

BCATS 2008 Abstract Book

Welcome

*to the ninth annual symposium on
Biomedical Computation at Stanford (BCATS).*

This student-run one-day symposium provides an open, interdisciplinary forum for students and post-docs to share their latest work in computational biology and medicine with others from Stanford and beyond. Since its inception in 1999, BCATS has seen growth and change in the field of biomedical computation. This year's schedule features diverse, cutting-edge research from the largest applicant pool ever.

We thank our keynote speakers, student presenters, judges, sponsors, and all 2008 attendees.

The BCATS 2008 organizing committee

Tiffany Chen, Biomedical Informatics

Adam Grossman, Bioengineering

Xuhui Huang, SimBios

Hedi Razavi, Bioengineering

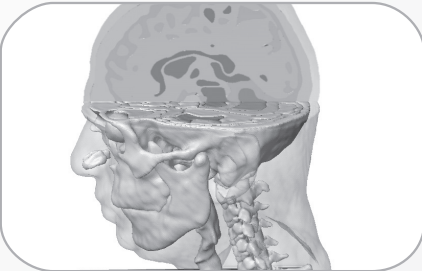
Jesse Rodriguez, Biomedical Informatics

Katherine Steele, Mechanical Engineering

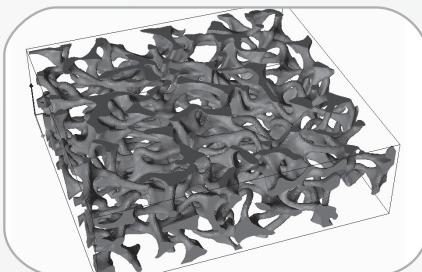
Rebecca Taylor, Mechanical Engineering

simpleware

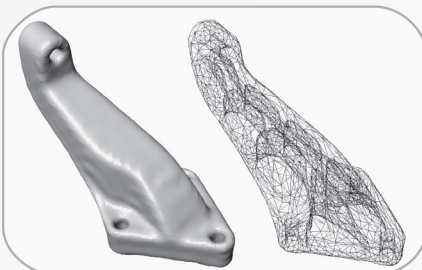
converting 3D images into numerical models



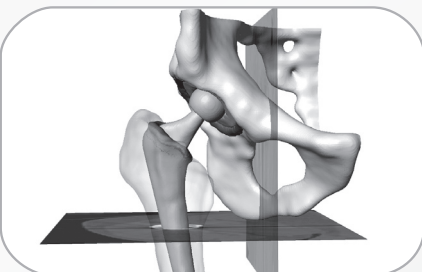
Biomechanics



Biomaterials Research



Reverse Engineering



Implant Design

From 3D scan to model... *the Simple Way*

New Version of Simpleware out now!

Simpleware offers the world's most advanced software solution for the conversion of 3D scan data (e.g. MRI, CT, MicroCT, etc.) into high-quality computer models for CAD, FEA/CFD, and Rapid Prototyping.

Simpleware provides a complete image processing package that will help you significantly accelerate the process of segmenting and meshing 3D data.

Simpleware exports full volumetric and surface meshes, including contact surfaces and material properties directly to:

- ABAQUS
- ANSYS
- COMSOL
- FLUENT
- I-DEAS
- LS-DYNA
- MSC.PATRAN
- STL and CAD

Simpleware also allows the import and interactive positioning of CAD files into the image data.

For more information and a **2-month free trial version** of our new release contact us at info@simpleware.com quoting *BCATS* or browse our website:

www.simpleware.com

Simpleware Ltd., Innovation Centre, Rennes Drive, Exeter, Devon, EX4 4RN, UK
Tel: +44 (0)845 508 7240 • Fax: +44 (0)1392 247949 • Email: info@simpleware.com

Interested in how biocomputation is changing biology and medicine?

Sign up for a free subscription at:
www.BiomedicalComputationReview.org



Want to develop, share or find biosimulation software or data?

Explore the biosimulation repository and development environment at: www.simtk.org

Looking for high performance tools for solving equations?

Download the open-source SimTK Core libraries at: www.simtk.org/home/simtkcore



Interested in collaborating on important computational biological problems?

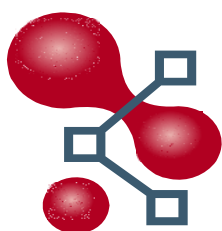
Visit us at: <http://simbios.stanford.edu>

Stop by our table at the industry reception

See how much we can speed up molecular dynamics via GPUs

Try OpenSim, our freely available musculoskeletal simulation software

See a coarse-grained molecular simulation using SimTK Core – and more



STANFORD CENTER FOR BIOMEDICAL INFORMATICS RESEARCH

Stanford's Center for Biomedical Informatics Research (BMIR) hosts prominent biological and medical researchers, computer scientists, physicians and students who develop new ways of acquiring, processing and analyzing biomedical and health data. We are committed to providing tools for a revolution in health care and biomedical research, ensuring the biomedical community is properly equipped for the information age. For more information on BMIR, visit <http://bmir.stanford.edu>.



The Biomedical Informatics Training Program aims to train the next generation of informatics researchers by educating students in clinical, public health, and bioinformatics. <http://bmi.stanford.edu>.

BMIR is host to the National Center for Biomedical

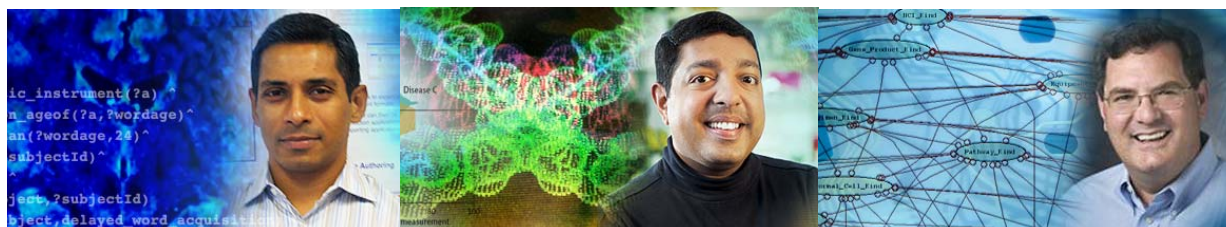


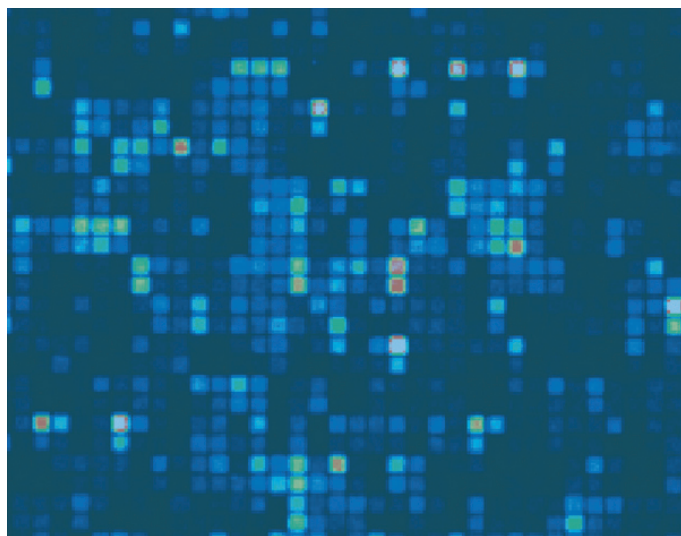
THE NATIONAL CENTER FOR
BIOMEDICAL ONTOLOGY

Ontology, one of the NIH's

seven National Centers for Biomedical Computing. The NCBO develops technologies that enable computer processing of biomedical information. Tools such as BioPortal, a Web-based application for accessing and sharing biomedical ontologies, give researchers new and powerful ways to describe, store and explore data.

<http://www.bioontology.org>





BECAUSE

Our CAUSE is Jennifer and her cancer.


For more than 30 years, Genentech has been at the forefront of the biotechnology industry, using human genetic information to develop novel medicines for serious and life-threatening diseases. Today, Genentech is among the world's leading biotech companies, with multiple therapies on the market for cancer and other unmet medical needs.

Our founders believed that hiring talented, enthusiastic people would make Genentech a success. Today, we still believe our employees are our most important asset.

Genentech's research organization features world-renowned scientists who are some of the most prolific in their fields and in the industry. Our more than 1,000 scientists and postdocs have consistently published important papers in prestigious journals and have secured approximately 6,900 patents worldwide (with another 6,000 pending). Genentech's research organization combines the best of the academic and corporate worlds, allowing researchers not only to pursue important scientific questions, but also to watch an idea move from the laboratory into development and out into the clinic. We are proud of our long history of groundbreaking science leading to first-in-class therapies, and we hope you will consider us as we continue the tradition.

Genentech is a proud sponsor of the 2008 Biomedical Computation at Stanford (BCATS) Conference & Symposium.

To learn more about our current opportunities, please visit careers.gene.com. Genentech is an equal opportunity employer.

 In January 2008, Genentech was named to FORTUNE's list of the "100 Best Companies to Work For" for the tenth consecutive year.



Genentech
IN BUSINESS FOR LIFE



BIO-X

Bio-X is a Stanford University program supporting interdisciplinary work related to biology. The program is a joint project of the Schools of Humanities and Sciences, Engineering, Earth Sciences, and Medicine. Physics, chemistry, engineering, and computer science are combined in approaching important basic and applied problems in biology and medicine.

Bio-X is designed to build community and communication.

The Bio-X Program which reaches across the university to nearly 400 faculty members in over 55 departments is facilitated by the James H. Clark Center, which was completed in 2003 thanks to the enormous generosity of Jim Clark and Atlantic Philanthropies.

The Clark Center comprises the equipment, resources and utilities required to conduct breakthrough research at the cutting edge of engineering, science and medicine.

**For more information on Bio-X,
visit our website at:**

<http://biox.stanford.edu>

BCATS 2008 SCHEDULE

8.00	On-Site Registration & Breakfast	<i>Outside</i>
9.00	Opening Remarks	<i>Hewlett 200</i>
9.15	Keynote Address: Andrew McCammon, PhD (pg 12) <i>Nasty Beasties: Computer-aided Drug Discovery for Infectious Diseases</i>	
10.00	Eduardo Abeliuk (pg 16) <i>Transcriptional pattern analysis of a microbial cell-cycle</i>	
10.15	Mitul Saha (pg 17) <i>Identifying conserved domains in cyro-EM maps of biological nanomachines</i>	
10.30	Relly Brandman (pg 18) <i>70S bacterial ribosome dynamics explored by mutual information in all-atom molecular dynamics simulation</i>	
10.45	Spotlight Presentations	
11.00	Poster Session I (odd-numbered posters)	<i>Packard</i>
12.00	Lunch	<i>Outside</i>
12.45	Keynote Address: Rowan Chapman, PhD (pg 13) <i>Investing in Personalized Medicine</i>	<i>Hewlett 200</i>
1.30	Johnathan Laserson (pg 19) <i>Bayesian population sequencing via the dirichlet process</i>	
1.45	Su-In Lee (pg 20) <i>Learning a priori on regulatory potential from eQTL data</i>	
2.00	Sarah J. Aerni (pg 21) <i>Analysis of gene regulation and cell fate from single-cell gene expression profiles in C. elegans</i>	
2.15	Spotlight Presentations	
2.30	Poster Session II (even-numbered posters)	<i>Packard</i>
3.30	Aaron S. Wang (pg 22) <i>Development of a bleed detection algorithm for vascular ultrasound</i>	<i>Hewlett 200</i>
3.45	Shahab Sheikh-Bahaei (pg 23) <i>Modeling the collective behavior of liver cells in clearing toxins</i>	
4.00	Jennifer L. Hicks (pg 24) <i>Predicting the outcome of musculoskeletal surgery in patients with cerebral palsy</i>	
4.15	Debashis Sahoo (pg 25) <i>MiDReG: Mining developmentally regulated genes using boolean implications</i>	
4.30	Keynote Address: Ron Kikinis, MD (pg 14) <i>Medical Image Computing: From Data to Understanding</i>	
5.15	Awards & Closing Remarks	
5.30	Reception	<i>Clark Plaza</i>

BCATS 2008 Posters

poster

page

SPOTLIGHT PRESENTATIONS

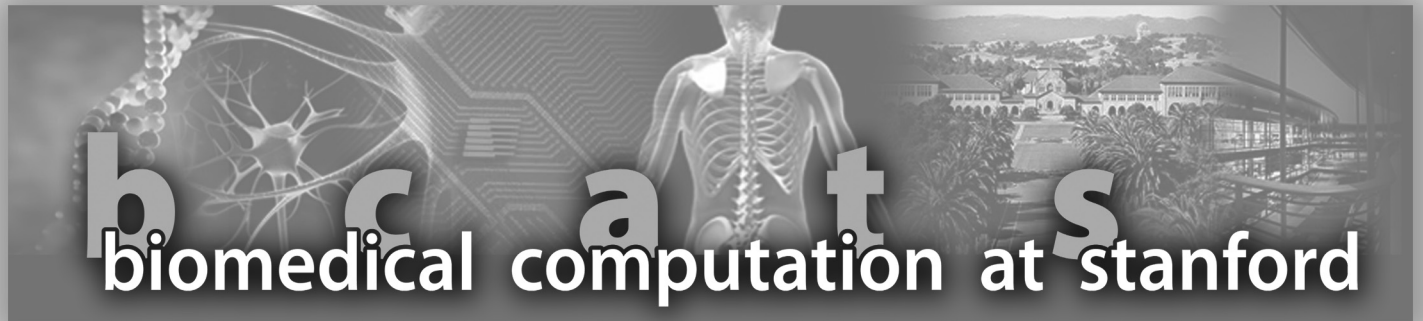
S1	Alexis Battle	Automated Reconstruction of Gene Networks from Quantitative Genetic Interaction Data	28
S2	Alistair N. Boettiger	Paused Polymerase Increases Synchrony of Gene Expression	29
S3	Andrea S. Les	Quantification of Hemodynamics in Abdominal Aortic Aneurysms During Rest and Exercise	30
S4	Dahlia Weiss	Computational Studies to Elucidate the Role of a Critical Histidine Residue in Transcription Elongation	31
S5	Jonathan R. Karr	Integrated Flux Balance Analysis Model of Escherichia coli	32
S6	Nikola Stikov	Relating Brain Diffusivity and Myelination using Quantitative Magnetic Resonance Imaging	33
S7	Timmy Siau	Robust Optimization of High Dose-Rate Brachytherapy	34
S9	Yuan Yao	Topological Methods for Exploring Biomolecular Folding Pathways	36
S10	Yuping Zhang	Prediction of Patient Outcomes from Longitudinal Microarray Data	37
1	Mona Ali	The ligand-free state of the TPP riboswitch	40
2	Sergio Bacallado	Bayesian comparison of Markov models of molecular dynamics	41
3	Mohsen Bayati	A Sequential Algorithm for Generating Random Graphs	42
4	Gregory R. Bowman	Adaptive Seeding: A New Method for Simulating Biologically Relevant Timescales	43
5	Sophie Candille	Genetic architecture of normal phenotypic variation in European populations	44
6	Sonny Chan	Toward Patient-Specific Simulation of Endoscopic Sinus Surgery	45
7	Rong Chen	FitSNPs: Differentially Expressed Genes are More Likely to have Variants Associated with Disease	46
8	Harpreet Dhatt	¹⁸ F-FDG Uptake within the Spinal Canal Follows a Predictable Pattern	47
9	Ariel Dowling	Increased Surface Friction Alters Risk of ACL Injury in Healthy Subjects	48
10	Samuel C. Flores	Fast modeling of RNA structure and dynamics using internal coordinates	49
11	Melanie D. Fox	Mechanisms of improved knee flexion after rectus femoris transfer surgery	50
12	Hong Gao	Increasing the Efficiency of Genomewide Association Mapping via Hidden Markov Models	51
13	Yael Garten	Prediction of novel pharmacogenomic interactions via text mining and network analysis	52
14	Samuel Hamner	Muscle Contributions to Propulsion and Support During Running	53
15	Lan Hua	A two-stage urea induced unfolding through its stronger dispersion interactions with proteins than water	54
16	Kathryn Hymes	Hamming Codes in the Context of DNA Self-Assembly	55
17	Matt Inlay	Identification of the branchpoint between B and T cell development through MiDReG	56
18	Ryan P. Jackson	Multiphoton and Transmission Electron Microscopy of Collagen in Ex Vivo Tympanic Membranes	57
19	Magdalena A. Jonikas	Bridging low and high resolution RNA structure modeling	58
20	Orr Keshet	Motion compensated volumetric modulated arc therapy (4D VMAT)	59
21	Hyun Jin Kim	Computation of coronary flow using a closed loop model of the cardiovascular system with a three-dimensional fluid-solid interaction model of the thoracic aorta	60

BCATS 2008 Posters

poster

page

22	Sean H. J. Kim	Simulation Modeling of Cultured Alveolar Type II Cells to Elucidate Principles of Pulmonary Alveolar Growth	61
23	Kranthi Kode	Parametric Modeling of Raman Spectra of Nanoparticles for Quantitative Unmixing	62
24	Daniel Korenblum	Capturing and Managing Biomedical Image Metadata	63
25	Ethan Kung	Re-creating Vascular Flow Impedance with a Physical Apparatus	64
26	Jennifer L. Lahti	Developing cystine-knot peptides as alternative scaffolds for therapeutic protein engineering	65
27	Jia-Ren Lin	How cancer evolves? Study the role of mutational regulation in somatic evolution by using digital organisms.	66
28	Ray Lin	Modeling the Onset of Fatal Metastases in Lung Cance	67
29	Tianyun Liu	Predict novel calcium-binding sites using FEATURE combined with modeling methods	68
30	Francesca Milletti	Engineering Protein Flexibility by Molecular Dynamics and Multivariate Statistics	69
31	Alexander A. Morgan	Hyper and Hypo Variability in Gene Expression Across Multiple Studies	70
32	Geoffrey Nelson	Small Animal Conformal Radiotherapy	71
33	Sarah Nelson	Testing the Importance of Surrounding Residues in Positioning Hydrogen Bond Donors in the Oxyanion Hole of Ketosteroid Isomerase	72
34	Hedi Razavi	Quantifying Changes in Pulmonary Vascular Morphology From Adolescence to Adulthood in Rats	73
35	Hersh Sagreiya	Evaluation of Clinical and Pharmacogenetic Factors Involved in Warfarin Dosing	74
36	Sriram Sankararaman	Inference of Ancestries in Admixed Populations	75
37	Florian F. Schmitzberger	Thin Client Architecture to Support Multicenter Radiological Trials	76
38	Jun Seita	Digital Pattern Analysis based on Microarray Dynamic-range Database for Complicated Developmental Models	77
39	Junhee Seok	Knowledge-based analysis of microarray for the discovery of transcriptional regulations	78
40	Shai S. Shen-Orr	Towards a Cytokine-Cell Interaction Knowledge Base of the Adaptive Immune System	79
41	Jessica Shih	Image-Based Modeling of Blood Flow in Human Coronary Arteries With and Without Bypass During Rest and Simulated Exercise	80
42	Lisa Singer	Selection of magnetic resonance imaging slice orientation in the calculation of volumetric breast density	81
43	Uygar Sümbül	Each pixel is a separate event: Application to dynamic MRI	82
44	Jian Sun	A Well-controlled Fast Clustering Method on Conformation Space of Biomolecules	83
45	Guillaume A. Troianowski	Hemodynamic simulations in the preoperative and virtual postoperative Fontan geometries	84
46	Rodney D. Wiersma	Real-Time Tracking of Implanted Fiducial Markers using Combined MV and kV Imaging	85
47	Guanglei Xiong	Simulation of Blood Flow in Deformable Vessels using Subject-Specific Geometry and Variable Mechanical Wall Property	86



BCATS 2008

Keynote Speakers

BCATS Keynote Speaker

Andrew McCammon, PhD

University of California, San Diego

Joseph E. Mayer Professor of Theoretical Chemistry Distinguished Professor of Pharmacology
University of California, San Diego

Biography

J. Andrew McCammon is the Joseph E. Mayer Chair Professor of Theoretical Chemistry and Distinguished Professor of Pharmacology at UCSD, and is an Investigator of the Howard Hughes Medical Institute. He received his B.A. from Pomona College, and his Ph.D. in chemical physics from Harvard University, where he worked with John Deutch (of MIT). In 1976-78, he developed the computer simulation approach to protein dynamics in Martin Karplus's lab at Harvard. He joined the University of Houston as Assistant Professor of Chemistry in 1978, and became the M.D. Anderson Chair Professor of Chemistry in 1981. He moved to UCSD in 1995. Professor McCammon has invented theoretical methods for accurately predicting and interpreting molecular recognition, rates of reactions, and other properties of chemical systems. In addition to their fundamental interest, these methods play a growing role in the design of new drugs and other materials. Professor McCammon is the author with Stephen Harvey of "Dynamics of Proteins and Nucleic Acids" (Cambridge University Press), and is the author or co-author of more than 550 publications in theoretical chemistry and biochemistry. More than 50 of his former students have tenured or tenure-track positions at leading universities or research institutes. In the 1980's, Professor McCammon guided the establishment of the computer-aided drug discovery program of Agouron Pharmaceuticals (now Pfizer Global Research and Development, La Jolla Laboratories), and contributed to the development of the widely prescribed HIV-1 protease inhibitor, Viracept (nelfinavir). The McCammon group's studies of HIV-1 integrase flexibility contributed to the discovery of the first in a new class of antiviral drugs by Merck & Co., named Isentress (raltegravir) and approved by the US FDA in 2007. Professor McCammon received the first George Herbert Hitchings Award for Innovative Methods in Drug Design from the Burroughs Wellcome Fund in 1987. In 1995, he received the Smithsonian Institution's Information Technology Leadership Award for Breakthrough Computational Science, sponsored by Cray Research. He is the recipient of the American Chemical Society's 2008 national Award for Computers in Chemical and Pharmaceutical Research. He is a Fellow of the American Academy of Arts and Sciences, the American Association for the Advancement of Science, the American Physical Society, and the Biophysical Society.

