BIOMEDICAL COMPUTATION @ STANFORD 2000 SYMPOSIUM PROCEEDINGS

BCATS 2000 SYMPOSIUM PROCEEDINGS

Printed in United States of America

Editors: *David Paik, Jonathan Dugan*
Associate Editors: *Brooke Steele, Olga Troyanskaya*

"Hands" artwork courtesy of *Biomedical Information Technology at Stanford (BITS)*

Ordering information: `bcats@email.com`
Web Site: `http://bcats.stanford.edu`

BIOMEDICAL COMPUTATION AT STANFORD 2000

Symposium Co-Chairs

Jonathan Dugan
David Paik
Brooke Steele
Olga Troyanskaya

Administrative Help

Stanley Jacobs
Kevin Lauderdale
Rosalind Ravasio
Darlene Vian

Symposium Volunteers

Michael Cantor
Jeffrey Chang
Carol Cheng
David Elgart
Valerie Favier
Yueyi (Irene) Liu
Jodi Elgart Paik
Bill Petitt
Rosalind Ravasio
Tomoko Shintani
Matt Stocksiek

Symposium Sponsorship

Biomedical Information Technology at Stanford (BITS)
& The National Library of Medicine
Northern California Pharmaceutical Discussion Group
DoubleTwist
InforMax
Incyte Genomics
GeneLogic
Skjervan, Morril, MacPherson, LLP
Genencor International
Guidant
SGI
Sun Microsystems

TABLE OF CONTENTS

I. Symposium Information…….. 1
   a. Acknowledgements
   b. Symposium Schedule and Map

II. Keynote Speakers……..……...… 5
   a. David Haussler, Ph.D.
   b. Richard Satava, M.D., F.A.C.S.

III. Abstract List……..…….. 9

IV. Scientific Talks Session I…….…….….17

V. Scientific Talks Session II……. 29

VI. Poster Session / Software Demonstration.…….…..…….... 39

VII. Symposium Participant List…….…….101

VIII. Symposium Sponsors……...……115

SYMPOSIUM INFORMATION

ACKNOWLEDGEMENTS

This symposium would not have been possible without help from many people and organizations, both financial and in the donation of peoples' time.

We'd like to acknowledge the Biomedical Information Technology at Stanford (BITS) faculty group for the initial idea of creating this symposium and for their suggestions in various aspects of the process of planning the symposium.

We'd like to thank Dr. Richard Satava and Dr. David Haussler for taking the time out of their very busy schedules to give the keynote speeches at the symposium.

We'd also like to acknowledge the National Library of Medicine for its financial support through the Medical Informatics Training Grant supplement T15-LM 07033.

We'd also like to acknowledge the Northern California Pharmaceutical Discussion Group, DoubleTwist, InforMax, Inc., Incyte Genomics, GeneLogic, Skjervan, Morril, MacPherson, LLP, Genencor International, Guidant, SGI, and Sun for their co-sponsorship of the symposium.

