bcats

biomedicalcomputationatstanford

2006

abstractbook

bcats 2006 is brought to you by
simbios • ncbi • bio-x
dell • genentech • stanford school of medicine • roche • applied biosystems
alloy ventures • affymetrix • agilent technologies
life sciences society • mathworks • bd biosciences • whole foods

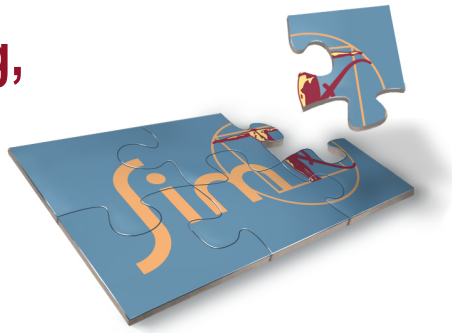
NIH Center for Biomedical Computation
Physics-based Simulation of Biological Structures

Interested in how biocomputation is changing Biology and Medicine?

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

Problems with developing, distributing, or finding biosimulation software?

Join the biosimulation supersite.
Register for a free account at: www.simtk.org



Looking for simulation software, models, and other bio-resources?

Find your answers at: www.simbiome.org

Interested in collaborating on important computational biological problems?

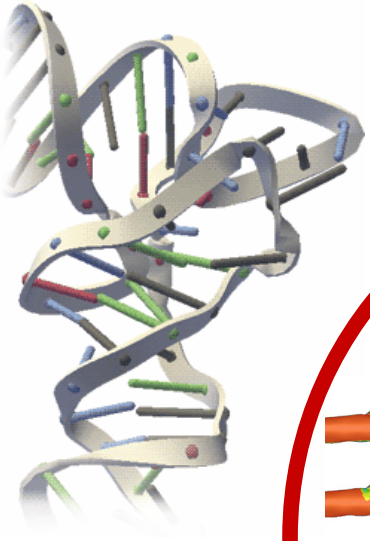
Visit us at: http://simbios.stanford.edu

NIH Roadmap Grant U54 GM072970

RNA regulates genes which affect cells, cancers, and autoimmune diseases.

See how Simbios is making fundamental discoveries about the structure and function of RNA.

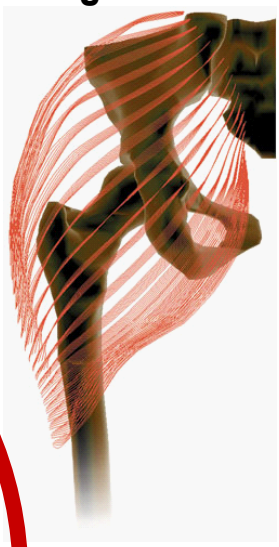
PI: Russ Altman & Dan Herschlag

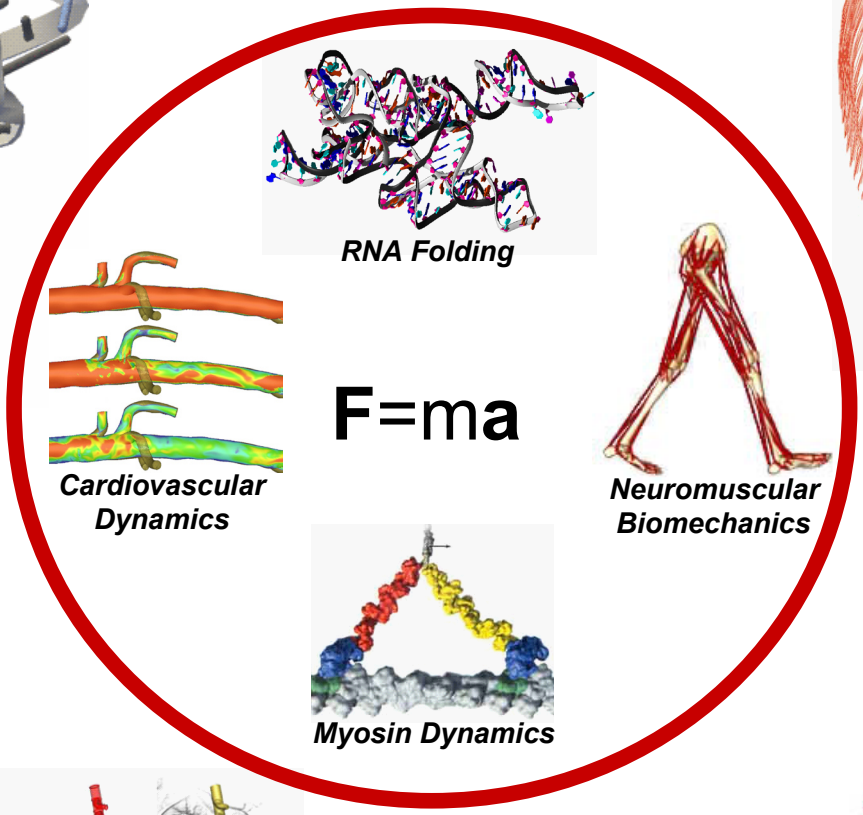


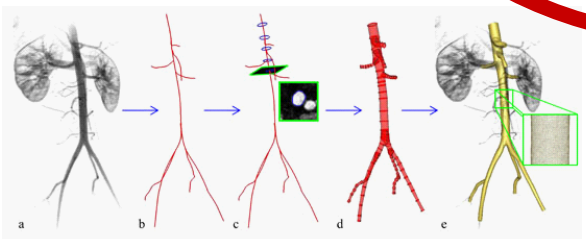
Neuromuscular biomechanics seeks to help people with cerebral palsy, Parkinson disease, birth defects, and spinal cord injuries.

See how Simbios is affecting physical therapies, orthotics, and surgeries.

PI: Scott Delp



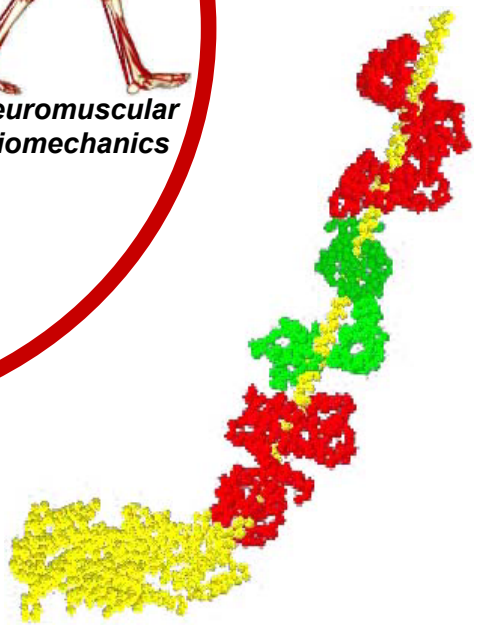




Cardiovascular disease kills more Americans than the next five diseases combined.

See how Simbios is impacting surgical, pharmacological, and catheter treatments.

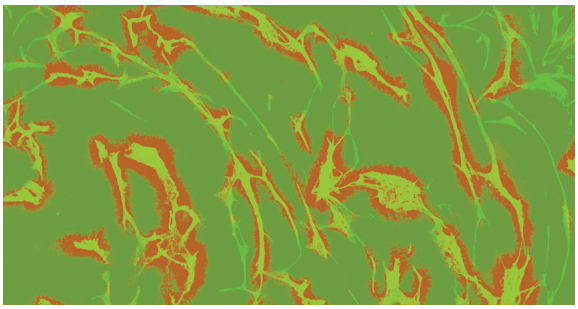
PI: Charley Taylor



Myosin is the fundamental source of motive force in living systems.

See how Simbios is deciphering this mechanical-chemical process.

PI: James Spudich



Darlene, Herceptin® Patient

BECAUSE

We are focused on truly innovative science.

For 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. The company is the leading provider of anti-tumor therapeutics in the United States.

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

Our more than 650 scientists have consistently published important papers in prestigious journals and have secured more than 5,500 patents worldwide (with an equal number 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 seeking Postdoctoral Fellows, Research Associates and Scientists in our South San Francisco headquarters in the following areas:

- Analytical Development
- Angiogenesis
- Antibody Engineering
- Assay Technology
- Bioinformatics
- Biomedical Imaging
- Cell Biology
- DMPK
- Immunology
- Medicinal Chemistry
- Molecular Biology
- Molecular Oncology
- Protein Chemistry
- Protein Engineering
- Translational Oncology
- Tumor Biology

Genentech is dedicated to fostering an environment that is inclusive and encourages diversity of thought, style, skills and perspective. Genentech offers an array of family-friendly programs, highly competitive health care benefits and a variety of onsite services for busy people with busy lives. To learn more about all of our current opportunities, please visit **www.gene.com/careers.** Please use "Ad - Stanford BioScience" when a "source" is requested. Genentech is an equal opportunity employer.

Genentech was named #1 on FORTUNE's 2006 "100 Best Companies to Work For" list.

Genentech
IN BUSINESS FOR LIFE

Welcome

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

This student-run one-day symposium provides an open, interdisciplinary forum for Stanford 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 a large applicant pool.

We thank our keynote speakers, student presenters, judges, sponsors, and all 2006 attendees.

The BCATS 2006 organizing committee

Samantha Chui, Biomedical Informatics
Chuong Do, Computer Science
Christie Draper, Mechanical Engineering
Nikesh Kotecha, Biomedical Informatics
Kaustubh Supekar, Biomedical Informatics

BCATS 2006 Schedule

8.00 am	**Onsite Registration and Breakfast**	*Sequoia Plaza*
8.45	**Opening Remarks**	*Hewlett 200*
	Scientific Talks Session I	*Hewlett 200*
9.00	**Padma Sundaram** *Colon polyp detection using smoothed shape operators*	*(page 16)*
9.15	**Gal Chechik** *Timed transcriptional control of metabolic transitions*	*(page 17)*
9.30	**Hyun Jin Kim** *Inflow boundary condition using a heart model for three-dimensional simulations of blood flow*	*(page 18)*
9.45	**Spotlight Presentations**	*(page 19)*
10.00	**Poster Session I** *(even-numbered posters)*	*Packard Lobby* *(pages 8–9, 30–84)*
11.00	**Keynote Address**	*Hewlett 200*
	Andrew McCulloch, PhD *Cardiac systems biology and multi-scale modeling*	*(page 12)*
11.45	**Lunch**	*Sequoia Plaza*
12.30 pm	**Keynote Address**	*Hewlett 200*
	David Lipman, MD *What sequence data tells us about seasonal influenza*	*(page 13)*

Scientific Talks Session II

Hewlett 200

1.15 **Anthony J. Sherbondy**

In vivo measurement of anisotropic tissue microstructure arrangement using most likely connections through diffusion imaging data *(page 20)*

1.30 **Julia Chen**

Relationship between periosteal residual stresses and strains and specific growth rates during development of the chick embryo Tibiotarsus *(page 21)*

1.45 **Ryan J. Tibshirani**

Automatic gating tools for the analysis of flow cytometry data *(page 22)*

2.00 **Spotlight Presentations** *(page 23)*

2:15 **Poster Session II** *(odd-numbered posters)*

Packard Lobby
(pages 8–9, 31–85)

Scientific Talks Session III

3.15 **Peter Kasson** *(page 24)*

Understanding membrane fusion through ensemble molecular dynamics simulation

3.30 **Evan C. Smith** *(page 25)*

Efficient auditory coding

3.45 **Su-In Lee** *(page 26)*

Identifying regulatory mechanisms associated with genetic variation

4.00 **Jeffrey A. Reinbolt** *(page 27)*

Investigating stiff-knee gait with subject-specific simulations

4.15 Industry Keynote Address

Hewlett 200

Cleve Moler, PhD *(page 14)*

Evolution of MATLAB

5.00 pm Awards and closing

Clark Center

BCATS 2006 Posters

Life Sciences Society now has the

Stanford Student Chapter

Who are we?

The Life Sciences Society (LSS), initiated in January 2005, is a non-profit organization, aiming at promoting the integration of *life sciences* with *computer science*, *mathematics*, and *engineering*, and encouraging new benefits through interdisciplinary advances involving biological and medical applications. Please visit www.lifesciencessociety.org**.**

The LSS Stanford Student Chapter is run by Stanford students. Besides providing an opportunity for students to know about different areas in life sciences, it also tries to build a bridge between the industry and the academia so that students can have a better vision of research direction and career plan, network with people with various backgrounds, and get more job information!

Why join us?

You will get all benefits that a national LSS member has, including:

- Discounts on all LSS activities, publications and at the LSS BioShop
- Complimentary subscription to The Scientist
- And a lot more… Please visit www.lifesciencessociety.org.

You will be invited to FREE monthly seminars. Speakers are outstanding professionals in various fields, from both academia and industry.

You will be informed of more job opportunities.

The membership fee is only US$20 annually for full-time students (while the national LSS membership fee is US$60)**!**

So join us NOW!

For more information or registration, please contact:

Peggy Yao (Peggy@LifeSciencesSociety.org)
LSS Stanford Student Chapter President

Please let us know if you are interested in being a committee member.

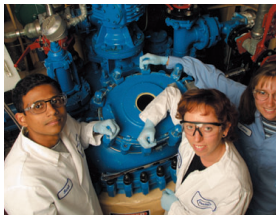
A Winning Formula

in Drug Discovery

At Roche Palo Alto, we are dedicated to discovering and developing new medicines that improve the lives of people affected with serious illnesses such as HIV/AIDS, Hepatitis C, Alzheimer's disease and arthritis.

Our Global Research Informatics Group is enabling the search for new medicines by providing information- and technology-based solutions to researchers around the world. Our areas of expertise include Bioinformatics, Discovery IM, Information Science, Infrastructure, NCD IM, and Scientific Computing.

Our park-like campus offers a wide array of professional and personal amenities designed to promote a healthy work-life balance. Our employees enjoy all the benefits of the San Francisco Bay Area, from access to leading academic and research institutions to beautiful year-round weather and a diverse intellectual community. We are pleased to provide a competitive compensation program, with generous vacation and holiday schedules, a comprehensive relocation program, and retirement and pension plans. To learn more about career opportunities at Roche Palo Alto, please visit our website at **http://paloalto.roche.com**.



Roche is a proud sponsor of BCATS!

As an equal opportunity employer, we are committed to workforce diversity.

BCATS Keynote Speaker

Andrew McCulloch, PhD

University of California, San Diego

Professor and Chair, Department of Bioengineering,
Whitaker Institute for Biomedical Engineering,
University of California, San Diego
http://cardiome.ucsd.edu/

Andrew McCulloch is Professor and Chair of Bioengineering at the University of California, San Diego, where he joined the faculty in 1987. He is a member of the UCSD/Salk Institute for Molecular Medicine, the California Institute for Telecommunications and Information Technology, a Senior Fellow of the San Diego Supercomputer Center, and a member of the Whitaker Institute for Biomedical Engineering, and the UCSD Center for Research on Biological Systems.

Dr. McCulloch was educated at the University of Auckland, New Zealand in Engineering Science and Physiology received his Ph.D. in 1986. Dr. McCulloch was an NSF Presidential Young Investigator and is a Fellow of the American Institute for Medical and Biological Engineering. He has served on the Board of Directors of the Bio-Medical Engineering Society, and is currently Associate Editor of the Journal of Biomechanical Engineering and co-Editor-in-Chief of Drug Discovery Today: Disease Models. He is on the editorial boards of the American Journal of Physiology: Heart and Circulatory Physiology and Computer Methods in Biomechanics and Biomedical Engineering. This year he has given the Konrad Witzig Memorial Lecture and the Donald Wassenberg Memorial Lecture. He co-founded Insilicomed, Inc. in 2000.

Dr. McCulloch's lab uses experimental and computational models to investigate the relationships between the cellular and extracellular structure of cardiac muscle and the electrical and mechanical function of the whole heart during ventricular remodeling and arrhythmia. Dr. McCulloch is a PI on the NCRR-supported National Biomedical Computation Resource, and has grants from the NHLBI, NSF and DOD on cardiac myocyte tissue engineering, the biomechanics of ventricular remodeling, signaling pathways in cardiac hypertrophy and failure, cardiac electromechanical interactions, and computational cardiac biology.

BCATS Keynote Speaker

David Lipman, MD

National Center for Biotechnology Information

Director, National Center for Biotechnology Information

Dr. David Lipman is currently the Director of the National Center for Biotechnology Information (NCBI), which is a division of the National Library of Medicine within the National Institutes of Health. NCBI was created by Congress in 1988 to do basic research in computational biology, and to develop computational tools, databases and information systems for molecular biology. After medical training, Dr. Lipman joined the Mathematical Research Branch of the National Institute of Diabetes, Digestive, and Kidney Diseases (NIDDK) as a Research Fellow. In his research on computational tools, he developed the most widely used methods for searching biological sequence databases. There are thousands of citations to Dr. Lipman's methods in papers which have used them to discover biological functions for unknown sequences and which have thereby advanced the understanding of the molecular basis of human disease. Since 1989, Dr. Lipman has been the Director of the NCBI, a leading research center in computational biology and one of the most heavily used sites in the world for the search and retrieval of biomedical information.

BCATS Industry Keynote Speaker

Cleve Moler, PhD

The MathWorks, Inc

Chief Scientist and Founder, The MathWorks, Inc

Cleve Moler is the original author of MATLAB and one of the founders of the MathWorks. He is currently chairman and chief scientist of the company. Before the MathWorks, he was a professor of math and computer science at the University of Michigan and the University of New Mexico. He got his PhD at Stanford in 1965 and has been a visiting professor at Stanford on three different occasions since then.

BCATS Talk Abstracts

session I
9:00am

Colon polyp detection using smoothed shape operators

Padma Sundaram
Christopher Beaulieu
Sandy Napel

Purpose

Computer-aided detection (CAD) algorithms identify locations in Computed Tomographic (CT) images of the colon that are most likely to contain polyps. All existing methods estimate a curvature-based feature (either directly or indirectly) at the isoboundary voxels. However, curvature is a smooth notion, while our data are discrete and noisy. Moreover, as a second order differential quantity, curvature amplifies noise, resulting in even noisier estimates. Our method (the Smoothed Shape Operator (SSO) method), overcomes these issues by using a geometry processing approach. Throughout, we use techniques explicitly designed for piece-wise linear surfaces. All our computations occur on the geometry, instead of in the voxel grid.

Methods

We start with a triangle mesh approximation to the mucosal surface. All of the remaining computation occurs on this triangulated surface. Our method of estimating curvature is based on deriving the Shape Operator (SO), also known as the second fundamental form. At each point on a smooth surface, this operator captures the rate of change of the normal in each direction. In order to account for noise, we smoothed the SO using a principled approach on the surface. We then computed curvature estimates from the smoothed SO and used them to cluster the vertices into candidate patches based on their surface connectivity. To identify polyp candidates, we developed a score based on the surface integral of the Gaussian curvature. The output of our-method is a list of candidate patches, sorted according to this score . We evaluated our method on a data set consisting of 35 patients containing a total of 122 polyps ranging in size from 1.8 mm to 27 mm. We evaluated our method over various polyp sizes. We compared our results to the performance of an existing method (Surface Normal Overlap (SNO) method).

Results

At a FP rate of 25 per case, the sensitivity of our method was 80.3% (all sizes), 60% ([0,5) mm), 89.6% ([5, 10) mm) and 100% (>=10 mm), compared to 38.5% (all sizes), 15.9% ([0,5) mm), 43.1% ([5,10) mm) and 75% (larger than 10 mm) for the SNO method.

