tured that the presence of a positive circuit is necessary condition for a gene network to admit several stationary states. By choosing a mathematical model for a gene network, Dr. Soule showed that the conjecture becomes a precise mathematical assertion. He also presented open questions about negative circuits.

#### Mathematical Problems in Cancer Modeling Natalia Komarova (UC Irvine)

Komarova's presentation covered two different topics: (1) Luria-Delbruck distribution, i.e. the distribution of the number of mutants in the colony that has undergone number of divisions, and (2) color categorization in people. Her interest is in modeling cancer, which is all about mutations; cancer arises by mutation and cancerous cells inside the tumor mutate as well. She pointed out that it is still unknown how cancer originates and how treatment failure can be prevented. She described the fixed-size problem vs. fixed time; knowing the number of mutants at a given time is not the same as knowing it at a given size. The second part of the talk concentrated around the idea of how different cultures have different color categorizations.

Combinatorics of Sequence Alignment Lior Pachter (UC Berkeley)

In his presentation, Pachter focused on three questions: (1) the structure of the set of possible "alignments" and how it can be effectively utilized for computations, (2) how we should define similarity and when is the similarity statistically significant, and (3) generalizations from the comparison of pairs of sequences to multiple sequences. His message was that the problem of sequence comparison continues to be central for molecular biology and questions about combinatorics of sequence alignment inspire new mathematics.

#### MONDAY: NOVEMBER 9, 2009

A Geometric Twist on Tactically-Driven Cell Migration Martin Wechselberger (University of Sydney)

A system of PDEs was constructed based on modeling the cellular migration into the wound space, in which a sharp interface was observed. The system may be transformed into conservation laws with small viscous perturbations. Instead of classical analysis of PDEs, Wechselberger provided an alternative approach by GSPT (geometric singular perturbation theory) to show the existence of a travelling wave solution with sharp interface.

Detection of Pairwise Gene Interactions in GWAS Data Using a Decomposition of the Pearson X2 Statistic on Contingency tables Glenn Tesler (UC San Diego)

The disease association test on gene interactions of large population generates a computation on a huge table for the markers, which is challenging for computation. A geometric transformation of this table was studied by Tesler and his collaborators. They used a decomposition of the Pearson X<sub>2</sub> statistics to cluster correlated markers efficiently.

#### Geometry of Fixed Points in Threshold-linear Networks Carina Curto (NYU)

Investigation of the relationship between network connectivity and the stimulus space leads to questions about stable fixed points associated. Curto discussed some recent results on the stable cliques which requires the stable submatrices of the threshold-linear networks, one particular kind of networks.

#### Stochastic Fluctuations in Biochemical Networks John Mattingly (Duke)

This talk focused on recent results on some typical types of networks, such as reaction chains with stochastic influx and side reactions are reviewed. Mattingly also talked about recent work of his team on developing second order numerical methods of stochastic integration.

Global Injectivity Criteria Inspired by Properties of Biochemical Networks Gheorghe Craciun (UW-Madison)

The injectivity, i.e. the non-existence of multiple equilibria of certain biochemical reaction networks has been widely investigated. Craciun showed a close relationship between the injectivity and the cycles of the netw orks. The conclusion leads to the Jacobian of the interactioneffect matrix, which is related to various problems such



as Jacobian conjecture and Bezier problem in computer graphics.

The Differential Geometry of Biological Growth Alain Goriely (University of Arizona)

The mechanism of residual stress developed during differential growth is hard to describe. A new method, called virtual configuration, which relieves the stress in a proper way, is considered. The configuration is then found to be associated with a metric and affine connection, which provides possible ways to compute some physical quantities and also helps understand the mechanism of residual stress.

#### **TUESDAY: NOVEMBER 10, 2009**

Topology and Topography of Biomedical Data Gunner Carlsson (Stanford)

Carlsson described how the homological methods of algebraic topology could be used to determine qualitative aspects of high dimensional data. He began with a study on diabetes that was done in the early 1970's, at a time when there was not a clear awareness of type I (juvenile) and type II as distinct diseases. The study represents an early attempt to use the geometry of data to understand qualitative features of the data; specifically, a three-dimensional projection of the data was drawn by hand, resulting in an object which looks like a ball with two "wings" that arise from two areas on the surface of the ball that are relatively far apart. The ball corresponded to the population of patients without diabetes, whereas the two wings corresponded to what are now understood to be type I and type II diabetes.

Carlsson then explained how homological methods can be used to obtain similar information about the "shape" of higher dimensional data. Often, clustering in data is determined by choosing a metric for the ambient space of the data, and then declaring clusters by associating all those data points that lie within a specified distance of each other. If instead this "cutoff" distance is viewed as a parameter, one can analyze the family of clustering regimes that arise from the metric to try to determine characteristics which seem to be inherent properties of the data rather than artifacts of the choice of cutoff value. By viewing each cutoff value as determining a triangulation of the data, where edges are inserted between data points that are "close", one can calculate quantities, called Betti numbers, that encode geometric properties of the spaces (in two dimensions, surfaces) corresponding to these triangulations, such as how many pieces make up the space, how many "holes" it has, etc. Then it can be determined which of these Betti numbers are robust with respect to the family of triangulated spaces.

In a similar way, by choosing a mapping between the ambient space of the data and a convenient topological space, one can use open coverings of the topological space as an alternative way to produce a triangulation by viewing sets of data points as points themselves and inserting edges between these if the corresponding sets in the covering intersect. Carlsson's talk ended by presenting a recent application of this covering method to data from a study on cancer, and showed how -- as in the diabetes study -- the resulting geometry seems to indicate the presence of a distinct class of patients that had not been previously identified, a class consisting entirely of patients who had survived the cancer, thus providing valuable insight into better understanding the disease.

#### Group Dynamics in Phototaxis

Doron Levy (University of Maryland, College Park)

Levy began his talk with a humorous introduction in which he gave general words of advice on the importance of working directly with biologists and of trying to understand how their perspective and priorities differ from those of mathematicians. Levy then gave a description of the behavior he is trying to analyze, namely the movement towards light that is exhibited by a particular species of cyanobacteria.

Roughly, a population of bacteria that is quiescent in the absence of light will, after a certain delay, begin to form finger-shaped groups which move in the direction of a light that has been turned on. How the bacteria sense the light and "decide" to move in its direction is not known, but Levy's effort was to model the collective movement alone. Levy contrasted the case of chemotaxis, where the Keller-Segel equations can be used, but which do not incorporate any delay in the initiation of movement. The "fingers" display a higher density on the boundary, and recruit bacteria that are close to the boundary, whereas bacteria that are in less dense areas tend to remain immobile. Levy chose to capture this recruitment by defining a variable that is meant to represent a level of excitation, and arranged for the level of excitation of a particular bacterium to move toward the average of that of the bacteria within a certain distance of the bacterium. In order to capture the delay, he also included the assumption that a bacterium would not begin to move until its level of excitation reached a particular threshold.