Keynote Address

Nasty Beasties: Computer-aided Drug Discovery for Infections Diseases

The selective character of the binding and reactivity of key biological molecules is essential for life. Properly understood, such selectivity can be exploited in the design of drugs, novel antibodies or enzymes, sensors, or a host of other materials or devices. This talk will provide a brief overview of how computer simulations are now enabling the discovery of new drugs for the treatment of infectious diseases. Particular emphasis will be placed on proper accounting of the flexibility of the receptor in the design of such drugs. The potential of new generations of computing hardware and methodology to dramatically transform this area of work will be emphasized.

BCATS Keynote Speaker

Rowan Chapman, PhD

Mohr Davidow Ventures

Venture Partner
Mohr Davidow Ventures

Biography

Rowan Chapman joined MDV in 2001. Through her relationships with industry and academic leaders advancing the fields of personalized medicine, drug discovery technology and regenerative medicine she both sources new life sciences investments and works to develop and grow MDV life sciences companies including Artemis Health, Pacific Biosciences, Adamas Pharmaceuticals and ParAllele BioScience (acquired by Affymetrix). Rowan is also a member of the Personalized Medicine Coalition.

Prior to joining MDV she held the position of director of business development at Rosetta Inpharmatics (acquired by Merck) where she established collaborative partnerships with organizations in the pharmaceutical, agricultural and biotech fields as well as among various research institutions. Previously, Rowan held the position of marketing manager at Incyte Genomics.

Rowan holds a Ph.D. in cellular and molecular biology from the MRC Laboratory of Molecular Biology in Cambridge University, United Kingdom, where she also earned a bachelor's degree with first class honors in Biochemistry. She served postdoctoral fellowships at the University of California, San Francisco, as well as at the MRC-Laboratory of Molecular Biology at Cambridge. Prior to her work in industry, she authored or co-authored more than a dozen articles and patents.

Keynote Address

Investing in Personalized Medicine

The molecular diagnostics industry is being driven by the availability of sensitive, high-throughput detection methodologies for genome-wide analysis of DNA, RNA, protein and metabolites. This presentation will focus on how investment firms such as MDV make investments in tools and molecular diagnostics that drive personalized medicine. Examples from portfolio companies such as Pacific Biosciences and Raindance Technologies will be discussed, posited on the basis of theoretical and computational predictions from mixture models, and subsequently validated experimentally.

BCATS Keynote Speaker

Ron Kikinis, MD

Harvard Medical School

Director of Surgical Planning Laboratory, Professor of Radiology
Department of Radiology, Brigham and Women's Hospital
Harvard Medical School

Biography

Dr. Kikinis is the founding Director of the Surgical Planning Laboratory, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, and a Professor of Radiology at Harvard Medical School. This laboratory was founded in 1990.

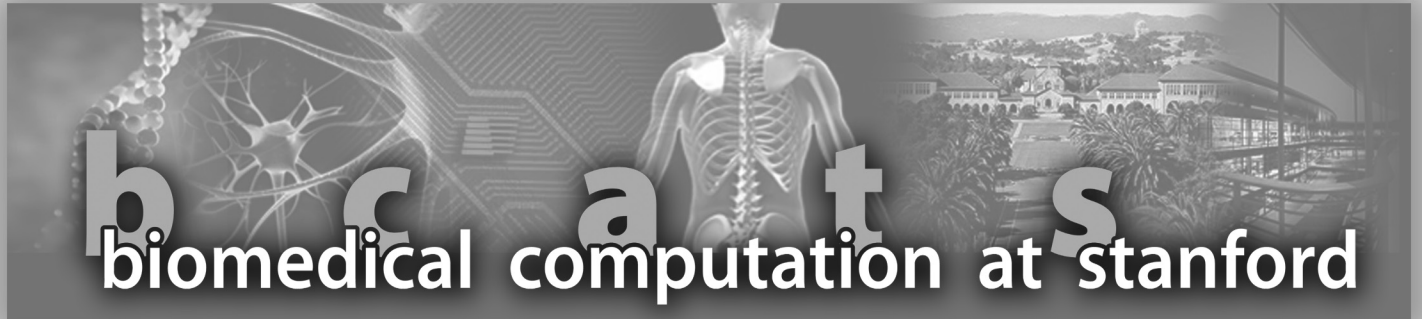
Dr. Kikinis is the Principal Investigator of the National Alliance for Medical Image Computing (NA-MIC, a National Center for Biomedical Computing, an effort which is part of the NIH Roadmap Initiative), and of the Neuroimage Analysis Center (NAC a National Resource Center funded by NCRR). He is also the Research Director of the National Center for Image Guided Therapy (NCIGT), which is jointly sponsored by NCRR, NCI, and NIBIB and co-director of the IGT program at CIMIT.

Dr. Kikinis has led and has participated in research in different areas of science. His activities include technological research (segmentation, registration, visualization, high performance computing), software system development (most recently the 3D Slicer software package), and biomedical research in a variety of biomedical specialties. The majority of his research is interdisciplinary in nature and is conducted by multidisciplinary teams.

Keynote Address

Computational modeling of biological tissues using mixture theory: Positing new hypotheses and testing them experimentally

The application of continuum mechanics to the analysis of biological tissues and cells continues to expand today, incorporating modern concepts of mixture theory to understand the complex interactions of solid, solvent and solute constituents. The complexity of these mixture models naturally increases with the need to describe complex mechanics phenomena at the tissue and cellular levels, such as the frictional properties of porous-permeable articular cartilage, partial volume recovery of cells to osmotic loading, and residual stresses in the arterial wall resulting from growth and remodeling. Analytical solutions to canonical problems, which are essential for understanding these mechanisms, become more difficult to formulate when modeling multiple mixture constituents and accounting for osmotic and electromechanical effects, due to their inherent nonlinear nature. Computational models which can address such coupled mechano-electrochemical phenomena remain in their infancy. Their widespread dissemination will bring new and essential tools for biomedical engineers in their effort to understand biological and physiological phenomena. These computational tools may help formulate important new hypotheses to describe observed phenomena hitherto considered too difficult to interpret. In this presentation examples will be provided where novel scientific hypotheses were posited on the basis of theoretical and computational predictions from mixture models, and subsequently validated experimentally.



BCATS 2008 Talk Abstracts

talk
1

RNA structure and dynamics during a microbial cell cycle

**Eduardo
Abeliuk**

Stephen G. Landt

Lucy Shapiro

Harley H. McAdams

Purpose:

The aquatic bacterium *Caulobacter crescentus* is a model organism for study of cell cycle regulation and the establishment of cellular asymmetry. We are exploiting the correlations between probe-level microarray signals across multiple experiments to identify previously unidentified RNA transcripts and novel regulatory mechanisms in *Caulobacter*.

Materials and Methods:

We use data from microarrays performed with a custom-designed tiled *Caulobacter* Affymetrix array that has 1 probe every 5 bp on both strands of every intergenic region and multiple probes for each coding region on both sense and antisense strands (Fig. 1). For each probe, RNA expression from multiple experiments comprises elements of a signal vector. Vectors from probes of a contiguous DNA segment will be correlated if an RNA transcript spans the segment. Examining the probe cross-correlation signals from different cell cycle time-windows, we can identify when the concentrations of the different transcripts are changing. To allow identification of RNA species we constructed a web-based browser that shows the probe expression correlations, gene locations on the chromosome, mRNA and protein cell cycle expression profiles, and other genomic features together on one display (Fig. 2). We can visually scan the genome using the browser to search for interesting correlations. To systematically search for transcripts, we assign a p-value to every DNA segment using a binomial distribution model to estimate the probability of the observed cross-correlation values between the probe pairs within the segment. A small p-value suggests a high likelihood that an RNA transcript is expressed from that segment (Fig. 3).

Results:

We identified and verified biochemically 27 novel *Caulobacter* intergenic non-coding RNAs (Fig. 4). Before this study, only four *Caulobacter* sRNAs were known. We have also shown that the cross-correlation signals can identify frequently overlooked features of transcriptional regulation such as multiple start sites (Fig. 5), multiple termination sites within operons (Fig. 6), and antisense RNAs (Fig. 7a). For example, our analysis surprisingly suggests that within a multi-gene transcript, an individual gene can be differentially regulated by the activation of an antisense RNA that appears to be post-transcriptionally regulating the concentration of the protein encoded on the sense strand (Fig. 7b).

Conclusions:

The work presented here has allowed us to predict many previously unidentified transcripts. These discoveries are suggesting new hypotheses for the control of gene expression during the cell-cycle. We believe this work is an excellent merger between bioinformatics and molecular biology.

URL: <http://omics.stanford.edu/bcats2008/>

Identifying Conserved Domains in Cryo-EM Maps of Biological Nanomachines

talk
2

Mitul
Saha

Michael Levitt
Wah Chiu

Purpose:

A major development in structural biology over the last 15 years has been the success of cryoEM in the enormously challenging task of determining the structures of large macromolecular assemblies (LMAs, such as ribosomes, chaperonins, viruses). CryoEM has emerged as a method distinctly suited for determining structures of LMAs in their near-native conditions and inferring conformation flexibility associated with their working mechanisms. However, unlike X-ray crystallography and NMR, cryoEM does not yield structures with atomic resolution. A significant portion of current cryoEM-based structure determination research is about bridging this resolution gap between cryoEM and conventional methods (X-ray crystallography, NMR) through computational means. Towards this end, we present a new, first-of-its-kind, fully-automated computational tool MOTIF-EM for identifying domains or motifs in large macro-molecular assemblies (such as chaperonins, viruses, etc.) that remain conserved upon conformation or evolutionary changes.

Materials and Methods:

MOTIF-EM takes cryoEM volumetric maps as inputs. The technique used by MOTIF-EM to detect conserved sub-structures is inspired by a recent breakthrough in 2D object recognition. The technique works by constructing rotationally-invariant, low-dimensional representations of local regions in the input cryoEM maps. Correspondences are established between the reduced representations (by comparing them using a simple metric) across the input maps. The correspondences are clustered using hash tables and graph theory to retrieve conserved structural domains or motifs.

Results:

MOTIF-EM has been used to extract conserved domains occurring in large macro-molecular assembly maps, (such as those of viruses P22 and epsilon 15, Ribosome 70S, GroEL, etc.) which remain conserved upon conformation changes or are repeated in the same structure. It has also been used to build atomic resolution models of certain maps. MOTIF-EM was also able to identify the conserved folds shared between dsDNA bacteriophages HK97, Epsilon 15, and ϕ 29, which in turn points to close evolutionary links in dsDNA phages.

Conclusion:

Tools like MOTIF-EM are useful as discovery of conserved structural entities can point to conserved active sites - indicating common function, evolutionary links, and drug binding targets for therapies.

In the future we would like to use MOTIF-EM to do structure comparison on a much larger scale.

Such as, comparing a given cryoEM map with all the available structures in various databanks (such as EBI cryoEM and PDB databanks) and discovering new links between all known macromolecular structures.

URL: <http://ai.stanford.edu/~mitul/motifEM>

talk 3

**Relly
Brandman**

*Relly Brandman
Yigal Brandman
Vijay S. Pande*

70S bacterial ribosome dynamics explored by mutual information in all-atom molecular dynamics simulation

Purpose:

The ribosome translates mRNA into protein by coordinated dynamics of hundreds of thousands of atoms and understanding these motions is a critical part of understanding ribosome function. We describe dynamics from an all-atom molecular dynamics simulation of the 70S bacterial ribosome and describe correlated motions between atoms in order gain insight into ribosome dynamics.

Materials and Methods:

The GROMACS software package¹ was used for molecular dynamics. We characterize coordinated motion between atoms by a new method for calculating the mutual information between atom motions.

Results:

We generated a 53 ns trajectory, the longest continuous all-atom molecular dynamics simulation of the ribosome to date, and calculate the mutual information between all pairs of backbone atom trajectories. In addition to corroborating previously described dynamics, we describe interactions between structural elements and larger-scale coordinated motion. We reveal that RNA helices in parallel orientations move in a tightly coupled fashion. Coupled motion is observed between critical non-bonded regions such as the beta hairpins of proteins L4 and L22 that define the narrowest region in the tunnel. Hinge functionality is revealed by looking for two coupled groups of atoms that move independently from one another; functionally important regions such as the L1 arm are shown as examples. Large-scale patterns reveal coupled motion between regions with similar functions, for example those associated with tRNA entrance. The catalytic peptidyl transferase center is shown to be composed of two independently moving A- and P-site regions corresponding to previously described symmetry-related regions. Motions in the small subunit suggest mechanical positioning of mRNA.

Conclusion:

Computational modeling can give insight into dynamics, however the computational power necessary to incorporate critical details such as side-chain, electrostatic and solvent interactions for biologically relevant timescales is currently beyond our reach. Our 53 ns trajectory likely reflects thermal motion around the crystal structure and local conformational dynamics. Because our analysis is consistent with previously described motions, we hypothesize that correlations in motion seen on short timescales may give some insight into motions on longer timescales. In addition to motions consistent with previously described dynamics, we begin to uncover how structural elements may move relative to one another and describe large-scale patterns of motion.

References:

1. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, & Berendsen HJ (2005) J Comput Chem 26, 1701-1718.

Bayesian Population Sequencing via the Dirichlet Process

Purpose:

New technologies for sequencing, 454 and Solexa, supply millions of noisy short reads from a DNA sequence. These technologies can also be applied to a population of DNA sequences, thus allowing reads from almost all the sequences in a specific environment. Population Sequencing is the task of reconstructing the original sequences from these reads. There are many cases in which the isolation of a specific DNA sequence is infeasible, and hence population sequencing must be applied[1]. Another application of population sequencing is in discovery of population structure[2] or in discovery of infectious agents in a population[3]. While current state of the art methods are geared towards single sequence assembly, our method is designed to solve population sequencing.

Methods:

We introduce a generative model for population sequencing. We view the reads as samples generated, with some noise, from a mixture of sequences. As the number of sequences in this mixture is unknown, we use a Dirichlet Process Prior over the number of sequences[4]. The model also accounts for the uncertainty in the length and the content of the sequences. We perform inference on this model using a novel MCMC algorithm that performs a combination of Gibbs moves, split-and-merge moves between sequences, and efficient alignment algorithms. This combination of methods allows us to jointly denoise the reads and assemble the sequences. This is in contrast to other sequencing algorithms which perform these tasks in a stage-wise manner.

Results:

Tested on synthetic data, our method perfectly reconstructed a genome of size 10K bases, from 30K noisy reads of sizes 80-120b (read noise was 1% for insertion, deletion and letter-change, reflecting the known noise model for 454). We compared against Velvet[5], a state of the art sequence assembly method. Velvet also reconstructed the sequence but with many errors - out of all 100-mers in the original sequence, nearly none survived in Velvet's reconstruction. We also compared our method to Velvet on real data gathered from bee population. Our algorithm put more than half of the reads in contigs of length 727 or more, compared to 315 in Velvet. Finally, we ran our algorithm on a dataset collected from 8 known variants of HIV. Our method reconstructed all the variants achieving a Jaccard score of 92% (original vs reconstructed set of all 100-mers). Our method performs as well as the state of the art methods on single sequence assembly, and significantly outperforms these methods when applied to population sequencing.

References:

1. Venter et al., Science 2004
2. Wang et al., Gen. Res. 2007
3. Cox Foster et al., Science 2007
4. Ferguson, Ann. of Stat. 1973
5. Zerbino and Birney, Gen. Res. 2008

*Jonathan
Laserson
Vladimir Jojic
Daphne Koller*

talk 5

Learning a Prior on Regulatory Potential from eQTL Data

Su-In Lee

Aimee M. Dudley

David Drubin

Pamela A. Silver

Nevan J. Krogan

Dana Pe'er

Daphne Koller

Purpose:

Expression data of genetically diverse individuals (eQTL) provide a unique perspective on the effect of genetic variation on cellular pathways, and help identify sequence polymorphisms with phenotypic effect. However, with data from relatively few individuals, it is difficult to distinguish true causal polymorphisms, especially in systems with significant linkage disequilibrium, where traits are often linked to large chromosomal regions.

Methods:

Researchers traditionally select among plausible polymorphisms by favoring those that are more conserved, that lead to significant amino acid change, or that are in genes whose function is related to the phenotype. But how do we know how much weight to attribute to these different features? **Lirnet** is a novel method that learns from eQTL data how to weight regulatory features, such as gene function and conservation, or type and position of genetic polymorphism, which are available for any organism. **Lirnet** thus induces a regulatory potential for candidate polymorphisms, allowing us to select among them. The weights determining the relevance of different regulatory features are learned automatically, simultaneously with the regulatory network.

Results:

We apply **Lirnet** both to the human HapMap eQTL data and to a yeast linkage eQTL data (1), and show statistically that it significantly improves the construction of regulatory networks. **Lirnet** shows better ability to recover previously known regulatory interactions than other methods (2,3), and outperforms other methods (3,4) in predicting causative sequence polymorphism within a large, linked chromosomal region. In one example, **Lirnet** uncovered a novel, experimentally validated connection between Puf3, a sequence-specific RNA binding protein, and P-bodies, cytoplasmic structures regulating translation and RNA stability, as well as the particular causative SNP in Mkt1 that induces the variation in the pathway.

Conclusion:

The increasing availability of eQTL data provides an opportunity to uncover the causal effects of by genetic variation. In this study, we propose a novel machine learning approach which learns to recognize causal polymorphisms, and thereby learn a better regulatory network. Unlike other methods, **Lirnet** can flexibly use any regulatory features, including sequence features that are available for any organism, and automatically learn their weights in a dataset-specific way. This ability will allow **Lirnet** to be used in detecting causal variations in mammalian organisms, where large linked regions are a pervasive problem.

References:

- 1 Yvert et al (2003) Nat Genet 35(1): 57-64.
- 2 Lee et al (2006) PNAS 103(38):14062-7.
- 3 Zhu et al (2008) Nat Genet 40(7):854-61.
- 4 Suthram et al (2008) Mol Syst Biol. 162.

Analysis of gene regulation and cell fate from single-cell gene expression profiles in *C. elegans*

Sarah J. Aerni

Purpose:

The goal of this project is to analyze gene expression data at the single cell level from *C. elegans*, which has a known cell lineage. We want to quantitatively assess molecular changes during development and elucidate repeating modules from the cell lineage. To do this, we developed a pipeline to extract and analyze expression data from 3D confocal images of *C. elegans*.

Materials & Methods:

We developed a cell lineage analyzer that can automatically name each nucleus and extract levels of expression of a fluorescent reporter from a 3D confocal image. The cell lineage analyzer can identify up to 363 nuclei with 86% accuracy. We used the pipeline to analyze 200 images showing expression of 51 different genes in 363 cells. From the expression of these genes, we created a molecular signature for every cell and then compared the signatures of different cells to each other. We developed a system to use observed levels of expression to predict expression patterns in ancestral cells in the lineage (similar to parsimony used in evolution). By tracing when genes begin to be expressed in the cell lineage, we generated a developmental activity map showing when changes in cell fates occur.

Results:

A surprising result from our analysis of molecular signatures is that cells with identical fates can express different sets of transcription factors, suggesting that different gene regulatory networks are used to generate the same cell. The developmental activity map showed specific cell divisions in the cell lineage where major differentiation events take place. Our methods allowed us to identify repeating modules within the cell lineage, including developmental clones (identical progeny generated from a founder cell) and sublineages (separate parts of the lineage that generate the same pattern of cells).

Conclusions:

We have generated a data table showing expression in specific cells, from individual worms, and from a large number of genes. In the future, the number of genes will increase and the pipeline will be applied to different developmental stages and different growth conditions, which will improve the accuracy and resolution of the molecular signatures. Analysis of this initial data set has already returned the surprising result that the same cell type can be generated via different gene regulatory networks. Unlike images that must be manually browsed, our expression data table is suitable for computational analysis allowing us to analyze expression from all genes and images in parallel. Similar computational approaches may be equally fruitful for other types of GFP or RNA in situ images.

References:

1 Long et al Automatic recognition of cells (ARC) for 3D images of *C. elegans*, RECOMB

Xiao Liu

Fuhui Long

Hanchuan Peng

Serafim Batzoglou

Eugene W. Myers,

Stuart K. Kim

talk 7

**Aaron S.
Wang**

Fritz Bech

Charles A. Taylor

David H. Liang

Development of a Bleed Detection Algorithm for Vascular Ultrasound

Purpose:

Most deaths from trauma can be prevented by rapid control of hemorrhage. While trained clinicians can perform quick diagnosis of vascular trauma using imaging ultrasound in the hospital, a portable ultrasound device with an automated algorithm to detect major bleeding would be useful for medics in the field. We characterized normal blood flow in the upper extremities with an image-based flow model and evaluated its ability to detect vascular abnormalities by looking for flows that deviate from the model.

Material and Methods:

The power-law model [1] states that blood flow Q is proportional to the vessel diameter d taken to a power k , $Q \propto d^k$, where k is defined by the bifurcation pattern: $d_0^k = d_1^k + d_2^k$ (0 = parent vessel, $1&2$ = children vessels). To adapt this law for the upper extremities, ultrasound was performed on the brachial, ulnar, and radial arteries of 28 healthy volunteers to measure diameters and flows during rest and exercise (2.7 watts, 30 lb grip). The best-fit k value for the brachial bifurcation was determined and then used in the power-law to predict normal flow splits in children vessels: (Q_1, Q_2) . In injury, the flow split deviates from the normal split by a Q_{dev} amount of flow: $(Q_1 - Q_{dev}, Q_2 + Q_{dev})$. Flow split deviation (FSD) quantified Q_{dev} as a percent of total flow: $FSD = Q_{dev}/(Q_1 + Q_2)$.