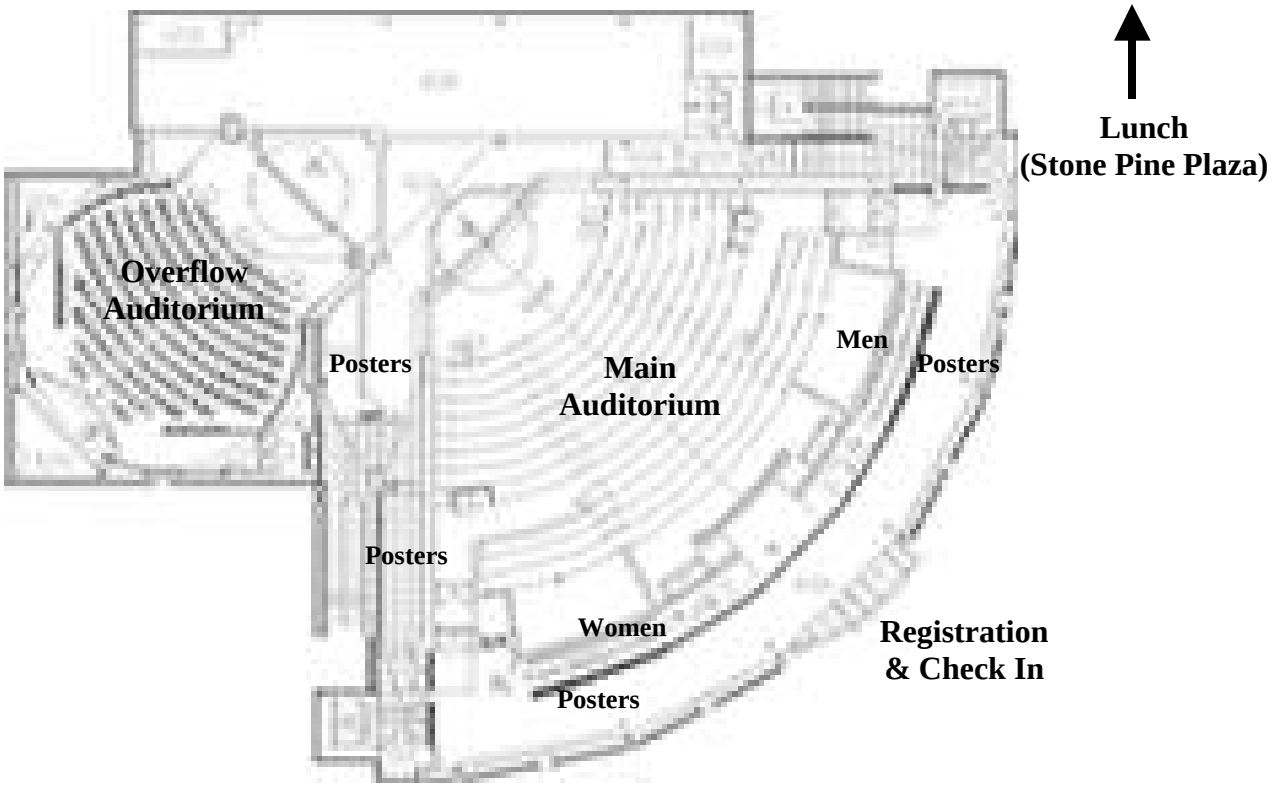
We'd also like to acknowledge the following people for their generous help in organizing the conference: Michael Cantor, Jeffrey Chang, Carol Cheng, David Elgart, Valerie Favier, Yueyi (Irene) Liu, Jodi Elgart Paik, Bill Petitt, Rosalind Ravasio, Tomoko Shintani, and Matt Stocksiek.

Last but not least, the co-chairs would like to thank their friends and family for being supportive and understanding when all we seemed to do and talk about was BCATS.

SYMPOSIUM SCHEDULE AND MAP

Saturday, October 28, 2000

8:00 am	**-**	**9:00 am**	On Site Registration and Badge Pickup Poster and Software Demonstration Setup
9:00 am	**-**	**9:30 am**	Opening Comments
9:30 am	**-**	**10:15 am**	Keynote Address I
10:15 am	**-**	**10:30 am**	Break
10:30 am	**-**	**12:00 pm**	Scientific Talks Session I
12:00 pm	**-**	**1:00 pm**	Lunch (Stone Pine Plaza)
1:15 pm	**-**	**2:00 pm**	Keynote Address II
2:00 pm	**-**	**2:15 pm**	Break
2:15 pm	**-**	**3:45 pm**	Scientific Talks Session II
3:45 pm	**-**	**5:15 pm**	Poster Session / Software Demonstrations
5:15 pm	**-**	**5:30 pm**	Closing Presentation and Awards



KEYNOTE SPEAKERS

David Haussler, Ph.D.
U.C. Presidential Chair in Computer Science
University of California Santa Cruz

A WORKING DRAFT OF THE HUMAN GENOME

We discuss the bioinformatic challenges in creating and using the current public working draft of the human genome, and look at what lies ahead as the genome is finished and comparisons with other vertebrate genomes are made. Working with Francis Collins and the major public sequencing centers, an international group led by Eric Lander and John Sulston has produced the initial working draft of the genome (`http://www.ncbi.nlm.nih.gov/genome/central/`). Jim Kent at UCSC and David Kulp at Neomorphic, Inc., among many others, made substantial contributions to this effort (`http://genome.ucsc.edu/`). We look at the current state of this draft genome, discussing assembly and genefinding methods, and methods for mapping sequences from other vertebrates onto the human genome. It is our hope that this work will soon lead to a significantly better understanding the functional organization of our genome.