Levy showed movies of the simulations generated by the corresponding system of stochastic ODEs (modeling the behavior of every individual bacterium), and these movies showed a striking resemblance to the movies of the behavior of actual bacteria. Levy then described how to produce a PDE model of the same behavior by choosing discrete stochastic representations of space and the processes of excitation and movement, and then passing to the continuum limit -- via the Ito calculus -- to arrive at the corresponding system of PDEs. Finally, he described a third model based on an extension of the Cucker-Smale equations used to model flocking behavior, and suggested that the average solutions to this model trace out curves for which there is no corresponding theory to account for them.

#### Mathematical Problems Arising in Phylogenetics John Rhodes (University of Alaska, Fairbanks)

In the final talk, Rhodes described how algebraic geometry can be applied to the statistical inference of phylogenetic trees by relying on the following basic model of how mutations occur across time in evolution. First, suppose that we are given a tree describing the descent of a group of species (or taxa) from a single common ancestor. For

example, suppose that a common ancestor o evolves into taxa 1 and 2, and that taxon 1 further evolves into taxa 3 and 4, leaving taxa 2, 3, and 4 as extant species; this is an example of a 3-taxon tree. Now further suppose that for a given site on a given gene, the occurrence of the bases A, C, T, and G in taxa 0-4 is the outcome of the following random process: First, a base for species o is drawn from a fixed (root) distribution; then bases for taxa 1 and 2 are drawn, respectively, from two fixed conditional distributions associated to the edges 0-1 and 0-2; and finally, the bases for taxa 3 and 4 are drawn, respectively, from two fixed conditional (edge) distributions associated to the edges 1-3 and 1-4. The resulting distribution of bases among the extant taxa can be described by a 4x4x4 array (or tensor) P indexed by base triples, where the entry at, for example, ACT is the probability that the bases for taxa 2, 3, and 4 are A, C, and T, respectively.

Under these conditions, the fundamental observation is that the tensor P is a polynomial function f of the probabilities associated with the root and edge distributions, so that the space of all possible base distributions for the extant taxa can be viewed as a subset of the complex algebraic variety V that is the image of f. Now, from algebraic geometry, we know that the variety V corresponds uniquely to the ideal I of all polynomial functions that vanish on V, and it was in terms of this ideal that Rhodes described several results regarding the computability of I and its potential role in inferring the tree corresponding to a given group of extant species. One important context for these results was the case where it is assumed that the collection of all sites of a given gene divides into n classes, where all the sites in a given class are assumed to mutate according to the same random process; this context is known as the *n*-class mixture model, and the case where n=1 is known as the general markov (GM) model.

Rhodes explained that in the case of the GM model, there are results showing that the ideal I could be computed explicitly from complete knowledge of the 3-taxon case. Then, as to the problem of inference, he explained that in the case of n-class mixtures, knowledge of the sequence data of a group of taxa generically determines both the tree and the associated root and edge distributions that gave rise to the taxa, providing strong information about the tree and relative times of divergence. Rhodes concluded by pointing out that this tree is in fact a function of the gene under consideration and that different genes could give rise to different tree, and suggested that an analogous strategy --- known as a coalescent model --may be used to determine the most likely species tree and times of divergence from knowledge of the corresponding gene trees.

### Biofilms and Infectious Disease March 22-25, 2010

#### ORGANIZERS

John Gunn, The Ohio State University Dan Wozniak, The Ohio State University Nick Cogan, Florida State University

Report by MBI postdocs Suzanne Robertson, Rebecca Tien, and Kun Zhao.

#### **OVERALL SUMMARY**

The workshop covered the important role that biofilms play in infectious diseases in humans. Talks focused on biofilms that are generated both in indwelling devices (such as catheters) as well as outside of indwelling devices such as in oral cavities, lungs, and the GI tract. One of the major problems in medicine is the persistence of these biofilms despite efforts to eradicate them from the patient. A particular focus of this workshop were the mechanisms for persistence and how mathematical models can be used to test various hypotheses about persistence as well as to suggest new experimental approaches to ridding a patient of persistent biofilms.

#### MONDAY: MARCH 22, 2010

What Can a Model of Persister Formation Tell You? Nick Cogan (Florida State University)

The main focus of Nick's talk was how to get rid of bacteria within a biofilm. The first part of the talk addressed what tolerance mechanisms can arise in bacteria living within a biofilm (physical, physiological, phenotypic or some sort of adaptive response). Most of the talk focused on phenotypic tolerance mechanisms. In phenotypic tolerance, a subpopulation of bacteria can develop phenotypes that are extremely resistant to antibiotics and biocides. These types of bacteria are referred to as persisters. This state of bacteria is temporary and reversible, so that a bacteria cell can move in and out of being a persister as the environment necessitates. There are many explanations of how persisters arrive. A simple model of persisters was discussed. The actual model is a system of ordinary differential equations (ODEs), which include both susceptible and persister classes existing on a substrate. There are no explicit equations for the nutrient or the antibiotic described. Susceptible bacteria grow according to Michaelis-Menton kinetics. Persister cells can revert to susceptible cells. This does not happen if there is antibiotic present. Persisters don't grow (this is the cost for the phenotype).

The next part of the talk looked at different things that can be done with a simple model. For example, hypothesis testing allows investigators to look at what happens to the population of persisters with alternating dosings of antimicrobials. Different dosing strategies will lead to a varying number of persisters in the population. Another way to use the model is for optimization of procedures such as dosing level (p) that will yield successful treatment. Parameter estimation methods can be used to get a sense of parameter values allowing for more biological realism in the models. The last thing that can be done is to compare the initial simple model to other explanations of persister formation such as senescence. One can then compare the rate at which persisters wash out of the system for each of the hypotheses generated. The two models end up with very different estimates.

#### TUESDAY: MARCH 23, 2010

The Pel Polysaccharide of Pseudomonas aeruginosa Matthew Parsek (University of Washington)

Dr. Parsek focused on answering the question of how *P. aeruginosa* build biofilm communities. He addressed some recent results on the regulation of pel polysaccharide gene expression and its relationship to the second messenger bis-(3' - 5')-cyclic dimeric guanosine monphosphate (c-di-GMP) in biofilm communities. Dr. Parsek and his collaborators used a tube biofilm culturing system to measure pel transcription in a biofilm. Data suggest that surface attachment induced the accumulation of c-di-GMP in the cell resulting in FleQ dissociation from the pel promoter region, activating pel transcription and biofilm formation. Different functionalities of Pel and another polysaccharide Psl in P. aeruginosa strains PA14 and



POA1 were also discussed. It is reported that Pel expression contributes to aggregate builder activity and Pel is required for cluster formation, mediating cell-to-cell interactions. Also, Pel protects PA14 biofilm from aminoglycosides. On the other hand, the importance of Pel in PAO1 is negligible. It is reported that PsI appears to be the dominant polysaccharide in PAO1 biofilms. Several questions are addressed during the talk such as "Is PsI promoting cell-to-cell interaction in PAO1?", "Are Pel and PsI functionally redundant?", "Is PsI always the predominant exopolysaccharide when PsI and Pel are present?" which lead to future research in the field.