Computational 3D blood flow simulations of the brachial bifurcation with different size punctures in the children vessels were used to correlate the amount of bleeding with the degree of flow split deviation. This method of using FSDs was evaluated in a clinical study to see if surgically placed wrist arteriovenous fistulas (AVFs) can be detected in dialysis patients. This method was further evaluated in in-vivo rabbit studies where specific bleed rates were created in the descending aortic bifurcation.

Results:

The best-fit k for the brachial bifurcation was invariant between rest and exercise and was determined to be 2.75. The correlation between predicted and measured normal blood flows was $R^2=0.95$. Simulations verified that larger punctures produced larger FSDs. In the clinical study, dramatic increases in FSDs easily differentiated between the arms before and right after AVF placement. Initial in-vivo studies showed that this method can detect moderate bleed rates.

Conclusion:

Normal arm blood flow, regardless of physiologic state, can be well characterized and predicted with a $k=2.75$ power-law. The strategy of identifying perturbations of flow at bifurcations may provide a rapid automated method for localizing active bleeding with ultrasound.

References:

[1] Zamir, M and Brown, N. Arterial branching in various parts of the cardiovascular system. Am J Anat. 1982;163:295-307.

Modeling the Collective Behavior of Liver Cells in Clearing Toxins

Purpose:

Hepatocytes, the parenchymal cells of the liver, express heterogeneous, location-dependent enzyme and transporter activities to detoxify compounds, apparently following an intrinsic agenda, the principles of which are not fully understood. This phenomenon is known as liver zonation.

In order to gain insight into this phenomenon, we developed and validated *in silico* agents that collectively are able to mimic hepatocytes' cooperative behavior. The agents self-organize their response to toxic shocks following simple rules.

Materials and Methods:

Using game theory, we created models of hepatic toxin elimination processes. The game is simple: each player (a cell or hepatic zone) can either eliminate or ignore an incoming, simulated compound. There is an immediate cost for elimination: resource consumption. On the other hand, if all players ignore a toxic compound, there is a subsequent, toxicity-dependent cost to all players. The players are arranged in an n by m grid space. Simulated compound objects move stochastically from perimeter towards the grid center. Each player uses the Q-learning algorithm to simultaneously minimize its energy consumption and toxicity costs.

Results and Conclusion:

At equilibrium, the players have cooperated by adopting heterogeneous location-dependent strategies. Downstream players commit most to toxin elimination. Ample literature evidence validates this strategy: xenobiotic metabolism is preferentially located downstream, in the perivenous region(1). The model predicts that upstream players focus on eliminating the highly toxic compounds. We found no literature evidence that directly validates this prediction.

The model also predicts that the location of hepatotoxicity depends on compound toxicity: less toxic compounds damage the downstream region, whereas compounds that are more toxic damage the upstream region. Supporting evidence for this observation is available. The zonal activity of some of several highly toxic compounds(2), such as TCDD (LD50=0.034 mg/kg), cyclochlorotine (LD50=2-3 mg/kg), and Gossypol (LD50=5 mg/kg) selectively damage the upstream zone. Alternatively, less toxic compounds such as Acetaminophen (LD50=1295 mg/kg) selectively damage the downstream zone.

Adaptive cooperation among hepatic cells may have been an important factor in the evolution of within organ, location-dependent strategies used by cells to deal with short- and long-term changes in each cell's environment as well as that of the organism. Simulations of the type used provide new insight into such processes.

URL: http://biosystems.ucsf.edu/research_zonation.html

References:

- 1 Jungermann K. Zonation of metabolism and gene expression in liver. *Histochem. & Cell Biol.* 95;103(2):81-91.
- 2 TOXNET US Nat'l Library of Medicine, 08. toxnet.nlm.nih.gov. Accessed 05/08

**Shahab
Sheikh-Bahaei**

C. Anthony Hunt

talk 9

**Jennifer L.
Hicks**

**Michael H.
Schwartz
Scott L. Delp**

Predicting the outcome of musculoskeletal surgery in patients with cerebral palsy

Purpose:

Cerebral palsy, a neurological disorder caused by an injury to the developing brain, affects approximately 3 out of 1000 children. These patients have poor motor control, muscle weakness, and bone deformities, which inhibits safe and efficient locomotion. Currently, surgeons must interpret an overload of information, including data about a patient's muscle activations, strength, and movement patterns, when designing a treatment plan. Analyzing this information is challenging and surgeons frequently recommend different treatments from the same set of clinical information. Consequently, while orthopedic surgeries are often performed to improve a patient's gait, interventions are not always successful.

The goal of this study was to develop an objective, statistical classifier to identify good candidates for a hamstrings lengthening procedure, one of the common musculoskeletal surgeries received by children with cerebral palsy.

Methods:

We identified 73 subjects who received hamstrings surgery, a procedure that lengthens one or more of the muscles in the back of the thigh. This surgery is intended to relieve tightness and enable the patient to walk with a less flexed knee. We classified a subject as "improved" after surgery if he or she exhibited a 25% or larger improvement in knee flexion. We next trained a Naïve Bayes Classifier to predict the outcome of the surgery based on pre-operative data. We identified the 30 variables that were significantly different between the two outcome groups and then tested subsets of 3 to 7 of these significant variables to find the subset with maximal prediction accuracy using 10-fold cross-validation.

Results:

We were able to predict whether subjects' knee flexion would improve after a hamstrings lengthening with 80% cross-validation accuracy, in contrast to the actual improvement rate of 55%. Our classifier used 6 pre-operative variables that combined information about the subject's motion at the hip and knee during gait, joint range of motion, muscular strength, and whether the patient received any concomitant procedures. The statistical model used in this study, while relatively simple, gives a set of weightings for each variable and an estimate of the probability that a subject will improve, both of which enhance the clinical accessibility of this approach.

Conclusions:

This study has demonstrated the potential for using predictive statistical models to help objectively guide the prescription of hamstrings surgery in patients with cerebral palsy. Further, this model can be readily modified to make predictions for other common treatments.

URL: <http://www.stanford.edu/~jenhicks>

MiDReG: Mining Developmentally Regulated Genes Using Boolean Implications

Debashis Sahoo

Jun Seita

Deepta Bhattacharya

Irv L. Weissman

Sylvia K. Plevritis

David L. Dill

Purpose:

Precursor cells differentiate to their terminal progeny through a series of developmental intermediates and a network of gene expression changes that gradually establish lineage commitment and the identity of the mature cell type. The identification of genes that are involved in this process has largely been dependent upon the physical isolation and characterization of gene expression patterns within these developmental intermediates. Thus, when intermediate steps are unknown for a particular cellular differentiation pathway, the identification of genes that are developmentally regulated in that pathway can be difficult. We present a new method to discover developmentally regulated genes using large repositories of publicly available microarray data.

Materials and Methods:

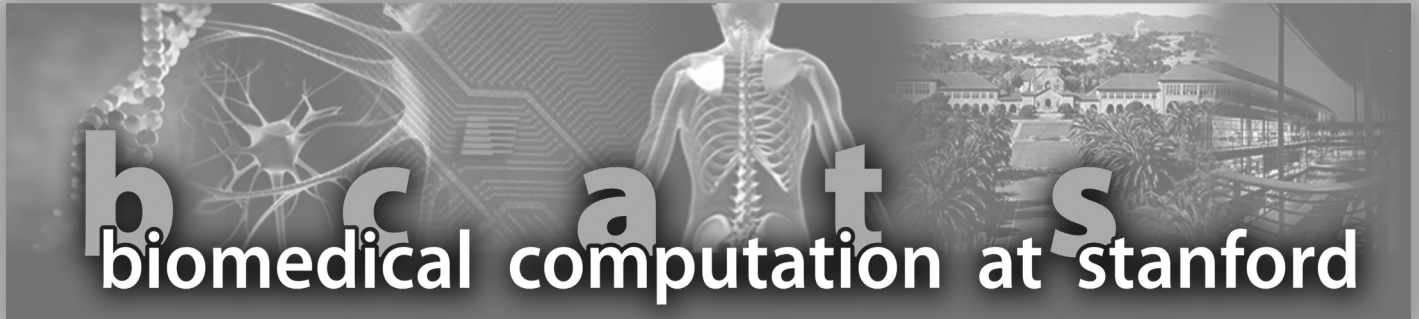
The method uses Boolean implications (if-then rules) mined from existing data and conservation across species to predict genes whose expression is established as precursor cells differentiate to their progeny in the absence of data from cells at intermediate steps of development, which have not been previously identified. In order to predict B cell precursor genes, two well known seed genes were considered: KIT and CD19. The candidate B cell precursor genes are computed by using following conserved Boolean implications in large repositories of gene expression datasets: KIT high ==> A low ==> CD19 low.

Results:

The algorithm was validated by applying it to B cell development, for which most developmental stages between hematopoietic stem cells and fully differentiated B cells are known and can be isolated. The algorithm predicted 19 genes that are expressed after the KIT+ progenitor cell stage and remain expressed through CD19+ mature B cells. These predictions were validated by sorting 12 mouse hematopoietic populations at different stages of development ranging from hematopoietic stem cells to germinal center B-cells using fluorescence activated cell sorting. Empirical measurement of the expression of 14 of these predicted genes using qRT-PCR confirmed that the expression of 10 of these genes is indeed stably established during B-cell differentiation. Through combined inclusion of CD19 and AICDA, we expanded our list of predicted developmentally regulated genes to 62. Thirty-three of these genes have been ablated in mice, and 18 of these mutants have been reported to have defects in B cell function or development.

Conclusion:

Our method have the potential to predict novel genes that are expressed early in other developmental pathways that are less well characterized than B cells.



BCATS 2008 Poster Spotlight Abstracts

poster
S1

Alexis Battle

Automated Reconstruction of Gene Networks from Quantitative Genetic Interaction Data

Daphne Koller

Purpose:

Genetic interaction (GI) assays measure the interaction between genes by comparing the phenotype of a double knockout organism to the phenotypes of the two single knockout organisms. Such assays shed light on underlying cellular pathways. For instance, common wisdom says that two genes in the same linear pathway will exhibit an alleviating interaction, where the double knockout exhibits a less severe phenotype than expected. We formalize such intuitions and provide the first method for automated reconstruction of large gene networks from GI data.

Materials and Methods:

Fluorescent microscopy allows measurement of quantitative cellular phenotypes such as the precise amount of a reporter protein. Technological advances have enabled large quantitative experiments which measure most pairwise interactions for a large set of genes[1]. We use a yeast GI dataset[2] that measures an unfolded protein response marker, using a set of genes related to the endoplasmic reticulum.

We represent gene networks as directed graphical probabilistic models, thus posing our problem as a Bayesian network structure learning task. We define a probability distribution over graph structures based on statistical properties of the pairwise measurements, formalizing standard biological interpretation of GI data. We then sample from this distribution to find high scoring graphs and quantify our confidence over structure choices, employing Annealed Importance Sampling[3] to improve the efficiency of producing independent samples.

Results:

Our methods allow us to efficiently reconstruct gene networks over hundreds of genes. On the yeast ER dataset, our learned network reflects known biology, including the structure of glycosylation and ERAD pathways. Our methods produced biologically useful hypotheses, including prediction of relationships for gene pairs that were not yet measured, and the network placement of uncharacterized genes.

Conclusion:

We present the first computational method for reconstructing cellular networks from GI data. We define a distribution over graph structures, from which we draw likely gene networks and compute confidence estimates for each structure choice. Applied to yeast data, our technique learned a network over hundreds of genes, correctly reconstructed known pathways, and produced novel biological hypotheses that could improve our understanding of protein folding and degradation.

References:

- [1] Schuldiner et al., Exploration of the Function and Organization of the Yeast Early Secretory Pathway through an Epistatic Miniarray Profile. *Cell*, 2005.
- [2] Jonikas et al., The building of a cellular folding compartment. in preparation.
- [3] Neal, Annealed Importance Sampling. *Statistics and Computing*, 2001.

Paused Polymerase Increases Synchrony of Gene Expression

**Alistair N.
Boettiger**

Michael Levine

Purpose:

Many biochemical sensing/signaling networks exhibit robust response behavior even at very low signal concentrations. As the number of molecules involved becomes small, the random nature of molecular collisions may constitute a significant amount of 'noise' in the signal [1-3]. We investigate 2 questions: (1) does molecular noise effect gene expression in metazoan life and embryonic development, (2) if so, what mechanisms exist in metezoa to filter molecular noise and achieve the observed synchronous, coordinated cell behaviors?

Materials and Methods:

We employ a cell culture assay to compare the time course of activation of two different cis-regulatory modules which respond to the same inducing factor. We also employ high resolution single cell in situ hybridization experiments in *Drosophila* embryos to examine the time course of activation of developmental genes in their native context. Both these experimental setups allow us to compare the individual properties among a large population of clonal cells. We analyze these images with custom written quantitative image processing scripts. We compare these measurements to theoretical predictions of our Markov Models of stochastic gene expression.

Results and Conclusions:

We demonstrate (a) molecular noise does effect gene expression in embryonic development and adult function and (b) several genes employ a newly discovered regulatory mechanism [4] to overcome this noise and achieve synchronous expression. The first conclusion follows from our observation that considerable cell-to-cell variation in induction statistics can easily be measured with our approach. Surprisingly, we discover that this measurable variability differs considerably, even between genes responding to the same signaling inputs. We find 8 genes that exhibit dramatically more synchronous expression than their partner genes (genes expressed in response to the same transcriptional inputs). Unlike their less synchronous counterparts, all eight of these genes are regulated at elongation, through release of a paused polymerase. We propose this mechanism has evolved in part to mitigate the negative effects of input noise. We employ new mathematical models of gene regulation to investigate how regulation through polymerase pausing achieves improved expression synchrony.

References:

1. Paulsson J, Berg OG, and Ehrenberg, M. (2000) PNAS 97, 7148-7153.
2. Elowitz MB, Levine AJ, Siggia ED and Swain PS. (2002) Science. 297, 1183-1186.
3. Spudich JL. and Koshland DE (1976) Nature 262 467-471.
4. Zietlinger J, Stark A, Kellis M, Hong JW, Nechaev S, Adelman K, Levine M. and Young, RA. (2007) Nature Genetics 39 1512-1516.

Quantification of Hemodynamics in Abdominal Aortic Aneurysms During Rest and Exercise

Andrea S. Les

C. Alberto Figueroa

Shawn C. Shadden

Ronald L. Dalman

Robert J. Herfkens

Charles A. Taylor

Introduction:

Rodent studies show that increasing infrarenal flow via a distal arterio-venous fistula slows the growth of experimental Abdominal Aortic Aneurysms (AAAs) [1]. We hypothesize that exercise may slow AAA growth in humans by attenuating pro-inflammatory hemodynamic conditions such as low and oscillatory shear stress and flow stagnation. This work combines magnetic resonance imaging (MRI) with computational fluid dynamics to study blood flow in human AAAs under rest and exercise conditions.

Methods:

Seven men and one woman (mean age 73.3 ± 5.70 years) with small AAAs were imaged in the supine position using a 1.5T GE Signa MR scanner. A 3D gadolinium-enhanced magnetic resonance angiography (MRA) sequence was used to image the aorta. A cardiac-gated phase contrast sequence (PC-MRI) was used to measure aortic blood flow velocity at the suprarenal (SC) and infrarenal (IR) levels. 3D geometric models were constructed from the MRA and then discretized into finite-element meshes (average size 9.0 ± 2.3 -million elements). The PC-MRI images were segmented at 24 time-points and then assembled into SC and IR volumetric flow waveforms. The SC waveforms were mapped to the inflow faces of the corresponding meshes using a Womersley velocity profile. At each outlet, a three-element Windkessel model consisting of a proximal (R_p) and distal (R_d) resistance and capacitance (C) was used to represent the vessels downstream of each outlet (Fig. 1 & 2). The resting R_p , C , R_d values were specified using the PC-MRI data, literature, and measured brachial blood pressures; these values were scaled to simulate moderate exercise associated with a 50% increase in heart rate. The Navier-Stokes equations were solved using a stabilized finite-element method and assuming rigid walls. Resultant mean wall shear stress (MWSS) and oscillatory shear index (OSI) were quantified in 1-cm strips at the SC, IR, largest AAA diameter, and suprabifurcation levels; turbulent kinetic energy (TKE) was also calculated.

Results and Discussion:

All simulated resting and exercise flows and pressures were within 5% of target values. MWSS increased (Fig. 3) while OSI decreased (Fig. 4) in all four locations from rest to exercise. Specifically, MWSS increased from 2.5 ± 2.1 to 21.7 ± 14.6 dynes/cm² and OSI decreased from 0.27 ± 0.06 to 0.21 ± 0.04 at the largest diameter of the AAA, eliminating areas of low, oscillatory shear within the AAA. For all patients, TKE increases drastically from rest to exercise, perhaps reducing flow stagnation within the AAA (Fig. 5).

References:

1. K. Hoshina et al., 2003, Wall shear stress and strain modulate experimental aneurysm cellularity, J Vasc Surg, vol. 37, pp. 1067-74.

Figures can be found at <http://www.stanford.edu/~asles/BCATS/>

Computational Studies to Elucidate the Role of a Critical Histidine Residue in Transcription Elongation

poster
S4

Dahlia Weiss

Xuhui Huang

Dong Wang

David A Bushnell

Roger Kornberg

Michael Levitt

Recently resolved X-ray structures of the RNA polymerase II transcription elongation complex show extensive interactions of a structurally conserved element, the trigger loop (TL), with a correctly matched nucleotide (NTP) in the active site. The interaction between a critical histidine residue, His1085, on the TL and the phosphate group of the correctly matched NTP is key for stabilization of the TL in the “closed” conformation. In order to understand the role of this histidine residue in stabilizing TL-NTP interactions, we performed all-atom Molecular Dynamics (MD) simulations of the RNA pol II transcription elongation complex in explicit solvent. In the wildtype (WT) MD simulations, RNA pol II transcription complex solvated in explicit solvent with a total of 370,000 atoms, is shown to be stable in the closed conformation. In addition to the WT complex, we also simulated single point mutations of His1085 to Phe and Tyr.

Our simulation results are consistent with existing experimental observations. Point mutation to Phe led to a total loss the interaction with the NTP, explaining loss of activity, while mutation to Tyr made the interaction less favorable compared to WT, explaining a decreased elongation rate. We also investigated different protonation states for His1085, inaccessible to experimental observation. We found that the protonated histidine (HIP) forms the most stable interaction with the phosphate group of GTP. Both hydrogen bond interactions and salt bridge interactions contribute to the stabilization. This stabilization is suggested to have catalytic significance.

Integrated Flux Balance Analysis Model of *Escherichia coli*

Jonathan

R. Karr

Markus W Covert

Nan Xiao

Tiffany J Chen

Purpose:

The effort to build a whole-cell model requires the development of new modeling approaches, and in particular, the integration of models for different types of processes, each of which may be best described using different representation. Flux-balance analysis (FBA) has been useful for large-scale analysis of metabolic networks, and methods have been developed to incorporate transcriptional regulation (regulatory FBA, or rFBA). Of current interest is the integration of these approaches with detailed models based on ordinary differential equations (ODEs).

Materials & Methods:

We developed an approach to modeling the dynamic behavior of metabolic, regulatory and signaling networks by combining FBA with regulatory Boolean logic, and ordinary differential equations. We use this approach (called integrated FBA, or iFBA) to create an integrated model of *Escherichia coli* which combines a flux-balance-based, central carbon metabolic and transcriptional regulatory model with an ODE-based, detailed model of carbohydrate uptake control. We compare the predicted *Escherichia coli* wild-type and single gene perturbation phenotypes for diauxic growth on glucose/lactose and glucose/glucose-6-phosphate with that of the individual models.

Results:

We find that iFBA encapsulates the dynamics of three internal metabolites and three transporters inadequately predicted by rFBA. Furthermore, we find that iFBA predicts different and more accurate phenotypes than the ODE model for 85 of 334 single gene perturbation simulations, as well for the wild-type simulations.

Conclusion:

We conclude that iFBA is a significant improvement over the individual rFBA and ODE modeling paradigms, suggesting that rFBA may now be thought of as a scaffold with which any ODE-based or other model that has an interface with metabolism may be integrated. This integration would allow processes which have been characterized and modeled in isolation to be re-evaluated in the context of their global effects. As more ODE-based models are developed and integrated into frameworks like that described here, it may eventually be possible to capture a majority of the known biological processes which occur in *E.coli* or other organisms in a single computational model.

URL: <http://covertlab.stanford.edu/projects/iFBA>

References:

Covert MW, Xiao N, Chen TJ, and Karr JR (2008) Integrated Flux Balance Analysis Model of *Escherichia coli*. *Bioinformatics*. 24(18): 2044-2050.

Relating Brain Diffusivity and Myelination using Quantitative Magnetic Resonance Imaging

Nikola Stikov

Lee M. Perry

John M. Pauly

Brian A. Wandell

Robert F. Dougherty

Purpose:

Magnetic resonance imaging (MRI) provides a diverse array of tissue contrast mechanisms. Here, we explore combining two quantitative MR contrast mechanisms in order to better understand the structure of white matter in the brain. Diffusion tensor imaging (DTI) is currently the only method available to identify and measure white matter fascicles of the brain in-vivo. Bound pool fractions (BPF) estimate the proportion of protons bound to macromolecules such as myelin. We used DTI to identify tracts in the corpus callosum (CC) and the optic radiation (OR) of a healthy volunteer, and we calculated the BPF along those tracts to study the relationship between myelination and the diffusion tensor.