Richard Satava, M.D., F.A.C.S.
Professor of Surgery
Yale University

THE BIOINTELLIGENCE AGE: MEDICINE AFTER THE INFORMATION AGE

The Information Age is NOT the Future, the Information Age is the present. There is something else in the future. As a place holder, the name BioIntelligence Age is suggested. The Information Age is already a century old and the biotechnology revolution is over 40-50 years old. The future will belong to the interdisciplinary sciences that are emerging at the interface of traditional sciences. Thus, nanotechnology (physical and information sciences), embedded biosensors (biologic and physical sciences) and rational drug design (biologic and information sciences) are paving the way. The most complex of the new technologies will incorporate all three sciences - biologic, physical and information. An example is tissue engineering which is beginning to grow synthetic organs. Thus the future belongs to the interdisciplinary team of researchers, the hallmark of the BioIntelligence Age. They will create a world that is populated by transparent, microscopic, ubiquitous sensors that are networked together to change our current dumb and unconnected world into a smart and networked one.

Even as we struggle to understand these revolutionary changes, there are words of caution from respected scientists. By 2030 or 2040, computers will have the same computational power of a human brain, but will such systems be intelligent, have emotions or even be controllable by humans? As Bill Joy suggests in “Why the Future Doesn't Need Us”, genetics, nanotechnology and robotics will become self assembling and self maintaining - thus they may not need the human species which created them. Scientists must proactively consider the consequences of their progress before entering into a Faustian bargain. The future is bright, but we must enter the BioIntelligence Age with our eyes wide open.

ABSTRACT LIST

SCIENTIFIC TALKS SESSION I

A First Look at the tRNA and snoRNA Genes Sets in the Human & Arabidopsis Genomes
Todd M. Lowe

Relational Data Mining: Using Probabilistic Relational Models to Discover Patterns in Epidemiological Data
Lise Carol Getoor, Benjamin Taskar, Jeanne Rhee, Peter Small, and Daphne Koller

Blind Prediction of Protein Structure: What Have We Learnt and Where Do We Go?
Ram Samudrala, Yu Xia, and Michael Levitt

A Fluoroscopic X-Ray Registration Process for Three-Dimensional Surgical Navigation
Hamid Reza Abbasi, Sanaz Hariri, Shao Chin, Robert Grzeszczuk, Daniel Kim, Gary Steinberg, and Ramin Shahidi

Evaluation of Musculoskeletal Models Derived from Magnetic Resonance Images
Silvia S. Blemker, Allison S. Arnold, Deanna S. Asakawa, and Scott L. Delp

Mathematical Analysis of Ensemble Dynamics: Making Simulations of Protein Folding Feasible
Michael Randall Shirts, and Vijay S. Pande

SCIENTIFIC TALKS SESSION II

Discovering DNA Motifs in Upstream Regulatory Regions of Co-expressed Genes
Xiaole Liu, Jun Liu, and Doug Brutlag

Developing and Evaluating Assessment Measures From a Simulation Tool: Tinkering Towards Utopia?
Carla Marie Pugh

A One-Dimensional Finite Element Method For Simulation-Based Medical Planning For Cardiovascular Disease
Jing Wan, Brooke Steele, Thomas J.R. Hughes, and Charles A. Taylor

Finding Distinctive Expression Patterns in Microarray Data with Independent Components Analysis
Joshua Michael Stuart, Soumya Raychaudhuri, Xiaole Liu, and Russ Altman

Automated Quantification of 4D Ultrasound for Carotid Artery Disease
Haobo Xu, David S. Paik, Barbara Ross, Thilaka Sumanaweera, John Hossack, R. Brooke Jeffrey, and Sandy Napel

Designing a Knowledge Base for Pharmacogenomics: an Ontology for Genetic Information
Daniel Rubin, and Russ Altman

POSTER SESSION / SOFTWARE DEMONSTRATION

01 Project Mothra: Designing a System for Video-based, Markerless, Human Motion Analysis in an Arbitrary Environment
Ajit M Chaudhari, Richard W Bragg, Eugene J Alexander, and Thomas P Andriacchi

02 Using Metacomputing Tools To Facilitate Large-Scale Analyses of Biological Databases
Allison Waugh, Glenn A. Williams, and Russ B. Altman

03 Cluster Comparisons
Alok Saldanha

04 Prediction of Novel Functional Domains Using Rates of Evolution
Alexander Simon

05 Recognizing Polyps From 3D CT Colon Data
Salih Burak Gokturk, Burak Acar, David Paik, Carlo Tomasi, Christopher Beaulieu, and Sandy Napel

06 Folding of a Coarse-grained Model of the Tetrahymena Ribozyme
Bradley J. Nakatani, and Vijay S. Pande

07 Use of Multiple Clustering Algorithms for Analysis of Human Lung Cancer Gene Expression Data
Jessica Ross, and Glenn Rosen