Pipe Biofilm Pathogens: Are They just a Nosocomal Risk? Nicholas J. Ashbolt (U.S. Environmental Protection Agency)

Dr. Ashbolt reported some recent development in the understanding of basic questions in pipe biofilm pathogens, in the area of respiratory infections from bacteria that develop within hospital drinking water systems. Although it is commonly regarded that the chance of an individual becoming infected by some reference waterborne pathogen present in the drinking water should be less than in any year, the instantaneous levels of risk to a water consumer vary over the course of a year, and waterborne disease outbreaks have been associated with shorter-duration periods of heightened risk. Performing probabilistic microbial risk assessments is becoming commonplace to capture the impacts of temporal variability on overall infection risk levels. A Bayesian analysis and some probability models are proposed to understand outstanding questions such as how much sporadic legionellosis and other biofilm-related disease occurs from non-institutional drinking water exposures. Great efforts are needed in developing new mathematical models in the field for future investigations.

Role of Microbial Biofilms in Device-related and Other Chronic Diseases

William Costerton (Allegheny-Singer Research Institute)

Dr. Costerton gave an overview about recent developments on the understanding of basic characteristics and effective treatments of biofilms in device-related and other chronic diseases. It is reported that after engineering methods were utilized in the field, which takes into account full complexity and dynamics of biofilms, the understanding of biofilms is driven to a whole new level compared with traditional thinking about biofilms, which is now considered naive. It was recently discovered that the genes of biofilms are unstable and have plasticity, which stimulate interest in the use of inhibiting signals to minimize biofilm formation, and even to stimulate the dissolution of existing biofilms by promoting detachment. The use of physical forces (e.g., DC fields, and ultrasonic waves) to disrupt the internal communications within biofilms is also discussed in the talk.

Communication and Cooperation in Bacterial Populations: Mechanistic and Evolutionary Perspectives Martin Schuster (Oregon State University)

Dr. Schuster reported recent progress made on the understanding of how bacteria communicate and cooperate to perform a variety of multicellular behaviors, including biofilm formation. He and his collaborators proposed an integrated methodology involving mechanistic and evolutionary approaches to investigate communication, also termed quorum sensing (QS), and cooperation in the model bacterium and opportunistic pathogen, Pseudomonas aeruginosa. On a mechanistic level, they utilized a variety of different approaches, including transcriptomics, ChIPchip, and mutagenesis, to identify directly and indirectly regulated genes, and to characterize additional regulators of the QS system. With respect to sociomicrobiology, they utilized in vitro evolution and analysis of natural P. aeruginosa populations to gain insight into the propensity of cheating in bacterial populations, which is a threat common to social systems across all domains of life. They also identified variants that ceased production of shared extracellular factors and took advantage of their production by the group. It is also reported that their evolution-in-a-test-tube experiment revealed a mechanism of cheater control, while the investigation of the underlying mechanism is underway. Comparing and contrasting current mechanistic and sociobiological views on biofilm formation are mentioned as an attempted extension of their current work.

Persister Cells and the Paradox of Relapsing Chronic Infection Kim Lewis (Northeastern University)

Dr. Lewis and his team discovered that pathogens responsible for chronic infections form small populations of dormant cells, called persisters, which are not killed by antibiotics. Antibiotics function only on active cells, so when treatment ceases, persisters grow and repopulate, causing a relapse. The protein identified by the Northeastern researchers, known as TisB, triggers a cellular response that leads to the development of persister cells. In this talk, Dr. Lewis reported recent discoveries on the molecular mechanisms of persister formation. By conducting research using E. coli, Dr. Lewis and his team discovered that while the antibiotic was killing most bacteria, it also induced production of TisB in some bacterial cells, which resulted in the formation of persisters. The role of the molecular mechanisms of persister formation in disease, such as biofilm infections of catheters, cystic fibrosis, and oropharyngeal candidiasis are discussed as well in the talk. It is also observed that isolation of hip mutants, which are reduced in persister formation, may provide insights into effective approaches to eradicate persisters.

#### WEDNESDAY: MARCH 24, 2010

#### Quorum Sensing and Biofilm Modeling Jack Dockery (Montana State University)

Common to all biofilm accumulation models are the following assumptions: 1) Substrate diffuses into a biofilm and is growth limiting; 2) Biofilm feeds on the substrate and expands. He gave a brief history of QS, showing that interest in the subject started in the late 1960s but that only more recently have models been applied to understand QS in biofilms.

Dockery's talk focused on modeling *P. aeuruginosa*. He spoke of two quorum sensing regulatory systems. If one of the systems is knocked out, a biofilm can still form but is structurally different. On a cellular level, autoinducer is produced by each cell and can move through diffusion across the cell membrane. A simple ODE model is used to describe how the molecule gets produced. The initial

model is a system of 8 equations but this system can be simplified in a couple of ways (setting some equations at their quasi steady-state, treat the slow equations as constant relative to the fast equations). These methods reduce the model to a 2D system describing the accumulation of the autoinducer intracellularly and extracellularly. Essentially the model is a simple way of looking at diffusive exchange. The model demonstrated that the coupling of diffusion with positive feedback can lead to hysteresis in the model.



He discussed translating individual-level behavior into population level modeling. This type of multiscale approach is important because it is numerically unfeasible to use only cell-based models because the number of total cells that need to be modeled is very large. There are a number of modeling approaches to accomplish this such as one-dimensional and multi-dimensional continuum as well as cellular automata and individual based models. The second two types of models have nice-looking results that are real-world looking, but the first two are much easier to model. A series of more complicated models were then discussed using the population level approaches described.

Biofilm Growth Models Hermann Eberl (University of Guelph)

Dr. Eberl gave a historical overview of modeling concepts

for biofilm growth, illustrating how modeling techniques for big ecological systems have already been applied to biofilm models. The biofilm growth models he presented all assumed attachment had already been made, and stopped before the detachment phase begins. The models focused on the intermediate growth phase, and one big test of these models is whether they could explain the common "mushroom" formation found in biofilms. Like many ecological, multi-scale and multi-physics systems, biofilms are very complex and the biology included in formulating mathematical models must be selective. One model cannot fit all biological purposes - one can take a primarily ecological or mechanical viewpoint to build a model. Dr. Eberl's focus is on the ecologically based models, making the point that the objective of a modeling study determines the length and time scales of interest.