Conclusion

Diffusion of SOs on the geometry, as opposed to smoothing scalar curvature values or smoothing raw CT images, results in increased coherency among local principal curvature directions. In general, the availability of high resolution data calls for a shift in the approach to polyp detection: from image processing to geometry processing. Our method is a step in that direction. Preliminary results indicate that our method may have the potential of acting as a reliable second reader. This could help help CT colonography evolve into an accepted procedure in the clinical setting.

session I
9:15am

Timed transcriptional control of metabolic transitions

Gal Chechik

Aviv Regev
Daphne Koller

Metabolic processes are tightly controlled and coordinated based on the cell's environment. Upon a sudden change in environmental conditions, a cell may have to drastically reconfigure its metabolism and re-route its fluxes. Recent works demonstrated that transcriptional control of metabolic enzymes plays an important role in the transition between metabolic states. In particular, it was shown that transcriptional timing of AA E.coli enzymes is tightly controlled, resulting in sequential activation of enzymes following the order of the linear pathway in which they participate. While this example is instructive, the role and timing of transcription in controlling metabolism has not been studied on a systems scale.

Here, we present a comprehensive approach to discover temporal patterns in the expression of metabolic genes that are significantly coupled to the relations between the corresponding enzymes within the metabolic network. Our approach relies on a novel parametric model of gene expression time courses that captures typical behaviors of transcription responses following environmental change. Our model maps a single gene's expression profile into a compact parametric representation reflecting onset and offset time with good temporal resolution. We next perform a systematic search over network motifs in the full metabolic network of S. cerevisiae, to detect ordered temporal patterns within motifs that are significantly over-abundant. Our analysis indicates several classes of temporal motifs including "just-in-time" activation in linear pathways, "fast shut off" of genes in reverse order, and "same time activation" in triplets of enzymes that have a logical-AND relation. Processes could exhibit different timing patterns to respond optimally to different environmental changes. For example, after a heat shock, Glycerol is produced in a just-in-time manner to reduce protein denaturation, and its production is slowly and efficiently stopped under de-heating. However when exposed to an oxidative agent, Glycerol production is quickly stopped (though in a wasteful manner).

Interestingly, temporal motifs typically extend across traditional pathway "boundaries", occur across multiple environmental conditions, and cluster to form "hyper-motifs" that densely cover particular sub-parts of the full metabolic network. We also find that temporally-ordered transcription can often be explained by gradual low-to-high binding affinities of a few transcription factors, suggesting a possible mechanism to account for fine timed transcriptional control.

These findings identify the central points of cellular control on metabolic re-programming, and provide a systems level view into its adaptive use.

session I
9:30am

Inflow boundary condition using a heart model for three-dimensional simulations of blood flow

Hyun Jin Kim
Irene E . Vignon-Clementel
Kenneth E. Jansen
Charles A. Taylor

Purpose

Arterial pressure and flow result from the interaction between the ventricle and the arterial circulation. However, previous simulations of blood flow in arteries has not considered the interaction between the ventricle and the aorta and instead utilized prescribed velocity profiles as the inflow boundary condition. In this work, we used a novel approach to assigning boundary conditions in blood flow simulations [1] to couple the inflow to a heart model to account for the interaction of the ejecting ventricle and the aorta.

Materials and Methods

We used a lumped-parameter heart model that couples the inlet arterial flow and the pressure in systole [2]. In diastole, the heart model is decoupled from the downstream vasculature and zero flow is prescribed at the aortic inlet. Moreover, during diastole the heart model is utilized to calculate the change in ventricular volume and pressure based on the myocardial relaxation and influx from the atrium. To simulate the contraction and relaxation of the ventricle in time, we scaled a normalized elastance function using physiologically relevant parameters [3]. A stabilized three-dimensional finite element method is used to compute the pressure and flow in the cardiovascular model originating from the left ventricle.

Results

The arterial pressure and flow waveforms were obtained for a simple tube model using this inflow boundary condition. The changes in cardiac properties were simulated to show that the heart and the arterial circulation are coupled and interact with each other.

Conclusion

We have shown that a lumped-parameter heart model can be coupled to the arterial system as an inflow boundary condition and can be solved implicitly by considering the interaction between the ventricle and the arterial circulation. Using this approach, we can simulate how the arterial flow and pressure are affected by the cardiac property changes and vice versa.

References

[1] Irene E. Vignon-Clementel, C. Alberto Figueroa, Kenneth E. Jansen and Charles A. Taylor, "Outflow boundary conditions for three-dimensional finite element modeling of blood flow and pressure in arteries," Computer Methods in Applied Mechanics and Engineering, 195:3776-3796, 2006.

[2] Patrick Segers, Nikos Stergiopulos, Nico Westerhof, Patrick Wouters, Philippe Kolh and Pascal Verdonck, "Systemic and pulmonary hemodynamics assessed with a lumped-parameter heart-arterial interaction model," Journal of Engineering Mathematics, 47:185-199, 2003.

[3] Hideaki Senzaki, Chen-Huan Chen and David A. Kass, "Single-beat estimation of end-systolic pressure-volume relation in human: A new method with the potential for noninvasive application," Circulation. 94:2497-2506, 1996.

session I
9:45am

Spotlight Presentations

9:45	**Dana Paquin**	
	Multiscale deformable registration of medical images	(page 44)
9:48	**Rachel Weinstein**	
	Impulse based PD control for joints and muscles	(page 40)
9:51	**Xing Chen**	
	Flux control in metabolic networks: optimality and robustness	(page 64)
9:54	**Irene Liu**	
	Building a nosology based on disease etiology	(page 66)

session II
1:15pm

In vivo measurement of anisotropic tissue microstructure arrangement using most likely connections through diffusion imaging data

Anthony J. Sherbondy

Michael Ben-Shachar
Robert F. Dougherty
Pat Hanrahan
Sandy Napel
Brian A. Wandell

Purpose

Diffusion imaging coupled with fiber tractography (DFT) measures the course of fibers through living tissue. These measurements have been demonstrated for central nervous tissue, cardiac muscle and skeletal muscle. Despite the success of current DFT methods, many limitations still exist. Our work focuses on the use of DFT for applications in which we are confident that two anatomical regions are connected and we wish to use DFT to identify the most likely pathways between these regions.

We introduce a probabilistic DFT algorithm (MetroTrac) to achieve this goal. We introduce the algorithm in three parts: a scoring procedure to measure the likelihood of a pathway, a sampler to systematically explore the set of possible pathways, and an inferential step that specifies how the sampled pathways can be used to derive a conclusion.

Material and Methods

We define a scoring procedure designed to reflect the anatomical plausibility of a pathway given the data. We introduce properties of symmetry and independence that we believe essential for such scoring, but are generally overlooked by tractography algorithms. The sampling algorithm explores the space of potential pathways that have endpoints within two end regions. Since the number of possible pathways is infinite, we develop sampling strategies that are efficient by focusing on high scoring pathways. We introduce one strategy based on the Metropolis algorithm and one based on importance sampling and resampling. The inference we make is that the highest scoring pathways are the most anatomically valid ones between any two endpoints. Thus, we derive a number of quantitative measures from the estimated pathways, including properties such as fiber group location, diffusion along fibers, fiber length, and so forth.

Results

In a series of experiments using the images from the brains of four human subjects, we show that MetroTrac estimates positions of known pathways that are consistent with those found using an existing popular tractography algorithm. We also demonstrate cases in which MetroTrac finds high scoring pathways that are missed by the previous tractography algorithm.

We perform similar measurements of fiber geometry in the gastrocnemius muscle in one human subject. Skeletal muscle measurement capability of the current system is compared with other deterministic tractography results.

Conclusion

MetroTrac offers a foundation for measuring the cellular geometry of living tissue by estimating the most likely pathways connecting anatomical regions within diffusion imaging data.

Relationship between periosteal residual stresses and strains and specific growth rates during development of the chick embryo Tibiotarsus

session II
1:30pm

Julia Chen

Betty Zhao
Michael Longaker
Jill Helms
Dennis Carter

Purpose

The periosteum is a thin elastic membrane that surrounds all bones and envelops osteogenic progenitor cells. Rapid bone growth during development may stretch the periosteum, creating biaxial tensile residual stress in vivo that may serve as bone-forming stimuli [1]. Growth-generated stresses and strains in the periosteum could modulate bone growth and morphology by affecting cell biology through mechanotransduction [2]. This study correlates residual stresses and strains in the periosteum with specific growth rates of the chick embryo tibiotarsus.

Materials and Methods

Tibiotarsi from embryonic days e11-20 were microCT scanned to determine specific growth rates from power-fitted growth curves. Longitudinal or circumferential periosteal incisions made during e14-20 were imaged with a dissection microscope to measure opening dimensions. Circumferential incisions released longitudinal residual strains and vice versa. An orthotropic finite element model (COMSOL) was created to determine residual strains from opening dimensions. The model was 2.25x12 mm (mean quarter geometry) with dimension effects analyzed. Circumferential (EC) to longitudinal (EL) moduli ratio was varied from 1-10, and Poisson's ratio (vCL) was estimated as 0.45 to account for anisotropy and non-linearity observed in periosteum [3,4]. Accuracy was verified by the Westergaard solution. Correlation coefficients were calculated for statistics.

Results

Longitudinal specific growth rate decreased from 0.17 percent/day to 0.09 during e11-20. Circumferential rate decreased from 0.14 to 0.08. Both were similar in scale and decreased by half. When EC/EL=10, longitudinal residual strains were high- 105.4% at e14 decreasing to 73.3% by e20. Circumferential strains decreased from 9.9% to 6.8%. Both showed significant negative correlation with age. Strain calculation was sensitive to EC/EL. Since material properties were unknown, relative residual stresses (SL/SC) were evaluated, averaging 1.14 MPa/MPa, or almost equal. Dimensional changes incurred errors of ~3%.

Conclusion

In a growing chick embryo tibiotarsus, the periosteum may experience equibiaxial tensile stress, as well as large biaxial tensile strains in vivo. Residual strains decreased with lower specific growth rates. Similar longitudinal and circumferential specific growth rates are the likely cause of equibiaxial tensile stress. High growth-generated residual stresses and strains may be potent bone-forming stimuli for periosteal progenitor cells. Further investigations are needed to accurately determine periosteum material properties for modeling purposes.

References

[1] Rooney and Archer, 1992 [2] Henderson and Carter, 2002 [3] Bertram et al, 1998 [4] Uchiyama et al, 1998

session II
1:45pm

Automatic gating tools for the analysis of flow cytometry data

Ryan J. Tibshirani

Jonathan Irish
Nikesh Kotecha
William Lu
Robert Tibshirani
Garry P. Nolan

Purpose

Flow cytometry measures multiple features of every cell in a sample and is widely used in immunology, cancer biology, drug discovery, and human disease diagnosis. Many steps in the analysis of flow data currently require manual gating, a time-consuming process wherein a scientist identifies subpopulations by drawing regions or "gates" based on fluorescence. This is a considerable bottleneck in two important analysis steps: compensation and fluorescent cell barcoding (FCB). The former corrects for spectral overlap between fluorescent markers and the latter allows multiple samples to be stained and collected in unison. We present algorithms for automated discovery of cell populations and apply them to automate gating in compensation and in FCB.

Materials and Methods

1. Automated compensation: We have created an algorithm that selectively gates the control populations used in compensation. For each control collected, the algorithm first determines the channel where the primary signal was measured (the user need not specify this) and draws rectangular gates around the control populations by examining local minimums in a computed density function.
2. Automated barcode gating: We have used FCB to encode and stain 96 populations in a single tube. This involves using additional fluorescence dimensions to create a 6 x 4 x 4 encoding system and subsequent decoding via gating. Customarily, these gates require the user to manually draw at least 30 gates and use combinations of them to identify the original 96 populations, a process that can take several hours. We have written an algorithm that uses a combination of k-means clustering and principal component analysis to automatically gate these 96 populations.

Results

Our testing shows that both algorithms are highly accurate. The automatic compensation algorithm had an average performance time of 0.0793 seconds across five experiments compared to the 10 minutes it takes to do manually. Our FCB gating algorithm averaged 0.2979 seconds across three experiments, a significant time savings over the 2 hours it takes to gate manually.
Both algorithms are deterministic, meaning that they will always return the same result given the same data. This is important because it ensures that two people analyzing the same experiment will always get the same gating result using our algorithms, and moreover it provides a standard for gating across experiments.

Conclusion

Our algorithms reduce analysis time and are as accurate as the manual gating process; they provide a standard for gating across experiments. Automatic discovery of subpopulations is a common analysis hurdle in many scientific disciplines. Thus there is likely to be applications of our algorithms that extend beyond flow.

session II
2:00pm

Spotlight Presentations

2:00 **Rashmi Raj**

Multi-relational data mining of time-oriented biomedical databases *(page 61)*

2:03 **David J. Carlson**

Effects of linear energy transfer (LET) on intrinsic radiation sensitivity – tests of the putative mechanisms underlying the cell killing effects of ionizing radiation *(page 81)*

2:06 **Gilwoo Choi**

Quantification of radial compression and deflection of superficial femoral artery due to musculoskeletal motion *(page 33)*

2:09 **Greg M. Goldgof**

CONTRAfold: RNA secondary structure prediction for single and multiple sequences without physics-based models *(page 55)*

session III
3:15pm

Understanding membrane fusion through ensemble molecular dynamics simulation

Peter Kasson

Vijay Pande

Purpose

Membrane fusion is a fundamental process in cell biology, responsible for many cellular functions including neurotransmitter release, peptide hormone secretion, and infection by enveloped viruses such as influenza or HIV. Understanding the fusion process will address a fundamental problem in biology, and the ability to manipulate fusion will enable novel therapies for diseases such as influenza and Parkinson's. Fusion is mediated by complex yet transient protein-lipid assemblies, posing a challenge to traditional structural and biochemical investigation. To date, no experimental approaches have yielded high-resolution structures of fusion intermediates, and details of the fusion mechanism remain a matter of much debate. We have developed novel simulation techniques to generate high-resolution structural predictions and quantitative mechanistic models of fusion consistent with available experimental data. We correlate these models with functional data on fusion protein mutants, generating novel structural hypotheses for fusion protein-lipid complexes.

Methods

Simulation of vesicle fusion on the needed timescales is beyond the reach of traditional supercomputers. Therefore, we use worldwide distributed computing to simulate thousands of individual fusion reactions via molecular dynamics. We then employ Markovian State Model analysis to build mechanistic models from the molecular dynamics data, allowing robust statistical comparisons of mechanism and kinetics over sub-millisecond timescales—100x longer than previously attainable.

Results

We report simulation of thousands of successful fusion events for a chemically detailed model of lipid vesicles. Systematic statistical analysis of these fusion events yields new predictions for membrane fusion mechanisms that have an enhanced ability to explain existing experimental data. Novel predictions for the structure and dynamics of fusion protein complexes help explain experimental data on fusion protein mutants. We also perform comparative analyses of how fusion mechanisms depend on the lipid environment, illustrating another means for cellular regulation of fusion.

Conclusion

We have developed powerful simulation methodology for generating quantitative and high-resolution mechanistic models of membrane fusion, a process essential to cellular function. These models provide new capabilities to integrate and explain experimental findings and to suggest novel hypotheses for how cells and viruses manipulate lipid membranes. Our work represents a generalizable advance in simulation technologies and analytic methods and also suggests specific mechanistic hypotheses regarding fusion.

session III
3:30pm

Efficient auditory coding

Evan C. Smith
Michael Lewicki

Purpose
The auditory neural code must serve a wide range of auditory tasks that require great sensitivity in time and frequency and be effective over the diverse array of sounds present in natural acoustic environments. It has been suggested (Barlow, 1961; Atick, 1992; Simoncelli & Olshausen, 2001; Laughlin & Sejnowski, 2003) that sensory systems might have evolved highly efficient coding strategies to maximize the information conveyed to the brain while minimizing the required energy and neural resources. Here we show that, for natural sounds, the complete acoustic waveform can be represented efficiently with a nonlinear model based on a population spike code.

Methods
In this model, idealized spikes encode the precise temporal positions and magnitudes of underlying acoustic features (Smith & Lewicki, 2005). These acoustic features are learned by adapting a dictionary of kernel functions to the statistics of a large database of natural sound signals using this spike code.

Results
We find that when the features are optimized for coding either natural sounds or speech, they show striking similarities to time-domain cochlear filter estimates, have a frequency-bandwidth dependence similar to that of auditory nerve fibers, and yield significantly greater coding efficiency than conventional signal representations (Smith & Lewicki, 2006).

Conclusion
These results indicate that the mammalian auditory code might approach an information theoretic optimum with respect to behaviorally relevant sounds from our natural environment. Furthermore, the acoustic structure of speech might itself be adapted to the coding capacity of our mammalian auditory system.

References
Atick, J. J. (1992). Could information-theory provide an ecological theory of sensory processing. Network, 3(2):213–251.
Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. In Rosenbluth, W. A., editor, Sensory Communication, pages 217–234. MIT Press, Cambridge.
Laughlin, S. B. and Sejnowski, T. J. (2003). Communication in Neuronal Networks. Science, 301(5641):1870–1874.
Smith, E. & Lewicki, M. (2006). Efficient auditory coding. Nature, 439(7079):800–805.
Smith, E. C. and Lewicki, M. S. (2005a). Efficient coding of time-relative structure using spikes. Neural Computation, 17(1):19–45.
Simoncelli, E. and Olshausen, B. (2001). Natural image statistics and neural representation. Annual Review of Neuroscience, 24:1193–1216.

session III
3:45pm

Identifying regulatory mechanisms associated with genetic variation

Su-In Lee

Dana Pe'er
Aimee M. Dudley
David Drubin
Pamela Silver
George M. Church
Daphne Koller

Purpose

One of the fundamental goals in biology is to identify the genetic variations that create the phenotypic diversity among individuals. There has been an intensive research effort on identifying the genetic sequence variations that change various phenotypes, including many genetic diseases. Recently, a new framework, treating an expression level of genes as 'expression phenotype' and identifying the causative genetic variations, started to shed light on molecular-level understanding of the consequence of genetic sequence variation to the phenotypic diversity. In this study, we propose a probabilistic model approach, called Geronemo (genetic regulatory network of modules) that can reveal regulatory mechanisms by which genetic sequence variations perturb the complex web of regulatory interactions. In this talk, we present that Geronemo shows stronger statistical power than traditional approaches and reveals two novel mechanisms associated with the sequence variations in yeasts, from two very different biological contexts: chromatin modification and mRNA degradation

Materials and Methods

Given the expression and genotype data from 112 genetically different yeast individuals, Geronemo automatically constructs a set of co-regulated genes (modules), whose regulation can involve both expression and genotype of regulators. By exploiting the modularity of biological systems, Geronemo reveals regulatory relationships that are indiscernible when genes are considered in isolation, allowing the recovery of intricate combinatorial regulation. It can also reveal an indirect effect of sequence variations to the expression variation through the expression of regulators.

Results

Geronemo reveals novel regulatory mechanisms caused by the genetic perturbations in the yeast regulatory network. We find a significant overlap between the modules with chromosomal characteristics and those regulated by chromatin modification proteins. Also, a large fraction of the variance in the expression can be explained by a small number of markers associated with chromatin modifiers. Our results suggest that a significant part of individual expression variation in yeast arises from evolution of a small number of chromatin structure modifiers. Geronemo reveals a novel regulatory mechanism in a totally different biological context. Our experiments support the novel regulatory connection, derived by Geronemo, among three mechanisms: regulation of translation initiation and RNA de-adenylation by Puf3; mRNA decapping by Dhh1; and the translation initiation repressor Gcn20.