08 In Vivo Validation of Cardiovascular Blood Flow Simulations
Joy Ku, Gregory Wai Mong Chan, Mary Draney, Frank Arko, Chris Zarins, and Charles Taylor

09 Computational Analyses of the Differences between Daily and Intermittent Alendronate Treatment
Christopher J. Hernandez, Gary S. Beaupre, Robert Marcus, and Dennis R. Carter

10 Mechanical Influences on Oblique Pseudarthrosis Formation
Elizabeth G. Loboa Polefka, Gary S. Beaupre, and Dennis R. Carter

11 Internal and Relative Structural Conservation of Discrete Protein Sequence Motifs
Steven Paul Bennett, and Douglas Brutlag

12 Constrained Global Optimization for Estimating Molecular Structure from Atomic Distances
Gl7enn A. Williams, Jonathan M. Dugan, and Russ B. Altman

13 Automated Individualized Decision Support
George Christopher Scott, Ross Shachter, and Leslie Lenert

14 A Comparative Statistical Error Analysis of Neuronavigation Systems in a Clinical Setting
Hamid Reza Abbasi, Sanaz Hariri, David Martin; Daniel Kim, John Adler, Gary Steinberg, and Ramin Shahidi

15 Neuronavigational Epilepsy Focus Mapping
Hamid Reza Abbasi, Sanaz Hariri, David Martin, Michael Risinger, and Gary Heit

16 Comparative Tracking Error Analysis of Five Different Optical Tracking Systems
Jeremy Johnson, Rasool Khadem, Clement C Yeh, Mohammad Sadeghi-Tehrani, Michael R Bax, Jacqueline Nerney Welch, Eric P Wilkinson, and Ramin Shahidi

17 Use of XML/RDF to Create Structured Metadata for Medical Images
John Joseph Michon

18 A Real-Time Freehand 3D Ultrasound System for Image Guided Surgery
Jacqueline Nerney Welch, Jeremy A. Johnson, Michael R. Bax, and Ramin Shahidi

19 Bridging the Gap: Simulated Dynamics of Lipid Bilayers at Boundaries
Peter M. Kasson, and Vijay S. Pande

20 KB-Driven Model Building: Challenges and Approaches
Mike Cantor, Peter Karp, and Masaru Tomita

21 MOTIFFEATURE: Automated Construction of 3D Models from Sequence Motifs
Mike Hsin-Ping Liang, and Russ Altman

22 Guideline Interchange Format: A Representation for Sharable, Computer-Interpretable Guidelines
Mor Peleg

23 A Finite Element Model of the Human Cornea
Assad Anshuman Oberai, Peter M. Pinsky, and Thomas A. Silvestrini

24 Mechanical Regulation of Growth Plate Morphology
Sandra Shefelbine, and Dennis R. Carter

25 The Importance of Swing Phase Initial Conditions in Stiff-knee Gait: A Case Study
Saryn Goldberg, Steven Piazza, and Scott Delp

26 A New Twist on the Helix-Coil Transition: A Non-biological Helix with Protein-like Intermediates
Sidney P. Elmer, and Vijay S. Pande

27 Sequence Analysis and Structure Comparison of the SH3 Domain Family
Stefan M. Larson, and Alan R. Davidson

28 Representing Contextually Changing Decision Making Behavior in Medical Organizations
Carol HF Cheng, and Raymond E Levitt

29 Medline Query-by-Example
Elmer Bernstam, Olga Troyanskaya, and Jeff Chang

30 Offline Testing of a Computerized Decision Support System for Management of Hypertension
Susana Martins, MK Goldstein, BB Hoffman, RW Coleman, SW Tu, R Shankar, M O'Connor, MA Musen, SB Martins, N Hastings

31 Comparison of Ribosomal Models to Experimental Data with the RiboWeb System
Michelle Whirl Carrillo, and Russ B. Altman

32 The Mouse SNP Database: Mapping QTLs in silico
Jonathan Usuka

33 A New Method for Determining Protein Function Similarity based on Keywords and Gene Ontology
Yueyi Liu, and Russ Altman

34 Optimizing Knowledge-based Energy Functions. From Lattice Study to Real Proteins
Yu Xia, and Michael Levitt

35 Monte Carlo Simulations of Folding of Simple Alpha Helices
Bojan Zagrovic, Jessica Shapiro, and Vijay Pande

36 Automatic Detection and Quantification of Abdominal Aortic Thrombus in CT Angiograms Based on Clustering and Global Geometric Information
Feng Zhuge, Sandy Napel, David Paik, and Geoffrey D. Rubin