Methods:

Measurements were performed on one human subject using a 1.5T GE Signa LX MRI scanner. The field of view covered the whole brain; voxel sizes ranged from 1.5mm to 2mm isotropic. We used cross-relaxation to estimate the BPF. Diffusion tensor tractography was used to segment the corpus callosum (CC) and the optic radiation (OR) pathways into 9 regions of interest (ROI) comprised of up to 1000 tracts. We compared the BPF and the DTI diffusivity parameters in those 9 ROIs.

Results:

There was a small but statistically reliable correlation between BPF and the DTI parameters. White matter voxels with higher BPF tended to have higher fractional anisotropy (FA) of the diffusion tensor ($r=0.16$), higher mean diffusivity ($r=0.10$), higher axial diffusivity (the largest eigenvalue of the DT; $r=0.18$) and lower radial diffusivity (mean of 2nd and 3rd eigenvalues; $r=-0.09$). All correlations were significant at $p<1e-8$. From the ROI analysis of FA vs. BPF, three patterns can be identified. The genu and splenium of the CC have high BPF (.13) and FA (.75), the body of the CC has low BPF (.115) and high FA (.75), and the OR has low BPF (.113) and low FA (.52).

Conclusions:

The small positive correlation between BPF and FA across all white matter is consistent with previous work suggesting that FA is somewhat sensitive to myelination. The differences in BPF values in the CC are consistent with previous work suggesting that fibers in the genu and splenium are more tightly packed compared to the CC body. We propose that BPF reflects the bulk myelin density within a voxel - a function of both the amount of myelin surrounding each axon and the fascicle packing density. Diffusivity measures are more sensitive to fascicle packing density and fascicle direction coherence within a voxel. Thus, BPF and DT are complementary measures that combine to provide a more complete insight into tissue microstructure.

References:

Yarnykh and Yuan, Neuroimage 23(2004); Aboitiz et al, Brain Res 598 (1992); Basser et al, Biophys J 66 (1994).

poster
S7

Robust Optimization of High Dose-Rate Brachytherapy

**Timmy
Siauw**

**Adam Cunha
Ron Alterovitz
Jean Pouliot
Alper Atamturk
Ken Goldberg**

Purpose:

High dose-rate (HDR) brachytherapy cancer treatments selectively irradiate cancerous tissue while sparing the surrounding organs. The procedure uses a short-range radiation source that is moved along catheters inserted into the body. The treatment plan is optimized to best conform to 3D tissue contours, which are identified by a physician using preoperative 3D imaging data. Discrepancies between the actual and intended dose may be caused by organ movement due to edema or involuntary movement, and imaging noise and ambiguity can cause contouring errors. Failure to take this uncertainty into account during dose optimization may decrease the treatment's effectiveness and increase the risk of side effects. In this work, we aim to find a treatment plan that is both clinically acceptable and stable with respect to uncertainty.

Materials and Methods:

The problem of optimizing the HDR brachytherapy plan can be formulated as a linear program (LP) [1]. Our uncertainty model assumes the true contours of the organs lie somewhere between inside and outside margins. We incorporate this model into a robust LP formulation which optimizes for the "worst-case" scenario [2]. We evaluate the standard treatment plan against the robust treatment plan on data from 19 patients. The dose-rates are calculated using the Ir 192 radioactive seed. All optimizations are performed using the MOSEK toolbox in MATLAB. We examine the variation of clinical metrics with respect to uniform perturbation (shrinking and growing) of the organ contours.

Results:

The robust plan on average maintains a clinically acceptable dosage to the prostate an additional +1.5mm of contour perturbation over the nominal plan. The robust treatment plan is also "hotter", delivering more radiation to the surrounding organs, which is undesirable. However, the treatments are clinically equivalent in the case of the bladder and urethra. The rectum shows noticeable discrepancies (favoring the nominal plan) and will have to be investigated further.

Conclusion:

Robust optimization can produce HDR brachytherapy treatment plans that are clinically acceptable, yet significantly more robust to contour uncertainty than standard optimization approaches. Future work will investigate target vs. organs at risk trade-offs as well more sophisticated uncertainty models that are less conservative.

References:

- [1] R. Alterovitz, E. Lessard, J. Pouliot, I-C. J. Hsu, J.F. O'Brien, and K. Goldberg, "Optimization of HDR brachytherapy dose distributions using linear programming with penalty costs," *Medical Physics*, v. 33, no. 11, pp. 4012-4019 (2006)
- [2] A. Ben-Tal and A. Nemirovski, "Robust convex optimization." *Mathematics of Operations Research* v. 23 no. 4 (1998)

poster
S9

Topological Methods for Exploring Biomolecular Folding Pathways

Yuan Yao**Jian Sun****Xuhui Huang****Pande Vijay****Leonidas Guibas****Gunnar Carlsson****Purpose:**

Characterization of transient intermediate or transition states is crucial for the description of biomolecular folding pathways, which is however difficult in both experiments and computer simulations. Such transient states are typically of low population in simulation samples. Even for simple systems such as RNA hairpins, recently there are mounting debates over the existence of multiple intermediate states. In this study, we develop a computational approach to explore the relatively low populated transition or intermediate states in biomolecular folding pathways, based on a topological data analysis tool, Mapper, with simulation data from large-scale distributed computing. Such an approach may provide a graphical representation of the free-energy landscape, and be reduced to a Markov state model with a Bayesian model selection which captures the long-term behavior on biomolecular folding pathways.

Materials and Methods:

The method is inspired by the classical Morse theory in mathematics which characterizes the topology of high dimensional shapes via some functional level sets. The function, called as filters here, is often chosen as a density or conditional density filter which helps us focus on the structures on pathways, followed by clustering analysis on its level sets, which helps separate low populated intermediates from high populated uninteresting structures. Transition or Markov state models can be built up based on metastates obtained from grouping clusters above.

Results and Conclusion:

One successful application of this method is given on a debating example, a RNA hairpin with GCAA tetraloop, where we are able to provide structural evidence from computer simulations on the multiple intermediate states and exhibit different pictures about unfolding and refolding pathways [1]. Another successful application is to the Alanine-dipeptide, where we discover metastable states used for Markov models with a much shorter time but a competitive accuracy than previous methods [2]. The method is effective in dealing with high degree of heterogeneity in distribution, capturing structural features in multiple pathways, and being less sensitive to the distance metric than nonlinear dimensionality reduction or geometric embedding methods. The methodology described in this work admits various implementations or extensions to incorporate more information and adapt to different settings, which thus provides a systematic tool to explore the pathways in complex biomolecular folding systems.

References:

1. Bowman et al., JACS 2008.
2. Chodera et al., JCP 2007.

Prediction of Patient Outcomes from Longitudinal Microarray Data

Purpose:

Time course gene expression profiling is increasingly applied in biomedical research to monitor the progression of diseases and effects of drug treatments. An important goal of the computational analysis is to predict clinical outcomes based on microarray data of patients over time. However, most existing approaches of prediction use data at one time point only. Here we propose a method for patient outcome prediction using longitudinal gene expression.

Materials:

A dataset of 454 arrays was used in this study, including 429 arrays of blood samples from trauma patients at days 0, 1, 4 post injury and 25 arrays from healthy controls. The variable for clinical outcomes is time of recovery (ToR).

Methods:

Given longitudinal microarray data and survival-type of outcome variable, the main objective is to utilize the temporal pattern of gene expression in feature selection for the prediction model. First, we find the projection of longitude gene expression most correlated with the outcome variable. Second, we compute the Cox score for each feature using the projected data. Third, we form a reduced data matrix consisting of only those features whose score exceeds a threshold. Fourth, we compute the first or first few principal components of the reduced data matrix to form the predictors. Finally, we use these reduced principal components in a Cox proportional hazards model to make predictions.

Results:

Our results on simulated data and real trauma data show that the use of the longitudinal structure in the data yields significant improvement in prediction. On simulated data, the method achieved a very significant log-rank test ($p\text{-value} < 0.001$) in prediction, which was much better than using individual time point data ($p\text{-value} \sim 0.1$). On the trauma dataset of three time points (days 0, 1, and 4) after injury, the log-rank $p\text{-value}$ of prediction was 0.004 using the proposed method. In comparison, the $p\text{-value}$ was about 0.02 using individual time point data. In addition, we observed that the longitudinal prediction method yields more stable results.

Conclusion: We have presented a method for utilizing temporal gene expression patterns for survival-type of prediction of patient clinical outcomes. In our analysis of simulated data and trauma patient data, we showed that better and more robust predictions of outcomes can be achieved using time course gene information.



BCATS 2008 Poster Abstracts

poster
1**The ligand-free state of the TPP
riboswitch*****Mona Ali******Jan Lipfert******Sebastian Doniach******Daniel Herschlag*****Purpose:**

Riboswitches are RNA molecules which occur in the 5' UTR region preceding a gene, and which control expression of the gene in response to the presence of a specific small metabolite [1]. Upon binding of the metabolite, the riboswitch undergoes a structural change which turns gene expression on or off via different mechanisms, including transcription termination, translation inhibition and splicing control. By definition, the structures of both the ligand-bound and ligand-free states of the riboswitch are essential to its functionality, and although the ligand-bound state of many riboswitches has been elucidated through x-ray crystallography, their ligand-free state has not been well characterized. Here, we present atomic models of the ligand free state of the Thiamine Pyrophosphate (TPP) riboswitch, stabilized by physiological concentrations of Mg^{2+} , derived from measurements using small angle x-ray scattering (SAXS).

Methods:

By measurement of SAXS from separate RNA fragments corresponding to the junctions of this riboswitch, we obtain structural models of these individual junctions. We stitch together the structures of the junctions to form atomic scale models of the ligand free state of the whole molecule.

Results:

We tested our atomic scale models of the whole molecule by comparing a theoretically derived SAXS profile for the models against experimental data that we collected for the riboswitch in the presence of Mg^{2+} [2]. Using SAXS as a filter, we are able to isolate a number of possible models for the ligand-free state.

Conclusions:

Application of geometrical constraints derived for base pairing in RNA molecules allows us to provide a model of the conformational changes which occur between the open state, which we characterize, and the TPP-bound state, the structure of which is known from x-ray crystallography [3].

References:

- [1] W. Winkler, A. Nahvi, R. R. Breaker, *Nature* 419, 952 (2002)
- [2] D. Svergun, C. Barberato, M. H. J. Koch, *J. Appl. Cryst.* 28, 768 (1995)
- [3] S. Thore, M. Leibundgut, N. Ban, *Science* 312, 1208 (2006)

Bayesian comparison of Markov models of molecular dynamics

poster
2

**Sergio
Bacallado**

Vijay Pande

Purpose:

Markov models are essential to the analysis of complex conformational dynamics. Building a Markov model requires a simultaneous discretization of time and space, after observing a number of trajectories. Much work has been devoted to the automatic discovery of metastable states, and to determining the correct lag time at which to coarse-grain the time domain [1,2]. However, the validation of these algorithms is based largely on consistency checks of the Markov property, which make use of various spectral functions of the transition matrix. These tests are sensitive to statistical error, and make it difficult to compare models with different numbers of states. We develop a probabilistic score to compare Markov models at a fixed lag time.

Materials and Methods:

To build Markov Models, state space must first be finely discretized into a large number of microstates. We used techniques developed in the references in order to lump the microstates into macrostates, such that the dynamics projected onto the macrostates are Markovian.

We define a hierarchical stochastic model for the generation of microstate sequences, given the macrostate definitions. We then compare different lumpings using Bayes factors. The hierarchical model depends on a prior for the parameters of the Markov chain. We compare the performance of a common prior for general Markov chains, to a novel prior for reversible chains introduced recently by Diaconis and Rolles [3].

Results:

We test the validity of our score in three different systems that exhibit metastable dynamics: (a) An illustrative Metropolis-Hastings simulation on an artificial 2D potential, (b) A molecular dynamics simulation of alanine dipeptide, and (c) A molecular dynamics simulation of the helical Fs-peptide. In each case, the models selected by our score are in accordance with previous predictions, or physical arguments based on the nature of the potential.

Conclusion:

We have introduced a method that can compare Markov models of metastable dynamics, even when they have different numbers of macrostates. The method incorporates the knowledge of detailed balance, which in some of the cases studied made a significant difference. Furthermore, it allows us to use highly parallel, non-equilibrium data for model validation, in contrast with consistency checks based on spectral properties of the transition matrix.

-References:

- [1] J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope, *Journal of Chemical Physics*, 126, p.155101 (2007).
- [2] W. Swope, J. Pitera, and F. Suits, *Journal of Physical Chemistry B*, 108, p.6571 (2004).
- [3] P. Diaconis and S. W. W. Rolles, *Annals of Statistics*, 34, p.1270 (2006).

poster
3

A Sequential Algorithm for Generating Random Graphs

Mohsen Bayati**Jeong Han Kim****Amin Saberi****Purpose:**

The focus of this research is to develop a new tool for detecting motifs in biological networks (e.g. transcription networks). Discovery of network motifs helps to find certain patterns of interconnects that serve as building blocks of the biological networks. The goal is to characterize the dynamics of the entire networks based on the dynamics of some smaller building blocks (see [1] for more details).

In order to detect meaningful patterns in a large transcription network, a standard approach is to compare the statistical significance of a pattern in the real network versus a randomized one [1]. Unfortunately, generating a randomized network that is very close to the real network can be a challenging task. For example, generating a random network that has the same degree sequence as the real network has been studied extensively as an interesting theoretical problem (see [5] for detailed surveys). The best algorithm for generating a graph with m edges and maximum degree d was given by McKay and Wormald [2] that produces random graphs with uniform distribution in $O(m^2 d^2)$ time. However, this running time can be slow for the networks with very large number of edges. This has constrained practitioners to use simple heuristics that are non-rigorous and have often led to wrong conclusions [3, 4].

Results:

Let d_1, \dots, d_n be non-negative integers that add up to $2m$. Our algorithm for generating a graph with degree sequence d_1, \dots, d_n works as follows: start with an empty graph and sequentially add edges between the pairs of non-adjacent vertices. In every step of the procedure, the probability that an edge is added between two distinct vertices i and j is proportional to $d'_i d'_j (1 - d_i d_j / 4m)$ where d'_i and d'_j denote the remaining degrees of vertices i and j . We will show that this algorithm produces an asymptotically uniform sample with running time $O(md)$ when $d = O(m^{1/4})$.

Conclusions:

We designed and analyzed a very efficient algorithm that can be useful for motif detection. We also believe that our approach can be useful to design random networks with further constraints. For example, networks with a fixed distribution of certain sub-networks.

URL: <http://simula.stanford.edu/~mohsen/papers/randgraphext.pdf>

References:

- [1] U. Alon, An Introduction to System Biology: design principles of biological circuits, 2006, Chapman & Hall/CRC.
- [2] B. McKay and N. C. Wormald, Uniform generation of random regular graphs of moderate degree, 1990, J. Algorithms 11, 1, 52-67.
- [3] R. Milo, N. Kashtan, S. Itzkovitz, M. Newman and U. Alon, On the uniform generation of random graphs with prescribed degree sequences, 2004,
- [4] R. Milo, S. ShenOrr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon, Network motifs: Simple building blocks of complex networks, 2002, Science 298, 824-827
- [5] N. C. Wormald, Models of random regular graphs, 1999, Surveys in combinatorics, London Math. Soc. Lecture Notes Ser., Vol 265. Cambridge Univ. Press, Cambridge, 239-298.

Adaptive Seeding: A New Method for Simulating Biologically Relevant Timescales

Gregory R. Bowman

Xuhui Huang

Yuan Yao

Jian Sun

Vijay S. Pande

Purpose:

Molecular Dynamics (MD) simulations may complement experiments by allowing the evolution of molecular systems to be monitored with atomic resolution. However, it is currently only possible to simulate a few dozen nanoseconds of molecular dynamics in a month whereas most biologically relevant processes take a few microseconds, or even seconds. Simulating a microsecond would take about a year so the utility of MD is limited. Our new adaptive seeding method, on the other hand, is able to completely sample processes that take tens of microseconds experimentally in a single week. We can then predict experimental observables like folding free energies and explain the chemical details underlying processes like folding. Thus, we are able to bring the explanatory and predictive power of MD to bear on biological phenomena in an unprecedented way.

Methods:

First, thousands of short Simulated Tempering (ST)[1] simulations are run in parallel. ST simulations perform a random walk in temperature space allowing broad and rapid sampling of the entire accessible space. These non-equilibrium simulations are sufficient to identify all the states the molecule visits by building a Markov State Model (MSM)[2] using topological methods developed in the computer science department[3]. Many short constant temperature MD simulations are then started from each state in a process called seeding and a new MSM is built that captures the true equilibrium distribution of the system. Convergence is ensured by comparing the state populations predicted by two independent sets of simulations with very different initial configurations. The entire procedure may be repeated until convergence.

Results:

We have applied the adaptive seeding method to a small RNA system that folds in about ten microseconds. One set of simulations was started from the folded state determined by NMR while another was started from a random coil. The models obtained from each set of simulations agree within error and, therefore, represent the true equilibrium distribution of the molecule. This sampling was achieved in about a week and identifies all the dominant folding pathways whereas traditional MD would have taken over a year to simulate a single folding event.

Conclusions:

We are now able to rapidly simulate experimental timescales, allowing MD to reach new levels of explanatory and predictive power.

References:

1. Marinari, E. and G. Parisi. Europhysics Letters, 1992. 19(6):451-458.
2. Noe, F. and S. Fischer. Current Opinion in Structural Biology, 2008. 18(2):154-162.
3. Bowman, G.R., et al. J Am Chem Soc, 2008. 130(30):9676-8.

poster
5

**Sophie
Candille**

Devin Absher

Mark Shriver

Rick Myers

Greg Barsh

Hua Tang

Genetic architecture of normal phenotypic variation in European populations

Purpose:

Human appearance is highly variable within and between populations but the genetic basis of this variation is mostly unknown. What is the genetic architecture underlying normal phenotypic variation? Polymorphisms in which genes determine this variation? Do these polymorphisms have large or small effects on phenotypic variation? Are the same polymorphisms fixed in all populations sharing similar traits or have similar traits arisen independently? Here I use a combination of high-throughput genotyping and statistical analyses to begin to answer these questions for human pigmentation and facial features.

Material and Methods:

DNA and phenotypic information were obtained from several hundred Europeans living in Ireland, Poland, Italy and Portugal. The phenotypic data include precise measurements of skin, hair and iris pigmentation as well as three-dimensional facial pictures. One hundred ninety facial distances were measured on each 3D picture and these distances were analyzed by principal components analysis (PCA). 180 women, 45 from each country, were selected to be genotyped at 317,000 single nucleotide polymorphisms. I performed quality control of the genotypic and phenotypic data, followed by a study of the structure of the population using PCA. Next I conducted genome-wide association studies (GWAS) to identify the genetic determinants of the variation in pigmentation and the first five principal components of facial distance variation in these four European populations.

Results:

I found that the first two components of genetic structure separate the 180 participants according to their country of origin. Facial features and even more so pigmentation phenotypes are stratified along the main axes of genetic variation. This result means that it will be necessary to carefully adjust for population stratification in the GWAS. Preliminary results for the GWAS of pigimentary traits indicate that both known and novel pigmentation genes may underlie pigmentation variation in Europe.

Conclusion:

This work makes progress towards the characterization of the genetic architecture of normal phenotypic variation in pigmentation and facial features. It also establishes pigimentary traits as ideal test cases for the study of complex traits whose genetics is confounded by group ancestry. I plan to use the European pigmentation phenotypes and the genotypes collected in this study to better characterize the bias introduced by population stratification in association studies.

Toward Patient-Specific Simulation of Endoscopic Sinus Surgery

Purpose:

Endoscopic sinus surgery is a technically challenging procedure with a high risk of complication due to the proximity of critical structures. A thorough understanding of the patient's specific anatomy is required to achieve a successful outcome. High-resolution, multi-planar computed tomography (CT) images can provide the surgeon with detailed anatomical information for pre-operative planning. However, the surgeon must make the conceptual leap from viewing cross-sectional images to the anatomy encountered in the surgical field, an inexact process even for experienced surgeons. Our goal is to provide a means for the surgeon to view and interact with pre-operative patient data in a surgically meaningful manner.

Materials and Methods:

A virtual surgical environment capable of directly loading pre-operative CT images to construct a virtual model of the patient was developed. Endoscopic views of the virtual patient from within the nasal cavity are rendered in real-time using a ray-casting approach implemented on commodity graphics processors. A hybrid rendering approach [1] incorporating both isosurface and direct volume rendering was chosen to reproduce the mucosal surface and the underlying bone and sub-mucosal tissue as accurately as possible. The surgeon manipulates the virtual endoscope or surgical tool intuitively using a Phantom Omni haptic interface device. Contacts between the tool and the patient's anatomy are detected, and resistive force feedback is also rendered in real-time [2].

Standard pre-operative sinus CT scans were obtained from two patients who underwent endoscopic sinus surgery. Virtual endoscopic images gathered by exploring the patient's nasal cavity in the simulation environment were compared to endoscopic views seen in the intra-operative video of the same patient for an initial validation of the tool.

Results:

The virtual surgical environment showed high-fidelity anatomical correlation between virtual and intra-operative findings for both patients. Structural similarities and correlation of anatomical landmarks were clearly identified in the comparison of the virtual and intra-operative endoscopic views. Haptic feedback constraining the position of the instrument was felt to be reasonably realistic in guiding the surgical tool.

Conclusion:

Endoscopic sinus surgery can be a formidable challenge for the practicing rhinologist and a daunting task for residents in training. The ability to plan and rehearse the procedure on an accurate virtual model of the patient can allow the surgeon to be better prepared and more confident in the operating room.

References:

1. Scharsach, H., et al. (2006). EuroVis, 315-322.
2. Salisbury, J.K. and Tarr, C. (1997). ASME, 61-67.

poster
7

Rong Chen

Atul J. Butte

FitSNPs: Differentially Expressed Genes are More Likely to have Variants Associated with Disease

Purpose:

With more than 200,000 microarray studies and more than 10,000 disease-associated DNA variants in the public repositories, we hypothesize that there is a general association between gene expression and disease-associated variants.