Dr. Eberl began by presenting one of the two earliest biofilm models, the Wanner-Gujer model from 1986. This 1dimensional model is the classical (wastewater) engineering biofilm model. It has been very successful in the field of environmental engineering, but few mathematical results have been published. Due to its 1-D construction, the model cannot explain heterogeneity in biofilms such as the mushroom structure formation. This prompted a search for new (multi-dimensional) biofilm growth models beginning more than 10 years ago in 1997 that could predict biofilm structure. Many different types of models have been developed, including individual based models and cellular automata (stochastic) models, as well as partial differential equation models (deterministic). These models often feature the Fickian diffusion of dissolved substrates, which are then consumed in the production of new biomass. This "biomass density" is commonly the one variable in the model, incorporating both bacteria and exopolymeric substances (EPS). The models differ in how they describe the spatial distribution of biomass. Eberl spent the remainder of his talk walking through the history and progress for the use of each type of model in describing biofilm growth.

Biofilm dispersion as a Novel Method to Control Chronic Biofilm Infections

David Davies (State University of New York at Binghamton)

Dr. Davies discussed the role of dispersion in a biofilm, which is strictly mediated by the resident organisms. The process can be influenced by fluid dynamics but doesn't originate there. Dispersion is an important process for the survival of the population within the biofilm and there are many mechanisms that allow for escape from the biofilm. In dispersion, free-floating cells within the matrix are released which allows for a reduction in crowding but also allow the organisms to spread to new sites.



David Davies (SUNY at Birmingham)

Most of the research discussed by Davies focused on P. aeruginosa biofilm and an inducer (cis-Decenoic Acid), which was identified as playing an important role in dispersion. Once this molecule was identified, a variety of experiments were done to understand the dispersion response of cells to increasing concentrations of the inducer molecule. He also looked at P. aeruginosa dispersion with different fatty acids to see if there were other molecules that could act as an inducer. The original molecule discovered appeared to have the most dramatic effect on dispersal. Davis' current understanding of the dispersion response mechanism is that within a biofilm, P. aeruginosa produces this inducer. In addition, it is known that protein synthesis is required for biofilm dispersion induced with cis-Decenoic Acid. Further research determined that cis-Decenoic Acid itself is probably not a dispersion inducer. Instead, it's responsible for inducing a transition in the cells from a sessile to a disseminating mode of existence. For example, it can assist in changing P. aeruginosa from existing in a chronic mode, to existing in an infectious mode.

The last area of focus was the role that cis-Decenoic Acid might play in biofilm control/management. He looked at the role that this molecule potentially played on other strains of common biofilm-forming bacteria other than *P. aeruginosa*. Of the 7 other strains looked at, all showed dispersion induced by cis-Decenoic Acid. This was true for both Gram-positive and Gram-negative bacteria as well as yeast.

Biofilm Formation in Vibrio cholerae Fitnat Yildiz (UC, Santa Cruz)

Vibrio cholerae, a natural inhabitant of aquatic ecosystems, is a facultative human pathogen and the causative agent of cholera, a disease responsible for killing 120,000



Bill Costerton (Allegheny-Singer Research Institute)

people per year and infecting many more. The formation of matrix-enclosed biofilms enhances the environmental survival, transmission and infectivity of V. cholerae. V. cholerae produces both smooth and rugose colony variants differing in biofilm-forming capacities and motility. The rugose form is wrinkled with decreased motility and has an increased ability to form biofilms. There is also increased vibrio-polysaccharide matrix production in the rugose variant. In this talk, Dr. Yildiz identified and characterized many structural components involved in biofilm formation and regulation in V. cholerae.

A Novel Signaling Network Essential for Regulating Pseudomonas aeruginosa Biofilm Development Karin Sauer (Binghamton University)

Dr. Sauer explored the regulatory events behind the formation of biofilms by *P. aeruginosa*. During the initial 8 hours of growth, there is a reversible attachment - *P. aeruginosa* can attach to a surface and leave. After 24 hours, bacteria have cemented themselves on a surface and begin to form clusters. This irreversible attachment is followed by 2 maturation stages, and a dispersion stage after 9 days. The biofilm exhibits phenotypic changes over time, with the sequential activation of quorum sensing regulons and the temporal and sequential production of proteins. Dr. Sauer's talk focused on the regulation of biofilm growth, and whether it is dependent on intrinsic

factors in addition to extrinsic factors such as nutrient availability and system hydrodynamics. If so, this would imply that biofilm formation is a developmental process and a regulatory cascade would be present with a hierarchical activation of regulatory proteins.

Dr. Sauer concluded that genetic pathways are hierarchically ordered, with hierarchical divisions serving as checkpoints enforcing the coordination and directionality of development, as the stage specific arrest of biofilm formation is possible. Also, the developmental pathways have evolved as dedicated systems for the regulation of biofilm formation, independent of nutritional conditions.

#### THURSDAY: MARCH 25, 2010

The Freter Model of Biofilm Formation Hal Smith (Arizona State University)

The work in this talk by Dr. Smith was inspired by the models of Freter, a microbiologist at the University of Michigan who was interested in colonization resistance in the gut. This talk switched gears from physical and growth processes inside biofilms to larger scale ecological interactions and population dynamics issues.

Dr. Smith first presented the Freter chemostat model, which described the colonization of a surface by a microbial population in a fluid environment. The microbial population consists of planktonic cells in the fluid and adherent cells on the surface. Dr. Smith proved that bacteria survive in the chemostat if their growth rate exceeds the dilution rate, or if the biofilm wall cell growth exceeds the dilution rate of the chemostat, i.e. the wall cells grow fast enough to exceed sloughing.

Next, he looked at a flow reactor with biofilm growth. He explored colonization resistance in the gut and looked at what happens when resident microflora are challenged by invading pathogens. The multiple species are now competing for sites, and the resident species occludes the invader if they are already colonizing most of the available wall sites. Two species can both survive, spatially segregated, if one specializes in growth and the other in attachment. Smith also presented a chemostat model of gene transfer in biofilms and was able to address the issue of phage therapy for a biofilm. Simulations suggest a phase can do serious damage to a biofilm.

Oral Biofilms: Biology to Models (and Back Again?) Robert Palmer (Natl. Inst. Dental Craniofacial Research -NIH)

According to Palmer, the oral environment is complex. There are many locations in the mouth where biofilms can form and one can think of the salivary proteins in the mouth as a conditioning film to which bacteria can adhere to form a biofilm. Saliva comes from 3 sets of glands, which secrete into different areas in the oral cavity and the saliva produced in each region is compositionally different. The rate of flow of saliva can vary with food intake, breathing, talking, etc. This means that the various regions where biofilms can form in the mouth can experience very different conditions for bacterial growth.



Palmer discussed the fact that within the oral cavity there is a large diversity of bacterial species that can persist in the mouth. There are about 52 different species existing on the tooth surface, 347 in the sub-gingival plaque and all together 700 in the entire oral cavity. This is compared to only about 182 on the skin and 395 in the intestine. The complexity of the oral environment is constantly altered by the host, based on what the host is consuming.