37 Quantification of the Hydrophobic Interaction by Simulations of the Aggregation of Small, Hydrophobic Solutes in Water
Tanya M. Raschke, Jerry Tsai and Michael Levitt

38 ViewFeature: Integrated Feature Analysis and Visualization
D. Rey Banatao, Conrad C. Huang, Patricia C. Babbitt, Russ B. Altman, and Teri E. Klein

39 Using Human Language Ability to Learn and Recognize Protein Folds
Neil F. Abernethy

40 Structure and Stability of Collagen
Sean Mooney, Teri Klein

41 Combining Kinetic Inference to Extract Parameters and Predictor-Corrector Method to Develop Genetic Regulatory Circuits that are Consistent with Heterogeneous Experimental Data
Nizar Batada, Mike Laub, Harley McAdams

Demonstrations:

42D An Interactive Biomechanical Model of the Human Hand
Robert Pao-Feng Cheng, Jean Heegaard, Parvati Dev, Sakti Srivastava, Leroy Heinrichs, and Tonia Sengelin

43D Implementation of a Radio-Frequency Intravascular Ultrasound System for Quantitative Tissue Characterization in Coronary Arteries
Brian Courtney, Abel L. Robertson, Paul G. Yock, and Peter J. Fitzgerald

44D Two Sided Clustering for Yeast Gene Expression Using Probabilistic Relational Models
Eran Segal, Ben Taskar, and Daphne Koller

45D Web Applications for Microarray Data Analysis and Presentation
Christian A. Rees, Charles M. Perou, Douglas T. Ross, Jonathan R. Pollack, J. Michael Cherry, Patrick O. Brown, and David Botstein

46D IRaCS: A Literature Mining Tool for Fast Interpretation of Microarray Data
Sep Kamvar, Eldar Giladi, Jeanne Loring, and Mike Walker

SCIENTIFIC TALKS SESSION I

A FIRST LOOK AT THE tRNA AND snoRNA GENES SETS IN THE HUMAN & ARABIDOPSIS GENOMES

Todd M. Lowe

Introduction

Transfer RNA (tRNA) genes make up one of the largest gene families in all organisms. Taking into account known types of "wobble" base pairings between the third position of the mRNA codon & tRNA anticodon, eukaryotes require just 46 different tRNAs in theory. In reality, they have many, many more. Previous research has suggested some reasons why eukaryotes require such high tRNA redundancy. Now that we have complete gene sets for organisms from four diverse eukaryotic phyla, we will be able to directly address these hypotheses. Relationships between tRNA gene copy number, intracellular tRNA concentration, and protein codon usage are examined.

SnoRNA genes are probably the second largest gene family in eukaryotes. These non-coding RNAs are required for processing and modification of ribosomal RNA, each one pairing with a particular site of post-transcriptional modification. snoRNAs can be grouped into two families, the H/ACA box family which direct pseudouridylations of rRNA, and the C/D box family, which guides 2'-O-ribose methylation of rRNA. Approximately 100 genes are anticipated in yeast, and over 200 in humans, based on the number of modifications found in the respective ribosomal RNAs. In a previous study, I computationally identified C/D box snoRNAs for nearly all of the 55 ribose methylation sites in yeast. Now, I seek to do the same for the human genome, a much more challenging task due to the 215-fold increased search space.

Materials and Methods

tRNAscan-SE v. 1.21 was used to scan all completed eukaryotic genomes, and the results were manually inspected for low-scoring pseudogenes. All results were deposited in an on-line database of tRNAs, the Genomic tRNA Database (`http://rna.wustl.edu/GtRDB/`), which I maintain.

The probabilistic snoRNA scanning program was re-trained on biochemically identified human or plant snoRNAs, and used scan the human and Arabidopsis genomes. The results were manually inspected to identify the highest scoring candidate snoRNAs for each known or phylogenetically inferred methylation site in ribosomal RNA.

Both tRNAscan-SE and the snoRNA scanning programs were designed and implemented as part of my graduate thesis.

Results

How many more tRNA genes would you guess humans need relative to "lower" eukaryotes such as the roundworm C. elegans? If you guessed 2-3 times as many, you'd be wrong. The current draft of the human (~90% completed) contains approximately 39 fewer tRNAs than the worm (540 human vs. 579 worm). The "completed" fly genome contains just 11 more tRNAs than the single celled baker's yeast (285 fly vs. 274 yeast). As the first completed genome of a plant, Arabidopsis shows the most tRNAs in any organism to date: 614 genes. Contributing reasons for these somewhat unexpected results will be discussed. For example, Arabidopsis contains a nearly complete complement of its mitochondrial tRNAs within the nuclear genome. Also, an array of 81 tRNAs was found in just 40 Kbp in a highly amplified 3-tRNA repeat region.