Results:

We analyzed all human microarray studies from Gene Expression Omnibus (GEO), identified differentially expressed genes, and compared them with a list of curated human genes with disease-associated variants. We found that 99% of disease-associated genes were differentially expressed in one or more microarray experiments with 14% specificity. Furthermore, the more experiments in which a gene was differentially expressed, the more likely it contained disease-associated variants. We then derived a list of functionally interpolating SNPs (fitSNPs), and successfully prioritized type 1 diabetes genes from the top 7 loci of Wellcome Trust Case Control Consortium genome-wide association study (GWAS) with 75% sensitivity and 89% specificity.

Conclusion:

Our study demonstrates that differentially expressed genes are more likely to have variants associated with disease. FitSNPs can serve as an effective tool to systematically prioritize candidate SNPs in the identified loci of more than 100 published GWAS.

URL: <http://fitsnps.stanford.edu/>

18F-FDG Uptake within the Spinal Canal Follows a Predictable Pattern

**Harpreet
Dhatt**

**Brian Kim
Jarrett Rosenberg
Edward Graves
Deepak Behera
Sandip Biswal**

Purpose:

Understanding physiologic spinal cord glucose metabolism may provide insight regarding spinal cord diseases such as myelitis, trauma, and malignancy. We have refined characterized of the spinal cord and cauda equina by taking measurements at every slice of the whole body PET/CT through the length of the spine.

Method and Materials:

Approval was attained by our IRB. A retrospective review of 13 negative whole body 18F-FDG PET/CT studies with non-CNS cancers was performed. Studies with vertebral marrow hyperplasia, severe spinal arthritis, cervical/thoracic kyphosis or motion artifact had been previously excluded. Using the transaxial CT to define the spinal canal, oval region of interests (ROIs) of fixed dimensions were placed within the canal and excluded the bony elements of the spine. Corresponding PET measurements were obtained at every slice of the study C1 to S1. ROI measurements included average standard uptake values (SUVavg), which were standardized against L5 where minimal neural tissue was present. Data was analyzed using RT Image analysis software and statistical package Stata. Significance was $p < 0.005$.

Results:

A significant decreasing cranial-caudal vertebral SUV trend with linear and quadratic effects relative to vertebral location has been identified. Standardized de-trended residuals from the regression were regressed on the individual vertebral segments, with L2 as the reference segment (because its mean residual was closest to zero). T12 was found to have a significantly different SUV.

Conclusion:

A predictable decreasing cranial-caudal pattern of 18-FDG uptake with a focal increase at T12 is observed in the spinal canal subjects. The nature of this pattern is unknown at this time, but may be related to relatively higher levels of grey matter and contributions from lower extremity peripheral nerves at this level.

poster
9

Increased Surface Friction Alters Risk of ACL Injury in Healthy Subjects

Ariel Dowling**Stefano Corazza****Ajit Chaudhari****Thomas Andriacchi****Purpose:**

Gender is a risk factor for ACL injury, as females have a 4 to 6 times greater rate of injury [1]. Specific kinematic and kinetic measures during landing, including a decreased knee flexion angle and an increased abduction moment, have been reported as increasing the risk for ACL injury [3]. Additionally, an increased coefficient of friction (COF) of the floor surface leads to increased incidence of ACL injury during sports involving run to cut maneuvers [5]. This study tested the hypotheses that increased floor friction would alter a subject's movement in the variables associated with risk of ACL injury, and these changes will be altered more in females than males.

Materials and Methods:

22 healthy subjects (11 male, 11 female) were evaluated in an IRB approved study with informed consent. Subjects performed a 30° cut off of their dominant leg under two conditions; on a low friction surface (COF = 0.38) and on a high friction surface (COF = 0.99). A markerless motion capture system and two force plates were used to measure full body kinematics and kinetics from video captured from eight cameras [2,4]. The data were examined at foot contact during the landing of the dominant leg.

Results:

Increasing the friction of the cutting surface caused a significant decrease in knee flexion angle at foot contact. Females maintained a significantly lower knee flexion angle on each surface compared to the males. Increased surface friction was also associated with an increase in the knee abduction moment. Men had a small increase and women had a large increase in abduction moment between friction surfaces.

Conclusion:

These results suggest that subjects anticipate the COF and change their landing mechanics accordingly. The subjects seem to be more at risk for an ACL injury on the high friction surface because they had a decreased knee flexion angle and an increased knee abduction moment. Therefore, the higher incidence of ACL injuries on high friction surfaces could be a result of these specific kinetic and kinematic changes.

Females appear to adopt a riskier landing strategy on the high COF surface compared to males because they had a greater increase in knee abduction moment. Additionally, females landed in a more extended position compared to males on both surfaces. Both of these results suggest that females are more at risk for an ACL injury on a high friction surface.

References:

1. Arendt E, Dick R. (1995) Am J Sports Med, 23(6): 694-701.
2. Corazza S et al. (2006) Annals Biomed Eng 34(6): 1019-1029.
3. Hewett TE et al. (2006) Am J Sports Med, 34(2): 299-311.
4. Muendermann L et al. (2006) J Neuroeng Rehab, 3(6).
5. Orchard J et al. (1999) Med J Aust, 170(7): 304-6.

Fast modeling of RNA structure and dynamics using internal coordinates

Purpose:

RNAs play an important, possibly dominant role in gene regulation. However their structure and dynamics have proven difficult to solve by experimental or theoretical means. Genetics and pharmacology come down to individual molecules, and therefore both the causes and the treatment of disease will be better understood when structure and dynamics are accurately modeled. In particular, there is a need for a simple tool to build up putative RNA structures from easily obtained information, and then determine their conformational dynamics.

Materials and Methods:

The recently announced Simbody code library permits unprecedented control over the physics, flexibility, forces, and constraints acting on molecules or parts of molecules. We have created an application using this library which uses the physics of RNA as well as user-provided base pairing and stacking interactions to determine three-dimensional structure and dynamics.

Results:

The method successfully reproduces the structure of transfer RNA and the conformational range of HIV Trans-activation Response Element.

URL: <http://www.simtk.org>

Conclusion:

The results suggest a wide variety of applications in the understanding of basic biology and disease.

*Samuel
Coulbourn
Flores*

*Chris Bruns
Peter Eastman
Russ Altman*

poster
11

Mechanisms of improved knee flexion after rectus femoris transfer surgery

Melanie D. Fox**Jeffrey A. Reinbolt****Sylvia Öunpuu****Scott L. Delp****Purpose:**

One of the most common gait problems in children with cerebral palsy is the inability to sufficiently flex the knee during the swing phase of walking, or “stiff-knee gait” [1,2]. Surgical transfer of the rectus femoris muscle is a common treatment; however, individual patient outcomes vary substantially [3]. It is difficult to improve surgical outcomes because the mechanism by which some patients improve is unknown. Traditionally, it has been thought that transferring the muscle converts it from a knee extensor, which inhibits knee flexion, to a knee flexor. However, experimental studies have found that the muscle produces a knee extension moment even after transfer [4], possibly due to scarring to underlying tissue [5]. The purpose of this study was to investigate the mechanism by which the transferred muscle improves knee flexion and to compare different surgical techniques.

Methods:

Muscle-actuated dynamic simulations were created of ten children with cerebral palsy and stiff-knee gait. These simulations were altered to represent surgical transfers of the rectus femoris to the sartorius and the iliotibial band. Rectus femoris transfers in which the muscle remained attached to the underlying vasti through scar tissue were also simulated. We compared simulated improvements in peak knee flexion in swing from different transfer simulations to determine how knee flexion is improved after transfer surgery.

Results:

Both transfer to sartorius and transfer to iliotibial band simulations predicted greater ($p < 0.001$) improvement in peak knee flexion ($32^\circ \pm 8^\circ$ and $28^\circ \pm 8^\circ$, respectively) than the scarred transfer simulation ($14^\circ \pm 5^\circ$).

Conclusion:

Simulations revealed that the mechanism for improvement in knee flexion after surgery is reduction of the muscle’s knee extension moment. Our results suggest that significant improvement in knee flexion can be attained even when the extension moment of the rectus femoris is only reduced by half, as may occur in a scarred transfer, supporting the finding that the muscle is not converted to a knee flexor after transfer. Our simulated knee flexion improvements after scarred transfer are consistent with reported surgical improvements following isolated transfer [6]. Simulated non-scarred transfers to iliotibial band and sartorius predict greater improvements in knee flexion, suggesting methods to reduce scarring may augment knee flexion.

URL: <http://stanford.edu/~melanief/BCATS08.pdf>

References:

1. Wren T, et al., 2005. J Ped Ortho 25, 79-83.
2. Sutherland D, Davids J, 1993. Clin Orth Rel Res 288, 139-147.
3. Yngve D, et al., 2002. J Ped Ortho 22, 672-676.
4. Riewald S, Delp S, 1997. Dev Med Child Neur 39, 99-105.
5. Asakawa D, et al., 2004. JBJS 86, 348-354.
6. Hemo Y, et al., 2007. J Child Orth 1, 37-41.

Increasing the Efficiency of Genomewide Association Mapping via Hidden Markov Models

Hong Gao

Hua Tang

Carlos D. Bustamante

Purpose:

With the rapid production of high dimensional genetic data, one major challenge in genome-wide association studies is to develop effective and efficient statistical tools to resolve the loss-of-power problem caused by performing multiple statistical tests across millions of SNPs simultaneously.

Methods:

We present a novel method that serves as an effective measure for balancing power and type I error in genome-wide association studies. The approach uses hidden Markov model and its derivate Markov hidden Markov model to estimate the posterior probabilities of a markers being in an associated state.

Results:

We conducted extensive simulations based on the human whole genome genotype data from the Glaxo-SmithKline project to calibrate the sensitivity and specificity of our method and compared to many popular approaches for correcting multiple tests including Bonferroni correction, False Discovery Rate or positive False Discovery Rate control and Q-value control. Our simulation results suggested that at comparable low false positive rates, our method is significantly more powerful than many other approaches, given moderate susceptibility of disease-predisposing variants. Application of our method to the data set generated by Wellcome Trust Case Control Consortium using 14,000 cases and 3,000 controls confirmed its powerfulness and efficiency under the context of the large-scale genome-wide association studies.

Conclusion:

Our method is much more powerful than many other approaches given moderate susceptibility of disease-predisposing variants.

poster
13

Yael Garten

Russ B. Altman

Prediction of Novel Pharmacogenomic Interactions via Text Mining and Network analysis

Purpose:

Pharmacogenomics studies the relationship between human genomic variation and the variation in drug response phenotypes. The field is rapidly gaining importance: it promises to personalize medicine through drugs targeted to particular subpopulations based on genetic background, and can be used to keep drugs in the development pipeline by targeting such subpopulations. The pharmacogenomics literature has expanded rapidly, but is dispersed in many journals. Due to this large amount of dispersed data and to the fact that pharmacogenomic associations are discussed in the literature in a variety of non-standardized ways, it is challenging to identify important associations between drugs and molecular entities--particularly genes and gene variants. Thus, these critical connections are not easily available to investigators or clinicians who wish to survey the state of knowledge for any particular gene, drug, disease or variant.

Natural Language Processing and text mining techniques can allow us to convert the free-style text to a computable, searchable format in which pharmacogenomic concepts (such as genes, drugs, polymorphisms, and diseases) are identified, and important links between these concepts are recorded. As it is unlikely to expect a shift towards structured reporting, free text is likely to continue to be the source of pharmacogenomic information as new discoveries are made.

Methods and Results:

In my research, I am creating a resource that will identify and gather information from the literature of relevance to pharmacogenomics, will automatically extract facts from the text (full text when available), and will allow large-scale analysis of the collective pharmacogenomic knowledge published in the various journals. This knowledge, in the form of interactions reported by researchers between genes, gene variants, drugs and diseases, will then be analyzed at a systems level. By creating networks of these interactions, we can then move forward to infer and predict additional connections that may exist.

Conclusions:

Analyses of the network may suggest new drug targets, new applications for existing drugs, and areas of Pharmacogenomics that appear under-investigated. We have the unique capability to evaluate our methods using high quality manually curated data sets.

Muscle Contributions to Propulsion and Support During Running

poster
14

**Samuel
Hamner**

**Chand John
Scott Delp**

Introduction:

Understanding how muscles propel our body forward and support our body weight during a movement such as running is challenging because important variables, such as muscle forces, are generally not measurable. Muscle-driven simulations of walking have been analyzed to gain insight into muscle actions [1,2]. However, simulating running presents new challenges due to high accelerations and forces. The purpose of this study was to create a three-dimensional, dynamic simulation of running and quantify how muscles contribute to propulsion and support (i.e., the forward and upward accelerations of the body mass center) during running.

Methods:

Motion and ground reaction force data were measured on an adult male subject running at 3.9 m/s on a treadmill. The subject was represented by a 31 degree-of-freedom musculoskeletal model with the legs and back driven by 92 muscles [3,4] and arms [5] driven by torque actuators. The muscle excitations needed to drive the model to track the measured motion were computed using OpenSim, dynamic simulation software described by Delp et al. [6]. To examine the contributions of individual muscles to propulsion and support, each muscle was independently perturbed during the simulated movement to determine its contribution to the acceleration of the subject's mass center [7]. We focused on the propulsive phase of running, which is defined as the time period when a foot is on the ground and the body mass center is accelerating forward.

Results:

The muscles contributing the most to both forward propulsion and upward support are gastrocnemius and soleus (i.e., the ankle extensors located in the calf). In contrast, the quadriceps muscle group acted to strongly resist forward progression, and hamstrings acted to pull the body mass center downward.

Discussion:

The ankle extensors are a key muscle group that both propel our body forward and support our body weight during the propulsion phase of running. Creating the first three-dimensional, muscle-driven simulation of running has allowed us to analyze specific muscle actions and answer the fundamental question of which muscles propel our body forward and support our body's weight. Future simulations could give us insights into better training techniques and help us better understand different running styles, such as sprinting.

URL: <http://www.stanford.edu/~samner/hamnerBCATS.html>

References:

1. Liu et al. (2008) J Biomech in press
2. Neptune et al. (2004) Gait Posture 19:194-205
3. Delp et al. (1990) IEEE Trans Biomed Eng 37:757-767
4. Anderson and Pandy (1999) Comp Meth Biomech Biomed Eng 2:201-231
5. Holzbaur et al. (2005) Annals Biomed Eng 33:829-840
6. Delp et al. (2007) IEEE Trans Biomed Eng 55:1940-50
7. Liu et al. (2006) J Biomech, 39:2623-2630

poster
15

Lan Hua

Ruhong Zhou

Dave Thirumalai

Bruce J. Berne

A two-stage urea induced unfolding through its stronger dispersion interactions with proteins than water

Purpose:

To explore the molecular mechanism by which urea denatures proteins and which of the forces is dominant and how denaturation occurs

Materials and Methods:

We use all-atom microseconds molecular dynamics simulations of hen egg-white lysozyme in 8M urea, generated on IBM BlueGene/L computer, to analyze the denaturation by urea which has been observed experimentally and in simulations.

Results:

We find that urea accumulates around lysozyme which expels water molecules from the first hydration shell of the protein. A two stage penetration of the protein is observed, with urea penetrating the hydrophobic core before water, forming a “dry globule”. The direct dispersion interaction between urea and the protein backbone and sidechains is stronger than for water, which gives rise to the intrusion of urea into the protein interior and also to urea’s preferential binding to all regions of the protein, making urea as a “surfactant” that further solubilizes the sidechains in water. This is augmented by preferential hydrogen bond formation between the urea carbonyl and the backbone amides which contributes to the breaking of intra-backbone hydrogen bonds.

Conclusion:

Our study supports the “direct interaction mechanism” whereby urea has a stronger dispersion interaction with protein than water, which implies a two-stage unfolding of proteins by urea.

Hamming Codes in the Context of DNA Self-Assembly

**Kathryn
Hymes**

**Tiyu Wang
Aditya Mittal**

Purpose:

The process of self assembly is ubiquitous in nature by the spontaneous formation of molecular structures from the reactions of smaller components: atoms assemble into molecules, which can further aggregate and react to form crystals, larger biomolecules and organisms. DNA Self-Assembly seeks to utilize the fundamental properties of DNA toward the development of methods in artificial self-assembly for applications in nanotechnology and DNA computing. In the context of DNA self assembly, error detection and correction in DNA tiling systems pose critical problems from the viewpoint of any practical implementation. Coding theory offers a myriad of error correcting schemes that can be used to leverage the combinatorial properties of DNA toward improved accuracy in self assembling systems. This work presents a DNA tiling system that performs error detection in a self-assembling system via an implementation of Hamming Code verification.

Materials and Methods:

Simulations were coded in XGrow4, a DNA tile assembly simulator, and corroborate the robustness of the developed tiling model. Concepts in Coding Theory, namely Hamming Codes, were used as a basis for error detection. Hamming codes are a class of binary linear codes that are prevalent in information transmission. While a parity check sum over a binary string is limited to single bit error detection, Hamming codes generalize the basic approach to enable correction of single-bit errors and the identification of two bit errors.

Results:

In our work, we construct and simulate a DNA tiling system capable of detecting errors in a given Hamming code represented as a DNA 'seed' column. Difficulties stem from a number of sources, such as complexity management for the distinction of data and parity bits and information propagation throughout the system. From a starter seed column of DNA tiles representing the Hamming codeword to be verified, our tile system can determine correctness using just $O(1)$ tile types. In order to avoid keeping track of parity and data bit positions in the DNA Hamming code separately, we exploit the correspondence between the parity bit positions in the Hamming code and the columns of the binary counter previously developed Goel et al. This results in a 32 tile system which assembles into a $\log(L - D) \times L$ rectangle at temperature two with only single strength glues. The correctness of each parity check from the DNA Hamming Code is shown as one of two glues on the northern edge of the system.

Conclusion:

We see this as a promising potential mechanism for error correction and reduction in combinatorial DNA tile systems. Developing new error detecting algorithms will be essential for the continuing development of DNA computing.

poster
17

Matt Inlay

Deepta Bhattacharya

Debashis Sahoo

Tom Serwold

David Dill

Irv Weissman

Identification of the branchpoint between B and T cell development through MiDReG

Early hematopoietic development proceeds in a stepwise manner through progenitor intermediates, which sequentially lose potential for alternate lineages. One such intermediate, the common lymphoid progenitor (CLP), can give rise to B, T, natural killer, and dendritic cells but not myeloid cells in vivo, but has been suggested to preferentially give rise to B cells. Thus it remains unclear whether CLPs represent a node where lymphoid fate decisions are made, an intermediate en route to B cell commitment, or a mixture of two such populations.

To determine if the CLP population contains a mixture of both B-lineage-committed and uncommitted lymphoid progenitors, we used MiDReg analysis (described in accompanying manuscript) of over 2000 publicly-available mouse expression arrays to predict surface proteins with differential expression during the transition from fully multipotent progenitors to committed B cells. One of these markers, Ly6d, bisects common lymphoid progenitors (CLPs) into two distinct populations. Upon adoptive transfer, Ly6d⁻ CLPs have robust B, T, natural killer, and dendritic cell output, while Ly6d⁺ CLPs produce almost exclusively B cells, even after intrathymic injection. E2A-deficient animals, which lack mature B cells, also lack Ly6d⁺ CLPs, yet retain significant numbers of the uncommitted Ly6d⁻ uncommitted CLP population. Thus, Ly6d expression marks the earliest identified step at which B cell commitment is established in vivo, with Ly6d⁻ CLPs residing at the branchpoint where the B cell lineage separates from the other lymphoid fates. These results highlight the utility of MiDReg to identify novel stages and mechanisms of hematopoietic development, and demonstrate the existence of a B cell-unbiased CLP.

Multiphoton and Transmission Electron Microscopy of Collagen in Ex Vivo Tympanic Membranes

poster
18

**Ryan P.
Jackson**

**Cara Chlebicki
Tatiana B. Krasieva
Reena Zalpuri
William J. Triffo
Sunil Puria**

Purpose:

We are investigating the transmission of high-frequency sound from the eardrum to the cochlea. It is believed that in larger mammals the malleus switches from a hinging motion at low frequencies to a torsional motion at high frequencies. If true, this would enable the efficient transfer of sound at high frequencies because the torsional mode has a lower moment-of-inertia, which overcomes attenuation due to mass. Corroborative morphological differences in eardrum symmetry and malleus-incus complex articulation across several mammalian species have been observed. We speculate that further important differences might be observed in the composite structure of the eardrum, which consists of highly structured collagen layers surrounded by soft tissue, and the connection of the eardrum with the malleus.

In this presentation, we will show our preliminary results using multiphoton microscopy (MPM) and transmission electron microscopy (TEM) to perform detailed imaging of the eardrum and its attachment to the malleus. Our goal is to further understanding of its composite-structural properties as they relate to a torsional mechanism. To achieve this, we will attempt to develop a three-dimensional map of collagen fiber micro- and ultra-structure in the eardrum through these imaging modalities.

Materials and Methods:

Ex vivo, human eardrums were acquired from the Palo Alto VA hospital within one-day post mortem. The excitation wavelength for all MPM imaging was 920 nm. This enabled simultaneous collection of two-photon fluorescence (TPF), isolated at 535-590 nm, and second harmonic generation (SHG), isolated at 435-485 nm, light for imaging. SHG can generate images of collagen molecules specifically, so imaging of the collagen layers versus the other biological material was possible. TEM imaging was carried out according to conventional protocols using glutaraldehyde fixed samples.

Results:

We have imaged ex vivo human eardrums by MPM (N=13) and by TEM (N=3). Collagen fibrils and fibers are clearly visible in many of the images and follow the essential patterns of collagen structure expected.