Several studies have looked at the transition between health and disease. For one experiment, patients were clustered into profiles based on their oral community. In looking at efficacy of therapies for different diseases within these groups, the group of individuals with potentially the most pathogenic community of bacteria responded best to therapy. One can think of the individual as an ecosystem where subjects respond differently to the same periodontal therapy depending not just on their ability to cope with the infection, but also on the initial state of the microbial community in their oral cavity.

The last topic discussed was co-aggregation of cells. Several studies have shown the importance of co-aggregation in structuring bacterial communities. Co-aggregation was defined by Palmer as cell-to-cell recognition/binding that brings different species of bacteria into direct contact. These interactions can be very specific and are thought to have arisen through co-evolution. The key point is that spatiotemporal interactions are highly important in the structuring of bacterial communities within the oral cavity.

Candida Biofilm: Journey from the Bench to the Patients Mahmoud Ghannoum (University Hospitals Case Medical Center)

Dr. Ghannoum focused on fungi in a biofilm. According to Ghannoum, any substrate you put inside a patient has the potential for biofilm formation. For example, infections from catheters are often a result of biofilm formation. When catheters are inserted into the bloodstream, organisms can come from various sources (e.g., skin). This could lead to quite serious diseases for the patient such as Candida, which has a 40% mortality rate. An important aspect of Dr. Ghannoum's research was to optimize a biofilm model to understand and quantify the optimal conditions for biofilm growth. Experiments were also conducted to investigate the structure and development of biofilms. The comparison of the model to *in vivo* conditions of biofilm development in Candida was quite similar.

Biofilms are notoriously difficult to treat. This is because they are tolerant to antimicrobials. This is true for a large number of antimicrobials such as biocides and antibiotics. The question is whether there is an antifungal that can work against biofilm. A few antifungals were found to be effective. The next step was to look at what the mechanisms of antifungal resistance were. Within the cell membrane of fungi is a sterol component called ergosterol (similar to human cholesterol), which is a major target for antifungals. At the mature phase, sterols play an important role in resistance.

One problem remains for eradicating biofilms in catheters, namely biofilms are caused not only by Candida, but also by bacteria. Therefore, treatment needs to include a broad-spectrum antimicrobial to address both problems. Through experiments, a solution (B-Lock) was found that is effective for removal of both fungi and bacteria in catheters.

# **EDUCATION PROGRAMS**

#### SUMMER PROGRAM IN MATHEMATICAL BIOLOGY FOR UNDERGRADUATES (JUNE 21 - JULY 2, 2010)

The summer of 2010 marked the MBI's fifth annual Summer Program for Undergraduates that includes a twoweek active survey of mathematical biology followed by a six-week Research Experience for Undergraduates (REU) program.

The first week of the program involved tutorials and hands-on computer labs in mathematical bioscience topics. The first day saw Dennis Pearl presenting key issues in statistical phylogenetics - aligning molecular sequences and inferring evolutionary trees. In the afternoon, David Gerard led a computer lab, giving students a chance to try out the ClustalX alignment program along with PAUP and MrBayes phylogenetics software. On Tuesday, Victor Jun lead a morning tutorial covering selected topics in bioinformatics such as using databases, and analyzing and visualizing microarray and ChIP data. That afternoon, Brian Kennedy guided the students in trying out online bioinformatics software. Wednesday saw Joe Verducci presenting issues in the quantitative analysis of chemogenomic data leading to an introduction to the SCOOP method (Shrunken Centroid Ordering by Orthogonal Projections) for selecting differentially expressed genes, while Yushi Liu supervised the afternoon computer lab using the R package and the BioCoductor program. Kate Calder presented a lively tutorial on statistical analysis of environmental data the following day and Jenny Brynjarsdottir led the afternoon computer lab using R. The week concluded with Joe Tien's tutorial on the principles of mathematical epidemiology focusing on the basic "SIR" (Susceptible/Infected/Recovered) framework and important modifications. MBI Postdoctoral Fellow, Suzanne Robertson, assisted by Jeff Dunworth, organized the afternoon computer lab that gave participants experience with the XPP and MatLab programs.

Dividing into teams, the students were given a chance to study a real problem in their chosen topic area during the second week. The two-week survey concluded with each of five teams participating in a mini-conference, making both poster and oral presentations on their projects. The mathematical epidemiology team (Kossi Folke, Sam Kank-



Tour of the Illumina Next Generation Sequencing Lab.

an, Youngmin Park, and Yi Zeng) presented their studies of disease dynamics comparing a simple SIR model to one that allows disease transmission to be dose dependent. The phylogenetics project team (Spencer Carran, Natalie Hood, Tanya Singh, and George Zhang) presented an analysis of the evolution of flightless birds and testing whether their divergence occurred before or after the separation of the continents. Next, the Environmental Statistics group (Xinru Cai, Jennifer McVearry, Nghi Le Phuong Nguyn, and Sujana Rajkarnikar) described their study of the impact of the Exxon Valdez oil spill on harbor seal populations. The bioinformatics project, presented by Rui Cao, Krupa Harishankar, Hongyang Pi, and Nathan Rapport, explored different methods for microarray gene expression normalization on TGFb mediated genes in ovarian cancer cells. Finally, the chemogenomics team of Clark Butler, Kate Groskreutz, Orrin Shindell, and Nicole Thurmond, examined the use of the SCOOP method to find genes associated with the early detection of breast cancer. The collaborative nature of all of these efforts was illustrated as each student presented a substantial part of their group's work.

During this two-week program, the students also toured labs that use quantitative methods in the biological and medical sciences. This included a tour of the Illumina Next Generation Sequencing Lab where Pearlly guided the students through the computationally intensive work of



All of the participating students in the two-week program.

the lab. Angelika Nelson gave the group a tour of Ohio State's Museum of Biological Diversity with its major acarology and plant (more that a half million specimens each), insect (over 3.5 million specimens), fish (1.5 million specimens), and mollusk (150,000 specimens) collections that are available for both teaching and research. As part of this tour, Joe Cora demonstrated the museum's insect bioinformatics system. In the final tour, the students paid a visit to Libby Marschall's Aquatic Ecology Laboratory where her team of graduate students showed off their work in studies of fish populations in Lake Erie.

At the conclusion of the two-week program, the REU component of the summer program then sponsored five students to spend six weeks going into much more depth in a research project in their chosen area. Yi Zheng's project involved both a theoretical and empirical analysis of a mathematical epidemiology model of waterborne pathogens that included terms for multiple transmission pathways and non-linear incidence. Clark Butler worked on combining SCCOP and Support Vector Machine methodologies for the early detection of breast cancers using only expression data from a non-invasive peripheral

blood assay. Tanya Singh worked on novel diagnostics for the convergence of Markov chain Monte Carlo runs in Bayesian phylogenetics and applied them to a study of the evolution of 78 strains of the Human Pappiloma Virus and their relationship with cervical cancer. Xinru Cai worked on an environmental statistical analysis of air pollution in Dallas County, Texas - focusing on spatial effects and the association with heart disease. Finally, Hongyang Pi described her bioinformatics project using information from both the JASPER and TRANSFAC transcription factor data bases to compare different techniques for normalizing and summarizing array data and making detection calls.