Conclusion

Our results show Geronemo's ability to reveal mechanisms associated with genetic variation, thereby showing potential to help understanding genetic diseases such as type II diabetes.

Investigating stiff-knee gait with subject-specific simulations

session III
4:00pm

Jeffrey A. Reinbolt

Melanie D. Fox
Scott L. Delp

Purpose
Stiff-knee gait is a symptom of spastic cerebral palsy characterized by diminished knee flexion during the swing phase of gait. This diminished knee flexion has been attributed to excessive excitation of the rectus femoris (RF), a knee extensor, during swing [1]. It has more recently been observed that many stiff-knee patients exhibit excessive knee extension moments prior to swing [2]. Our first aim evaluated whether abnormal RF excitation prior to swing or during swing has a greater influence on peak knee flexion.

Rectus femoris transfer surgery, a common treatment for stiff-knee gait, reattaches the distal tendon to a new site, such as the sartorius muscle. Some patients show dramatic improvement after this surgery while others suffer further impairment. Our second aim evaluated the utility of computer simulations to determine the potential efficacy of RF transfer.

Materials and Methods
This study included ten cerebral palsy patients who exhibited stiff-knee gait and underwent RF transfer. Five patients were classified as "good outcomes" and five as "poor outcomes" based on measurements of postoperative knee flexion [2]. We generated subject-specific simulations of each patient using a musculoskeletal model with 21 degrees-of-freedom and 92 muscles. We used computed muscle control [3] to solve for muscle excitations that result in a simulation consistent with the measured preoperative gait. The first aim was investigated by eliminating RF excitation separately prior to swing and during swing. The simulated effects on peak knee flexion were compared for each subject, considering excitation magnitudes during each phase. The second aim was investigated by simulating RF transfer to the sartorius. The simulated effects on peak knee flexion were compared between good and poor outcome groups.

Results
Peak knee flexion was influenced more by abnormal RF excitations prior to swing compared to those during swing, given the excitation magnitudes during each phase. This finding redefines indications for RF transfer which traditionally focus solely on swing. Simulations of RF transfer show a significant difference ($p = 0.0296$) in peak knee flexion improvement between good (40 deg) and poor (28 deg) outcome groups. This difference illustrates the utility of subject-specific simulation for investigating treatments for stiff-knee gait.

Conclusion
Subject-specific simulation is a powerful tool for identifying the biomechanical causes of an individual's gait abnormalities and has the potential to improve treatment planning.

References
1. Perry J, 1987. Dev Med Child Neurol, 29, 153-8.
2. Goldberg S, et al., 2006. J Biomech, 39, 689-98.
3. Thelen D, et al., 2003. J Biomech, 36, 321-8.

We're not Thor

We just make his hammer

Developing a device that connects the world. Testing the water to make sure it stays clean. Discovering a cure that keeps the world safe. They all require the same thing: the right tools.

With Agilent we make sure you have them. Our experience in the fields of electronics, communications, life science and chemical analysis gives us a unique perspective shared by no other company in the world. And we build that expertise into every product we make, from integrated test and measurement solutions to advanced technologies and breakthrough components.

So whether you're a titan of industry or on the verge of becoming one, trust the high-tech toolmakers at Agilent. We'll help make you stronger.

www.agilent.com

Agilent Technologies

dreams made real

BCATS Poster Abstracts

poster 0

Evaluation of hemodynamic efficiency in a new "Y-Graft" design for the fontan operation

Adam Bernstein
Alison L. Marsden
Irene E. Vignon-Clementel
Jeffrey A. Feinstein
V. Mohan Reddy
Charles A. Taylor

Purpose

Computational fluid dynamics can be used to evaluate the performance of new surgical procedures at no risk to the patient. In this study, we used fluid simulations to redesign the total cavopulmonary connection (Fontan procedure), an operation used to treat single ventricle congenital heart defects. These patients, representing 20 per 100,000 live births, have severe underdevelopment of one side of the heart resulting in poor circulation, low oxygen levels, and eventual heart failure. During the Fontan procedure, surgeons connect the inferior vena cava and superior vena cava end-to-side to the left and right pulmonary arteries, resulting in a t-shaped junction. Our previous studies have shown high-energy losses and inefficiencies in this geometry, particularly during exercise. In the current study, we hypothesize that redesigning the Fontan procedure using a Y-shaped inferior vena caval graft will result in a more energy efficient system.

Materials and Methods

The new Y-graft Fontan model was constructed by modifying patient specific MRI image data using known Y-graft dimensions. Resistance boundary conditions were enforced on the outlets of the pulmonary vasculature using pressure data from patient's cardiac catheterizations. Respiratory waveforms were constructed using flow data from Phase Contrast MRI. Time-dependent 3-D blood flow simulations were then performed using a finite element solver. Pressure and energy losses were analyzed during rest and exercise conditions ranging between 2-4 times resting flow rates.

Results

Under resting conditions, simulation results corresponded with expected pressure drops obtained during cardiac catheterization. During both rest and exercise cases, we observed moderate gains in efficiency using the Y-graft model compare to the original Fontan geometry.

Conclusion

We have demonstrated a moderate increase in energy efficiency in the new Y-graft model during both rest and exercise flow conditions. Our virtual redesign of the Fontan procedure marks the first time that a computer-generated model has been utilized in surgical innovation for congenital heart disease.

poster 1

An algorithm for generating muscle-actuated simulations of long-duration movements

Chand T. John
Frank C. Anderson
Eran Guendelman
Allison S. Arnold
Scott L. Delp

Purpose

How does the human body generate movement? This question must be answered to optimize athletic performance, understand muscle function, and treat movement disorders. Experiments alone cannot capture the cause-effect relationships leading to movement generation, in part, because important variables, such as muscle forces, are not measurable. However, physics-based simulation, in combination with experiments, can elucidate the complex interactions between the elements of the neuromusculoskeletal system leading to movement generation.Recent application of robotics-based control has enabled simulation of walking and other movements from a dynamic model of the musculoskeletal system and motion capture data [3]. However, dynamic inconsistencies between the measured body motions and measured external forces [2] have limited the durations of these simulations to about half a second. Previous approaches to generate longer simulations have either applied large external forces called residuals, or dramatically altered the body motions to eliminate the need for external forces [3].

Materials and Methods

We have developed a new method called the residual reduction algorithm (RRA) that enables coordinated simulation of long-duration movements by altering the body motions in such a way that the residuals are reduced rather than eliminated. We have implemented RRA in SimTrack [1], a software framework for generating simulations of movement. RRA works by stepping forward in time in small increments, solving an optimization problem to compute the residuals and joint torques needed to drive the dynamic model to follow the measured body motions. Because the residuals and joint torques do not perfectly reproduce the forces applied by the subject's muscles during motion capture, the dynamic model's motion may differ slightly from the measured body motion. We used SimTrack to generate a simulation of normal walking over 2 seconds long for one subject.

Results

The muscle excitations in the simulation were consistent with the EMG recorded for the subject. The simulation matched the original joint angles to within 1 degree. Application of RRA dramatically reduced the residuals.

Conclusion

RRA has enabled the generation of longer-duration simulations than were previously possible in our framework. We believe RRA will enable the generation of simulations of a variety of long-duration movements and yield a new way of investigating a wide range of fundamental questions in biomechanics.

References

[1] Delp et al. (in press). IEEE Trans Biomed Eng. [2] Kuo (1998), J Biomech Eng 120, 148-59. [3] Thelen and Anderson (2006), J Biomech 39, 1107-15.

poster 2

A dynamic three-dimensional finite element model of the patellofemoral joint

Edith M. Arnold

Thor F. Besier
Scott L. Delp
Gary S. Beaupre

Purpose

Patellofemoral pain (PFP) is a common and debilitating disorder of the knee. The pathway to PFP is complex and involves many complicated structures. A possible mechanism for pain is increased stress in the cartilage at the patellofemoral joint that triggers nerve endings in the underlying bone. Finite element analysis has become a powerful means for understanding patellofemoral mechanics. Although these models have provided valuable information about cartilage stress they represent quasi-static loading scenarios. Patellofemoral pain is often associated with dynamic events that involve large loads, such as running or jumping. Previous models also neglect many of the surrounding soft-tissue structures of the knee. We developed a dynamic three-dimensional finite element model of the patellofemoral joint that will enable us to examine contact during 0-30° of flexion and transient stress during dynamic events.

Materials and Methods

Three-dimensional models of the knee were developed from MR images of live subjects. Our models utilized an explicit integration scheme, which enabled computationally inexpensive formulations to resolve deformation and generalized contact between many structures with few convergence problems. The patellar and femoral cartilage were modeled as deformable, eight-node linear brick elements. The patella, femur, and tibia were represented by rigid four-node quadrilateral elements. Quadriceps tendons, the patellar ligament, and other "soft-tissue" structures were composed of four-node, three dimensional membrane elements. Muscle forces were applied via connector elements attached to the quadriceps tendons. The material properties of the cartilage were assumed to be linear elastic, homogenous, and isotropic. Tendons were assigned tension-only, linear elastic, homogenous, isotropic properties. Bone was constrained to be rigid since it is so much less deformable than the cartilage.

Results

Preliminary simulations run successfully and show plausible motion of the patella during knee flexion. Cartilage stress distributions and contact pressure patterns resemble those calculated in quasi-static simulations. Modeling the quadriceps as membrane elements with general contact produces the expected wrapping around the femur during deep flexion.

Conclusion

This model provides novel and useful information about cartilage stress during dynamic events at very low computational expense. It has the potential to enable several studies which are not possible with current models. These include modeling soft tissues to determine their contribution to patellar motion, determination of cartilage stress during 0-30° of flexion, and investigation of dynamic stresses during running.

poster 3

Quantification of radial compression and deflection of superficial femoral artery due to musculoskeletal motion

Gilwoo Choi
Christopher P. Cheng
Ga Young Suh
Robert J. Herfkens
Charles A. Taylor

Purpose

The cyclic deformations in stents due to musculoskeletal movement [1] have been hypothesized to cause the high prevalence of stent fractures [2] in the superficial femoral artery (SFA). Our goal was to quantify in vivo radial compression and deflection (buckling) of the SFA due to hip and knee flexion [3].

Methods

Seven male healthy volunteers (56±5 years) were imaged in the supine and bent leg positions (hip flexion angle 39±6°, knee flexion angle 86±6°) using a time-resolved 3D gradient-recalled MRA sequence (TRICKS). Geometric models of the SFA were created using custom software. Radial compression of the SFA was quantified by changes in the ratio of major and minor axes (circularity index) of the best-fit ellipse to the cross-section of the lumen. Deflection was defined as the maximum out-of-axis distance for every 5 cm segment of the SFA. The SFA was divided into three equal parts (top, middle, and bottom) for comparison.

Results

From supine to bent leg position, the SFA compressed radially by 5.4±6.3% ($p<0.05$) on average. Deflection of the SFA increased from the supine to the bent leg position by 5.7±4.0mm ($p<0.001$) with a maximum range of 15 mm in a distal segment

Conclusions

All segments of the SFA experience radial compression whereas only the proximal and distal segments exhibit significant buckling with hip and knee flexion. These findings may aid in developing pre-clinical tests for SFA stents.

References

[1]Smouse, H.B., et al., Changes in major peripheral arteries during joint movement before and after stent placement in the cadaver model, Transcatheter Cardiovascular Therapeutics 2004, Washington, D.C.
[2]Scheinert, D., et al., Prevalence and clinical impact of stent fractures after femoropopliteal stenting. J Am Coll Cardiol, 2005. 45(2): p. 312-5.
[3]Arena, F.J., Arterial kink and damage in normal segments of the superficial femoral and popliteal arteries abutting nitinol stents--a common cause of late occlusion and restenosis? A single-center experience. J Invasive Cardiol, 2005. 17(9): p. 482-6.

poster 4

Three-dimensional analysis of respiratory motion of renal artery

Ga Young Suh
Mary T. Draney
Gilwoo Choi
Charles A. Taylor

Purpose

Daily respiration induces bending and twisting of renal arteries. This motion can be the possible cause of renal stent fracture and thrombosis. Moreover, this is important factor for renal artery stent design. However, renal artery motion during breathing in three-dimension is not quantified yet. In this study, we analyzed the complex motion of renal artery and kidney using MRA-based three-dimensional model.

Methods

Eight male volunteers (age 57±16) provided image data with contrast-enhanced magnetic resonance angiography (MRA) using a 1.5T magnet. For the respiratory motion, we acquired two phases of the MRA in separate breath-holds during contrast injection (inspiration breath-hold) and after contrast (expiration breath-hold). Using custom software ASPIRE2, we constructed three-dimensional model of renal arteries. Deflection, curvature and torsion were calculated for each inspiration and expiration phase. Also, we measured the transition of ostia as well as the position of kidney to quantify the proximal and distal translation or renal artery.

Results

Comparing inspiration and expiration phase, we found that the left renal artery has more curvature change near ostia than the right side. Also, on both sides, the displacement of ostia has been observed as well as translation of kidney.

Conclusions

We have shown that renal artery experiences complex bending and twisting during respiration cycle. Also, kidneys are moving in medial-lateral direction as well as superior-inferior direction. This respiratory motion may affect the renal stent inducing tension, compression and torsion and contributing to stent fracture.

References

[1] Draney, M.T., et al., Three-dimensional analysis of renal artery bending motion during respiration. J. Endovasc. Ther. 2005. 12: p380-386
[2] Bessias, N., et al., Renal artery thrombosis caused by stent fracture in a single kidney patient. J Endovasc Ther. 2005. 12: p516-520
[3] Ahmad, N.R., et al., Respiration-induced motion of the kidnys in whole abdominal radiotherapy: implications for treatment planning and late toxicity. Radiotherapy and Oncology. 1997. 42: p87-90

poster 5

Evaluating the effect of a tibial torsion bone deformity on muscle function during gait

Jennifer Hicks

Allison Arnold
Frank Anderson
Michael Schwartz
Scott Delp

Purpose

Children with cerebral palsy often walk with excessive hip and knee flexion, a movement pattern known as crouch gait. Tibial torsion, a rotational deformity of the bone in the lower leg, may contribute to this abnormal gait pattern. The goal of this study was to determine, using a musculoskeletal model, how tibial torsion leads to crouch gait and when surgery should be considered. First, we tested the hypothesis that tibial torsion inhibits the capacity of the ankle plantarflexor muscles to support the body by reducing their moment arms. We also tested the hypothesis that tibial torsion alters the joint axis orientations, therefore reducing the capacity of muscles to extend the hip and knee by altering the dynamic interactions between muscles and the skeleton during gait.

Materials and Methods

Multi-segment, 3D models of the musculoskeletal system [1] were created with tibial torsion deformities ranging from 0° to 60° (Figs 1,2). First, moment arms of the plantarflexors were calculated for each model. Second, the capacity of muscles to extend the joints during gait was determined for each model. At every 2% of the single limb stance phase of gait, a unit muscle force and its corresponding ground reaction force [2] were applied to the model. The resulting accelerations of the hip and knee were calculated, which gave a measure of the muscle's capacity to extend the joints during gait in the presence of excess torsion.

Results

Tibial torsion had a minimal effect on the moment arms of the plantarflexors, as the percent decrease was only 3% for the largest deformity tested (Fig 4). However, tibial torsion reduced the hip and knee extension capacity of nearly all the muscles responsible for supporting the leg during single limb stance (Fig 5). With a tibial torsion deformity of 30°, the capacities of soleus, gluteus medius, and gluteus maximus to extend both the hip and knee were reduced by over 10%. With a 60° tibial torsion deformity, the extension capacities were reduced by up to 50%.

Conclusions

The modeling results suggest that excess tibial torsion may contribute to crouch gait, especially when the deformity is 30° or larger. This conclusion is supported by clinical findings for a large group of patients with cerebral palsy (Fig 6). The deformity had the greatest impact on the plantarflexor and gluteal muscles, suggesting that correcting tibial torsion may be important if these muscles are weak. The analysis technique developed in this study also provides a powerful framework for examining how bone geometry affects muscle function during a broad range of movements.

References

[1] Delp et al, IEEE Trans Biomed Eng 1990, 2:46-55 [2] Anderson et al, Gait Posture, 2003, 17:159-69.

poster 6

Computation of cartilage material properties with an interpolant response surface

Kathryn Keenan
Lampros Kourtis
Thor Besier
Derek Lindsey
Garry Gold
Scott Delp
Gary Beaupre

Purpose

Indentation testing is used to evaluate cartilage material properties (1). One established method for indentation testing uses a flat, porous indenter (2). Using a biphasic material model for cartilage, the determination of three material constants (aggregate modulus, Poisson's ratio, permeability) requires a biphasic finite element model and optimization to determine the solution that best matches the experimentally determined creep curve (3). The purpose of this study was to develop a web resource to calculate best-fit biphasic constants based on input from a standardized creep indentation test of human cartilage. This was achieved with a multi-dimensional cartilage interpolant response surface, created using a set of finite element solutions for a range of cartilage properties.

Methods

Cartilage was modeled as a homogenous, isotropic, poroelastic material. The nonlinear, time-dependent finite element model was solved using ABAQUS (ABAQUS Inc). Loading consisted of a 12s linear ramp to 0.35N followed by a creep phase (deformation over time under constant load) to a total time of 5000s.

We first generated an initial coarse set of 1,728 finite element solutions, and then used piecewise cubic Hermite polynomial interpolation (MATLAB, Mathworks Inc.) to create the response surface (958,230 curves). A range of values for Poisson's ratio (nu), Aggregate modulus (Ha), Permeability (k), and Cartilage Thickness were used to create the response surface. The established methodology compares the final 30% of total displacement of experimental and model curves (4).

The user provided the results of a creep test, and a least-squared search was used to find a curve within the response surface that best fits the experimental data.

Results and Conclusion

The response surface consistently returned displacement-time curves that closely matched experimental data; average normalized root mean square error was 0.0107. RMSE was normalized by peak displacement and calculated by:

RMSE = sqrt[(1/n)*sum((model(t)-experimental(t))^2)]

The material properties were comparable to the literature (3) (e.g. Ha = 0.375MPa, nu = 0.0, k = 1.80E-15 m^4/Ns).

The response surface is a valuable resource to researchers who require cartilage material properties, but do not have expertise in optimization and finite element modeling required to determine material properties from experimental tests. The resource is available at https://simtk.org/home/va-squish.

poster 7

Low resolution 3-D models from small-angle X-ray scattering as input for Poisson-Boltzmann calculations

Jan Lipfert

Vincent B. Chu
Yu Bai
Daniel Herschlag
Sebastian Doniach

Purpose

Nucleic acids are central to the storage, transmission, and control of genetic information. As DNA and RNA are highly negatively charged, electrostatics and the associated counterion cloud play a dominant role in their function and stability. Poisson-Boltzmann (PB) theory has been used to (semi-) quantitatively model the thermodynamics of ion binding to RNA using atomic resolution models derived from crystallography or NMR spectroscopy for RNA [1]. Here, we seek to extent the applicability of PB theory to systems and conformational ensembles for which no atomic resolution data is available. We demonstrate that reconstruction algorithms can be used to obtain low resolution 3-D electron density maps of nucleic acids from 1-D small-angle X-ray scattering (SAXS) data and that the resulting models can be used successfully for electrostatics calculations to determine the extent of ion binding.