Conclusions:

We obtained images of collagen structures in ex vivo human tympanic membrane samples through MPM and TEM imaging. Our results appear to conform to the small number of comparable images in the literature.

URL: <http://www.stanford.edu/~puria1/BCATS08.html>

References:

- Fay, J., Puria, S., and Steele, C. R., "The Discordant Eardrum," Proc. Nat. Acad. Sci. USA, 103(52), 19743-19748 (2006).
- Lim, D. J., "Structure and Function of the Tympanic Membrane," Acta Oto-rhino-laryngologica Belg., 49, 101-115 (1995).
- Raub, C. B. et al, "Noninvasive Assessment of Collagen Gel Microstructure and Mechanics Using Multiphoton Microscopy," Biophys. J., 92, 2212-2222 (2006).

poster
19

Bridging low and high resolution RNA structure modeling

Magdalena A. Jonikas

Russ B. Altman

Purpose:

Large RNA molecules, such as ribozymes and the ribosome, can adopt 3D structures that perform catalysis of chemical reactions. Understanding the functions of these complex molecules depends critically on understanding their structures. Creating 3D structure models of RNA molecules is a significant challenge that has been approached from two directions. Modeling RNA structures by coarse graining simplifies the computational complexity of the problem, which allows modeling of large RNA molecule with limited accuracy [1]. In contrast, using full atomic representations allows physically accurate modeling that is severely limited by computational complexity [2]. We present a method that bridges these two approaches by creating full atomic structures from coarse grain models. This allows output from coarse grain modeling to be used as input for full atomic modeling.

Methods:

Starting with a coarse grain model of an RNA molecule, we break the structure into fragments. We search for fragments with similar coarse grain geometry in a library of full atomic RNA structures, resulting in an ensemble of full atomic matches to each fragment. We randomly select one match for each fragment and combine them to make a full atomic structure. These structures have a high probability of containing collisions between atoms, so we use a Metropolis Monte Carlo algorithm to search for a combination of fragments with no collisions. The final structure has full atomic resolution and is nearly identical to the input structure at the coarse grain level.

Results:

We successfully created full atomic structures from coarse grain RNA crystal structures stripped of their full atomic detail. We match the original full atomic structures with a root mean squared deviation (RMSD) of less than 3.0 Å for five structures ranging in size from 76 to 244 residues. We also successfully generated full atomic models from coarse grain predictions of two RNA molecules. The full atomic models remain within 1 Å RMSD of their coarse grain template, and the RMSD values to the crystal structure do not change by more than 1 Å RMSD with the addition of full atomic resolution. We also demonstrate that the full atomic structures we generate can be successfully used as input for full atomic physics based molecular dynamics.

Conclusion:

We have developed a method that bridges coarse grain and full atomic approaches to RNA structure modeling. Our method will allow RNA modelers to take advantage of the strengths of both approaches and combine them for improved RNA 3D structure modeling.

References:

- [1] Jonikas, M.A. et al., submitted to RNA.
- [2] Major, F. et al., PNAS, 1993. 90(20).

Motion compensated volumetric modulated arc therapy (4D VMAT)

Orr Keshet

*Louis Lee
Yunzhi Ma
Lei Xing*

Purpose:

4D CT provides temporal information that can be explicitly incorporated in modulated arc therapy dose optimization to accommodate temporal changes in anatomy that occur during treatment. Here we present a concept and proof of feasibility of organ motion compensated VMAT (4D VMAT) optimization.

Methods:

A cylindrical digital phantom with a moving target was designed. The involved structures move cyclically in 10-phase with the target movement range and period of 1.5 cm and 6 sec, respectively. The 4D VMAT plan was constrained to a single pass of 360 degree gantry rotation. The target cyclic motion trajectory was replicated to form a 'breathing' curve to which a correlation with the full gantry rotation was established. A reference phase was chosen to which all spatial and dosimetric information from other phases were mapped onto through the use of a deformable registration model, and the inverse planning goals were defined in this reference phase. The full rotation was initially modeled as a series of fixed source positions for every 5 degrees starting from 0 degrees. The initial aperture at each gantry angle was chosen to be the beam's eye view of the target at the phase determined from the phase/gantry-angle correlation. An aperture-based optimization using simulated annealing algorithm was adopted to optimize a weighted least-squares cost function with volumetric constraints. The dose deposited for each angle was mapped to the reference phase for the calculation of cumulative dose and assessment of the cost value. The process terminated when the cost function converged. The resultant 4D dose from the complete rotation was calculated by mapping the dose deposited at various angles (correlated to different phases) onto the reference phase. The 4D VMAT plan and dose volume histograms were studied and compared with those of the 3D VMAT plans optimized with the target volume encompassing the whole motion range. Two lung cases with 4D CT scans were also selected and tested using the developed 4D VMAT procedure.

Results:

A motion compensated VMAT plan optimization framework has been formulated that enables arc delivery with a 100% duty cycle. The 4D VMAT plan of both the phantom and lung cases yielded significantly improved coverage of the target and sparing of the surrounding normal tissues when compared to the 3D VMAT plan.

URL: http://mips.stanford.edu/research/lab?lab_id=2789

Conclusions:

4D VMAT is feasible based on (1) the aperture-based optimization that takes 'time' as a parameter expressed in a form of phase/gantry-angle correlation and (2) the spatial/dosimetric mapping between phases using deformable registration. Real-time image guided adaptation of the 4D VMAT plan to account for a possible breathing irregularity is in progress. The current study helps in assessing the degree of improvement of 4D VMAT over its 3D counterpart and providing a theoretical benchmark for 4D VMAT.

poster
21**Hyun Jin
Kim****J. Shihj
I. E. Vignon-
Clementel
K. E. Jansen
C. A. Taylor****Computation of coronary flow using a closed loop model of the cardiovascular system with a three-dimensional fluid-solid interaction model of the thoracic aorta****Purpose:**

Unlike flow in the other arterial system, coronary blood flow is low in systole when the ventricles contract, which increases intramyocardial pressure. Coronary flow increases in diastole when the ventricles relax, thereby decreasing the intramyocardial pressure. We developed a closed loop system comprised of a three-dimensional finite element model of the aorta with coronary arteries and major upper branch vessels, lumped parameter heart models representing the left and right sides of the heart, a three-element Windkessel model representing the pulmonary circulation, and computed the left and right ventricular pressures and coronary flow and pressure by considering the interactions between these computational and analytic models. We represent the intramyocardial pressure of the coronary vascular beds using the computed left and right ventricular pressures. Additionally, we couple a lumped parameter coronary vascular bed model to each coronary outlet to represent the impedance of the downstream coronary vasculature.

Materials and Methods:

The Coupled Multidomain Method [1] was used to couple a lumped parameter coronary vascular bed model [2] to each coronary outlet. Parameter values of the model were adjusted to represent subject-specific coronary flow and pulse pressure. The intramyocardial pressure was approximated with either the left or the right ventricular pressure which was computed from the closed loop system. To model the contraction and relaxation of the ventricles, time varying elastance function which was adjusted from a normalized elastance function [3] to represent subject-specific heart rate, cardiac output and pulse pressure was utilized. The shapes of the inlet, major upper branch vessels and the aortic outlet were constrained using an augmented Lagrangian method to achieve a robust algorithm [4]. A stabilized finite element method was used to solve for blood velocity, pressure and the wall displacements of the three-dimensional model of the thoracic aorta, coronary arteries and major upper branch vessels.

Results:

Computed coronary flow and pressure as well as the ventricular pressure and the aortic flow and pressure are physiologically realistic. In particular, coronary flow show low flow in systole when the intramyocardial pressure is high, and high flow in diastole when the intramyocardial pressure is low.

Conclusion:

We have shown that we can compute physiologically realistic coronary flow and pressure by considering the interactions between the cardiac properties, coronary impedance and arterial systems. Using these interactions, we can study how the changes in the cardiac properties, coronary arteries or arterial system influence each other.

References:

- [1] I. E. Vignon-Clementel et al., CMAME, 195:3776-3796, 2006.
- [2] S. Mantero et al., J. of Biomed. Engng., 14:109-116, 1992.
- [3] H. Senzaki et al., Circulation, 94:2497-2506, 1996.
- [4] H. J. Kim et al., CMAME, submitted.

Simulation Modeling of Cultured Alveolar Type II Cells to Elucidate Principles of Pulmonary Alveolar Growth

Purpose:

We have detailed descriptions of most biochemical and cellular processes involved in alveolar development, but lack insight into the elaborate, cohesive, epigenetic principles of operation underpinning systemic phenomena. What are the generative principles? How do they come about? We created and studied in silico cell mimetic analogues that reproduced basic alveolar growth attributes of cultured pulmonary alveolar type II (AT II) cells. The results [1] represent an early, but significant step towards answering those questions.

Methods:

We created AT II cell mimetic models within the MASON agent-oriented simulation framework. The models use discrete event representation methods. Cultured AT II cells were represented abstractly using quasi-autonomous agents that adhere to a small set of axiomatic operating principles. Other culture elements were conflated and represented using two types of passive objects: matrix (e.g., culture substrate) and media (e.g., cyst lumen) devoid of cells and matrix. Cell axioms defined preconditions, based on the local extracellular environment, and corresponding cell actions including migration, cell-cell adhesion, and intercellular rearrangement. Cell culture simulation starts with random plating of simulated cells, and then allowing them to act autonomously according to each cell's action logic.

Results:

During execution, simulated cells developed stable structures that closely approximated alveolar-like cysts in vitro. Cell axioms enabled disarrayed cell clusters to consistently organize into coherent multicellular cysts. As in vitro, final cyst size was dependent on the initial cell density. Simulated cyst growth was sensitive to changes in cell migration mechanics. Cell migration along a cell density gradient, rather than random motility, led to rapid, efficient cyst growth. Cell speed had material effects but played a relatively minor role in cyst growth.

Conclusion:

A small number of axiomatic operating principles gave rise to self-organizing developmental behaviors that mimicked those of cultured AT II cells. While it is premature to assign a specific biological mapping to the in silico operating principles, the findings enforce the idea that complex alveolar morphogenetic phenomena are a consequence of adherence to a small set of epigenetic principles. More advanced adaptations are expected to stand as scientifically useful, plausible representations of alveolar epithelial cell operating principles during lung development and regeneration.

URL: http://biosystems.ucsf.edu/research_epimorph.html

References:

[1] Kim SHJ et al. (2007) Systems Modeling of Alveolar Morphogenesis In Vitro. Proceedings of ISCA 20th International conference on computer and applications in industry and engineering.

poster
23

**Kranthi
Kode**

**Sailaja Elchuri
Cathy Shachaf
David Paik**

Parametric Modeling of Raman Spectra of Nanoparticles for Quantitative Unmixing

Purpose:

In recent years, Raman spectroscopy has gained increasing interest as a tool for rapid and nondestructive analysis of biological samples. Raman spectroscopy can differentiate the spectral fingerprint of many molecules, resulting in very high multiplexing capabilities[1]. However, one of the problems in quantitative unmixing of Raman spectra is related to the presence of a relatively high fluorescence background[2]. Also, changes in experimental conditions result in widening of Raman peaks and/or shifting of peaks which make quantitative unmixing very difficult. The objective of our study was to apply mathematical techniques on parametric models of the Raman spectra of nanoparticles to unmix the contributions from multiple nanoparticles to allow for simultaneous multiplexed quantitation of nanoparticle concentrations for tumor cell characterization.

Materials and Methods:

Gaussian and lorentzian functions were used to model the individual Raman peaks and the broad fluorescence components of COINs (Composite Organic - Inorganic Nanoparticles) in solution. During modeling, parameters such as the peak height, peak location and full width at half maximum (FWHM) of Raman peaks as well as the background were optimized simultaneously. These parameters were extracted from the raw spectra using a peak detection algorithm and were used as initial guess for the optimization process.

Results:

We modeled individual Raman peaks and the fluorescence background to accurately represent the Raman spectra of COINs. The model has been tested using direct linear least squares. The fluorescence component was subtracted from the model to obtain background-free Raman spectra.

Conclusion:

The accurate parametric modeling of Raman spectra opens up new perspective on unmixing algorithms. Many more options are made available for dimensionality reduction in unmixing techniques by our work. With the accurate background subtraction as opposed to blind baseline correction methods and the new unmixing algorithms, quantitative unmixing of Raman spectra will be much more reliable resulting in huge potential for application of Raman spectroscopy in the analysis of nanoparticles.

URL: <http://stanford.edu/~kranthi/Ramanmodels.html>

References:

- [1] S. Keren, C. Zavaleta, Z. Cheng, A. de la Zerda, O. Gheysens, and S. S. Gambhir, Proceedings of the National Academy of Sciences, 2008.
- [2] N. N. Brandt, O. O. Brovko, A. Y. Chikishev, and O. D. Paraschuk, Applied Spectroscopy, Vol 60, # 3, 2006.

Capturing and Managing Biomedical Image Metadata

Daniel

Korenblum

Cesar Rodriguez

Chris Beaulieu

Daniel Rubin

Sandy Napel

Purpose:

Modern medical imaging devices such as computed tomography (CT) and magnetic resonance imaging (MRI) are capable of generating thousands of images in a given patient examination. Images are interpreted by diagnostic radiologists who generate free text reports describing the imaging findings & diagnosis, but there is no explicit linkage between image content & the text used in a report. Millions of images are produced daily without the ability to systematically “mine” the image content for clinical, research or educational purposes. Compared with non-medical image search technology, this is an enormous lost opportunity.

Methods:

Our approach for meeting this challenge is to provide physicians a way to explicitly associate their descriptions; e.g., “a non-enhancing mass in the right lobe of the liver”, directly to each image region & a way to retrieve images based on these descriptions or more technically, based on metadata. We are developing two software tools, the Physician Annotation Device (iPAD) & the Biomedical Image Metadata Manager (BIMM), to provide the capabilities described above. Our approach incorporates the Digital Imaging & Communications in Medicine (DICOM) standard to represent medical images, the Radiology Lexicon (RADLEX) for describing the content in radiologic images, & the Annotation & Image Markup (AIM) information model for structuring image annotations & other metadata. RADLEX is a controlled terminology for radiology, which provides a consistent set of terms for radiologic descriptions. The AIM information model provides a standard structure for organizing the components of a radiologic description; e.g., “the anatomic entity in the description is the right lobe of the liver”.

Results:

We used iPAD to capture & upload anonymous CT images to BIMM. ROI Pixel data was validated against original DICOM image data using Matlab. iPAD usage time was recorded for later comparison with “regular” annotation times.

Conclusions:

Using our tools, physicians describe a finding in a biomedical image using controlled terms. The descriptions, image pixels, & other metadata are stored in a relational database organized according to the AIM information model. On the BIMM server, users can then retrieve images & regions in the images based on metadata queries. For example, a query like “give me all images with a hypodense mass in the right lobe of the liver” is possible. Our tools enable the community to build large databases of images & metadata, permitting physicians & machine learning tools to compare new cases against previous ones for content-based image retrieval & decision support applications. This emerging approach changes the present focus on pure image storage to “image mining.”

URL: <http://rufus.stanford.edu>

poster
25

**Ethan
Kung**

**Charles A. Taylor
Christopher Zarins**

Re-creating Vascular Flow Impedance with a Physical Apparatus

Purpose:

The eventual goal is to build a deformable-wall vascular phantom that can be used to do validation work for FEM codes currently used in cardiovascular system modeling. In an FEM computer simulation, boundary impedance conditions at the outlets of a domain are modeled as lumped parameter circuits. The construction of a physical apparatus that can precisely produce desired boundary impedance conditions is an important step towards completing a physical phantom to be used for the validation studies of the FEM code.

Materials and Methods:

An RCR lumped parameter circuit is typically used as boundary conditions in the FEM simulations conducted in the Taylor lab. In this presentation I investigate how to build a flow resistance and capacitance with accurate and appropriate values over the flow & pressure operation ranges of physiological simulation.

I will present design considerations for the flow resistance, which mainly focuses on the necessity of keeping the flow laminar. The use of a fine-pore honeycomb is best suited for the resistance module construction. Appropriate length and diameter of the honeycomb block determines the resistance and the operating range (in terms of flow rate) of the resistance module.

The flow capacitance can be re-created with pockets of compressible gas (ie. air) in the flow system. The main design considerations involve keeping the variation of the capacitance small during its operation. The amount of air and its initial pressure define the acceptable operating range of the capacitance module in terms of pressure or total amount of inflow into the capacitor. The operating range is defined by an acceptable percent variation in the capacitance value during its operation.

Results:

I will present measured experimental results characterizing the flow resistance and capacitance modules I have built.

Conclusion:

The characteristics of the resistance and capacitance modules within their specified operating range make these modules suitable for use at the outlets of a vascular phantom as the impedance boundary conditions.

Developing cystine-knot peptides as alternative scaffolds for therapeutic protein engineering

poster
26

**Jennifer
L. Lahti**

**Adam P. Silverman
Jennifer R. Cochran**

Purpose:

Protein scaffolds are structural frameworks used to present variable polypeptide regions that can be engineered to bind molecular targets. Cystine-knot peptides (knottins) are promising protein scaffolds since they are thermally stable, resistant to proteolysis, and non-immunogenic. Further, knottins have multiple disulfide-bonded loops capable of displaying conformationally-constrained variable polypeptides for molecular recognition. In this work, we assess the structural tolerance of amino acid sequence diversity in knottin loops. We present a combinatorial approach to explore the structural tolerance of knottin loops to sequence diversity by using yeast display, directed evolution, and bioinformatics.

Materials & Methods:

Our studies focus on EETI, a model member of the knottin family that naturally binds trypsin. Libraries of EETI variants with randomized loop sequences were displayed on the surface of yeast[1]. Clones that retained the native knottin fold were selected by fluorescence-activated cell sorting against Alexa488-labeled trypsin. Isolated clones were sequenced and then analyzed using RELIC2 bioinformatics tools. Positional pair-wise covariance analysis [3,4] was performed on the variable loop sequences of isolated clones and predictive pairs of amino acids at covarying loop positions were identified. From these collective analyses, we rationally designed EETI variants predicted to retain the knottin fold.

Results:

Sequence analysis of EETI variants revealed sequence-structure relationships of the knottin fold. Sequence diversity was best tolerated in the C-terminal loop of EETI, where we identified a consensus glycine-tyrosine doublet at the final two loop positions and three covarying positional pairs. Remarkably, all of the EETI variants that were rationally designed using these data retained the predicted knottin fold, verifying the structural importance of the observed sequence trends.

Conclusion:

EETI is generally tolerant of sequence diversity in its loop regions, suggesting it has potential to be a useful molecular scaffold. Sequence trends governing structural integrity described here can guide future protein engineering efforts with the EETI knottin. Importantly, our combinatorial approach can be used to probe structure-function relationships of proteins while accounting for compensatory mutations and the influences of the local amino acid environment.

References:

- (1) Boder, E. T.; Wittrup, K. D. *Nat Biotechnol* 1997, 15, 553-7.
- (2) Mandava, S. et al. *Proteomics* 2004, 4, 1439-60.
- (3) Fodor, A. A.; Aldrich, R. W. *Proteins* 2004, 56, 211-21.
- (4) Yip, K. Y. et al. *Bioinformatics* 2008, 24, 290-2.

poster
27

Jia-Ren Lin

Jia-Hao Fan
Charles Ofria

How cancer evolves? Study the role of mutational regulation in somatic evolution by using digital organisms.

Purpose:

Cancer is considered as an evolutionary disease, not only because the “mutator” phenotype is commonly found in neoplastic or malignant tumors but also the genomic instability (or chromosome shuffling) serves as the hallmark of all cancer. Understanding these mutational events is crucial for both preventing cancer (mutations in tumor suppressors and oncogenes) and cancer treatment (mutations in drug-resistant genes). However, the evolutionary aspect of cancer is usually overlooked because the experimental approaches are limited in this field. Therefore, we perform in silico experiments here to investigate the underlying mechanism of mutational regulation in somatic evolution, on the other word “caner”.

Materials and Methods:

Avida, an artificial life system was developed by Dr. Adami and colleagues in Caltech. The Avida system combines 2D cellular automata and genetic programming to simulate different aspects in evolution. This tool has been extensively used in evolutionary studies, including long-term adaptation, quasi-species dynamics, epistatic interactions and mutational robustness. The current version of Avida is maintained by Dr. Ofria in MSU. Here, we have modified the original Avida to incorporate several features of cancer evolution (ex: chromosome shuffling, aging related mutagenesis and stress-induced mutation). The hardware used in this study included one Intel Core Due 1.86 GHz laptop (2GB RAM) and one AMD Athlon64 3500+ Desktop (2GB RAM). The data analysis was done using Perl scripts and Microsoft Excel.

Results:

The preliminary results from analyzing the interaction between point mutation and chromosome shuffling reveal that both types of mutation are required for fast evolution. Moreover, the chromosome shuffling would act as the primary trigger for exponential growth of fitness. In addition, the flexible mutation rate would provide an evolutionary advantage in both stressed and un-perturbed conditions. Finally, the simulation of age-dependent mutagenesis demonstrates an underlying mechanism of aging induced carcinogenesis.

Conclusion:

This research reveals a potential use of digital organisms in studying cancer evolution. The intriguing findings on the interaction between different types of mutation might open a new avenue to suppress malignant evolution of cancer.

References:

Merio et al. 2006 Nature Review Genetics Vol. 6 pp924-935
Bielas et al. 2006 PNAS Vol. 103 pp18238-18242
Adami-C 2006 Nature Review Genetics Vol. 7 pp109-118
Ofria-C and Wilke-CO 2004 Artificial Life Vol. 10 pp191-229

Modeling the Onset of Fatal Metastases in Lung Cancer

Ray Lin

Bronislava M. Sigal
Sylvia K. Plevritis

Purpose:

Evaluating of the effectiveness of CT screening on lung cancer mortality has become a critical issue given divergent interpretations of the current evidence [1,2]. The goal of this project is to predict the benefit of CT screening on reducing lung cancer mortality using a model of the natural history of lung cancer. This model is developed to predict the tumor size of lung cancer transition to lethal disease from metastases.