SnoRNA searches of the Arabidopsis genome turned up over 60 new snoRNA gene predictions, in addition to the 21

previously identified genes. In many cases, these gene predictions were supported as they appear to be part of polycistronic arrays of multiple snoRNA genes. From these results, we can predict with some certainty the existence of 50+ corresponding ribose methylations in ribosomal RNA.

SnoRNA searches of the human genome turned up strong candidates for nearly all of the 100 known ribose methylation sites. These results are still currently being analyzed. As expected, most snoRNAs occur in multiple copies, spread widely across the genome in many cases. A collaborating experimental snoRNA lab is in the process of verifying these 55+ new gene candidates.

Web Page

`http://rna.wustl.edu/GtRDB/`

RELATIONAL DATA MINING: USING PROBABILISTIC RELATIONAL MODELS TO DISCOVER PATTERNS IN EPIDEMIOLOGICAL DATA

Lise Carol Getoor, Benjamin Taskar[*]*, Jeanne Rhee, Peter Small, Daphne Koller*

Biological data sets are often characterized by their rich relational structure. Such a data set might contain: demographic, clinical, and genomic information about patients; genomic and drug-resistance information about infectious agents; drug treatment history; and epidemiologic contact tracing for patients. Traditional approaches to statistical data analysis often have difficulty dealing with such complex structured datasets. Probabilistic relational models (PRMs) are a recent development that extend the standard attribute-based Bayesian network representation to incorporate a much richer relational structure. A PRM specifies a template for a probability distribution over a relational database. It specifies, for each type of entity in the domain a dependency model for each attribute in that table. This model encodes the way in which the attribute of an object in that table depends on other attributes, including those of related objects.

In our work, we have developed algorithms for learning PRMs directly from structured data. Our methods build on the work in learning Bayesian networks, and provide a powerful and flexible method for learning from relational data. Our algorithm takes as input a relational database and tries to detect the most significant direct correlations in the data. It performs a heuristic search over the space of possible dependency structures using a Bayesian scoring function.

We have applied our algorithm to a database of epidemiological data gathered at the San Francisco Tuberculosis Clinic (1991-1999), containing 1843 patients and their approximately 21,000 contacts. The database contains patient demographic and clinical attributes. Additionally, sputum samples are obtained for each patient, and undergo a genetic marker analysis determining the strain of Mtb that is causing disease in the patient. A contact investigation is performed for each patient to identify persons with whom the patient has been in contact during his/her infectious period. Data for each contact include the relationship of the contact to the case (e.g., family member, co-worker) and the contact's age.

The learned PRM contains rich dependency structure both within classes and between attributes in different classes (see `http://www-cs/~getoor/tb.ps`). The domain experts who developed the database found the model interesting, and most of the dependencies quite reasonable: the dependence of age at diagnosis on HIV status --- typically, HIV-positive patients are younger, and are infected with TB as a result of AIDS; the dependence of the contact's age on the type of contact --- contacts who are coworkers are likely to be older than contacts who are school friends. There are also dependencies that indicate a bias in the TB control procedures: contacts who were screened at the TB clinic were much more likely to be diagnosed with TB and receive treatment than those screened by their private medical doctor.

There are also correlations that are clearly relational, and that would have been difficult to detect using a non-relational learning algorithm. For example, there is a dependence between the patient's HIV result and whether he transmits the same strain to a contact: HIV positive patients are much more likely to transmit the disease. Another example is the correlation between the ethnicity of the patient and the number of patients infected by the strain: Asian patients are more likely to be infected with a strain

which is unique in the population, whereas other ethnicities more often have strains that recur in several patients. The reason is that Asian patients are more often immigrants, who arrive at the U.S. with a latent strain of TB, whereas other ethnicities are often infected locally.

Our learning algorithms for PRMs provide a powerful technique for discovering the statistical dependencies in a relational domain. These methods are particularly well suited to data encountered in many biomedical domains, where our goal is scientific discovery from a rich relational dataset.

BLIND PREDICTION OF PROTEIN STRUCTURE: WHAT HAVE WE LEARNT AND WHERE DO WE GO?