Materials and Methods

We use the reconstruction algorithm DAMMIN[2] to reconstruct 3-D low resolution models of RNA from 1-D small-angle X-ray scattering data[3]. For electrostatic calculations we use routines based on the adaptive Poisson-Boltzmann solver (APBS)[4,5].

Results

We show for three nucleic acids of known atomic resolution structure that low resolution 3-D electron density maps can be reliably obtained from 1-D SAXS data. Furthermore, it is demonstrated that for these test systems the extent of ion binding computed from PB calculations using the low resolution models agrees favorably both with calculations using atomic resolution models and with experimental results.
Furthermore, we use SAXS to characterize the conformational landscape of a glycine riboswitch from Vibrio cholerae [6] as a function of Mg2+ and glycine concentration. PB calculations using 3-D low resolution models are used to rationalize the salt dependence of the different conformational transitions of this RNA of unknown atomic resolution structure.

Conclusion

Our results suggest that low resolution 3-D models obtain from SAXS data can be used successfully as input for PB calculations of nucleic acids. This extends the applicability of PB theory to systems or conformations for which no atomic resolution structures are available.

References

[1] Misra & Draper, J. Mol. Biol. 299:813-825 (2000)
[2] Svergun, Biophys. J. 76:2879-2886 (1999)
[3] Lipfert, Chu, Bai, Herschlag & Doniach, J. Appl. Cryst., in press
[4] Baker, Sept, Joseph, Holst & McCammon, PNAS 98:10037-10041 (2001)
[5] Chu, Bai, Lipfert, Herschlag & Doniach, submitted
[6] Mandal, Lee, Barrick, Weinberg, Emilsson, Ruzzo & Breaker, Science 306:275-279 (2004)

poster 8

Novel sampling strategies in structural biocomputing

Peter Minary

Michael Levitt

One of the major challenges of modern computational biology is the effective search over the vast amount of conformations a protein could assume. Although, significant simplification of this problem could be achieved by introducing coarse grained protein models, the major bottlenecks are still present and more efficient sampling algorithms are needed. Here, we demonstrate the potential of some promising algorithms, which could have an impact on this rapidly expanding field. First, it is shown that using non linear variable transformations designed to warp conformational space could lead to substantial gains in sampling efficiency. Second, the application of an energy domain based multiple replica Monte-Carlo sampling method is presented for both conformational and sequence sampling. The performance of these algorithms are benchmarked and compared to widely used conformational sampling methods such as parallel tempering.

poster 9

Dependence of knee kinetics on inertial model of the shank during run-to-stop movements

Andy Nnewihe

Ajit Chaudhari
Stefano Corazza
Lars Muendermann
Thomas Andriacchi

Purpose

Many knee injuries occur during movements with rapid decelerations, so understanding the loading conditions at the knee during these movements is important in preventing these injuries[1]. In high-speed activities, the accelerations and velocities may get large enough to require a more accurate representation of the moments of inertia. This study investigated the importance of using subject-specific inertia tensors and center of mass locations of the anatomical segments to calculate loading at the knee joint.

Materials and Methods

Markers were placed on fifteen subjects according to the Point Cluster Technique (PCT) protocol [2], and an opto-electronic motion capture system with a force plate was used to collect motion data. There were 10 female and 5 male subjects [mean age: 23.9 +/- 4.5 years, mean BMI: 22.3 +/- 2.8]. Each subject performed a run-to-stop maneuver, landing on a single leg on a force plate. A laser scan with an accuracy of 1 mm was taken of each subject in order to acquire the outer surface of the shank. The slender rod model and a full inertial model (Fig. 1) were used to derive the inertia tensor (Table 1) for the shank, and the kinetics were calculated in MATLAB using force plate data and inverse dynamics[3]. For comparison, the kinetics were also calculated assuming the inertial tensor of the shank has all zero entries.

Results and Conclusion

The results of this study have shown that a slender-rod approximation of segment inertia appears to be valid for the calculation of knee flexion/extension and adduction/abduction moments during a run-to-stop within a 1% accuracy. Depending on the tolerance level, a more precise model might be needed to obtain acceptable internal/external rotation moments (Fig. 2) for which a 4% average difference was found. Future research can be performed using the novel full inertial model to calculate full body kinetics.

References

1. McLean S. et al (2003) J Biomech Eng, 125(6):864-74.
2. Andriacchi T.P. et al (1998) J Biomech Eng, 120(6):743-9.
3. Clauser C. et al (1969) AMRL-TR-69-70.

Acknowledgements

We thank Lise Leveille and Vicky Chi for collecting the gait data.

poster 10

Impulse based PD control for joints and muscles

Rachel Weinstein

Eran Guendelman
Ron Fedkiw

Purpose

We propose a novel approach to proportional derivative (PD) control that exploits the fact that such equations can be solved analytically for a single degree of freedom. We then extend this method through an optimization procedure to determine muscle actuations needed to achieve desired joint angles.

Materials and Methods

The analytic solution tells us what the PD controller would accomplish in isolation without interference from neighboring joints, gravity and external forces, outboard limbs, etc. Our approach to time integration uses an inverse dynamics style formulation that automatically incorporates global feedback so that the per joint predictions are achieved. Stiffness is decoupled from control without the need for estimating external forces so that we obtain the desired target regardless of a joint's stiffness, which merely determines when a target angle is hit. Whereas PD is typically applied via torques allowing drift, we follow [1] working with impulse and velocity as opposed to force and acceleration. This also allows for robust incorporation of collisions and contact. In particular, we use the framework of [2] making heavy use of post-stabilization to implement our PD control method. Besides obtaining smooth motion, an advantage of dynamic (versus kinematic) controllers is response to unanticipated forces. For example, a simulated subject should easily move around his own limbs, but struggle to lift a heavy object.

Results and Conclusion

We created a skeleton from the Visible Human data set and animated the skeleton flailing in a net and swinging a mace. The skeleton joint movement is defined by an analytic function targeted via our PD control, while the mace and net are freely-moving unactuated joints. We also demonstrate an optimization-based method for inverse muscle actuation with an example of our skeleton performing a series of push-ups. The push-up motion was created using analytic functions, and our optimization method computes the muscle actuation required to achieve the desired joint angles throughout the body. Unlike torque actuation, muscle actuation is restricted to lie along the muscle's line of action, and inequality constraints are used to enforce bounds on muscle force. Our method trivially handles muscles spanning multiple joints. Even with the additional calculation of the muscular impulses, this simulation only required three minutes a frame. Examples with lower degrees of freedom ran in interactive time.

References

[1] Guendelman, E., Bridson, R., and Fedkiw, R. Nonconvex rigid bodies with stacking. SIGGRAPH, 22(3):871–878, 2003.

[2] Weinstein, R., Teran, J., and Fedkiw, R. Dynamic simulation of articulated rigid bodies with contact and collision. IEEE TVCG, 12(3):365–374, 2006.

poster
11

Hybrid modeling and digital abstraction of the Caulobacter regulatory network

Xiling Shen

Harley McAdam
Mark Horowitz

Purpose

The top-level master regulatory proteins in the bacterium Caulobacter crescentus form a central control circuitry that coordinates genetic modular processes that implement the cell cycle (e.g, replicate the chromosome). The complexity of the modular processes hinders the traditional bottom-up way of modeling. A top-down holistic simulation model with various abstraction techniques is constructed to provide insights into the properties of the regulatory network.

Method

We borrowed the concept of state machine from engineering to model the modular processes abstractly. Discreet states are assigned to processes, which are combined with the continuous protein levels from the regulatory network to construct a hybrid simulation model. Implemented in Simulink and Stateflow, the simulation encapsulates all aspects of the cell cycle by showing how the cell controls the cell cycle and responses to external signals in both normal and mutant cells.

Results

The simulation results are used to verify existing hypothesis as well as help guide biologists' future experiments. The simulation also confirmed the observations that regulatory networks were in general very robust. Its functionality is insensitive to the precise values of the parameters in the model, even though the exact timing of the protein levels and the states of the modules is not. By further abstracting the ODE models into digital gates with thresholds and delays, we discovered that the regulatory network forms an asynchronous fundamental mode machine, whose operation is such that the regulatory network has to converge into a unique steady state to trigger the waiting modular processes to transition forward. Adopted by electrical engineers, this kind of asynchronous circuits is robust towards timing variations by avoiding racing conditions. Therefore, this design scheme has been chosen by evolution to regulate Caulobacter's cell cycle robustly.

The digital gate abstraction of the ODEs has also helped identified various motifs such as auto-regulation and feed-forward loops, whose characteristics have been shown to help the robustness of the overall system.

Conclusion

Recent discoveries on regulatory networks are key to understand how living organisms work from the system level. A holistic modeling approach has been developed using hybrid control concepts to simulate the Caulobacter cell cycle regulatory network with complex modular processes. The design of the regulatory circuitry is identified as an asynchronous fundamental mode machine, a type of circuit often designed for robust asynchronous operations. The digital gate abstraction helps explain the roles of the various motifs found in the circuitry.

poster 12

Intracochlear pressure and derived quantities from a three-dimensional linear model

Yong-Jin Yoon

Sunil Puria
Charles R. Steele

The measurements of gerbil intracochlear pressure (Olson, 1998, 2001) offer an unusual opportunity for validation of model calculations. Presently, we extend the macro-mechanical cochlear model for the chinchilla anatomy (Yongjin et al, 2006 ORL) to the the gerbil anatomy. The BM properties are physical, with orthotropic elastic properties and no fictitious mass or damping. Hence there are no free unjustified parameters for adjustment to fit experimental results.

Intracochlear pressure in the ST is obtained by adding the fast wave to the traveling pressure wave. From the intracochlear pressure simulation, derived quantities including (i) BM velocity, (ii) pressure difference across OC, and (iii) the OC impedance in the base are calculated by following Olson's procedure (1998; 2001). These quantities are compared with animal measurements and show excellent agreement. By comparing exact and estimated OC impedances, we find that the fast wave component in the estimated OC impedance causes phase fluctuation out of the reasonable range (negative real part even for the passive cochlea).

The comparison of animal measurements and model results of derived quantities (i – iii) is promising, but not fully satisfactory. The CF-to-place map in the passive model and frequency responses of BM velocity magnitude and intracochlear pressure are in close agreement with those observed in animal measurement. The feed-forward linear active model shows excellent agreement with experimental data in the BM relative velocity magnitude. However, the calculated phase shows a larger roll-off at the CF by two and half cycles for the frequency dependence at a fixed point for both the active and passive cases. Several parameter variations were explored to determine the cause of this discrepancy. This includes scala area and duct area tapering rate variations.

poster 13

Lung nodule CAD false positive reduction using a novel shape analysis approach

Arpit Aggarwal

G. D. Rubin

D. S. Paik

Purpose

Existing Computer Aided Diagnosis (CAD) algorithms may produce many false positives in order to achieve a high sensitivity for detecting small lung nodules (~ 3 mm). The purpose of this work was to develop an algorithm to automatically filter out as many false positive CAD detections as possible without adversely affecting the sensitivity, thus creating a composite algorithm with a much improved performance for nodules of all sizes.

Method

We have built a new method for shape analysis to distinguish between lung nodules and typical CAD false positives using a non-parametric approach. Localized shape estimation was done by combining ray casting with geometric measures of isointensity contours. A statistical analysis of these measures was then used to construct features for a logistic regression classifier. For evaluation, the method was given our previously published SNO CAD algorithm's detections as input. 20 chest CT scans were performed and there were a total of 290 nodules with 95 being less than 3 mm in size. 5-fold cross-validation along with agglomerative clustering was performed on the dataset to assess the FP reduction algorithm when combined with SNO CAD.

Results

The algorithm achieved an average area under the ROC curve value of 0.985 in classifying the SNO CAD output. In particular, the combined FP reduction algorithm with SNO CAD was able to attain a false positive rate of only 5 per patient with a sensitivity of 82.06%, as compared to SNO CAD alone, which gave a false positive rate of 540 per patient for a similar sensitivity.

Conclusion

Our novel algorithm was able to considerably improve the performance of an existing CAD algorithm by successfully filtering out most of its false positives and correctly classifying almost all of its detections. High false positive rate in CAD algorithms limits their use by radiologists but this algorithm significantly reduces the findings they have to review, without leaving out smaller nodules. Being modular in nature it can also be added on top of any primary CAD algorithm.

References

1. Paik DS, Beaulieu CF, Rubin GD, et al, Surface normal overlap: a computer-aided detection algorithm with applications to colonic polyps and lung nodules in helical CT. IEEE Trans Med Imaging 2004; 23:661-675.

2. Lorensen WE, Cline HE, Marching cubes: A high resolution 3D surface construction algorithm. International conference on Computer Graphics and Interactive Techniques 1987; 163-169.

poster 14

Multiscale deformable registration of medical images

Dana Paquin

Doron Levy

Lei Xing

Introduction

Image registration is the process of determining the optimal spatial transformation that maps one image to another. Image registration is necessary, for example, when images of the same object are taken at different times, from different imaging devices, or from different perspectives. The two images to be registered, called the fixed and moving images, are the input to the registration algorithm, and the output is the optimal transformation that maps the moving image to the fixed image. Applications of image registration include image-guided radiation therapy, image-guided surgery, and tumor detection, as well as many non-medical applications, such as computer vision, and pattern recognition.

Purpose

In the context of medical imaging, the goal of the registration process is to remove artificial differences in the images introduced by patient movement, differences in imaging devices, etc., but at the same time, to retain real differences due to actual variations of the objects. Medical images, however, often contain significant levels of noise due to instrumentation imperfections, data acquisition techniques, image reconstruction methods, transmission and/or compression errors, and other factors. Although numerous successful image registration techniques have been published, we have shown that ordinary image registration algorithms can fail to produce meaningful results when one or both of the images to be registered contains significant levels of noise. Thus, we have developed a technique that enables successful image registration when one or both of the images to be registered is noisy.

Methods

Using the hierarchical multiscale image decomposition of Tadmor et al., we decompose the noisy images to be registered into a series of coarse and fine scales in which each scale resolves increasingly more detail than the previous scales. Using free-form deformation B-spline registration techniques, we then iteratively register the successive scales of the noisy images with one another.

Results

With numerous image registration experiments in both two and three dimensions, we demonstrate that our multiscale deformable registration algorithm accurately registers medical images that contain noise levels significantly greater than the levels at which ordinary deformable registration techniques fail.

Conclusion

Our multiscale deformable registration technique is particularly well-suited for the problem of image registration in the presence of noise.

poster 15

Volume features for semi-rigid segmentation and registration of low-resolution electron density maps

Emilio Antunez

Leonidas Guibas

A protein's function is often determined by the molecules to which it can and cannot bind. This specificity is largely determined by the shape and motion of rigid protein domains. Techniques for visualizing these rigid subunits therefore provide key insights into the behavior of the molecule.

Cryo electron microscopy (EM) has become an important tool in such analyses, because it can capture a protein in several natural conformations. The resulting electron density maps are generally too low-resolution for atomic structure determination. Still, a human observer can identify rigid domains and their relative motion from this data. As the number of available cryo EM datasets is increasing rapidly, the need is growing for routines to perform this structural analysis automatically.

Existing techniques for automation rely on matching molecular surface features. While this approach is effective when the particle is limited to large hinge motions, these features are often distorted, obscured or insufficiently descriptive for some complex motions, particularly if most of the activity is in the interior of the molecule.

We present a technique that utilizes distinctive features contained in the interior of the molecular volume. Preliminary tests indicate that high-density filaments (corresponding to large amino acids and segments of the protein backbone) are better preserved and more easily observable over a wide range of motions. These filaments therefore provide a better basis for motion and structural analysis than surface features alone.

poster 16

Error prediction and performance evaluation of eigenshape based algorithm for reconstructing missing segments vascular centerlines in CT angiography (CTA)

Tejas Rakshe

Dominik Fleischmann
Jarrett Rosenberg
Justus Roos
Sandy Napel

Purpose

To build a statistical model for error and evaluate performance of a PCA based method for reconstructing missing vascular centerlines, using a database of centerlines from other patients.

Materials and Methods

A database was constructed consisting centerlines of femoropopliteal artery from CTA scans of 30 subjects without peripheral arterial occlusive disease. Leave-one-out cross validation of a eigenshape based algorithm was performed on the database by simulating occlusions of various lengths and reconstructing them using the algorithm. Point-wise maximum departure (MD) for each case was used as the error metric. Regression analysis was performed on MD with the length and location of the occlusion, age and sex of the subject, and estimation error in the neighborhood of the occlusion (NE). The results were compared with the results obtained by a minimum mean squared error (MMSE) estimate.

Results

The 50th percentile of MD and the 95% confidence interval upper bound for each occlusion length were: (Occ. length : 50th percentile MD: 95% upper bound, all in mm) 10:0.14:0.15, 25:0.40:0.42, 50:0.81:0.88, 75:1.32:1.46, 100:1.76:2.06. The most important effect on error was that of occlusion length ($p<0.0001$), with a (log-log) coefficient of 1.06 (95% Confidence Interval (CI): 1.04-1.08) (each 10^1 increase in length results in a 10^1.06} increase in MD. A simple linear regression of just log-MD on log-Occl. Length accounts for 50% of the variability in the latter. There was a smaller effect of log-NE ($p<.001$), with a coefficient of 0.08 (95% CI: 0.04-0.12). There was no significant effect of age or sex of the patient. The random-intercept effect of individual arteries accounted for 28% of the remaining variability. The eigenshape based methods produced more accurate results than MMSE method, with a Wilcoxon signed rank test of the distribution of maximum error between the two methods having a p-value <0.00001.

Conclusions

The eigehshape based algorithm reconstructs missing centerlines with acceptable accuracy up to 100mm OL. Longer occlusions may benifit by simple user input constraining the path through a small number of points. The expected error can be predicted to a significant level of accuracy. The method may provide efficient generation of curved planar reformations through arteries of patients with occluded segments, greatly simplifying their assessment from cross-sectional imaging.

poster 17

Transparent rendering of intraluminal contrast for 3D polyp visualization at CT colonography

Rong Shi

Sandy Napel
Jarrett Rosenberg
Ajay Jayant Joshi
Pankhudi Pankhudi
Christopher F. Beaulieu

Purpose
Patient acceptance of CT imaging for the detection of colonic polyps, precursors to colon cancer, is limited by the need for cathartic colon cleansing. Recently, oral tagging material has been introduced that would allow water and fecal material to be subtracted from the images prior to evaluation. However, this doubles the size of the dataset and may introduce artifacts. To address these issues, we developed a classifier that permits transparent rendering of both tagging material and air without subtraction.

Materials and Methods
The algorithm couples machine learning and a painting metaphor to allow interactive classification and rendering. The user initially paints a subset of image voxels (tagging material, air, and soft tissue). Then an iterative training algorithm classifies the volume based on intensities and gradients, and 3D subvolumes of interest in the volume are displayed. The algorithm was tested on 26 tagged CTC cases (courtesy of Dr. Richard Choi, Walter Reed Army Medical Center), containing 49 polyps: 25 adjacent to air, and 24 either completely under or coated with tagging material. Non-polyp control locations (49 total) included 25 areas against air and 24 against tagging material. Three radiologists, without knowledge of presence of tagging material, independently viewed the 98 subvolumes for (a) presence or absence of a polyp, and (b) overall image quality.