All of the students taking part in the MBI undergraduate summer program were exposed to new areas of scholarship and appeared to gain an increased appreciation for the mathematical biosciences. The PowerPoint presentations from both the tutorials and mini-conferences are viewable on the MBI web site at http://www.mbi.osu. edu/eduprograms/undergrad2010.html.

# COLLOQUIA & SEMINARS

### **MBI COLLOQUIUM**

**Avner Friedman, Mathematics, Ohio State University** Mathematical Models May Lead to Better Treatment of Chronic Wounds (October 12, 2009)

Clay Marsh, Vice Chair for Research, Interim Director, Pulmonary and Critical Care, Internal Medicine, College of Medicine, OSU

Exploring Systems Approaches to Understand Molecular Networks in Health and Disease (October 19, 2009)

Noel Cressie, Program in Spatial Statistics and Environmental Statistics, Statistics, Ohio State University Biometry, R.A. Fisher, and Statistical Science (October 26, 2009)

**Tina Henkin, Microbiology, Ohio State University** *RiboswitchRNAs: Sensing Metabolic Signals withRNA Transcripts* (November 16, 2009)

## Peter Abrams, Ecology & Evolutionary Biology, University of Toronto

DensityDependenceandtheStructureofEcologicalTheory (November 23, 2009)

Larry Schlesinger, Samuel Saslaw Professor of Medicine, Director, Division of Infectious Diseases and the Center for Microbial, Interface Biology, Director, Medical Scientist program, Ohio State University

Unraveling the Molecular Events in the Lung Innate Immune Response to Tuberculosis (January 4, 2010)

## Jian-Qiu Wu, Molecular Genetics, Ohio State University

Molecular Mechanism of Contractile-Ring Assembly in Cytokinesis (January 11, 2010)

#### Cynthia Kenyon, Biophysics and Biochemistry, University of California, San Francisco

Genes and Cells that Influence the Rate of Aging in C. elegans (January 20, 2010)



Wolfgang Sadee (Ohio State University)

De Witt Sumners, Mathematics, Florida State University

DNA Topology (February 1, 2010)

#### Joyce R. Mclaughlin, Ford Foundation Professor of Mathematics, Inverse Problems Center (IPRPI) Director, Rensselaer Polytechnic Institute

Biomechanical Imaging: Viscoelastic Models, Algorithms, Reconstructions; Application to Breast, Prostate and Brain (February 8, 2010)

#### Zena Werb, Anatomy, University of California, San Francisco

Of Mice and Women: How Studying Development Gives Us Insights into Cancer (February 15, 2010)

## Rafael Irizarry, Johns Hopkins Bloomberg School of Public Health

Stochastic Epigenetic Variation in Evolutionary Adaptation and Common Disease (March 1, 2010)

## Michael Waterman, Biological Sciences and Mathematics, University of Southern California

Reading DNA Sequences Along Eulerian Paths (March 15, 2010)

#### David Terman, Mathematics, Ohio State University

Does Neuroscience Need Mathematics? And Vice-versa? (April 5, 2010)

#### Bin Yu, Statistics, University of California, Berkeley

Sparse Modeling: Unifying Theory and Human Visual Pathway (April 19, 2010)

## Wolfgang Sadee, Pharmacology, Ohio State University

Genetics of Human Phenotypic Variability - Searching for Disease Risk Factors along Evolutionary Paths (April 26, 2010)

#### Chris Johnson, Scientific Computing and Imaging Institute, University of Utah

Image-Based Biomedical Modeling, Simulation and Visualization (May 3, 2010)

#### Lazlo Szekely, Mathematics, University of South Carolina

The amount of Information Needed for Phylogeny Reconstruction (May 10, 2010)

### **POSTDOC SEMINARS**

#### Julia Chifman, MBI, Ohio State University

Phylogenetic Invariants (September 24, 2009)

#### Rebecca J. Tien, MBI, Ohio State University

Modeling Predator-prey Coevolution with Variable Cost of Prey Defense (October 1, 2009)

#### Suzanne Robertson, MBI, Ohio State University

Spatial Patterns in Stage-Structured Populations with Density Dependent Dispersal (October 15, 2009)

#### Judy Day, MBI, Ohio State University

Analysis of Transient Dynamics Motivated by a Mathematical Model of the Inflammatory Response (October 22, 2009)

#### Marisa Eisenberg, MBI, Ohio State University

Modeling Feedback Regulation in the Human Hypothalamic-Pituitary-Thyroid Axis (October 29, 2009)

#### Shu Dai, MBI, Ohio State University

Dynamics of an Amplitude Equation for Cardiac Alternans in One Dimension (November 12, 2009)

### Yunjiao Wang, MBI, Ohio State University

Oscillations in NFxB Signaling Pathway (November 19, 2009) **Kun Zhao, MBI, Ohio State University** 2D Boussinesq Equations with Partial Viscosity (December 3, 2009)

**Chirove Faraimunashe, MBI, Ohio State University** Models Incorporating HIV Infection, Treatment and Viral Mutations (December 7, 2009)

#### Najat Ziyadi, MBI, Ohio State University

A Model of Drug Resistance with Infection by Health Care Workers (December 10, 2009)

**Erik Bloomquist, MBI, Ohio State University** Hierarchical Models in Molecular Evolution (January 14, 2010)

#### Sam Handelman, MBI, Ohio State University Classes of Reciprocal Sequence Homologs (January 21, 2010)



#### Rasmus Hovmoller, MBI, Ohio State University

Phylogenetic Analysis and Phylogeography of H5N1 and H1N1 Influenzas (February 3, 2010)

#### Chuan Xue, MBI, Ohio State University

Modeling Ischemic Cutaneous Wounds (February 11, 2010)

#### Ahmet Ay, Mathematics and Quantitative Biology, Michigan State University

Deciphering a Transcriptional Regulatory Code: Modeling Short- range Repression in the Drosophila Embryo (February 17, 2010)

# COLLOQUIA & SEMINARS



Uma Abbas (Clevelan Clinic Foundation)

#### Harsh Jain, MBI, Ohio State University

The Molecular Basis of Synergism between Carboplatin and ABT-737 in Ovarian Carcinomas (February 18, 2010)

## Samir Ghadiali, Biomedical Engineering, Ohio State University

Influence of Cytoskeletal Mechanics and Networks on Lung Injury, Inflammation and Repair (March 4, 2010)

#### Kevin Passino, Electrical & Computer Engineering, Ohio State University

Modeling and Analysis of Swarms (April 1, 2010)

**Tao Shi, Statistics, Ohio State University** Multi-Sample Data Spectroscopic Clustering of Large Datasets using Nystrom Extension (April 8, 2010)