Ram Samudrala, Yu Xia, Michael Levitt

The Critical Assessment of protein Structure Prediction (CASP) methods conference was instigated to ensure that protein structure prediction approaches are tested rigorously without advance knowledge of the experimental answer. We have made predictions at all three CASP meetings, each time improving upon previously developed methodologies. In the recent CASP3, we made *ab initio* predictions based on a lattice-based exhaustive enumeration technique to sample protein conformational space, and an all-atom conditional probability discriminatory function to select native-like conformations. Using this approach, we were successfully predict the topology of small proteins, or fragments of a protein (up to ~60 residues), for more than 50% of the sequences modeled. The results represent a marked progress in bona fide *ab initio* prediction since the first CASP in 1994. We have taken the methodologies one step further by using predicted structure to predict function and guide experimental work for a 67-residue fragment of the DNA polymerase alpha-associated protein. A discussion on the utility of our approach for solving relevant biological problems will be presented, as well as the new approaches we have implemented at the fourth CASP, which recently ended.

A FLUOROSCOPIC X-RAY REGISTRATION PROCESS FOR THREE-DIMENSIONAL SURGICAL NAVIGATION

Hamid Reza Abbasi, Sanaz Hariri (CandMed), Shao Chin, Robert Grzeszczuk, Daniel Kim, Gary Steinberg, Ramin Shahidi

Back pain has a lifetime incidence of about 80% and is the 2nd leading reason why Americans see physicians. Causing suffering and stress, it costs as much as $50 billion a year for medical care, workers compensation payments, and time lost from work. Surgical procedures are performed to alleviate pain and neurological deficits; accurately placed transpedical screws may allow secure and reliable fixation of an unstable spine. However, often the surgeon has no direct visual guidance during the procedure. Alignment of the drill and the decision to proceed to a certain depth depends on the skill of surgeons. The surgeon must infer the 3D positions and dimensions of critical anatomic structures based on their relationship to exposed anatomical landmarks aided by 2D imaging data (e.g. plain films, fluoroscopy, and ultrasound). Studies show that incorrect placement of screws range from 10% to 40%. Neurologic complications due to imprecisely placed screws range from 1.5%% to 6%; inadequate biomechanical fixation is reported in up to 31% of cases.

Traditional cranial neuronavigation systems are inappropriate for use in spinal surgeries because the marker-to-bone relationship changes significantly in the spinal region from the time of preoperative CT image collection to the time of intraoperative marker registration. There is thus a need for intraoperative 3D real-time visualization of spinal anatomy. To fill this technological gap while addressing the unique constraints of spinal anatomy, the Stanford Image Guidance Laboratory (IGL) has developed a surgical navigation registration algorithm to allow the surgeon to precisely locate surgical tools with respect to the patient's anatomy during spinal surgeries using a computed tomography (CT) scanner pre-operatively and a C-arm fluoroscope intra-operatively. Use of this algorithm involves three steps: calibration, tracking, and registration.

The IGL spinal registration algorithm (SRA) addresses the problem of fine registering such a dynamic region (course registration being similar to the cranial registration). The SRA uses the original 2D axial planes of the CT scans to create a 3D reconstructed image of the patient. The SRA can now act as a virtual fluoroscope, obtaining virtual 2D fluoroscopic images from this 3D reconstruction in any plane (e.g. lateral, AP). These virtual fluoroscopic images are called digitally reconstructed radiographs (DRRs). Intra-operatively, two real oblique fluoroscopic images of the patient are obtained. The SRA matches the 2 real fluoroscopic images with the 2 DRRs. This match enables the navigation system to assign spatial positions acquired from preoperative CT images to the actual anatomical position of the patient through the following logic: a) The relationship of the camera to the real fluoroscope is known (the tracking step). b) The relationship of the real fluoroscope (i.e. OI) to the CT is known through the virtual fluoroscope (fine registration). The fine registration step is repeated each time the patient is moved and each time a large piece of equipment is moved in the OR (since this changes the magnetic field and thus changes the image).

IGL is currently in the process of testing the algorithm's accuracy using a phantom patient vertebra. Ultimately, utilization of a spinal navigation system with this noninvasive registration method will provide greater surgical precision in spine procedures, especially in more sensitive

anatomic areas such as the cervical and upper thoracic spine.

EVALUATION OF MUSCULOSKELETAL MODELS DERIVED FROM MAGNETIC RESONANCE IMAGES

Silvia S. Blemker, Allison S. Arnold, Deanna S. Asakawa, Scott L. Delp

Introduction

The medial hamstrings and psoas muscles are frequently lengthened surgically in an attempt to improve walking in children with cerebral palsy. Previous studies have suggested that analysis of muscle lengths during gait may be helpful in deciding when a muscle should be surgically lengthened (Hoffinger *et al*. 1993, Delp *et al*. 1996). These studies have relied on a computer model of the lower limb that represents the musculoskeletal geometry of an average-sized adult male. It is not clear how variations in subject size or the presence of musculoskeletal deformities may affect the accuracy of the muscle lengths estimated using the average-sized model. Therefore, techniques to accurately and non-invasively characterize muscle lengths of individual subjects must be developed to test the results of previous simulation studies.