Results
For polyp detection, readers achieved 96.6% sensitivity and 89.1% specificity (area under receiver operator characteristic curve {AUC} 0.98). Image quality ratings between polyp and control locations were not significantly different overall ($p < 0.258$). Image quality ratings were significantly lower overall for tagged versus untagged locations ($p < 0.001$) and for two readers (both $p < 0.001$), but not for the third reader ($p < 0.272$). However, readers' accuracy was equivalent between tagged (AUC = 0.99) and untagged (AUC = 0.97) locations overall ($p < 0.05$) and for each individual reader ($p < 0.05$).

Conclusion
Our method for rendering tagging material transparent resulted in a high level of polyp classification accuracy, without decreased accuracy compared to polyps in air. This new approach enables accurate 3D depiction of polyps under iodinated contrast. Further image quality improvement may be possible by exploiting more features used in classification therefore optimizing the algorithm.

poster 18

Lung quantization of mice to analyze the behavior of oncogene inactivation on RAS/ Myc tumors

Shradha Budhiraja

P. K. Bendapudi
D. S. Paik

Purpose

Lung carcinomas are one the most deadly forms of human cancers, with five-year survival rates in the range of 10-15%. It has previously been shown that the inactivation of the MYC proto-oncogene can lead to tumor regression and differentiation in murine models of T cell lymphoma, hepatocellular carcinoma, and osteosarcoma. The purpose of this work is to develop a method to follow the growth of Ras/Myc tumors and the consequences of oncogene inactivation on mice lungs. Calculating the Ras/Myc tumor size, shape and volume with existing methods was tedious and inaccurate. Six to ten hand drawn outlines were used to manually interpolate tumor volumes on commercially available software.

Materials and Methods

We developed a new and convenient method to measure the consequences of oncogene inactivation on lung tumors in mice. Since the CT images were not accurate, they were first calibrated to obtain the correct air and lung intensities. The lung was first segmented and then split into left and right lung respectively by a split plane given by the user. Each lung was then quantized, by measuring the total volume as well as the equivalent air volume by taking into account the partial volume effect. These volumes were then compared in different mice scans to study the effect of oncogene inactivation. For evaluation, the method was used on seven different mice scan sets. A murine lung tumor model was developed where MYC overexpression was conditionally controlled by a lung-specific promoter. Each mouse's health was monitored over four months and CT images were taken after every two weeks.

Results

The method was able to demonstrate the regression of Myc/Ras tumors upon inactivation. The expected behavior was observed in almost all the cases. The method overcame the earlier problems of chest wall invasion, inaccurate CT slices and incapability of easily defining diffusely infiltrating tumors

Conclusions

Our method was a significant improvement over the earlier method. The measurement of total and equivalent air volumes in the lungs accurately demonstrated the expected regression of both Ras and Myc tumors.

Web Page: http://shradha.budhiraja.googlepages.com/

poster 19

Discovering and validating disease subtypes using microarray datasets

Amy V. Kapp
Stefanie S. Jeffrey
Robert Tibshirani

Purpose
Traditionally the subtypes of a disease have been defined by differences in the symptoms experienced by patients. With the rising popularity and declining cost of microarrays, however, defining subtypes based upon molecular profiles is increasingly practical. In addition, defining the subtypes of a disease using molecular profiles has the potential to lead to a more sophisticated understanding of the disease.

Disease subtypes defined by molecular profiles typically have been identified using a single clustering of the patients and are validated biologically only, not statistically. We present a procedure for discovering and statistically validating disease subtypes using microarray datasets without requiring a priori knowledge of the disease's biology. Moreover, we apply this procedure to breast cancer tumor tissue microarrays and describe our results.

Methods
Five independent breast cancer tumor tissue cDNA microarray datasets (a total of 599 arrays) were used to characterize and validate the subtypes. The public domain statistical program R version 2.2.0 was used for every analysis. Statistical validation was done according to the procedure described in Kapp and Tibshirani (2006) (an implementation of the algorithm is in the R package clusterRepro).

Results
Using our novel approach, we found evidence in support of the most consistently identifiable subtypes of breast cancer tumor tissue microarrays being: ESR1+/ERBB2-, ESR1-/ERBB2-, and ERBB2+ (collectively called the ESR1/ERBB2 subtypes). We validated all three subtypes and found the subtype to which a sample belongs is a significant predictor of overall survival and distant-metastasis free probability.

Conclusions
We have developed an effective method for using microarray datasets to discover and validate disease subtypes. Our method can be applied to any disease for which at least two independent microarray datasets are available for analysis. Explicit details of our method [1] as well as a link to the clusterRepro package [2] are available on-line.

References
[1] Kapp AV, Jeffrey SS, Langerød A, Børresen-Dale A-L, Han W, Noh D-Y, Bukholm IRK, Nicolau M, Brown PO, and Tibshirani R. Discovery and validation of breast cancer subtypes. BMC Genomics (2006) 7:231.
[2] Kapp AV and Tibshirani R. Are clusters found in one dataset present in another dataset? Biostatistics (2006) in press; epub.

poster 20

Higher performance algorithms for learning signaling pathways using Bayesian networks

Byron Ellis

Purpose

Network models, particularly Bayesian networks has become a popular method for modeling causal cellular systems such as gene regulatory networks [1] and cell signaling networks [2]. Extending this model has two primary fronts, improvement to the statistical models used in the sampling and/or optimization of the cellular model and the improvement of algorithms required to deal with larger models in reasonable time, particularly in the face of missing data which generally requires the introduction of iterative techniques beyond those used in the model sampling process.

Methods

These sorts of models seem particularly well-suited to single-cell measurements such as multiparameter flow cytometry, which has been the focus of our efforts. However, as a technology, it is limited to a finite number of parameters which can be simultaneously probed, while the system of interest may be much larger. To address this we have begun to develop improved Bayesian Network models using a systems approach to modeling signaling pathway kinetics [3].

Results

The key component of this research has been the development of the Order-Graph Sampler, an extension of the Friedman-Koller Order Sampler [4]. This sampler reduces computation time to minutes from hours while providing improved estimates of Bayesian Network structure under unbiased testing results. Combined with our systems-based models, allowing for the use of flow cytometry data without discretization, we also perform well under cross-validation types of analysis when compared with other algorithms.

References

[1] Friedman, N., Inferring Cellular Networks Using Probabilistic Graphical Models, Science, 2004.
[2] Sachs, K., Perez, O., Pe'er D., Lauffenburger, D., and Nolan, G., Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data, Science, 2005.
[3] Zhang, X., De Cock, K., Bugallo, M., and Djuric P., Computational of probabilities for first-order chemical reactions using system theory, J. Chemical Physics, 2004.
[4] Friedman, N., Koller, D., Being Bayesian About Network Structure: A Bayesian Approach to Structure Discovery in Bayesian Networks, Machine Learning, 2003.

Use of a graphical model to integrate prior biological knowledge with microarray data

poster 21

Bernie Daigle

David Heckerman

Russ B. Altman

Purpose

DNA microarrays have become a popular tool for analyzing gene expression and its regulation in a single condition relative to a control. Though the high dimensionality of the data generated is an advantage of the technology, it also poses a formidable challenge: can the biologically relevant genes with the largest effect on the experimental phenotype be isolated from the vast majority of genes that are unchanged or have small secondary expression effects? Traditional analyses of differentially expressed (DE) genes often treat all genes independently, and prior biological knowledge, if used at all, is incorporated retrospectively. As there is no guarantee that the most biologically relevant genes are those with the most extreme expression values, these methods often discard signal with noise. On top of this, microarray experiments are often highly variable among both technical and biological replicates. It has become clear that informatics methods are needed to circumvent these aforementioned difficulties, and we propose one such method here.

Methods and materials

We design and implement an undirected graphical model to integrate prior biological knowledge with gene expression data. By incorporating knowledge from sources including gene ontology (GO) terms, biological pathways, and conserved DNA sequence motifs, we strive to identify biologically relevant DE genes more accurately than would be possible using any data source alone. A simulation study is used to determine the graph structural features of prior knowledge that give the best integration performance. We then validate our method by predicting target genes of S. cerevisiae Hsf1p, using an hsf1 mutant microarray dataset [1], conserved yeast DNA sequence motifs [2], and a gold standard based on Hsf1p chromatin immunoprecipitation assays [3].

Results

Simulations indicate that graph structures with a low clustering coefficient and high relative connectivity of DE genes to DE<->non-DE genes give the best performance. Validation results show that for a large range of parameter values, our method is able to more accurately identify Hsf1p target genes in terms of area under the ROC curve.

Conclusions

We have successfully implemented a method to incorporate diverse forms of biological knowledge with gene expression data, and we've shown in one case that the method can make better predictions by integration than by using data sources alone. Future work will focus on automatic selection of model parameters and application of the method to novel datasets.

References

[1] Yamamoto et al., J Biol Chem. 2005 Mar 25;280(12):11911-9.
[2] Kellis et al., Nature. 2003 May 15;423(6937):241-54.
[3] Hahn et al., Mol Cell Biol. 2004 Jun;24(12):5249-56.

poster 22

CBS+: A protege-based tool for literature curation at PharmGKB

Daniel Holbert

Amy Hodge
Teri Klein
Russ Altman

Purpose

CBS+ is a new tool for literature curation at PharmGKB. Its goal is to allow a team of curators to work efficiently in tandem, annotating pharmacogenomic literature in a structured, efficient way.

Materials and methods

CBS+ is built on top of a multi-user Protege knowledge base, using a literature-specific ontology and several novel plugins. The ontology stores basic information, such as an article's title and date, along with more complex information, including the relationships between genes, drugs, and diseases described in the article. The new plugins have several functions. One set of plugins enables efficient storage and browsing of very large controlled vocabularies within Protege. Another plugin imports data from an external database into a Protege project. Finally, a third set of plugins allows curated data to be exported from Protege in a variety of formats.

Results

CBS+ is currently under development. It can already represent all of the knowledge that was curated by scientists at PharmGKB using the previous curation methods. Moreover, the controlled vocabulary plugins have been shown to bring a huge speedup when compared to vocabularies stored in standard Protege classes.

Conclusion

CBS+ is a useful tool for curators of pharmacogenomic literature. Moreover, it includes several novel plugins which are useful in a broader sense as extensions to Protege.

References

Hewett, M., et al., "PharmGKB: the Pharmacogenetics Knowledge Base." Nucleic Acids Res, 2002. p. 163-5.
Noy, NF., et. al., "Protege-2000: an open-source ontology-development and knowledge acquisition environment." AMIA Annu. Symp. Proc., 2003. p. 953.

poster 23

Independent component analysis for automatic separation of biological calcium dynamics from in vivo cerebellar imaging data

Eran A. Mukamel
Axel Nimmerjahn
Mark J. Schnitzer

Purpose

Advances in two-photon imaging of calcium sensitive fluorescent probes have enabled simultaneous measurement of neuronal and glial cell activity in functioning brain circuits [1]. A major challenge in analyzing the resulting data is to separate independent biological signals arising from individual neurons and glia in an automated manner. Independent component analysis (ICA) is well suited to such blind source separation [2], and automatically rejects motion artifacts.

Materials and Methods

Neurons and glial cells in intact cerebellar cortex of anesthetized mice were labeled by local injection of fluorescent calcium indicator dye [1]. Using two-photon microscopy, spontaneous fluorescence-encoded calcium changes in individual cells in the molecular and Purkinje cell layers were recorded at a high temporal resolution of up to 53 frames per second. We pre-processed movie data using singular value decomposition for dimensional reduction. Using an ICA algorithm, we determined linear combinations of pixels (ICA filters) maximizing a measure of statistical independence [2]. Temporal deconvolution was applied to improve spike detection in neuronal signals [3].

Results

ICA successfully identified up to 20 independent signals per recording. Neuronal and glial cell traces were distinguished by their characteristic temporal dynamics [4] and by the shape and orientation of ICA filters [1]. Neuronal filters were parallel bars, consistent in shape and orientation with Purkinje cell dendritic arbors. The corresponding temporal traces showed fast (decay time, ~150 ms) repetitive (frequency, 0.3-1 Hz) calcium transients consistent with climbing fiber evoked complex spikes [1]. Cross-coherence analysis of neuronal spike trains revealed statistically significant synchrony among cells separated by ~100 um. A second class of ICA signals resulted from perpendicularly oriented filters, consistent with Bergmann glia fiber morphology and orientation. These signals showed slow, long lasting calcium elevations (several seconds) characteristic of glial calcium dynamics.

Conclusions

Our results show the utility of ICA for extracting functional cell signals from optical imaging data sets, complementing previous neuroscience applications of ICA to MEG and EEG data. The algorithm is computationally efficient for the analysis of large data sets. Taking advantage of spatial information while maximizing signal independence may further extend the utility of this algorithm for optical imaging applications.

References

1. Sullivan, M.R. et al. (2004), J. Neurophysiol. 94:1636. 2. Hyvärinen, A. & Oja, E. (2000), Neural Netw. 13: 4113. Yaksi, E. & Friedrich, R.W. (2006), Nat. Methods 3: 377 4. Nimmerjahn, A. et al. (2004), Nat. Methods 1: 31

poster 24

NetworkFinder, a weighted co-occurrence method to extract gene interaction and to construct biological network from Medline abstracts

Jia-Ren Lin

Tobias Meyer
Karlene A. Cimprich

Purpose

For biologists, literature mining currently means a keyword search in PubMed. However, it's usually not an easy task to extract useful information from tons of references. Therefore, methods for automatically extracting biomedical facts from the scientific literature have increasing demands in recent years. Moreover, owing to the increasing content in Medline database, literature mining is also becoming useful for both hypothesis generation and biological discovery. Therefore, we designed a text-mining program, NetworkFinder, which identifies biological relationships based on the textual co-occurrence of gene/protein terms in abstracts. Moreover, the NetworkFinder has been demonstrated to be capable to recapitulate the biological network and discover the signaling modules automatically. This system can also be used for discovering novel connection between genes based on the previous studies.

Materials and Methods

NetworkFinder Combines the power of Perl in text-processing and the web-query protocol of Medline service, it is able to automatically search PubMed literature about a given term and find out related genes/proteins. In addition, based on a unique weight system by considering the importance of each interaction (using citation impact factors), the all interactions can be summed & sorted. The strong interacting partners can be used to extent the search, as well as to finding out the implicit interactions betweens genes by using Swanson's ABC model in literature-based discovery with its unique ranking/weighting system.

Results and Conclusion

NetworkFinder successfully achieved the three goals: the literature reviewing, the network reconstruction, and the hypotheses generation. The unique feature of NetworkFinder in using citation-impact-factor weighted co-occurrence frequency improved the accuracy of retrieve gene interactions from PubMed abstracts. Finally, the literature-based-discovery (LBD) function of NetworkFinder was demonstrated successfully in this study. Although only partial (about 10%) predicted target genes were recalled, the results were much better than these form random pickup genes. The 5.9-41.5 folds increment of NetworkFinder LBD from random process proved that NetworkFinder can be used in hypotheses generation.

CONTRAfold: RNA secondary structure prediction for single and multiple sequences without physics-based models

poster 25

Gregory M. Goldgof
Chuong Do
Serafim Batzoglou

Purpose

RNA secondary structure prediction has become a central problem in computational genomics because of its applications to genome-wide ncRNA detection, automated functional classification, and basic RNA research. Since their introduction over twenty years ago, physics-based approaches such as Mfold have achieved the highest accuracy predictions. These programs identify a structure of minimum free energy using thermodynamic parameters that have been experimentally determined. The weakness of this approach is that these parameters are difficult to measure, overlap with each other, and do not represent all thermodynamic factors. Additionally, there may be non-thermodynamic factors that influence RNA folding that these methods ignore entirely. CONTRAfold is the first probabilistic RNA secondary structure prediction method to significantly outperform existing physics-based approaches. Whereas CONTRAfold uses similar parameters to physics-based approaches, the values of these parameters are derived by a fully-automated statistical learning algorithm.

Materials and Methods

Conditional log-linear models (CLLMs) are used to learn CONTRAfold's parameters, which are similar to those employed by Mfold. Example parameters include base-pairings, hairpin lengths, helix lengths, and internal loop asymmetry. A dynamic programming algorithm is used to determine the structure with the maximum expected accuracy.

Results

Using cross-validation on published structures taken from Rfam, CONTRAfold obtains the highest single sequence prediction accuracies, achieving 6% higher sensitivity and 3% higher specificity than Mfold, the best currently available technique. CONTRAfold also has the highest multiple sequence prediction accuracy, achieving 4% higher sensitivity and 7% higher specificity than Alifold, the leading multiple sequence predictor.

Conclusion

CONTRAfold demonstrates that statistical learning techniques are an effective alternative to physics-based models for both single and multiple sequence RNA secondary structural prediction. Because it provides higher accuracy than existing programs, CONTRAfold is a practical tool for biologists studying RNA molecules. In addition, we expect CONTRAfold to become a central component of future projects to perform genome-wide non-coding RNA detection as well as automated ncRNA classification.

References

CB Do, GM Goldgof, S Batzoglou. Multiple sequence RNA secondary structure prediction without physics-based models. Genome Informatics 2006.

CB Do, DA Woods, S Batzoglou. CONTRAfold: RNA secondary structure prediction without physics-based models. Bioinformatics, 22(14): e90-e98, 2006.

[2] G. F. Schroeder, U. Alexiev, and H. Grubmuller. Simulation of fluorescence anisotropy experiments: Probing protein dynamics. Biophys. J., accepted.

poster 26

A confidence score calculator for evaluating the metabolic pathways predicted by PathoLogic

Jing Shi

Michael G. Walker
Peter D. Karp
Stanley N. Cohen

Purpose

There are approximately 160 computationally derived pathway-genome databases in the BioCyc collection (http://www.biocyc.org) that were created using the PathoLogic software. How can we evaluate the validity of these predicted metabolic pathways? Previous research has used manual curation based on published literature and clinical tests data to validate the pathways. However, many genomes have very limited literature to start with, and the manual validation is labor-intensive.

Methods

we describe a Bayesian algorithm to calculate a confidence score for each predicted metabolic pathway by incorporating genome sequence, functional annotation, and gene expression data. We developed the algorithm using the literature-based pathways for E. coli in the EcoCyc database. The positive training data are the known pathways; the negative training data are randomly generated combinations of genes. For each pathway, we calculate a confidence score. We then evaluate the algorithm by determining its ability to distinguish the positive and negative training data (true pathways vs. randomly generated pathways).

Results

The algorithm separates the true pathways from the randomly generated ($p = 5e\text{-}22$, t-test). The area under the ROC curve for the algorithm is 0.95 (where the area under the ROC for a perfect classifier is 1.0 and 0.5 for a classifier that fails to separate the two groups. We then applied this algorithm to StreptoCyc, a pathway-genome database for Streptomyces coelicolor. We evaluated the algorithm's ability to distinguish predicted pathways in StreptoCyc from randomly generated combinations of genes. The area under the ROC curve is 0.94 and the p-value is 8e-16 (t-test) using the confidence scores to separate the predicted pathways from the randomly generated pathways in StreptoCyc.

Conclusion

We have developed a Bayesian algorithm that calculates confidence scores for the validity of predicted metabolic pathways by PathoLogic. The ROC area indicates that the algorithm performs well.

poster 27

Identifying changes in transcriptional control during mouse aging

Lucinda Southworth

Stuart Kim

Purpose

Aging is a complex biological process that largely escapes our detailed understanding. Although many physiological mechanisms have been implicated in the aging process, little is know about changes in transcriptional regulation. We believe gene modules that have changes in coexpression with age can ultimately elucidate differences in transcriptional control that contribute to or result from the aging process.