#### Chenglong Li, Medicinal Chemistry and Pharmacognosy, Ohio State University

Multiple Ligand Simultaneous Docking (MLSD): Orchestrated Dancing of Ligands in Binding Sites of Protein (April 22, 2010) Shili Lin, Statistics, Ohio State University Space Oriented Rank-based Data Integration (April 29, 2010)

**Ching-Shan Chou, Mathematics, Ohio State University** Systems Biology of Cell Signaling -- Spatial Dynamics (May 6, 2010)

#### **Mark Foster, Biochemistry, Ohio State University** Structure, Dynamics and Oligomerization in Ligand-Mediated Regulation of Gene Expression (May 13, 2010)

**Dan Siegal-Gaskins, MBI, Ohio State University** Emergence of Switch-like Behavior in a Large Family of Simple Biochemical Networks (May 20, 2010)

#### Joe Tien, Mathematics, Ohio State University Modeling Waterborne Diseases (May 27, 2010)

#### Joe Verducci, Statistics, Ohio State University

Exploiting Covariance Structure to Aid in the Discovery of Differentially Expressed Genes (June 3, 2010)

#### Huseyin Coskun, MBI, Ohio State University

Mathematical Models for Single Cell Motion and Model Based Inverse Problems (June 10, 2010)

#### Deena Schmidt, MBI, Ohio State University

Network Structure and Dynamics of Sleep-wake Regulation (June 17, 2010)

### **SPECIAL SEMINARS**

#### Linda Petzold, Mechanical and Environmental Engineering, Computer Science, University of California, Santa Barbara

Discrete Stochastic Simulation of Spatially Inhomogeneous Biochemical Systems (September 4, 2009)

#### Sze-Bi Hsu, National Tsing-Hua University

On a system of Reaction-diffusion Equations Arising from Competition of Phytoplankton Species (September 21, 2009)

Ume Abbas, Cleveland Clinic Foundation, Infectious Disease

Modeling the Potential Impact of ARV-based PrEP on HIV Prevention and Drug Resistance (December 11, 2009)

#### Christian Laing, New York University, Courant Institute of Mathematical Sciences

A Geometric Analysis to RNA Structure and Tertiary Mo-

tifs (February 2, 2010)

**Cecilia Diniz Behn, University of Michigan** Networks, Nullclines, and Narcolepsy: A Mathematical Investigation of Orexin Neurons (February 9, 2010)

Adriana Dawes, University of Alberta Maintenance of Protein Asymmetry in the Early Embryo of the Nematode Worm C. elegans (February 11, 2010)

# **PUBLIC LECTURE**

### IAIN COUZIN, DEPARTMENT OF ECOLOGY AND EVOLUTIONARY BIOLOGY, PRINCETON UNIVERSITY APRIL 28, 2010

lain is an Assistant Professor in the Department of Ecology and Evolutionary Biology and Adjunct Faculty in the Program of Applied and Computational Mathematics at Princeton University. Previously he was a Royal Society University Research Fellow in the Department of Zoology and Junior Research Fellow in the Sciences at Balliol College, University of Oxford. He did his Ph.D. at the University of Bath, UK. His work aims to reveal the fundamental principles that underlie evolved collective behavior and consequently his research includes the study of a wide range of systems from cellular aggregates to insect swarms, fish schools and human crowds. He is a member of the Faculty of 1000 Biology and in recognition of his research he was a recipient of a Searle Scholar Award in 2008 and the Mohammed Dahleh Award in 2009.

#### **The Perfect Swarm**

Collective organization is everywhere, both around us and within us. Our brains are composed of billions of interconnected cells communicating with chemical and electrical signals. We ourselves are integrated in our own collective - our human society. Elsewhere in the natural world hundreds of thousands of blind army ants coordinate a massive raid across the rainforest floor, a flock of birds arcs and ripples while descending to roost and a fish school convulses, as if one entity, when attacked by a predator. How can animal groups move in unison? How does individual behavior produce group dynamics? How do animal groups make informed unanimous decisions? From locust swarms to bird flocks, from consensus decision-making in fish to that among humans, I will discuss how, and why, coordinated collective behavior is so pervasive within the natural world.



lain Couzin

# FUTURE PROGRAMS

Evolution, Synchronization, and Environmental Interactions: Insights from Plants and Insects August 2010 - July 2011

#### **Organizing Committee**

- Vincent Gutschick (Biology, New Mexico State)
- Daniel Forger (Mathematics, Michigan)
- Mark Lewis (Biological Sciences, Alberta)
- Scott Nuismer (Biological Sciences, Idaho)
- David Rand (Warwick Systems Biology Centre)
- David Sumpter (Mathematics, Uppsala)

Myriad influences shape the patterns of evolution, timing, behavior and ecology of living organisms. These influences range from biochemical cues to configurations of temperature, space and light, to interactions with other organisms. This one-year program focuses on connecting influence to pattern for processes involving plants and insects.



Venus Fly Trap.

How do biotic and abiotic influences affect patterns of plants and insects? We investigate this complex question quantitatively, by focusing on specific areas where there has been recent growth, simultaneously in mathematical and statistical theories and in biological data and experiment. We propose to couple the mathematics and biology in new ways, allowing for innovative growth of both science and mathematics.

The year is based around the following workshops: (i) Mathematical modeling of plant development, (ii) Circadian clocks in plants and fungi, (iii) Insect self-organization and swarming, (iv) Ecology and control of invasive species, including insects, and (v) Coevolution and the ecological structure of plant-insect communities. Our mathematical investigation of these processes will rely upon a diverse array of quantitative theory, including geometry, control, optimization, pattern formation, spatial dynamics, evolution and data-model interaction.

The plant development workshop will connect biochemical mechanisms to geometric patterns, while simultaneously investigating the selection pressure for the geometric patterns. Circadian clocks will be evaluated both from the perspective of design features for feedback and control, and of robustness of these features to perturbation. Insect self-organization and swarming will employ dual perspectives of emergent self-organization properties arising from individual interactions, and optimal design of artificial swarms using diffuse (decentralized) information with implications for robotics and decentralized computer algorithms. Biological invasions will be understood, not only in terms of predictable forecasting of future invasions, but in terms of optimal control of the invasion processes. Finally, the physical and behavioral mechanisms involved in coevolution of plant-insect communities will be understood in terms of fitness advantages incurred evolution and adaptation.

Thus the underlying feature throughout the workshops is simultaneous investigation of mechanism and optimality: What mechanisms give rise to observed patterns? What is the fitness or optimality associated with observed patterns? It is through this simultaneous study of mechanism and optimality in plants and insects that the workshops will provide general insight to the processes of evolution, synchronization and environmental interactions.

The goals of the year program are (i) to develop, analyze and apply new mathematical models for processes of evolution, timing, behavior and ecology of living organisms that are tailored to investigate both mechanisms underlying the processes and optimality of associated patterns; and (ii) train interdisciplinary quantitative researchers at a variety of levels (graduate, postdoctoral and faculty) in the area of evolution, synchronization and environmental interactions for biological systems.