Materials and Methods:

We developed a stochastic model of the onset of fatal metastasis of lung cancer based on patient survival data. Growth of the primary tumor and the fatal metastatic burden are modeled as functions of time. The tumor size at the onset of fatal metastasis and the metastatic burden that will cause clinical detection and patient death are formulated in parametric forms. Model parameters were estimated from 30,691 Non Small Cell Lung Cancer (NSCLC) cases in the Surveillance Epidemiology and End Results (SEER) database. The model was calibrated to tumor size and stage at clinical detection and patient survival. This model is more stable than existing models under the potential phenomenon of stage migration in lung cancer with advances in detection technology.

Results:

The model demonstrated good fit to the distribution of tumor size and stage at clinical detection and patient survival. The estimated median tumor diameter at onset of fatal metastasis is around 1cm and the estimated survival time from onset of fatal metastasis is 3.6 years. We estimated that 52% of lung cancer cases were detected due to the symptoms of primary tumors whereas the rest were due to metastases.

Conclusion and Future Work:

Our estimates show that we would need to detect NSCLC tumors below 1cm to provide patients with a cure. Since CT is able to detect tumors around 5mm, it may potentially reduce lung cancer mortality significantly. We will conduct simulated clinical trials based on Monte Carlo techniques to further quantify the benefit of CT in the future.

Acknowledgements:

The authors would like to thank Dr. Amar K. Das of Biomedical Informatics Training Program for his insightful discussions about the design and the interpretation of the model.

References:

- [1] P.B. Bach et al., "Computed tomography screening and lung cancer outcomes," JAMA: The Journal of the American Medical Association, vol. 297, Mar. 2007, pp. 953-61.
- [2] C.I. Henschke et al., "Survival of patients with stage I lung cancer detected on CT screening," The New England Journal of Medicine, vol. 355, Oct. 2006, pp. 1763-71.

poster
29

Tianyun Liu

Russ B Altman

Predict novel calcium-binding sites using FEATURE combined with modeling methods

Purpose:

Predicting calcium-binding sites in proteins are essential for understanding the roles of calcium in biological systems. There are no universally applicable sequence motifs for calcium binding, so they must be recognized with structure-based methods. Ligand-binding sites in their apo states are often unstructured and their 3D structures cannot be determined experimentally. More than 2/3 of the PDB structures contain unstructured regions. This frustrates the recognition of ligand-binding sites by structure-based methods. We have previously presented FEATURE[1], which calculates physical and chemical properties around a potential binding site, with a Bayesian score to predict its likelihood of ligand binding. In this study, we combine previously published modeling methods with FEATURE to predict calcium-binding sites in unstructured regions.

Methods:

We validated the combined method to recover known calcium-binding sites. We then created a candidate dataset using structures from SCOP families that are known to bind calcium. We collected unstructured loop regions in these structures and built loop conformations using RAMP program suite[2]. FEATURE scan was applied to these loop conformations.

Results:

For validation, FEATURE identifies 75% these calcium-binding sites correctly when structures of binding sites in the apo state are rebuilt using modeling methods. We hypothesized that some of the unstructured regions in proteins from SCOP families that are known to bind calcium may themselves be cryptic calcium binding loops. We built the missing loops for 2746 unstructured regions and evaluated them for potential calcium binding. We made 102 predictions that achieve high FEATURE scores. Ten predicted sites are confirmed by experimental evidence; 14 predicted are supported by indirect experimental evidence; 78 sites are novel predictions. In the 78 novel predictions, a large number of predicted sites adopt anti-parallel beta structures, which share structural similarity with the calcium-binding C2 domain. 61 of the 78 predictions are in proteins that bind to other ligands, suggesting a role of calcium in regulating ligand binding.

Conclusion:

Combining FEATURE with modeling techniques enables us to discover potential ligand-binding sites in unstructured regions. We made 102 predictions of calcium binding and found direct and indirect experimental evidence for 24 predictions. The other 78 predictions are novel, some with intriguing potential biological significance.

References:

1. Wei L, Altman RB. Pac Symp Biocomput 1998:497-508.
2. Hung LH, Ngan SC, Liu T, Samudrala R. Nucleic Acids Res 2005;33:W77-80.

Engineering Protein Flexibility by Molecular Dynamics and Multivariate Statistics

**Francesca
Milletti**

**Christopher McClendon
Matthew P. Jacobson**

Purpose:

The dynamics of proteins influences protein folding, recognition and function. Molecular dynamics simulations can be used to assess atomic fluctuations and provide insights on protein flexibility. Predicting which point mutations might alter protein dynamics, and hence function, is a challenging task. In this study, we focus on HIV-1 Nef, a small protein critical for progression to AIDS. Nef, a promising target in anti-HIV therapy, promotes immunodeficiency by interacting with multiple cellular partners. Because the HIV virus evolves rapidly, Nef is highly polymorphic. Here, we identify point mutations that significantly alter Nef dynamics and relate flexibilities of Nef variants to clinical progression to AIDS. This study provides insights on ways to engineer protein flexibility by molecular dynamics and multivariate statistics.

Material and Methods:

We used Partial Least Squares (PLS) statistics to identify relationships between 70 diverse sequences of HIV-1 Nef and their atomic r.m.s. fluctuations obtained from 3.3 ns molecular dynamics simulations. The 3D models of each Nef sequence were obtained by homology modeling using the NMR structure of Nef as template. PLS results were used to assess the type and position in the sequence of residues expected to be major drivers for Nef flexibility.

Results:

We found that two out of the three point mutations identified as most important for flexibility are known to lead to loss of function in Nef. A relationship was established between the degree of flexibility of the core loop of Nef (residues 148-180) and Nef activity. Surprisingly, we found that two residues, H89 and F191, are critical for the flexibility of the core loop, although they are located in a region far from the loop itself. These results not only highlight a link between Nef dynamics and Nef function, but they also show how multivariate analysis of protein sequences and their dynamics can be used to identify features critical for function.

Conclusion:

We have developed a model that identifies which amino acids contribute most to Nef flexibility in various HIV strains. Our findings, which provide a possible link between Nef function and Nef flexibility, imply that it is possible to engineer proteins that have the desired degree of flexibility and offer the potential for applications in multiple fields, such as crystallography, protein engineering, and drug discovery.

**poster
31**

Hyper and Hypo Variability in Gene Expression across Multiple Studies

**Alexander
A. Morgan****Atul J. Butte****Purpose:**

Very high throughput measurements of mRNA expression such as DNA microarrays have given the biomedical research community an amazing ability to obtain a “bird’s eye view” of processes at a molecular level. The subsequent collection of the results of many of these experiments in repositories such as the Gene Expression Omnibus (GEO) affords us an opportunity to seek an even loftier view, what might be jocularly termed a “spy satellite view” of the variations in mRNA expression across all genes and all measured conditions. Here we investigated the properties of this very high dimensional space by looking at those genes whose mRNA varies the most (hyper-variable genes) and least (hypo-variable genes) and examining the physiological and functional associations with high and low variation.

Materials and Methods:

We use over 29,000 microarray results from GEO and explore the associations between three simple, non-parametric summary statistics: experimental rank variation ratio (varying experimental conditions showing differential expression between conditional groups), rank width (interquartile range of ranks), and rank median.

Results:

We show that variation caused by changing experimental conditions is independent of the variation seemingly inherent in expression at the level of individual genes in contrast with the assumptions made when using ‘housekeeping’ genes for normalization. These high level types of variation are conserved across species both phylogenetically close (rat and mouse) and distant (human and rodent). We also show that these measures of overall variation are associated with the likelihood of disease association and in their extreme values are enriched for a range of well-defined functional properties.

Conclusions:

In this study we develop methods of examining/summarizing overall gene expression variation and demonstrate their application in a meta-analysis involving over 29,000 microarrays. We show conservation in overall patterns of gene expression levels of variation changes across three species (human, mouse and rat) and show associations with known disease related genes. We also show the physiological importance of these measures of expression variation in association with the functional properties of the genes. These results can influence thinking about the selection of genes for microarray design and how to analyze measurements of mRNA expression variation in a global context of expression variation across many conditions.

Small Animal Conformal Radiotherapy

Small animal irradiation is commonly performed in the laboratory using single field techniques employing lead blocks for beam shaping. These procedures cannot duplicate the level of spatial precision and dosimetric accuracy that is regularly achieved in the clinic. Investigation of radiotherapy in deep-seated models of cancer, as well as application of radiotherapy regimens to animals that more closely resemble human treatments, requires a more precise delivery of radiation. An attractive solution for precision irradiation of small animals is to add the capability to an existing small animal microCT scanner. We have investigated application of a GE microCT scanner towards this goal by establishing the dosimetric performance of the scanner and developing hardware and software for variable collimation of the X-ray beam. We have measured and calculated the percent depth dose curves and the geometric accuracy of the therapeutic beam. We are able to deliver therapeutic doses to mice conformal to within less than 2mm.

poster
32

***Geoffrey
Nelson***

*Hu Zhou
Manuel Rodriguez
Rahil Jogani
Paul Keall
Edward Graves*

poster
33

**Sarah
Nelson**

**Paul Sigala
Daniel Herschlag**

Testing the Importance of Surrounding Residues in Positioning Hydrogen Bond Donors in the Oxyanion Hole of Ketosteroid Isomerase

Purpose:

Enzymes are central to biology and attractive targets for pharmaceuticals. While the chemical steps of many enzymatic reactions are known, the physical basis of enzyme rate enhancement is not yet understood. One long-standing hypothesis is that enzymes are electrostatically and geometrically complementary to substrate transition states and that the resulting transition state stabilization contributes to rate enhancement. Work with the bacterial enzyme ketosteroid isomerase (KSI), a model system in which to test this hypothesis, has suggested only a modest contribution to catalysis by charge complementarity [1]. It is therefore of interest to test the role of shape complementarity in KSI catalysis.

Materials and Methods:

In the KSI active site, residues Tyr57, Phe86, and Met116 may help position hydrogen bond donors Tyr16 and Asp103, both of which appear pre-positioned to preferentially stabilize the substrate tetrahedral transition state relative to the planar ground state. The magnitude to which mutations in Y57, F86, and M116 are able to reduce KSI catalysis could therefore suggest whether geometric complementarity plays a significant catalytic role. A series of kinetic assays that measure the conversion of substrate to product via UV absorbance spectrophotometry was used to examine the effect of each mutant on the catalytic rate constants k_{cat} , K_m , and k_{cat}/K_m .

Results:

The magnitudes of k_{cat}/K_m for single and double mutations made in Y57, F86, and M116 were at most 50-fold lower than that of wild-type. In general, effects on k_{cat} and K_m were also modest. Y57 and M116 mutations resulted in greater perturbations in k_{cat} , K_m , and k_{cat}/K_m than F86 mutations, suggesting a lesser role for F86 in KSI catalysis.

Conclusions:

The modest effects on k_{cat}/K_m observed by all mutations suggest either (1) minor roles for Y57, F86, and M116 in positioning Y16 and D103 and therefore in geometric complementarity or (2) shape complementarity makes no contribution to catalysis – e.g., mutations could cause small changes in the positioning of active site residues other than Y16 and D103. Mutants that combine mutations at positions 16 and 103 with mutations at positions 57, 86, and 116 will better distinguish between these possibilities. NMR studies of mutants will be used to reveal changes in hydrogen bonds and therefore alterations in residue positioning and interactions in the active site. X-ray crystallography will also help demonstrate any changes in the active site environment due to mutations. Together, both present and future studies should help to clarify the role, if any, of geometric complementarity in enzymatic catalysis.

[1] Kraut et al., PLoS Biol. 2006

Quantifying Changes in Pulmonary Vascular Morphology from Adolescence to Adulthood in Rats

Purpose:

Congenital heart defects (CHD) affect nearly 1 out of 100 births [1]. Pulmonary artery stenosis (PAS), a common form of CHD, is a local narrowing of the pulmonary arteries. Since it is difficult to know a priori which stenoses are hemodynamically significant, computational simulations of blood flow provide a means to predict outcomes of potential interventions [2]. In such modeling approaches, results of the governing equations of blood flow are greatly affected by the imposed outflow boundary conditions. To date, only generic idealized models of adult morphological networks have been used as boundary conditions. To determine the effects of a pulmonary stenosis on vascular morphology, we developed methods for casting, imaging and analyzing the microvasculature in small animals at different ages.

*Shahrzad Zarafshar
Chengpei Xu
Jeffrey A. Feinstein
Charles A. Taylor*

Methods:

The methodology developed was applied to 4 male Sprague-Dawley rats at 54, 110, 190 and 409g. The lungs were injected with a silicone polymer at constant flow rates. The filling pressures were recorded throughout the casting procedure and excess volume was allowed to flow back. The lungs were then imaged using microCT at 49 and 12.5-micron resolutions. Growth trends were obtained from a tree structure of the vasculature. Morphologic data was then classified using the diameter-defined Strahler system [3].

Results:

Methodology development resulted in optimal castings under polymer viscosity of 25cP, flow rate of 0.025 ml/min and weight-dependent polymer volumes. 49-micron scans confirmed uniform and consistent castings of vessels at all weights. Arterial diameters, maximal lung lengths and total heart and lungs tissue volumes increased with age. An increase in the total number of vessels was also observed with weight. Connectivity matrices describing branching patterns were calculated for each sample.

Conclusions:

Our casting, imaging and analysis methods enabled accurate quantitative characterization of pulmonary vascular beds to the level of pre-capillary arterioles. Comparison of vascular morphology has enabled accurate description of growth patterns from adolescence to adulthood. The same methods can be applied to determine the effects of a pulmonary artery stenosis on vascular morphology. Future steps include the application of findings to patient data for better prediction of pulmonary flow characteristics.

References:

- [1] Brickner ME, et al. Congenital heart disease in adults, The New England Journal of Medicine, 1/27/2000
- [2] Taylor CA, et al. Finite element modeling of blood flow in arteries, Comput. Methods Appl. Mech. Engrg. 158:1998
- [3] Jiang ZL, et al. Diameter-defined Strahler system and connectivity matrix of the pulmonary arterial tree. J. Appl. Physiol. 76(2):1994

**poster
35****Hersh
Sagreiya****Russ B. Altman**

Evaluation of Clinical and Pharmacogenetic Factors Involved in Warfarin Dosing

Purpose:

Warfarin, an oral anticoagulant, can be difficult to administer due to dramatic inter-individual variation in therapeutic dose. A high dose can lead to hemorrhage, while a low dose will not sufficiently protect against thromboembolic events. It is one of the leading causes for ER visits by the elderly and is traditionally dosed by trial-and-error [1]. Both genetic factors, particularly the VKORC1, CYP2C9, and CYP4F2 genes, as well as clinical factors, such as weight and age, have been shown to affect its dosage [2]. Dosing algorithms for warfarin have been proposed, but their relative performance needs to be evaluated.

Materials and Methods:

The first study evaluated the clinical factors involved in warfarin dosing. 1472 discharge summaries were queried for chart review using the STRIDE database. The following search terms were used: height, weight, age, race, and warfarin/coumadin. Approximately 350 of these records were selected because they contained sufficient information to determine that a stable dose was achieved. The second study examined both the genetic and clinical factors involved in dosing and was performed at the Stanford Oral Anticoagulation Clinic. Complete history and pharmacy records were obtained via the CoagClinic database, additional information was gathered with a questionnaire, and patient DNA was collected using Oragene•DNA saliva kits. These samples were then genotyped at 30 SNPs relevant to VKORC1, CYP2C9, and CYP4F2.

Results:

In the first study, four predictive values collectively explained about 15.5% of the variation in therapeutic dose. In order of importance, these factors were age ($R^2 = 9.5\%$, $p < 10^{-8}$), weight ($R^2 = 6.1\%$, $p < 10^{-5}$), height ($R^2 = 4.5\%$, $p < 10^{-4}$), and race ($R^2 = 2.4\%$, $p < .01$). The second study is currently enrolling patients.

Conclusion:

The first study demonstrates that the relationship between clinical variables and warfarin dose is sufficiently robust to appear in patient discharge summaries. A prior study found that clinical variables accounted for 21.5% of the variation in dose, which compares with 15.5% in our study [3]. The second study will determine the local performance of various pharmacogenetic dosing algorithms in the literature, it will assess the additional impact of CYP4F2, and it will analyze the effects of certain medications (i.e. amiodarone) and clinical factors (i.e. vitamin K supplementation). Finally, the data from this study will be posted on PharmGKB.

URL: <http://www.pharmgkb.org/>

References:

- [1] Budnitz et al. (2005), Ann Intern Med, 147(11), 755-65.
- [2] Caldwell et al. (2008), Blood, 111(8), 4106-12.
- [3] Gage et al. (2008), Clin Pharmacol Ther, 84(3), 326-31.

Inference of Ancestries in Admixed Populations

Sriram

Sankararaman

Srinath Sridhar

Gad Kimmel

Eran Halperin

Michael Jordan

Purpose:

Large-scale genotyping of SNPs (Single Nucleotide Polymorphisms) has shown great promise in identifying regions in the human genome that could be linked to diseases. One of the major obstacles involved in performing these studies is that the underlying population substructure could produce spurious associations. Population substructure can be caused by the presence of two distinct sub-populations or a single pool of admixed individuals (e.g. Hispanic or African Americans). The latter is significantly harder to detect in practice. Due to recombination events, the genomes of these admixed individuals consist of a mosaic of short segments that have originated from the different ancestral populations.

Methods:

Current methods for inference of population substructure in an admixed population are inaccurate even in relatively simple scenarios. Several of these methods are based on Markovian models. We introduce a new method, SWITCH [1], based on an augmented hidden Markov Model. SWITCH infers the ancestry of each individual at every SNP. A crucial component in accurate inference of ancestries is the initialization procedure. SWITCH is based on a fast, accurate initialization method, LAMP[2]. LAMP computes the ancestry structure for overlapping windows of contiguous SNPs and combines the results using a majority vote.

Results:

SWITCH can infer the ancestries even when the genotypes from the pure ancestral populations are unknown or unavailable. Our empirical results show that SWITCH is significantly more accurate and more efficient than existing methods for inferring locus-specific ancestries, enabling it to handle large-scale datasets.

We can also use SWITCH to answer other questions of biological interest in admixed populations, such as the inference of historic recombination events and the inference of alleles in ancestral populations that are unknown or extinct.

References:

- [1] Sriram Sankararaman, Srinath Sridhar, Gad Kimmel and Eran Halperin. Estimating local ancestry in admixed populations. *The American Journal of Human Genetics*, Vol 82(2), Feb 2008, pp. 290-303.
- [2] Sriram Sankararaman , Gad Kimmel, Eran Halperin and Michael I. Jordan. On the inference of ancestries in admixed populations. *Genome Research*, March 2008 (<http://genome.cshlp.org/cgi/content/full/18/4/668>)

URL: <http://lamp.icsi.berkeley.edu/lamp/>

poster
37

Thin Client Architecture to Support Multicenter Radiological Trials

Florian F.

Schmitzberger

Justus E. Roos

Sandy Napel

Geoffrey D. Rubin

David S. Paik

Purpose:

Geographically distributed radiological trials pose considerable logistical challenges, primarily in (1) transferring large datasets, (2) displaying interactively rendered image data, (3) gathering specific feedback on presented features (4) collecting and analyzing resulting data from a large number of sites.

The purpose of this study was to implement a software platform providing informatics support for multi-center radiological trials using state of the art networking and visualization technologies.

Methods and Materials:

An ideal solution to the difficulties associated with geographically distributed trials requires (1) no on-site support necessary, (2) no shipping of high end computer hardware or large datasets, (3) centralized collection of results, (4) radiological image visualization emulating a clinical environment.

Several approaches were considered: (1) traditional approach of sending hardware and software to each participant, (2) fat clients that run local copies of the rendering and questionnaire software and (3) a thin software client-server architecture connecting the client to a centralized server.

Results:

Results closest to an ideal solution could be achieved only with a thin client approach. A network-enabled client-server architecture was implemented that is capable of both providing image-rendering functionality (2D, 3D, MIP) as well as compiling feedback on each presented feature from the participant. Using the Fovia, Inc. rendering software, all images are rendered on the server and sent out to the client while maintaining full interactivity with frame-rates dependent on network latency and bandwidth. The user is presented with a familiar 2D and 3D interface as well as a questionnaire window. Responses are automatically collected by a central server, no data has to be stored with the user at any point and no on-site data coordinators are needed.

Conclusion:

Fulfilling all the requirements outlined above, our novel thin-client solution enables much larger radiological trials with less logistical overhead and expenses, removing a significant constraint of currently conducted trials in radiology.

URL: <http://www.stanford.edu/~florians/BCATS/>

Digital Pattern Analysis based on Microarray Dynamic-range Database for Complicated Developmental Models

Debashis Sahoo

Sylvis K. Plevritis

David L. Dill

Irving L. Weissman

Purpose:

The current methodology of microarray data analysis is to compare two or more samples for significantly differentially expressed genes. Unfortunately, this method typically disregards the dynamic range of measurements of the genes in question, and may thus fail to identify the appropriate genes. The possible range of measurements of a gene depends not only on biological factors, but also relies on the particular sequence of the probes, which affects the efficiency of hybridization of the mRNA on the array. For example, if expression levels of gene x may significantly differ between sample A and B, but both expression levels reside in the lower half of the dynamic range, the biological significance of this difference would decrease.

Likewise, if expression levels of gene y are not significant between sample A and B, but both are near its maximum possible value, gene y might be important for both sample A and B. To address this issue, we developed Digital Pattern Analysis based on Microarray Dynamic-range Database.