Materials and Methods

Microarray data was compiled for mice of ages 16 and 24 months across 16 different tissues. Five biological repeats for each of the sexes yield 160 arrays for each age group. Spearman correlations were calculated for each pair of the 8932 genes in the array. A Fisher transformation was applied to the correlations to make them approximate a normal distribution. We measured the changes in correlations between age groups in three ways. First, we identified pairs of genes whose correlation changed the most with age by finding the pairs with the largest absolute difference in correlation. Next, we looked at correlations of genes in Gene Ontology (GO) categories to find functionally related groups of genes that change with age. Finally, we clustered genes using complete linkage hierarchical clustering and identified clusters of genes whose correlation changed the most with age.

Results

The mean correlation for all pairs of genes between the young and old mice decreases slightly with age. When looking at the most significant differences in correlation (> 5 standard deviations from the mean), 2.1 times more genes went down with age than went up with age. Similarly, out of 636 GO categories, 15.53% show a significant decline in correlation. Fewer categories, 10.53%, show a significant increase with age ($p<0.05$). Among the significantly decreasing categories are those associated with oxidative stress (calmodulin binding and manganese ion binding) and cell adhesion. Both of these mechanisms have been implicated in aging. Finally, hierarchical clustering of the microarray data has identified clusters of genes whose coexpression declines with age. Further analysis may identify clusters of genes more highly correlated in young than old mice.

Conclusion

We have found age associated decline in coexpression for both individual gene pairs and groups of genes. Groups of genes may loose coordinated coexpression if they share a common, age-regulated transcription factor. A future search for shared promoter motifs or known transcription factor binding sites among genes showing a decline in correlation may lead to the identification of such factors.

References

Zhang JZ et al. Am J Physiol. 1999;276.
Culotta VC et al. Eukaryot Cell. 2005;4(7).
Andersson AM et al. Biochem J. 1993; 290.

poster 28

Chemogenomic profiling reveals gene and drug relationships in yeast

Maureen E. Hillenmeyer

Russ B. Altman
Ronald W. Davis
Corey Nislow
Guri Giaever

Purpose

Chemical genomics, the measure of the genomic response to a chemical compound, can be used to elucidate relationships among genes, among drugs, and between genes and drugs. High-throughput chemical genomic assays are facile in the single-celled yeast, where growth phenotype (fitness) is a simple, robust readout.

Methods

We measured responses of yeast knockouts to several hundred drugs and other stress conditions. We defined "co-fitness" interactions by deletion strains that exhibit similar fitness profiles across conditions. We measured the accuracy of these interactions to reconstruct known interactions, as compared with other high-throughput datasets such as co-expression, protein-protein interaction, and synthetic lethality. We also defined "co-inhibition" interactions by compounds that inhibit similar deletion strains and measured the relationship between these interactions to (a) chemical structure similarity and (b) shared mechanism of action in human.

Results

Co-fitness interactions accurately reflect the known interactions, and they interrogate different biological processes than other high-throughput networks. Co-inhibition is related to chemical structure in the heterozygous (single-copy) knockouts and mechanism of action in the homozygous (both-copy) knockouts.

Conclusion

The chemogenomic fitness profiling data comprise new and complementary information about gene function and drug mechanism of action.

poster 29

Predicting functional binding sites in proteins

Megan So

Joseph Chan
Samantha Chui

Purpose

FEATURE is a method developed by Bagley and Altman in 1995 that builds statistical models characterizing the biochemical and biophysical microenvironment surrounding a functional site on a protein. Currently, FEATURE relies solely on three-dimensional data from protein structures to build its models. We propose to augment FEATURE's performance by incorporating more readily obtainable sequence information from proteins for which structures have not yet been determined.

Materials and Methods

This project consists of two parts: (1) finding FEATURE models in which sequences provide more information than structures alone, and (2) updating the FEATURE program itself to incorporate this information. In the project described here, we have completed the first stage. In determining the differences between sequences and structures, we looked at the residue distributions of the parts of the proteins used by FEATURE. To do this, we took 50 active site motifs from PROSITE, and separated the corresponding proteins into two groups, those with structures, and those without. We performed a structure alignment on the proteins with structural data, and aligned the sequence-only proteins (using CBA and MUSCLE, respectively). Then, we aligned the two resulting alignments, and counted the residue distributions in each column of the alignment which was within the 7.5 Å radius which FEATURE uses to build its model.

Results

While a few of the models showed no difference between the residue distributions of the sequences and structures, many had an appreciable difference. Models built with fewer structures were more likely to have different distributions, which makes sense statistically.

Conclusions

Since there were a good number of models for which adding sequence information would improve the predictive power of FEATURE, we can conclude that this will be a beneficial addition to FEATURE. This will especially improve the quality of models built with a small number of structures.

References

1.Bagley S.C., Altman R.B. (1995). Characterizing the microenvironment surrounding protein sites. Protein Science. Vol. 4: 622-635.
2.Edgar R. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. Vol. 32, No. 5.
3.Sigrist C.J.A., et al. (2002). PROSITE: a documented database using patterns and profiles as motif descriptors. Brief Bioinform. Vol. 3: 265-274.
4.Valdar W. (2002). Scoring residue conservation. Proteins. Vol. 48, No. 2: 227-241.

poster 30

Latent covariate detection and verification

Patrick Perry

Art Owen

Purpose

We are given microarrays measuring the expression levels of about 9000 different genes in 40 different mice. We know both the ages and the genders of the mice, and we would like to determine which genes are significantly related to age. By postulating the existence of an unobserved covariate and then fitting this latent term we can sometimes greatly reduce our estimate of noise variance and thus obtain tighter confidence regions for the model parameters. Adding an extra term to the model will always produce a better fit; the statistical challenge is to assign the correct number of degrees of freedom and test for the significance of the latent term.

Materials and Methods

Our microarray data comes from Kevin Becker of the National Institutes of Aging. After performing standard linear regressions on the observed covariates, we can easily fit a latent factor term by taking the best rank-1 approximation of the residual matrix. This is similar to principal component analysis and can be gotten using the singular value decomposition.

Standard tests for the significance of a latent factor are either ad-hoc, or else they assume that the model residuals are uncorrellated. In our setting, assuming uncorrelating residuals is like assuming that all genes behave independently of each other, which is clearly incorrect. We instead make a weaker assumption, that the gene-gene population covariance matrix is sparse. With this assumption, if we take a small subset of genes, they are likely to be uncorrelated. We look at many such small subsets and then apply tools from random matrix theory to help us assign degrees of freedom to the latent term.

In some cases despite explaining a significant portion of the error, the structure of the estimated latent term may not be close to the population latent term. We propose a measure of interestingness based on random rotations of the estimated latent term and a measure of nongaussianity originally developed for exploratory projection pursuit.

Results

We ran our analyses on sets of data from 16 different types of tissue. The majority of the data sets were shown to have significant latent structure, and in many cases we were actually able to estimate potentially biologically-relevant latent structure.

poster 31

Multi-relational data mining of time-oriented biomedical databases

Rashmi Raj

Martin J. O'Connor
Amar K. Das

Background
Multi Relational Data Mining (MRDM) extends association rule mining to search for interesting patterns among data in multiple input tables (relations) rather than in one input table. Researchers have successfully applied MRDM in bioinformatics , but MRDM is limited in handling time-course and longitudinal data, which are commonly found in biomedical databases. MRDM cannot search for patterns involving the comparison or classification of temporal data, which are needed to study causal or dynamic phenomena.

Materials and Methods
To address these issues, we have developed a new MRDM method called ChronoMiner. The underlying algorithm uses as input a hierarchical view of relations, instead of the set view used in standard MRDM. In the hierarchical view, attributes of relations are related through "parent" and "child" relationship. ChronoMiner searches for interesting multi-relational patterns by partial or complete traversal of the virtual tree structure of the database relations. The hierarchical search facilitates the coupling and decoupling of new attributes for temporal pattern discovery. The mining of interesting patterns starts from the root and proceeds using top-down induction, allowing for comparison along the time dimension at every level of abstraction. We evaluated the algorithm by applying it to Stanford HIV Database (hivdb.stanford.edu) to mine associations between newly arising mutations in the HIV genome and past drug regimens containing protease inhibitors (PI).

Results
The database contained 4271 subjects who had a regimen containing protease inhibitors. In searching for new mutations that arose after the administration of a drug or drug category, ChronoMiner confirmed previously known associations. At the drug category level, for PI, it found 63P, 36I, 41K, 93L, 35D as the most frequent mutation occurrences. Traversing one level deeper, at each drug level, it could verify 41L, 67N, 70R, 210W, 215Y as the most frequent mutations for the drug AZT. We also found mutations, such as 122E for AZT, which our domain expert (Dr. Bob Shafer) viewed as novel and clinically meaningful.

Conclusion
Our research extends MRDM to include temporal comparisons and hierarchies in searching for patterns of interest in a biomedical database. The initial evaluation of the ChronoMiner algorithm provides promising results for the HIV drug resistance research domain. After further testing, we plan to extend this work to the discovery of novel time-oriented patterns in other biomedical genomics databases.

poster 32

Extracting binary signals from microarray time-course data

Debashis Sahoo

David L. Dill
Rob Tibshirani
Sylvia K. Plevritis

Purpose

StepMiner is a tool for identifying temporal expression patterns of genes that experience a binary transition. The primary goal of StepMiner is to assist biologists who are interested in understanding the temporal progression of genetic events and processes following a stimulus.

Materials and Methods

StepMiner extracts three categories of temporal binary signals: the first category contains genes whose expression level undergoes a single transition (increase or decrease) ; the second category contains genes whose expression undergoes two transitions, by either turning "on" then "off" or "off" then "on"; and the third category contains genes whose expression fits neither of the first two categories. StepMiner uses F-statistic and comparative F-statistic in an adaptive regression scheme with appropriate degrees of freedom to find the best category according to a user-specified p-value threshold. The p-value threshold can be adjusted using an acceptable false discovery rate (FDR). StepMiner demo version is available at http://verify.stanford.edu/~sahoo/public/StepMiner . Once the temporal groupings are identified, StepMiner provides GO annotations fully automatically.

Results

Simulation studies show that StepMiner performs best for time courses with ten to thirty time points. StepMiner is very fast: it can can process 15 microarrays of 40,000 genes each in less than 15 seconds. However, FDR calculation in StepMiner for this microarray data takes around 12 minutes. StepMiner was applied to a publicly available time course of microarrays monitoring gene expression levels in yeast during the diauxic shift in a glucose-limited culture. The heat map generated by StepMiner identifies two critical transitions in gene expression, occurring, at 8.25 hours and 9.25 hours. These time points correspond to observed changes in the growth rate of the yeast around 9 hr. GO annotations have lower p-values compared to a published analysis by Brauer et al. which relied on hierarchical clustering. The lower p-values suggest that the gene groups obtained by StepMiner are atleast as effective as hierarchical clustering.

Conclusion

StepMiner enables a user to identify the binary temporal behavior of genes in response to a stimulus. Compared to hierarchical clustering, StepMiner appears to produce heatmap that is arguably easier to interpret.

poster
33

Identifying relationships between gene co-evolution and metabolic function

Adam Sadovsky

Gal Chechik
Daphne Koller

Purpose

The purpose of this project is to study the interplay between two types of relationships between genes: (1) how they co-evolve over time and (2) how their enzymes co-function in the metabolic network. To quantify the degree of co-evolution between sets of genes, we model gene evolution as a continuous-time Markov process and use expectation maximization (EM) techniques to compute maximum-likelihood rate parameters for independent and dependent models; the difference in likelihood between these two models provides a good measure for co-evolution. To identify enzymes that potentially co-function, we search the S. cerevisiae metabolic network for sets of enzymes that conform to various wiring patterns (e.g. pairs of enzymes where one's products are the other's substrates). We then relate gene co-evolution to co-function by examining the distribution of co-evolution scores for different sets of genes. Our results show that the rate of co-evolution between genes tends to decay as the distance between them in the metabolic network is increased. In addition, some specific types of gene wiring patterns exhibit significantly higher rates of co-evolution than the average pair of neighboring genes. Finally, we found that several labeled functional subnetworks from the full metabolic network exhibit relatively high levels of pairwise gene co-evolution, while 43 of the 67 subnetworks contained no strongly co-evolving gene pairs. We therefore conclude that there is a strong relationship between the degree of co-evolution between genes and their functional relatedness.

Poster Description

My poster gives an overview of metabolic networks and co-evolution and then proceeds to describe in depth both the types of metabolic network wiring patterns we identify and our method for measuring gene co-evolution. The results section shows various co-evolution score distribution comparisons along with the conclusions that can be drawn from these comparisons.

poster 34

Flux control in metabolic networks: optimality and robustness

Xing Chen

Gal Chechik

Daphne Koller

Purpose

A metabolic network can be viewed as a graph, where compounds are nodes, and reactions are edges. In this study, we examine the control f metabolic fluxes, or rates of reaction in the Saccharomyces cerevisiae (yeast) metabolic network. We are interested in the cell's metabolic control "strategy," which determines how it sets fluxes. It is often assumed that a cell sets fluxes to maximize its biomass growth. However, recent studies have shown this to be inaccurate; for instance, certain wild-type organisms such as B. Subtilis have been shown to have a suboptimal flux distribution. Here, we explore the cell's metabolic strategy for setting fluxes by modeling metabolic network control using different optimization techniques, and comparing with biological flux data.

Material and Methods

1) Using linear optimization, we attempt to model flux control in the cell. Given a measured distribution of fluxes, we can look at the growth "landscape" for the region around this measured point. To do this, we model the cell setting different target flux levels using a LP by permuting the fluxes around the measured flux point, and attempting to maximize growth given stoichiometric (mass balance) and environmental constraints.

2) We test robustness in the cell's reactions by fixing growth above a baseline (90% of its measured value), forcing one flux f* below its measured value, and minimizing using quadratic programming the Euclidean distance between the other fluxes and their original measured values. Thus, we observe how far fluxes must move to compensate for a decrease in f*.

Results and Conclusion

Our preliminary results from the above experiments reveal two things. First, we see that the cell tends to set fluxes far from the edge where growth begins to drop. Thus, we theorize that cells maintain a robust flux distribution well above the minimal flux levels needed to achieve the measured growth. Furthermore, we see that even when the flux is reduced to low levels, the cell is able to almost always able to maintain the original growth flux; thus, the cell's original flux distribution was indeed built with high levels of redundancy. We hypothesize that the cell maintains high robustness to ensure survival given inaccurate noisy control and environmental perturbations.

References

Blank LM, Kuepfer L, Sauer U: Large-scale 13C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. Genome Biol 2005, 6:R49.

Fischer E, Sauer U: Large-scale in vivo flux analysis shows rigidity and suboptimal performance of Bacillus subtilis metabolism. Nature Genetics 37, 636 - 640 (2005)

poster 35

Incorporating comparative genomics to improve phosphorylation prediction

Samuel Pearlman

Zach Serber
James E. Ferrell

Purpose

To incorporate evolutionary data into existing protein phosphorylation prediction algorithms.

Materials and Methods

Eukaryotic protein phosphorylation is a ubiquitous post-translational mechanism for the temporal and spatial regulation of proteins. While it is estimated that one-third of human proteins are phosphorylated, only a few thousand phosphorylation sites have been experimentally verified, leaving many thousands of sites undiscovered. These proteins and their kinases are potential targets for drug therapies, but are costly to find experimentally. Thus it is of great interest to try to narrow the search by computationally predicting the positions of novel phosphosites within human proteins. Current methods of phosphosite prediction rely almost exclusively on primary sequence motifs, generating many false positives.

Molecular biologists have known for years that they can sometimes mimic the phosphorylated state of a protein by substituting an acidic residue, either aspartate or glutamate, for the phosphosite. Nature may have been employing a similar trick for far longer, evolving some genes to encode a phosphosite where once there had been an acidic residue. This substitution likely provides a selective advantage by converting a protein that was previously regulated only by synthesis and destruction into one that is switchable by a fast, reversible enzymatic reaction. By examining the amino acid replacement rates in multiple sequence alignments of our phosphoproteins and their most closely related homologs, I will determine if there are differences in the rates at which amino acids replace phosphorylated vs. unphosphorylated (control) serine residues.

Results

Evaluation of the amino acid replacement rates at over 3600 phosphorylation sites shows that acidic residues are enriched in replacing phosphorylated serines compared to our control serines. Additionally, phosphorylated serines are conserved significantly more than control serines, and there are fewer gaps at the position of phosphoserines. We also detected residues that are conserved near the phosphosite when there is a phosphorylatable residue at that site.

Conclusions

By incorporating rates of replacement of serine, threonine or tyrosine residues by aspartate or glutamate among homologous proteins, I believe we can find families of proteins containing phosphosites that evolved in this manner, and improve prediction by using a combination of primary sequence and evolutionary information. I also explore additional features that may inform a machine learning approach to prediction, including the concurrent evolution of kinase recognition motifs, which appear to arise in conjunction with the phosphosite.

poster 36

Building a nosology based on disease etiology

Irene Liu

Paul H. Wise
Atul J. Butte

Purpose

Nosology (In Greek Nosos==Disease) is the branch of medicine that deals with classification of diseases. Such classifications are often based on disease symptoms or the organ systems affected. We aim to build a nosology that reflects the current understanding of disease etiology.

Material and Methods

The etiological factors of various diseases are identified by examining the Medical Subject Headings (MeSH) annotations of articles in the MEDLINE database. In general, if an article discusses the etiological factors of a disease as one of its main points, its MeSH annotation will include informative subheadings such as "etiology" (for the disease) and "adverse effect" (for the etiological factors). Thus, examining the MeSH annotations of MEDLINE articles is an efficient and accurate way to obtain disease-etiological factor relationship.

Results

From the MEDLINE articles that were published from 1996 to 2006, we identified 4744 etiological factors associated with 3564 diseases. A nosology was then built by hierarchical clustering of the diseases based on the etiological factors. Disease mortality and mobility data from the CDC can also be tied in to the nosology. This nosology is, to our knowledge, the first that was derived from comprehensive review of disease etiology. It offers new insights into the classification of diseases. Identification of the etiological factors that contribute to diseases that affect the largest number of people may result in health policies that target such factors.

Subtyping and detecting recombination in HIV sequences with Bayesian models and signature patterns

poster 37

Wenqi Shao

Purpose

This project addresses the problem of detecting recombination and subtyping HIV DNA Sequences using two methods: Phyml and Marc Suchard's DualBrothers Multiple Change-Point Model. Subtyping DNA sequences will allow doctors to develop more effective HIV treatments. Increasing recombination in sequences creates the need for innovative techniques. Additionally, R was used to identify signature patterns within an DNA sequence. This could greatly hasten subtyping procedures.

Methods

Our first method used Phyml and R to visually identify the subtype of an HIV DNA subsequence. Phyml estimates large phylogenies by maximum likelihood. Next, Marc Suchard's DualBrothers Software was chosen to determine HIV DNA sequence subtypes and identify recombination. The technique simultaneously inferences recombination, crossover points (COPs), and parental representatives. The dual MCP model introduces two a priori independent change-point processes to describe spatial phylogenetic variation. The data set consists of N reference sequences and one unknown subtype DNA sequence. The command file includes parameters for the MCMC and hyperprior mean number of topology and substitution break-points. To obtain signature patterns, R was used to analyze those positions that differ above a certain threshold amongst the sequences. A classification tree for each position is produced by binary recursive partitioning.