### Workshops

- Workshop 1: Mathematical Modeling of Plant Development (September 27 - October 1, 2010)
- Current Topic Workshop: Bootcamp on Cancer Modeling (September 7-10, 2010)
- Workshop 2: Circadian Clocks in Plants and Fungi (October 25-29, 2010)
- Current Topic Workshop: Blackwell-Tapia Conference (November 5-6, 2010)
- Workshop 3: Ecology and Control of Invasive Species, Including Insects (February 21-25, 2011)
- Workshop 4: Insect Self-organization and Swarming (March 14-18, 2011)
- Current Topic Workshop: New Developments in Dynamical Systems Arising from the Biosciences (March 22-26, 2011)
- Workshop 5: Coevolution and the Ecological Structure of Plant-insect Communities (April 4-8, 2011)
- Current Topic Workshop: Modeling and Computation of Biomolecular Stucture and Dynamics (April 25-29, 2011)
- Workshop 6: Ocean Ecologies and their Physical Habitats in a Changing Climate (June 20-July 1, 2011)

### Stochastics in Biological Systems August 2011 - July 2012

### **Organizing Committee**

- Linda Allen (Math & Stats, Texas Tech)
- Richard Durrett (Mathematics, Cornell)
- **Timothy Elston** (Pharmacology, Georgia Tech)
- Thomas Kurtz (Math & Stats, Wisconsin-Madison)
- **Reinhard Laubenbacher** (Mathematics, Virginia Tech)

Stochasticity is fundamental to biological systems. While in many situations the system can be viewed as a large number of similar agents interacting in a homogeneously mixing environment so the dynamics are captured well by ordinary differential equations or other determinis-



Normal and mutant Pseudomonas (Credit: Hemantha Kulasekara)

tic models. In many more situations, the system can be driven by a small number of agents or strongly influenced by an environment fluctuating in space or time. Stochastic fluctuations are critical in the initial stages of an epidemic; a small number of molecules may determine the direction of cellular processes; changing climate may alter the balance among competing populations. Spatial models may be required when agents are distributed in space and interactions between agents form a network. Systems evolve to become more robust or co-evolve in response to competitive or host-pathogen interactions. Consequently, models must allow agents to change and interact in complex ways. Stochasticity increases the complexity of models in some ways, but may smooth and simplify in others.

#### Workshops

- Workshop 1: Issues in Probability Theory Arising from Biology (September 12-16, 2011)
- Workshop 2: Intracellular Networks (TBA)
- Workshop 3: Robustness in Biological Systems (February 6-10, 2012)
- Workshop 4: Evolution and Spread of Disease (March 19-23, 2012)
- Workshop 5: Spatial Models of Micro and Macro Systems (April 16-20, 2012)
- Workshop 6: Algebraic Methods in Evolutionary and Systems Biology (May 7-11, 2012)

# POSTDOCTORAL FELLOW PUBLICATIONS

Bloomquist EW, Suchard MA. "Bayesian Phylogenetics." In Frontiers of Statistical Decision Making and Bayesian Analysis. Edited by Chen MH, Dey DK, Muller P, Sun D, Keying Y. pages 316-324. Springer, NY 2010.

Bloomquist EW, Lemey P, Suchard MA. Three Roads Diverged? Routes To Phylogeographic Inference. Trends in Ecology & Evolution 25 (2010), pp. 626-632.

Kubatko LS, Gibbs HL, Bloomquist EW. Inferring Species-Level Phylogenies and Taxonomic Distinctiveness Using Multi-Locus Data in Sistrurus Rattlesnakes. In Press. Systematic Biology. 2010.

Dong Li, Tong Li and Kun Zhao, On a hyperbolic-parabolic system modeling repulsive chemotaxis. Mathematical Models and Methods in Applied Sciences (2010), in press.

Tong Li and Kun Zhao, On a quasilinear hyperbolic system in blood flow modeling. Discrete and Continuous Dynamical Systems-B (2010), in press.

Kun Zhao, 2D inviscid heat conductive Boussinesq system in a bounded domain. Michigan Mathematical Journal 59 (2010), no. 2, 329-352.

Ming-Jun Lai, Ronghua Pan and Kun Zhao, Initial boundary value problem for 2D viscous Boussinesq equations. Archive for Rational Mechanics and Analysis (2010), in press.

Kun Zhao, Global regularity for a coupled Cahn-Hilliard-Boussinesq system on bounded domains. Quarterly of Applied Mathematics (2010), in press.

Kun Zhao, On the isothermal compressible Euler equations with frictional damping. Communications in Mathematical Analysis 9 (2010), no. 2, 77-97.

Tong Li and Kun Zhao, Global existence and long time behavior of entropy weak solutions to a quasilinear hyperbolic blood flow model. (2010), submitted to Network and Heterogeneous Media.

Kun Zhao, Large time behavior for Cahn-Hilliard-Boussin-

esq equations on bounded domains. (2010), submitted to Electronic Journal of Differential Equations.

Dong Li, Yuan Lou and Kun Zhao, Blow up phenomena for reaction-cross diffusion models in population dynamics. (2010), in preparation.

John Lowengrub, Edriss Titi and Kun Zhao, Mathematical analysis of the Cahn-Hilliard-Hele-Shaw system. (2010), in preparation.

Tong Li, Ronghua Pan and Kun Zhao, Global existence and long time behavior for a chemotaxis model on bounded domains with large data. (2010), in preparation.

Dong Li and Kun Zhao, Quantitative decay of a one-dimensional hyperbolic-parabolic chemotaxis model with large data. (2010), in preparation.

Siegal-Gaskins D, Mejia-Guerra MK, Smith GD, and Grotewold E. 2010. "Emergence of switch- like behavior in a large family of simple biochemical networks". Under review.

Fiebig A, Castro Rojas CM, Siegal-Gaskins D, and Crosson S. 2010. "Interaction specificity and toxicity in a paralogous set of ParE/RelE-family toxin-antitoxin systems". Mol Microbiol 77: 236-251.

Siegal-Gaskins D, Grotewold E, and Smith GD. 2009. "The capacity for multistability in small gene regulatory networks". BMC Syst Biol 3:96.

Siegal-Gaskins D, Ash JN, and Crosson S. 2009. "Modelbased Deconvolution of Cell Cycle Time-series Data Reveals Gene Expression Details at High Resolution". PLoS Comput Biol 5 (8): e1000460.

Schmidt, D., Best, J., and Blumberg, M.S. Random graph and stochastic process contributions to network dynamics.

Day, J., Rubin, J., Chow, C.C., 2009. Competition between transients in the rate of approach to a fixed point. SIAM J. Appl. Dyn. Syst. 8(4), 1523-1563. (Nov 2009)

Day, J., Schlesinger, L.S., Friedman, A., 2010. Tuberculosis research: Going forward with a powerful "Translational Systems Biology" approach. Tuberculosis, (Edinb), 90(1):7-8. (Jan 2010)