Materials and Methods:

Estimating the true dynamic range of genes requires inspecting the measured values on a large set of highly diverse arrays. Fortunately, large amounts of array data are now publicly available. We compiled over 2000 mouse microarray datasets (on an Affymetrix platform) and normalized them all together using a standard algorithm to construct a Microarray Dynamic-range Database. Then a threshold separating "low" from "high" values was computed using our StepMiner algorithm.

Results:

Using the dynamic range and threshold computed from this much larger data set, we have been able to analyze gene expression based on a digitalized pattern (high/low) of expression from a complex model involving about a dozen populations from a developmental pathway with multiple branching points. We established the Digital Pattern Analysis Platform by integrating our Dynamic-range Database, the Gene Ontology Database, and our microarray datasets of highly purified hematopoietic stem/progenitor populations. We present data illustrating these applications in the identification of known and unknown differentiation pathway-specific genes.

Conclusion:

Digital Pattern Analysis based on Microarray Dynamic-range Database improves the significance of microarray data analysis, and expands the applicable scope of microarray data analysis toward complex biological models.

poster
39

**Junhee
Seok**

**Ronal W. Davis
Wenzhong Xiao**

Knowledge-based analysis of microarray for the discovery of transcriptional regulations

Purpose:

A knowledge-base, accumulated previous knowledge of gene-gene regulatory interactions, is important in the discovery and understanding of gene regulation in biological systems. Early methods for discovery of transcriptional regulation relationships from microarray data often focused on the high throughput experimental data alone. More recent methods have found success by incorporating prior knowledge of gene interactions; however, the specific properties of knowledge-bases that provide the most useful value in analysis of high throughput genomic data have not yet been characterized. In this work, we systematically evaluated the utility of knowledge-based analysis in the identification of regulatory relationships between transcription factors (TFs) and target genes (TGs).

Material and Methods:

We evaluated the usefulness of knowledge-bases for finding new regulatory pairs on 643 microarrays of yeast [1]. As for knowledge-bases, we utilized the gene interaction pairs of transcriptional regulation from YPD [2] with 3,043 pairs between 523 TFs and 919 TGs. We encoded the regulatory signal of a TF with gene expressions of its TGs in the knowledge-base, and compared with the naïve regulatory signal encoded with gene expression of the TF. Prediction for regulatory relations was done based on features of TF regulatory signal and TG expression. Correlation cutoff method and a support vector machine were used for classification.

Results:

We observed that a TF regulatory signal encoded with knowledge-base has significantly higher correlation with gene expression of its TGs. Predictions with knowledge-bases showed much improved sensitivity and specificity in inferring new transcriptional interactions than from microarray data alone, and comprehensive, direct and high quality knowledge-base provides better prediction performance. We extended our knowledge-based methods to predict genome-wide regulatory pairs and demonstrated that these pairs are reasonable candidates for biological verification with independent datasets of ChIP-chip binding and yeast growth fitness.

Conclusion:

Overall the results showed that knowledge-driven analysis significantly helps in the interpretation of high-throughput genomic data.

References:

- [1] Stuart JM, Segal E, Koller D, et al., Science, 302, 249-55 (2003)
- [2] Csank C, Costanzo MC, Hirschman J, et al., Methods Enzymol., 350, 347-73 (2002)

Towards a Cytokine-Cell Interaction Knowledgebase of the Adaptive Immune System

Purpose:

The immune system of higher organisms is, by any standard, complex. To date, using reductionist techniques, immunologists have elucidated many of the basic principles of how the immune system functions, yet our understanding is still far from complete. In an era of high throughput measurements, it is already clear that the scientific knowledge we have accumulated has itself grown larger than our ability to cope with it, and thus it is increasingly important to develop bioinformatics tools with which to navigate the complexity of the information that is available to us.

Methods:

Here, we describe ImmuneXpresso, an information extraction system, tailored for parsing the primary literature of immunology and relating it to experimental data. The immune system is very much dependent on the interactions of various white blood cells with each other, either in synaptic contacts, at a distance using cytokines or chemokines, or both. Therefore, as a first approximation, we used ImmuneXpresso to create a literature derived network of interactions between cells and cytokines.

Results and Conclusions:

Integration of cell-specific gene expression data facilitates cross-validation of cytokine mediated cell-cell interactions and suggests novel interactions. We evaluate the performance of our automatically generated multi-scale model against existing manually curated data, and show how this system can be used to guide experimentalists in interpreting multi-scale, experimental data. Our methodology is scalable and can be generalized to other systems.

poster
40

Shai S.

Shen-Orr

Ofir Goldberger

Yael Garten

Yael Rosenberg-Hasson

Patricia A. Lovelace

David L. Hirschberg

Russ B. Altman

Mark M. Davis

Atul J. Butte

poster
41

Jessica Shih

Hyun Jim Kim

Charles A. Taylor

Image-Based Modeling of Blood Flow in Human Coronary Arteries With and Without Bypass During Rest and Simulated Exercise

Purpose:

As the number of patients with coronary artery disease continues to rise, knowledge of the flow features within the coronary vascular system could prove useful in predicting flow changes due to coronary interventions. In this study, we used image-based modeling and computational fluid dynamics to calculate flow dynamics for a patient with coronary artery bypass grafts.

Materials and Methods:

A CT dataset from a patient who had previously undergone triple bypass surgery was acquired. Three-dimensional geometric models were constructed of the relevant vasculature using custom software. A 75% area reduction stenosis was artificially constructed in the areas suspected to have coronary artery disease. An additional model with the bypass vessels removed represents a case of completely occluded bypass grafts. We used a lumped-parameter heart model to calculate left ventricular pressure which was then incorporated into the coronary artery boundary conditions. We also implemented different lumped parameter models to represent the downstream vasculature at the other outlets. Under rest conditions, the parameters of the heart model were chosen to give realistic left ventricular and aortic pressures. Literature data provided the mean flow to the branches of the coronary arteries and the branches of the aortic arch [1,2]. The model parameters were also modified to reflect exercise conditions. Pulsatile blood flow simulations were computed using a stabilized finite element method under rest and exercise conditions.

Results:

We successfully performed a three dimensional blood flow simulation using a patient specific coronary bypass model and produced realistic coronary flow waveforms. Under exercise conditions, the mean coronary flow is higher in both models, but the model without bypass has lower mean coronary flow compared to the case with bypass, indicating that the bypass grafts increase perfusion to the heart muscle under exercise conditions. The coronary flow waveforms reveal decreased flow in systole and increased flow in diastole and a change in the shape of the flow profile from rest to exercise.

Conclusion:

This study shows that it is possible to simulate physiologically realistic coronary flow and pressure. Patient specific 3-D flow simulations could be used in the future to predict the outcomes of coronary interventions and to inform treatment for coronary artery disease.

References:

1. Kalbfleisch, H., and Hort, W., 1977, "Quantitative study on the size of coronary artery supplying areas post-mortem." Am Heart J, 94(2), 183-8.
2. Williams, L. R., and Leggett, R. W., 1989, "Reference values for resting blood flow to organs of man." Clin Phys Physiol Meas, 10(3), 187-217.

Selection of magnetic resonance imaging slice orientation in the calculation of volumetric breast density

poster
42

*Lisa
Singer*

*Nola Hylton
Catherine Klifa*

Purpose:

High breast density is associated with an increased risk of breast cancer [1]. Mammography, the standard modality for measuring breast density, is less sensitive in women with dense breasts [2]. Magnetic resonance (MR) imaging provides 3D images of breast structures that allow for the quantification of volumetric breast density (VBD) with improved accuracy due to the high contrast between breast tissue and fat. MR VBD is defined as the volume of breast tissue divided by the total breast volume. Our lab previously developed a semi-automated tool to calculate VBD from MR images [3]. This 3D technique requires the delineation of breast volume by a user and then utilizes an automated algorithm for segmentation of breast tissue from fat. The semi-automated processing requires significant time and labor, limiting its use to research. Full-automation of our technique could reduce analysis time, thus facilitating its use in the clinic. We explored the use of different slice orientations for VBD calculation. Due to the anatomy of the breasts, we hypothesized that volume delineation would be better suited to full-automation on coronal than on axial MR images. In a first step towards full-automation, the aim of this study was to compare VBD results using coronal versus axial MR data.

Materials and Methods:

Coronal and axial MR studies were obtained from 198 healthy pre-menopausal women. A subset of low, intermediate, and high breast density cases were selected on the basis of MR VBD previously measured from axial MR data. All images were processed in MATLAB using a semi-automated technique for volume delineation and an automated segmentation based on fuzzy-c-means for VBD calculation. Measurements were compared numerically, using actual VBD values, and categorically, using the VBD-based classification of density as low (VBD values $<.20$), intermediate ($.20-.40$), or high ($>.40$).

Results:

Coronal and axial-based VBD measurements differed numerically, but not categorically. Women were comparably classified as having low, intermediate, or high breast density in all cases thus far studied.

Conclusions:

The correspondence between breast density classifications using coronal and axial images encourages further development of a fully-automated coronal-based VBD method. While differences in VBD measurements did not affect the classification of VBD as low, intermediate, or high, the ability to correct for these differences must be further explored.

1. Boyd, N.F., et al., Cancer Epidemiol Biomarkers Prev, 1998. 7(12): p. 1133-44.
2. Kolb, T.M., J. Lichy, and J.H. Newhouse, Radiology, 2002. 225(1): p. 165-75.
3. Klifa, C., et al., Conf Proc IEEE Eng Med Biol Soc, 2004. 3: p.1667-1670.

poster
43

Each pixel is a separate event: Application to dynamic MRI

Uygar
Sümbül

John M. Pauly

Purpose:

Pixels of an image are modeled as separate stochastic events, which may be correlated. This view is particularly useful in time-resolved applications. An important special case is to process the time course of each pixel independently. This suboptimal solution enables fast reconstructions, and even real-time imaging when causality is satisfied. Our specific problem is to increase the frame rate of cardiac videos when data is acquired via dynamic magnetic resonance imaging(MRI).

Materials and Methods:

MRI data correspond to the Fourier transform of the underlying image. Consider the following dynamic system:

$$S_t = S_{t-1} + U_t$$

$$X_t = H_t S_t + W_t$$

Here, S_t is a vector of true image pixels at time t , and H_t is the transfer function of the MRI experiment. In the case of Cartesian acquisitions, H_t is an underdetermined Fourier transform operator. In MRI, non-Cartesian coverage of the Fourier image is sometimes desirable[1] and H_t is modified accordingly. U_t is the vector stochastic process corresponding to the change in the pixel values from time $t-1$ to t . W_t is the observation noise, which is white and Gaussian in MRI experiments. X_t denotes the raw data obtained at time t . The problem is to obtain the image pixels at time t using all the past and present incomplete observations. The model is linear and we apply the classical Kalman filter[2]. The second moments of the noise processes(U_t and W_t) are obtained from separate initialization scans.

Due to the large number of pixels, a direct application of the Kalman filter is computationally infeasible for practical applications. By imposing diagonality on the relevant matrices, we obtain a fast and causal reconstruction. This corresponds to uncorrelated pixels. Moreover, it is possible to reflect all the approximations to the covariance matrix of U_t and analyze the system as if the system noise covariance estimation is the only source of imperfection. For the behavior of the Kalman filter under incorrect noise estimates, see [3]. Data is acquired by a 1.5 T GE Signa scanner using an 8-element cardiac array. No ECG-gating or breath-holding was used.

Results:

We compared our results to the results obtained by a sliding window reconstruction, which is a basic real-time reconstruction algorithm. The overall image quality of the two reconstructions are comparable. However, the Kalman reconstruction provides a sharper depiction of the fast moving parts such as the cardiac valve.

Conclusions:

We presented a dynamic system framework for time-resolved imaging applications and applied a Kalman filter to the accelerated dynamic MRI problem using this framework. In many applications, fast reconstruction is vital. To this end, we imposed diagonality on the model. The results confirm that the resulting algorithm achieves better reconstructions at a comparable computational complexity.

URL: <http://www.stanford.edu/~uygar/bcats.jpg>

A Well-controlled Fast Clustering Method on Conformation Space of Biomolecules

Jian Sun

Purpose:

Representing the free energy landscapes of bio-molecular folding is difficult since they are the rugged and high-dimensional. Projecting the free energy landscape onto a few order parameters may obscure the free energy barriers. An alternative approach is to decompose the conformation space into clusters by grouping close conformations into the same state. Typical employed clustering algorithms such as K-means minimize the cluster variances. Thus, they tend to divide densely sampled regions, i.e. free energy basins, into many finer clusters, while fails to identify low density regions corresponding to intermediate or transition states of interest. Conformations within the same free energy basin normally inter-convert quickly, therefore, they should belong to the same state, and it is a waste to split it further.

Xuhui Huang

Yuan Yao

Gunnar Carlsson

Vijayv Pande

Leonidas J. Guibas

In this study, we introduce a clustering method called k-center which can control the cluster radius, so that conformations within the same free energy basin are grouped together.

Materials and Methods:

We use a simple and fast 2-approximation of the k-center algorithm, which provides the guaranteed performance, i.e. generates the clusters of radius all within the given threshold. By utilizing the triangular inequality, the algorithm can be further speeded up to generate thousands of clusters from millions of conformations within several hours using a single PC. Given the parameter k , k-center clustering algorithm computes a clustering of the conformations that minimize the diameter of the clusters.

Results and Conclusion:

We applied this clustering algorithm on two simulation data: Alanine-dipeptide [1] and Fs-peptide [2]. First, this algorithm is extremely fast. E.g. we generate 4K clusters from 195K conformations within three minutes for the Alanine-dipeptide, while it takes hours for typical k-mean clustering. Second, all clusters are of similar radius and thus

cluster population directly indicates the density of the conformation space, while clusters generated from K-means clustering often varies a lot in the radius, and cluster population is not related to the density of the phase space. Finally, clusters from our algorithm can be lumped into meta-stable states using spectral clustering. The efficiency of this approach is shown to be comparable to a much more complicated algorithm involving 10 iterations of splitting and lumping based on clusters generated by k-means.

References:

- 1 Chodera et al., JCP 2007
- 2 Sorin and Pande, Biophysical J. 2005
- 3 Dasgupta and Long, J. Computer and System Science 2005

poster
45

**Guillaume A.
Troianowski**

Alison L. Marsden

Charles A. Taylor

Jeffrey A. Feinstein

Irene E. Vignon-Clementel

Hemodynamic simulations in the preoperative and virtual postoperative Fontan geometries

Purpose:

Congenital heart defects (CHD) affect 1 in 100 newborns.

While most of them can be treated without leaving much sequelae, some require intensive surgery and gravely affect the child's circulatory system. The single-ventricle defect is one such defect and leaves the child with only one operational ventricle, requiring the systemic and the pulmonary circulations to be placed in series through several operations performed during young childhood. The last surgery, the Fontan operation connecting the inferior vena cava (IVC) and the pulmonary trunk with a Gore-Tex tube (W.L. Gore & Associates), is thought to have a huge impact on the short- and long-term outcomes of the procedure on the patient. The goal of this project is to simulate the hemodynamic conditions in the pre-operative geometry and in a "virtual" postoperative geometry using pre-operative data only, in order to predict the changes of hemodynamics that occur due to the surgery and analyze them. This could be used in the long term to adapt the position and the shape of the graft specifically to each patient.

Methods:

We constructed a patient-specific 3D model of the pulmonary arteries and vena cava of a 4 year-old child from MRI data taken before the operation. We then converted this model into a postoperative model by artificially inserting the "Fontan Graft" as would be done in the surgery. The boundary conditions were derived from catheterization and PC-MRI data. This allowed us to run simulations of the Navier-Stokes equations on the pre- and post-operative models and to analyze qualitative modifications in the blood flow characteristics in order to evaluate the "virtual surgical operation" performed. We then compared the mean central venous pressure to clinical data.

Results:

The results of the simulation before the "virtual surgery" fit well the medical data and despite a difference between the simulated and actual post-operation central venous pressure, the results of the simulation could provide relevant qualitative effects in terms of velocity. Furthermore, we could observe areas where the mean wall shear stress was lower after the virtual surgery than before.

Conclusion:

So far, the method we developed provides a way to compare different Fontan configurations and thus, in the long term, adapt it to the patient's specific vessel configuration before the operation actually takes place. Furthermore, the patterns in the wall shear stress distributions suggest a possible cause of long term thrombus formation, one of several complications which these patients are later susceptible to.

Real-Time Tracking of Implanted Fiducial Markers using Combined MV and kV Imaging

poster
46

**Rodney D.
Wiersma**

**Nadeem Riaz
Wu Liu
Lei Xing**

Purpose:

In the presence of organ motion, geometric target uncertainty can hamper the benefits of highly conformal dose techniques such as IMRT. A critical step in dealing with intra-fraction tumor motion is the real-time monitoring of the tumor position. The aim of this study is the first time demonstration of a real-time 3D internal fiducial tracking system based on onboard kV diagnostic imaging together with a MV electronic portal-imaging device (EPID).

Methods:

A Varian Trilogy radiotherapy system equipped with both kV and MV imaging systems was used in this work. A hardware frame grabber was used to capture both kV and MV video streams simultaneously at 30 fps. An in house built support vector machine (SVM) classifier tool using prior CT based knowledge was used to locate gold cylindrical markers in the kV/MV frames. Controlled kV beam switching, synced with the 'step' part of a step-and-shoot IMRT delivery, was investigated in allowing continuous 3D tracking in the presence of beam interruption. A correlation/prediction algorithm was used to buffer lost geometric marker information during kV/MV synchronization. The geometric tracking capabilities of the system were evaluated using a pelvic phantom with embedded fiducials placed on a 3D moveable stage.

Results:

The maximum 3D tracking speed of the kV-MV system is approximately 9 Hz. The geometric accuracy of the system is found to be on the order of less than 1 mm in all three spatial dimensions. Synchronized kV/MV switching is found to reduce MV scatter interference on kV imaging and to reduce the overall kV diagnostic dose needed for continuous tracking.

Conclusions:

A real-time 3D fiducial tracking system using combined kV and MV imaging has been successfully demonstrated for the first time. The technique is especially suitable for RT systems already equipped with on board kV and EPID imaging devices.

poster
47

**Guanglei
Xiong**

**Alberto C.
Figueroa
Nan Xiao
Charles A. Taylor**

Simulation of Blood Flow in Deformable Vessels using Subject-Specific Geometry and Variable Mechanical Wall Property

Purpose:

Recent advances in medical imaging techniques, such as MRI and CT, can provide exquisite anatomical information of the vasculature. However, in vivo hemodynamics is difficult to measure noninvasively. Simulation of blood flow based on computational fluid dynamics (CFD) provides useful information for understanding the formation and progression of cardiovascular diseases, planning their interventions. Previous works on blood flow simulation mostly relied on idealized geometry or subject-specific geometry obtained by tedious manual processes. In addition, they often assumed rigid vessel walls or deformable walls with constant mechanical property, which is not clearly true in reality.

Material and Methods:

We propose we proposed a new workflow to construct geometric model using 3-D level set segmentation by level set method, as well as subsequent geometric processing operations including smoothing, trimming, and pruning. This workflow greatly reduces human labor and increases the objectivity of the constructed model. We then assign variable wall properties, namely Young's modulus and thickness, on the constructed model through novel centerline-based and surface-based methods. The novel centerline extraction algorithm from the model is presented as well. We extend our previous coupled momentum method to solve the problem of deformable wall simulation with variable wall property. The output of our workflow is made to be closed linked with this method.

Results:

From the medical images, on three vascular models with abdominal (from MRI), thoracic (from CT), and cerebral (from CT) aneurysms were constructed and wall properties from the literature were then assigned onto them. Appropriate boundary conditions and constraints were used to ensure realistic and stable solutions. We demonstrate the simulation pressure, velocity, wall shear stress, and displacement on these three models.

Conclusion:

The new workflow was successfully applied and we believe our simulation is much closer to the reality of blood flow in humans than previous studies by taking advantages of both 3D-based subject-specific model and variable mechanical wall property.

URL: http://www.stanford.edu/~glxiong/blood_simulation/

References:

- [1] C.A. Figueroa, I. Vignon-Clementel, K.C. Jansen, T.J.R. Hughes, C.A. Taylor. "A Coupled Momentum Method For Modeling Blood Flow In Three-Dimensional Deformable Arteries". *Computer Methods in Applied Mechanics and Engineering*, 195(41-43), 5685-5706, 2006.
- [2] G. Xiong, A. C. Figueroa, N. Xiao, C. A. Taylor. "Simulation of Blood Flow in Deformable Vessels using Subject-Specific Geometry and Variable Mechanical Wall Property", in preparation for IEEE transaction on Biomedical Engineering.

bcats 2008 thank you

Guidance and Help

Blanca Pineda
Judy Polenta
David Dill
Jeanette Schmidt
Scott Delp

Previous Organizers

Annie Chiang
Yael Garten
Jenniffer Hicks
Marc Schaub
Marina Sirota

2008 Organizing Committee

Tiffany Chen
Adam Grossman
Xuhui Huang
Hedi Razavi
Jesse Rodriguez
Katherine Steele
Rebecca Taylor

Symposium Sponsors

Platinum Sponsors

Simbios
Bio-X

Gold Sponsors

Genentech
Stanford BMIR
Simpleware
School of Medicine Career Center

Silver Sponsors

Agilent Technologies
Butte Lab
Stanford Computer Forum

Other Sponsors

Zazzle
Stanford Bookstore



Agilent Technologies

Genentech
IN BUSINESS FOR LIFE

simpleware



STANFORD CENTER FOR
BIOMEDICAL INFORMATICS
RESEARCH



STANFORD
COMPUTER FORUM



BIO-X



**STANFORD
BOOKSTORE**

Office of
TECHNOLOGY
LICENSING



zazzle



Butte Lab
Stanford Center for Biomedical Informatics Research



STANFORD UNIVERSITY
SCHOOL OF MEDICINE CAREER CENTER