Results

Using R and Phyml, one can visually identify to which subtree the unknown DNA sequence belongs. A phylogenetic profile by DualBrothers consists of a top plot of the marginal posterior probabilities of the different po ssible topologies for each site along the alignment and a graph of the expected average branch length μ for each site. From 37 reference subtype RT sequences from the LANL HIV database, the bifurcation tree formula grouped various positions into nodes. A table detailing each node is produced.

Conclusion

For pure strains, Phyml often allocated the unknown DNA sequence into an obvious subtree. However, there was no reliability or accuracy score. The DualBrothers multiple change-point model realistically models spatial phylogenetic variation using a parsimonious number of parameters and incorporates uncertainty in topologies and locations of recombination sites. It is realistic and easy to implement. However, it is a little computer intensive. The use of R and ClustalW to find signature patterns goes beyond the simple alignment score cutoff used by VESPA. It successfully creates a classification tree of each nonunanimous position. Further research remains.

Reference

Suchard, Marc A., "Dual multiple change-point model leads to more accurate recombination detection," Bioinformatics 21(13), (2005): 3034–3042.

poster 38

Automatic identification of pharmacogenomic literature using two-level text classification

Rong Xu
Yael Garten
Daniel L. Rubin
Alan M. Garber
Russ B. Altman

Purpose

Pharmacogenomics is the study of how variation in human genetics leads to variation in response to drugs. Researchers in this field must stay abreast of the massive amounts of literature published in this field, however efficiently searching the 16 million references in PubMed is a challenging task. Automatic text classification methods exist to distinguish relevant articles from within a collection of articles. We used two level hierarchical text classification to (1) identify articles with pharmacogenomic relevance from the general corpus that comprises PubMed, and then (2) to classify these relevant articles into different subcategories which portray the various interest of researchers.

Materials and Methods

We performed hierarchical text classification: In our first step, we used 1400 manually curated papers from PharmGKB as our positive training set, and1400 randomly selected human-related publications from PubMed as our negative training set. We experimented with two different feature sets: (1) words in the title and abstract of the publication and (2) MeSH headings. Our second step was the subcategorization of pharmacogenomics-related papers into one or more of five categories (Genotype, Pharmacokinetics, Pharmacodynamics, Clinical Outcomes, Functional Assay, abbreviated as GN, PK, PD, CO, and FA respectively). We performed this step using a classifier trained on the manually classified articles currently existing within PharmGKB (several hundred articles per category). For example, to identify genotype-related papers, we trained a two-class classifier whereby the positive training set was the 812 articles classified as GN in PharmGKB, and the negative training set was comprised of all the rest of the 588 papers in PharmGKB. We experimented with the same features as were described in our first step. For both steps, we experimented with the following classification methods: Naïve Bayes, Maximum Entropy, and Decision Trees.

Results

For the first step, our best results were obtained by using the Maximum Entropy classifier trained on title and abstracts. This classifier yielded F1 scores of 94.1%, with 92.8% precision and 94.7% recall (the F1 score is a statistical measure of accuracy which combines precision and recall). Training on MeSH headings gave comparable results (91.1% F1, 91.8% precision, and 88.3% recall) .In the subsequent step of identifying the five subcategories, we trained Maximum Entropy classifiers using abstracts or MeSH headings. The F1 scores for the five subcategories are as follows, when using the title and abstract words, and MeSH headings, respectively: 1) GN subcategory: 74.2% and 74.7%, 2) PK subcategory: 89.2% and 85.2%, 3) PD subcategory: 76.8% and 76.4%, 4) CO subcategory: 78.9% and 77.6%, 5) FA subcategory: 84.9% and 81.4%.

Conclusion

Two level hierarchical text classification using abstracts or MeSH headings can accurately identify pharmacogenomics-related articles and can further classify these articles into subcategories of interest. We are currently experimenting to further improve the performance by semantically tagging abstracts with relationship and entity categories using a pharmcogenomics-related ontology (Pharmspresso project). We are also working on the identification of pharmocogenomic-related literature by using the full text of articles, in addition to using abstracts.

Combining text classification and hidden Markov modeling techniques for structuring randomized clinical trial abstracts

poster 39

Rong Xu
Kaustubh Supekar
Yang Huang
Amar Das
Alan Garber

Purpose

Randomized clinical trials (RCT) papers provide reliable information about efficacy of medical interventions. Current keyword based search methods to retrieve medical evidence, overload users with irrelevant information as these methods often do not take in to consideration semantics encoded within abstracts and the search query. Personalized semantic search, intelligent clinical question answering and medical evidence summarization aim to solve this information overload problem. Most of these approaches will significantly benefit if the information available in the abstracts is structured into meaningful categories (e.g., background, objective, method, result and conclusion). While many journals use structured abstract format, majority of RCT abstracts still remain unstructured.

Methods

3,896 structured RCT abstracts published from 2004 to 2005 were randomly selected and parsed into 46,370 sentences. Each sentence was a labeled input to a multi-class (Introduction, Objective, Method, Result and Conclusion) text classifier. To pick the best method to represent a sentence, we compared the results of text classifiers when the sentence was presented as a (1) N-gram, (2) bag-of-words with stemming (3) bag-of-words with no stop words, and (4) unprocessed bag-of-words. Performance was measured by classification precision, recall and F1 measure (a composite measure of classification precision and recall). In addition, to exploit the sequential ordering of sentences in an abstract, we used HMM to label sentence types. We have transformed the sentence categorization problem into a HMM sequence alignment problem. The HMM states correspond to the sentence types. Labeling sentences in an abstract is equivalent to aligning the sentences to the HMM states. There are five states in our HMM model: Background, Objective, Method, Result and Conclusion. The transition probabilities between these states were estimated from the training data by dividing the number of times each transition occurs in the training set by the sum of all the transitions. The state emission probabilities were calculated from the score output that were reported by the multi-class classifiers. Given the HMM model, state emission probabilities, and the state transition probabilities, Viterbi algorithm was used to compute most-likely sequence of states that emit all the sentences in the abstract. Subsequently, the state associated with the sentence was extracted from the most-likely sequence of states.

Results

We have evaluated the performance of classifying sentences in RCT abstracts using three text classification algorithms, namely, Naïve Bayes (NB), Maximum Entropy (ME) and Decision Tree (DT), basic (without modifications). Precision, recall, and F1 are compared across five type of abstract sections with and without HMM augmentation. HMM significantly improves the classification in all five types of sentences, with average of precision of 0.94 and recall of 0.93, compared with text classification alone (precision of 0.81 and recall of 0.80).

poster 40

Sequence variation and constraint in the human genome

David Goode

Gregory Cooper
Jeremy Schmutz
Ming Tsai
Richard M. Myers
Arend Sidow

The goal of this study is to examine how genetic variation within human populations is related to the level of evolutionary constraint acting on the genome. To assess genetic variation, 422 human subjects were resequenced at ~68 kb of the human genome enriched for sequences conserved across the mammalian lineage. Measures of evolutionary constraint were obtained using the Genomic Evolutionary Rate Profiling (GERP) program. Our set of SNPs is enriched in rare alleles, with 298 SNPs (out of 762) having one instance of the minor allele and a median minor allele frequency of 0.1%. A significant under-representation of SNPs in constrained sequence elements was found and SNPs in constrained elements tended to have lower minor allele frequencies and hetereozygosity than those outside of constrained elements.

poster 41

Investigating 4-dimensional structure-based function prediction

Dariya Glazer

Russ Altman

Elucidating the most basic physical principles of molecular interactions is the key to understanding and manipulating molecular function. Small structural changes within molecules can engender their function and specificity. Recapitulating this function, as essential to many engineering fields, requires thorough understanding of the relationship between the fine scale structural changes and functional disturbances. To date, there are about 40,000 structures available in Protein Data Bank (PDB). Most of those have been obtained by X-ray crystallography, which provides single static 3-dimensional (3D) structures. Many of these structures do not have a function assigned to them, and some of the structures with experimentally known functions fail to be assigned as such by the existing structure-based function prediction techniques. Artifacts introduced into the molecules as part of crystallography are a good example of why such failures occur. Furthermore, molecules are not frozen in space and time, but are naturally dynamic entities.

Molecular dynamics (MD) simulations use the experimentally solved static images and physical properties, such as Newtonian equations of motions and Wan der Waals forces, to computationally simulate the motion of and within the molecule through time. I want to investigate how dynamics within the molecule influence function over time, using GROMACS, a software suit enabling MD simulations, and FEATURE, a tool which creates models of physico-chemical characteristics, features, of 3D environments of interest as compared to some background. In particular, using pairs of structures for the same proteins, I built 4 types of physics-based MD simulations of calcium (Ca2) binding proteins: 1. protein with Ca2 bound (holo), 2. protein with Ca2 originally bound but removed for the simulation (holo- Ca2), 3. protein without Ca2 bound (apo), and 4. protein without Ca2 bound originally but added for the simulation (apo+Ca2). Correspondence of results from simulations 1 & 4 and 2 & 3 is expected, as scored by FEATURE Ca2 binding model. Furthermore, employing MD as a feature in assigning function to structure will be evaluated. Observations of improved FEATURE scores, regarding Ca2 binding, during the course of the simulation with structures for which a function has been assigned tentatively or not at all yet will encourage addition of MD as another feature in FEATURE. A positive feedback from experimental analysis of these results and translation of the method to other 3D environments will establish the first 4D structure-based function prediction technique.

poster 42

Protein denaturation in a nanodroplet

Gaurav Chopra

Michael Levitt

Purpose

Protein folding, in which a disordered polypeptide chain spontaneously rearranges its shape to become an ordered three-dimensional structure, is a very important process in biology. A wide variety of diseases are linked to the unfolding and subsequent misfolding of proteins and peptides. Even with the largest of supercomputers, the protein folding problem still remains unsolved. To understand the protein folding problem, many have studied the inverse problem, i.e. the protein "unfolding" problem, in which the ordered structure breaks down. It is seen that water plays a major role in structural transitions of alpha-helix unfolding mechanism [1], but the simulations with thousands of explicit water molecules in the system make it difficult to discover exactly what role the water plays. Does water make non-bonded interactions stronger through the hydrophobic effect? Does water break interpeptide hydrogen bonds by offering alternative hydrogen bonding partners. In an attempt to clearly delineate the role of water in alpha-helix denaturation, our research simulates an alpha-helix in a nanodroplet containing different numbers of water molecules. By varying the number of water molecules from zero to thousands, we expect to find the minimum number of water molecules, which makes the system behave like it is in bulk water.

Methods

The peptide used in these simulations is ALA15, a 15 residue chain of Alanine, initially forming an alpha-helix. Several molecular dynamics simulations are run using our ENCAD (Energy Calculation and Dynamics) code [2] by varying the number of water molecules in the solution. Water molecules were added around the protein as a spherical shell in random order. Several simulations are run varying the parameters of the initial seed value of the system and the randomness seed for adding water molecules to get better statistics for the data. Varying radius of the water shell around the protein defines the nanodroplet. The initial work presented here studies this system using only the Encad potential at different temperatures. For a complete picture we plan to repeat these studies with a number of other popular all-atom energy functions including OPLS, Gromos and Amber.

Results and Conclusions

We calculate the rate of unfolding by plotting the percent alpha-helix as a function of simulation time. Runs are repeated at different temperatures and with different numbers of water molecules in the nanodroplet. Comparison of these plots shed light on the role of water in alpha helix denaturation.

References

1. Daggett V., Levitt M., J. Mol. Biol. (1992)
2. Levitt M, ENCAD (1990)

poster 43

Exploring the folding free energy landscape of a helical peptide via serial replica exchange method

Xuhui Huang
Vijay Pande
Bruce Berne
Michael Levitt
Sanghyun Park
Dan Ensign

Purpose

The purpose is to compute the accurate folding free energy landscape of a 21-residue arginine-substituted alpha-helical peptide at different temperatures by all-atom simulations using serial replica exchange sampling method. The thermodynamics properties are measured and compared with the experimental values.

Methods

Replica exchange method (REM) is a powerful method for speeding up the sampling of conformational states of proteins with rough energy landscapes, where stable conformational states can be separated by large energy barriers. The usual REM requires that different replicas run synchronously. Serial replica exchange (SREM), a method that is equivalent to standard REM in terms of efficiency, yet runs asynchronously on a distributed network of computers. This unique feature enables it to study the protein folding on Folding@Home, a worldwide distributed computing environment. We apply SREM to a 21-residue alpha-helical peptide. All atom simulations are used and the size of the system is about 10000 atoms. The total simulation time is about 3 microseconds which is order of magnitude longer than the experimental folding time (~20ns).

Results and Conclusions

The convergence of Potential Energy Distribution Functions at different temperatures shows that our extensive sampling results in the complete convergence to the ensemble equilibrium. The thermodynamic properties such as helical content, helix radius, rise per reside and twist per reside and equilibrium torsion angle distributions are measured at different temperatures and in general agreement with experimental data. Folding free energy landscape shows a two-state folding for this peptide without any intermediate states. The "folded" state is slighted more stable (1 to 2 Kcal/mol) than the "unfolded" state at 300K.

References

1. Eric J. Sorin and Vijay S. Pande. Biophysical Journal , (88), 2516-2524,2005.
2. Byungchan Kim, Pu Liu, Morten Hagen, Richard A. Friesner, and B. J. Berne, J. Phys. Chem (B), in review, 2006

poster 44

Recognizing complex ligand binding sites using multiple local structural models

Jessica Ebert

Russ Altman

Purpose

While most functional prediction efforts operate at the sequence level, both the increasing number of protein structures with unknown functions produced by structural genomics efforts and the steady improvement of structure prediction algorithms has created an opportunity for the use of methods that operate at the level of tertiary structure. We present an extension to the FEATURE algorithm (Wei and Altman 1998) that uses multiple local models to characterize an active site or ligand binding site. FEATURE models a site of interest by identifying physicochemical attributes in a series of concentric shells around the active site that are over- or underrepresented with respect to the background. This process produces a model that can be use calculated the likelihood that a prospective environment in a protein structure performs the biochemical function represented by the model. Because the physicochemical attributes are averaged in each spherical shell, orientation of the training sites is unnecessary and variation in the placement of atoms within each shell is permitted.

Materials and Methods

For larger ligands, a single spherical environment may not be sufficient to capture all of the features relevant to binding and substrate specificity. Hence, we have developed an automated algorithm to select multiple centers on the ligand at which FEATURE models are built. The selected centers are those which best distinguish the ligand binding site from decoy sites at similar atom densities. In order to increase the specificity of the search, one may require a high scoring match to each model, while the sensitivity may instead be increased by allowing a strong match to one model to compensate for a weak match to another.

Results

We evaluate this method by applying it to ATP binding sites and show that the combination of two models centered at the C1 carbon of the ribose and the gamma phosphate atom outperforms either individual model. Furthermore, our modular approach allows us to recognize ATP binding sites even in apo proteins.

References

1. Wei L, and Altman RB.1998. Recognizing protein binding sites using statistical descriptions of their 3D environments. Pacific Symposium on Biocomputing. 497-508.

poster 45

Knowledge-informed coarse-grained modeling of RNA structure

Magda A. Jonikas

Randall J. Radmer
Alain Laederach
Russ B. Altman

Purpose

There exist many approaches that probe RNA structure at nucleotide resolution, including chemical footprinting, small angle x-ray scattering (SAXS) and FRET measurements. While these methods provide significant information for modeling the structure of RNA molecules, their incorporation into structural models is challenging and requires significant user time and expertise. Furthermore there are often varying amounts of structural information for RNA molecules. For example, subdomains of the Tetrahymena thermophila group I intron have been solved by crystallography individually. However, reconstructing a fully self-consistent structural model that incorporates all structural evidence remains a significant modeling challenge.

Materials and Methods

To address this problem, we have developed the Nucleic Acid Simulation Tool (NAST), a software package for coarse-grained simulations of RNA with the ability to incorporate low-resolution data from experiments, as well as different levels of structural information. By reducing each nucleotide to a single pseudo-atom, the computational complexity of modeling an RNA molecule is significantly decreased. To model these systems, we apply a knowledge-based potential derived from the statistics of known RNA structures. These potentials include terms for base-base distances, three-base angles, four-base torsions, and tertiary contacts. The coarse-grained resolution used in our system allows incorporation of low-resolution experimental data into the structure model. Additionally, NAST allows the user to specify different types of structural information throughout the molecules, including solved crystal structure and predicted structures. These structural data are used to apply appropriate potentials to each part of the molecule.

Results

We demonstrate that NAST can provide useful structural information by modeling the P4-P6 subdomain of the T. thermophila group I intron using no crystallographic information and comparing results to the solved structure. We also use NAST to refine the existing model of the entire T. thermophila group I intron for which several subdomains have not been solved by x-ray crystallography. The NAST software package allows us to predict the structure of these unsolved subdomains. NAST is available for download from https://simtk.org/home/nast.

poster 46

An efficient algorithm for computing conformations of protein loops

Guanfeng Liu

R. J. Milgram
A. Dhanik
J. C. Latombe

Purpose

The goal of this paper is to develop an efficient algorithm for sampling the self-motion manifold of protein loops.

Materials and Methods

Loop closure in proteins requires computing the values of the inverse kinematics (IK) map for a backbone fragment. It occurs in a variety of contexts, for example, structure determination from electron-density maps,loop insertion in homology-based structure prediction, backbone tweaking for protein energy minimization, and study of protein mobility in folded states. In this paper the structure of the inverse kinematics map of such fragments is studied based on a complete analysis of the singularities of the corresponding forward kinematics map. This structure, combining with convex optimization techniques, yields an efficient recursive algorithm to sample IK solutions for such fragments, by identifying the feasible range of each successive torsional dof. A numerical homotopy algorithm is then used to deform the IK solutions in accordance with the variance of geometric parameters.

Results

The successful rate of our algorithm for protein loop conformation space sampling is about 100 times higher than the RLG algorithm.

Conclusion

In this paper, a novel and efficient algorithm for computing conformations of protein loops is proposed that can be combined with collision detection algorithms to study and sample the clash-free subset of the self-motion manifold of a given fragment.

References

1. J. Cortes, T. Simeon, M. Renaud-Simeon, and V. Tran. Geometric algorithms for the conformational analysis of long protein loops. J. Computational Chemistry, 25(7):956-967, 2004.
2.E.A. Coutsias, C Seok, M.P. Jacobson, and K.A. Dill. A kinematic view of loop closure. J. Computational Chemistry, 25:510-528, 2004.

poster 47

mdMotif: De novo protein design using motif libraries

Ryo Shimizu

Purpose
mdMotif is a tool that allows the user to design new proteins and obtain a putative sequence that can fold into such a structure.

Materials and Methods
The basic premise of the application is that one can build complex protein structures using structural motifs as building blocks. mdMotif is dependent on a library of structural motifs. I-sites1, as provided by Chris Bystroff's lab at RPI, is one such library.

mdMotif has the following use-case in mind.

1. The user constructs a target structure using structural motifs (such as the I-sites library) building blocks.

2. The applications will simulate molecular dynamics on the target structure to see how "off target" the structure may become.

Assuming the user is happy with the final structure, the putative sequence that will produce a protein is simply the join of the motif sequence pattern (with or without loops inserted).

Results
The project is still in its proof-of-concept stage, and a sample application has been created.
The following is planned for the future.
1. An improved simulation engine
Currently, the molecular dynamics simulation is a naïve implementation of a mass-spring system. The bonds between atoms are modeled as stiff springs (ks = 4000), and the weaker, non-bonded interactions are modeled as weak springs (ks= 10 to 300). This would need to be improved to user more sophisticated energy function involving bond energy, non bond energy, van der Waals forces, etc. Additionally, by using implicit integration techniques, it may be possible to make the timestep large enough that the application can run reasonably on a PC.
2. Improve the user interface
Currently, there is no way for the user to orient the various structures in the three dimensional canvas—the orientation of the motif is the orientation as it came in the PDB file of the I-sites library. This will be improved so that the user can freely move individual motifs around and orient them in the way the user desires. In addition, the motif library should be readily accessible from the graphical user interface as palettes from which the user can pick and choose and place it onto the canvas—like a drawing program.
3. Validation
A method for validation is to be determined.

poster 48

Structural properties of proteins in minimalist lattice models

Sergio Moreno

Michael Levitt

Purpose

We are interested in finding the most characteristic structural properties of proteins considered as a whole. Our system of study is not any particular protein but the ensemble of all proteins. First we want to use a computer model of proteins. Due to the great computational cost of enumerating all the possible proteins we will use minimalist models, which are extremely simple but that still show protein-like features. Next we want to see if what is found using these models is also valid when considering real proteins.

Material and Methods

We use the "2D HP lattice model" that considers protein chains of a given length (that we vary from 16 to 27 residues) with only two types of amino acids: hydrophobic (H) and polar (P). The chains are placed on a 2-dimensional lattice and are self-avoiding. Any residue interacts only with residues that are closest in space but that are not adjacent along the chain. The election of the energy interaction terms specifies a particular model. The ensemble of all possible proteins in a model is found by considering each possible HP sequence and determining if there is a single structure whose energy is lower than if the sequence adopts any other structure (i.e. if the protein has a well defined native state).

Results

(1) In our minimalist models we have studied the different protein "shapes". Very few (~200) different shapes occur and those (~ 1-10) with the maximum possible number of contacts are adopted by most (~ 80%) of the proteins. Furthermore, the fraction of possible shapes that is adopted by proteins diminishes dramatically as the number of contacts decreases. (2) We have also considered the different protein HP "arrangements". In particular we have looked at the HP burial or exposure of the most popular arrangements, which is indicative of how protein-like models are. This is the biggest qualitative difference among models so this is helpful to choose a model to work on. (3) By considering a fixed shape and looking at the most popular arrangements we have found that those that bury the H residues as much as possible are popular, but not the most popular ones. The same is true if both the shape and the number of H residues are fixed. Therefore partial H exposure makes proteins more protein-like.

Conclusion

Our findings give a deeper understanding of the structure of proteins in minimalist lattice models. Future work will aim at studying if the same trends that are observed in our models also occur in real proteins. If that is the case that would prove that these models can be used as a starting point to gain more insight about real proteins.1989.

poster 49

Conservation and divergence of yeast transcriptional networks

Xiao Liu

Purpose

Variation in gene expression plays an important role in phenotypic diversity of related organisms1. To investigate genetic basis underlying the variation in gene expression systematically, genome-wide prediction is needed to distinguish targets from non-targets of a transcription factor in multiple, even distantly related, species. Hence I developed a target prediction approach by accounting for both binding site affinity and its conservation between species. Based on precise target prediction, one can compare regulons of transcription factor orthologs in different organisms and then investigate evolution of binding sites, functional drift of transcription factors and plasticity of regulatory networks.

Materials and Methods

My algorithm assumes that transcription factor orthologs in even distantly related species still recognize similar sequence motif. I used binding motif of a transcription factor determined in S. cerevisiae to approximate those in other species. Then, every gene was assigned a "Motif Score" based on matching of binding sites in its promoter to motifs. Then the "Motif Score" is adjusted according to how conservative it is among given set of species. The distribution of "Conservation-Adjusted Motif Score" (CAMS) was simulated using pseudo-motifs so that one can identify genes whose binding sites are more conserved than by random. Target prediction of CAMS was tested by Gene Ontology analysis and targets determined by published ChIP-chip experiments.

Results

This study focused on eight Hemiascomycetes, among which C. albicans diverged from seven Saccharomyces species ~ 400 myr ago. Through the comparision of predicted targets between genomes, transcription factors Ume6 and Cbf1 show dramatically changes, supporting the fact that they have diverged biological function between S. cerevisiae and C. albicans. On the other hand, Fkh1 and Gcn4 show highly conserved regulons. Between these extremes, Mcm1 regulon is only moderately conserved. However, species-specific targets are enriched in similar function. For example, Mcm1 is predicted to regulate different subunits of MCM2-7 complex in different species.

Conclusion

Precise target prediction enables one to compare regulons between even distantly related species. It suggests that regulon evolution of a transcrion factor is correlated with its functional change. Transcription factors are also observed to have multiple ways to regulate same biological process during evolution.

References

1.Ihmels, J. et al. Rewiring of the yeast transcriptional network through the evolution of motif usage. Science 309, 938-40 (2005).

poster 50

Quality assessment of microarray data and optimal filtering criteria

Takashi Kido

Janos Demeter
Gavin Sherlock

Purpose

The quality assessment of microarray data is an important issue, to prevent the risk of analysis of poor quality data. However, the methods for standard quality assessment have not yet been established. In addition, researchers use adhoc filtering criteria to select data for analysis, because there is little or no guidance on optimal filtering parameters. The goal of this project is (1) to generate quality metrics for individual microarrays, and microarray datasets, and (2)to learn how best to apply quality filters to microarray data using these metrics.

Materials and Methods

We define the "Q-Score" as a quality assessment tool using replicated spots. In order to assess and improve the quality of microarray data with Q-Score, we take the following approach:
1. Evaluate the "Q-Score" dynamics with several data sets and several filters.
2. Investigate the correlations between several spot metrics.
3. Understand the "Q-Score" dynamics and several filter correlations in multi-dimensional space.
4. Based on the empirical observations, generate quality metrics and develop algorithms for finding the optimal combination of quality filters.

Results

Experimental evaluation of Q-Score dynamics and experiment on filter correlations are shown in the poster presentation.

Conclusion

1. The Q-Score is reasonably consistent with the visual spot evaluations.
2. Inflection points of Q-Score may be useful for determining the optimal cutoff for filtering data.
3. Filtering with Ch2 normalized Intensity/median background and Regression Correlation improves the Q-Score the most.
4. Spot metrics can be clustered. Correlation matrix among dozens of filters is very robust; it may be useful for choosing orthogonal filters.

poster 51

Effects of linear energy transfer (LET) on intrinsic radiation sensitivity – tests of the putative mechanisms underlying the cell killing effects of ionizing radiation

David J. Carlson

Robert D. Stewart

Purpose

Beams of protons and more massive ions have the potential to improve the efficacy of radiation therapy for the treatment of cancer because ions such as these can produce more favorable dose distributions than those achievable with low LET radiations. The initial kinetic energy of the incident ion can also be adjusted to modulate the relative biological effectiveness (RBE) of a beam.

Methods

Several possible mechanisms may contribute to the increased RBE of high LET radiations. The local complexity of the DNA damage formed by ionizing radiation tends to increase with increasing LET, and if some forms of damage, such as complex double strand breaks (DSBs), are intrinsically unrejoinable and lethal, RBE will increase as the LET increases. Alternatively, a subset of the initial DSBs may be misrepaired or become unrejoinable because of random-seeming biological processes (e.g., damage fixation). Pairs of DSBs formed by the same or different tracks can also interact in pairwise fashion to form exchanges, i.e., the intra- and inter-track pairwise DSB interactions sometimes known as binary misrepair. Pairwise DSB interactions may increase with increasing LET because of proximity effects (formation of regionally multiply damaged sites) or because the overall DSB yield tends to increase with increasing LET. We used Monte Carlo methods to estimate the initial yield of various classes of simple and complex DSBs for selected low- and high-LET radiations. The DSB yields were then used in conjunction with the linear-quadratic (LQ) survival model to investigate the contribution to cell killing of unrejoinable DSBs, misrepaired and fixed DSBs, and exchanges formed through intra- and inter-track DSB interactions.

Results

Our analysis of published cell survival data for human kidney cells suggests intra-track DSB interactions are negligible for low-LET radiations. The analysis also indicates that all classes of DSBs are potentially rejoinable, even very complex DSBs composed of 10 or more lesions, and no class of DSB is intrinsically unrejoinable. The importance of intra-track pairwise DSB interactions increases with increasing LET but misrepaired and fixed damage still contribute significantly to one-track cell killing mechanisms. For x-rays (~1.9 keV/micron), up to 99.3% of the one-track killing is attributed to misrepaired and fixed DSBs. The remaining 0.7% of the one-track killing is attributed to lethal intra-track DSB interactions. For comparison, 54.1% of the one-track killing caused by 5.1 MeV alpha particles (~88 keV/micron) is attributed to misrepaired and fixed DSBs and 45.9% is due to lethal intra-track DSB interactions.

poster 52

Quantitative measures of finger, limb and postural movement velocities in very early stage, untreated Parkinson's Disease

Mandy Miller Koop

Nicole Shivitz

Helen Bronte-Stewart

Purpose

Parkinson's disease (PD) currently affects over 1 million Americans (1). A cardinal movement symptom is bradykineia (slowness in movement). In the early stages of the disease, patients almost always state that only one side of their body is affected. We were interested if quantitative motion analysis tools could measure bradykinesia in finger, limb, and/or postural movement in very early stage PD.

Materials and Methods

14 PD patients and 23 age-matched, control subjects consented to participate in this study. In PD patient group, earliest symptom recollection ranged from 6-22 months prior. No patient wanted symptomatic pharmacological treatment. All patients believed that only one side of the body was affected, thus measurements were defined as the affected (A) side or the non-affected (NA) side. Measurements included the velocity of finger movement during a repetitive alternating finger tapping task on a MIDI-interfaced keyboard,(Quantitative DigitoGraphy, QDG(2)), the root mean square angular velocity (Vrms) of the arm during repetitive, wrist pronation-supination, using an angular velocity sensor (Motus Bioengineering Inc (3)), and the velocity of postural movement during a simple reaction task of leaning in different directions, using computerized posturography (NeuroCom Inc(4)).

Results

Velocities of finger and arm movements were significantly slower on the "affected" side in PD patients compared to that of the non-dominant side of control subjects ($p<0.05$ and $p<0.001$ respectively). The velocity of postural movement in any direction was not significantly slowed in PD patients versus controls. There was a significant difference between hands of PD patients for arm velocity ($p<0.02$) but not for finger velocity. This was because finger bradykinesia was more evident in the NA hand compared to controls than limb bradykinesia (mean finger velocity of controls = 105.6, of the PD NA side = 66.2, and of the PD A side = 36.5; mean Vrms of controls = 700 deg/sec, of the PD NA side = 674 deg/sec, and of the PD A side = 442 deg/sec).

Conclusion

Reduced movement velocities (bradykinesia) are measurable in very early stage, untreated PD in finger and limb movements, but not in postural movements. Our results suggests that measurements of bradykinesia of finger movements will be the most useful to detect early stage PD.

poster 53

Dynamics and impact of the immune response to chronic myelogenous leukemia

Peter S. Kim

Christiane I-U. Chen
Holden T. Maecker
Doron Levy
Peter P. Lee

Purpose

The role of the host immune response in relation to current therapies in controlling cancer remains unclear. We hypothesize that novel molecular targeted therapies, such as imatinib for CML, may render leukemic cells immunogenic as patients enter remission.

Materials and Methods

To detect and quantify anti-leukemia immune responses in an antigen unbiased way, we utilized cryopreserved autologous pre-treatment leukemic blood samples (representing predominantly leukemic cells) as stimulators to detect anti-leukemia T cell responses in blood samples from CML patients in remission on imatinib. We studied patients over time to address the dynamics of such responses and developed mathematical models to gain insights into the dynamics and impact of these responses on the control of CML.

Results

Our experimental data show that anti-leukemia T cell responses develop in the majority of CML patients in remission, and that CD4+ T cells producing TNF-alpha represent the dominant response. However, these responses wane over time. Mathematical modeling suggests that the anti-leukemia T cell responses may play a critical role in maintaining CML patients in remission under imatinib therapy. Our model proposes a novel concept of an 'optimal load zone' for leukemic cells in which the anti-leukemia immune response is most effective. Imatinib therapy may drive leukemic cell populations to enter and fall below this optimal load zone too rapidly to sustain the anti-leukemia T cell response.

Conclusion

Our model shows that vaccination approaches, in combination with imatinib therapy, may optimally sustain the anti-leukemia T cell response to potentially eradicate residual leukemic cells for a durable cure of CML.

poster 54

Pharmspresso - A tool for semantic search in full-text articles to support curators and researchers

Yael Garten

Russ B. Altman

Purpose

We have developed Pharmspresso, a specialized search engine which supports querying of pharmacogenomic-related information, to assist researchers and curators of PharmGKB. Pharmspresso is an adaptation of Textpresso [1], which has been widely used by the model organism communities as a search engine enabling extraction of facts from literature. Pharmspresso allows searching the full text of literature for information about human genes, drugs, diseases, polymorphisms, and the relationships between them. The engine allows more sophisticated search than simple keyword searches, answering more challenging queries.

Materials and Methods

We use natural language processing (NLP) techniques combined with an ontology-based approach to parse the full text PDFs of hundreds of articles. Each article is converted into plain text, words are tokenized, and an ontology is used to mark words as belonging to the categories of the ontology. The ontology we use expands the Textpresso ontology which was manually created for use in searching literature related to model organisms. We have adapted the ontology for use in the domain of human pharmacogenomic-related publications. The XML-based markup of each article is indexed and used in subsequent queries submitted by users.

Results

The strength of the search engine is in its ability to search through the full text of articles published in hundreds of different journals, and to highlight sentences within these articles that contain facts of interest to the researchers. The tool automatically extracts facts from the literature, and visually enhances the search through many articles and thus can assist in meeting the goals of researchers and curators more efficiently.

Reference

[1] Muller HM, Kenny EE, Sternberg PW. Textpresso: an ontology-based information retrieval and extraction system for biological literature. PLoS Biol. 2004 Nov;2(11):e309.

poster
55

Modeling hydrophobic collapse in aromatic rich short peptide segments using molecular dynamics

Gaurav Chopra

Xuhui Huang
Michael Levitt

Purpose

Water is widely known as the 'basis of life'. Proteins are even more important (in Greek, protas means "of primary importance"). Consisting as they do of chains of 20 naturally occurring amino acids arranged to encode a 3-D structure, proteins provide the structure and function of all living systems. Although it has been long known that hydrophobic interactions are the major driving force for protein folding, there is no satisfactory model to describe hydrophobic clustering and collapse in proteins. In the last stage of the folding of certain protein oligomers, there is a drying (de-solvation) induced hydrophobic collapse if the buried protein surfaces are very hydrophobic [1]. We derive the order of occurrence between desolvation and hydrophobic collapse. This research combines methods from science and engineering to understand how burial of hydrophobic residues might produce a "hydrophobic collapse".

Materials and Methods

We have developed a method to find the most hydrophobic peptides (from a total of 3706), and have run 100 different examples for 10 ns each using GROMACS with OPLS force-field. The short peptide segments (30 residues long sequences) are extracted from the sequences of natural proteins. These proteins have known 3-D structures and contain peptides that are very hydrophobic and have many hydrophobic contacts. To study whether desolvation induces hydrophobic collapse of short peptide segments, the core radius of gyration and the water number close to the hydrophobic core were monitored. By the combination of these methods we can create a good model of collapse in hydrophobic peptides.

Results

We observe that some peptides remained close to their native X-ray or NMR structure throughout the simulations. Others changed conformation leading to large root mean square coordinate deviations from the starting structures. The simulation of these peptides showed that the number of hydrophobic residues or number of hydrophobic pair-wise contacts did not have any effect on collapse. Moreover, a high ratio of hydrophobic contacts to hydrophobic residues did not cause collapse. When the hydrophobic residues are far apart they tend to cluster to form a collapsed peptide with big cRMS (unstable). The hydrophobic collapse of these "unstable" peptide segments show that in most of the trajectories hydrophobic core collapse and de-solvation are concurrent. The stable peptides do not collapse, as the hydrophobic residues are already clustered in the initial configuration and thus movement of hydrophobic residues is small. Thus a wide range of stability of hydrophobic cores can be derived from same protein sequence.

References

1. Pu Liu, X. Huang, R. Zhou, Bruce J. Berne, Nature (2005).

Notes

Notes

Notes

ALLOY VENTURES

Combining Experience, Connections, and Capital to Create Winning Companies in both Information and Life Sciences

ADEZA · KALOBIOS · Avidia · Novasys Medical® Improving Women's Lives® · SERIOSITY · ALEXZA molecular delivery · CODON DEVICES · RESTORATION ROBOTICS · Fluidigm · Pharsight · Applied Biosystems · Arbor Surgical Technologies, Inc. · CVT | CV Therapeutics · LivingMicroSystems, Inc. · LABCYTE · Adiana OPTIONS FOR WOMEN · NuGEN www.nugeninc.com · CAMBRIOS · BÂRRX MEDICAL · AMGEN · Molecular Imprints · FAVRILLE · forteBIO · OPTIMEDICA · MACUSIGHT

Capital commitments since 1982 in excess of $1 Billion

General Partners

Craig C. Taylor
John F. Shoch
Douglas E. Kelly, M.D.
J. Leighton Read, M.D.
Daniel I. Rubin
Michael W. Hunkapiller
Ammar H. Hanafi
Tony Di Bona

Venture Partners

David W. Pidwell
Peter N. Loukianoff
David J. Ezequelle
W. Ferrell Sanders

www.alloyventures.com

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 Program is facilitated by a new Center (The James H. Clark Center), which was completed in Summer of 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 2006 Sponsors

bcats2006
thankyou

Administrative Help and Guidance

Blanca Pineda (Stanford)
Jonathan Dugan (Stanford)
Michael Alvarez (Stanford)
Vicky Markstein (Life Sciences Society)
Heideh Fattaey (Stanford)
Nancy Riewerts (Stanford)
Karen Kellner (Stanford)
Carol Maxwell (Stanford)
Suzanne Frasca (Stanford)

Previous Organizers

Maureen Hillenmeyer
Lior Ron
Lucinda Southworth
Ryan Spilker
Beverly Tang

Symposium Volunteers

Sarah Aerni
Jatinder Arora
Keith Callenberg
Chen Cao
Dave Chen
Jerry Chen
Annie Chiang
Bernie Daigle
Libby Day
Yael Garten
Daniel Kim
Queenie Lee
Ray Lin
Andrew Lookingbill
Alex Morgan
Masumi Patel
Rashmi Raj
Nigam Shah
Marina Sirota
Aaron Tam
Rebecca Taylor
Abhi Vase
Persis Wadia
Guanglei Xiong
Peggy Yao

Symposium Sponsors

Platinum Sponsors

Simbios
Bio-X

Gold Sponsors

Genentech
Alloy Ventures
Agilent
Roche
Life Sciences Society

Silver Sponsors

Stanford University School of Medicine Career Center
Affymetrix
Applied Biosciences
BD Biosciences

Other Sponsors

Dell
Mathworks
NCBI
Whole Foods

biomedicalcomputationatstanford