The goals of this study were to: (i) develop methods to construct subject-specific biomechanical models from magnetic resonance (MR) images, (ii) create models of three lower extremity cadaver specimens, and (iii) test the accuracy of muscle lengths and moment arms estimated using these models. To test the accuracy of the models, the hip and knee flexion moment arms estimated from models of the three specimens were compared to the moment arms determined experimentally on the same specimens. Because a muscle's moment arm determines its change in length with joint rotation, these comparisons also tested the accuracy with which the models could estimate muscle lengths over a range of hip and knee motions.

Methods

Models of three lower limb cadaver specimens were constructed from six series of T1-weighted spin-echo images (Fig. 1). Boundaries of the bones and muscles were outlined manually in the two-dimensional (2D) images (Fig. 1-A), and 3D surface models were created from these boundaries for each series. The surfaces from overlapping series were registered (Fig. 1-B), generating an accurate representation of the musculoskeletal anatomy at a single limb position (Fig. 1-C). To estimate the muscle moment arms for a range of limb positions, models of hip and knee kinematics were scaled to the specimens' bones (Fig. 2-A). The hip was assumed to be a ball-and-socket joint, and the hip center was estimated by fitting a sphere to the femoral head using a nonlinear least-squares algorithm. The knee model was based on published 3D measurements of tibiofemoral kinematics (Walker *et al*. 1988, Nisell *et al*. 1986). The musculotendon paths were derived from the 3D muscle surfaces (Fig. 2-B), and ellipsoidal wrapping surfaces were defined for each muscle to simulate wrapping over underlying structures (Fig 2-C).

The moment arms estimated from the models were compared to the moment arms determined experimentally on the same specimens using the tendon displacement method (An *et al.* 1984). The specimens were mounted in a jig that provided control of hip flexion, adduction, rotation, and knee flexion. Joint angles were monitored by tracking the locations of infrared emitters that were fixed to the bones. For each muscle, a wire was connected to the tendon, routed through a suture anchor at the muscle origin, and attached to a position transducer. Fourth order polynomials were fit to the tendon excursion vs. flexion data, and the hip and knee flexion moment arms were obtained from the first derivatives of the polynomial fits averaged over multiple trials.

Results

The moment arms estimated from the models compared favorably with the experimental data (Figs. 3 and 4). For the psoas (Fig. 3-A), the average errors between the experimentally determined hip flexion moment arms and the calculated moment arms ranged from 1.1 mm, or 5% of the experimental moment arms, to 2.7 mm (8%). For the medial hamstrings (Fig. 3-B), the hip extension moment arm errors ranged from 1.0 mm (2%) to 3.8 mm (9%). The average knee flexion moment arm errors for the medial hamstrings (Fig. 3-C) ranged from 0.1 mm (<1%) to 3.9 mm (9%). We also determined that the models could accurately estimate muscle lengths during walking (not shown, Arnold *et al*., 2000a).

Conclusion

Generic musculoskeletal models that compute the lengths and moment arms of soft tissues have been used to study the treatment of wide range of movement abnormalities and to plan orthopaedic surgical procedures. However, before any generic model can be used to guide patient-specific treatment decisions, the accuracy of the model must be tested. This study demonstrates that the combination of MR imaging and graphics-based musculoskeletal modeling is a promising approach for accurately estimating muscle lengths and moment arms in vivo (errors within 10%). Using the methods presented in this study, MR-based models of children with cerebral palsy have been developed and used to examine the causes of movement abnormalities (Arnold *et al*., 2000b).

References

1. An KN, Takahashi K, Harrigan TP, Chao EY: Determination of muscle orientations and moment arms. J. Biomech. Eng. 106: 280-282, 1984.
2. Arnold AS, Asakawa DJ, Delp SL: Do the hamstrings and adductors contribute to excessive internal rotation of the hip in persons with cerebral palsy. Gait and Posture, Awarded Best Paper of 1999, vol. 11, pp. 181-190, 2000a.
3. Arnold, AS, Salinas S, Asakawa DJ, Delp SL: Accuracy of muscle moment arms estimated form MRI-based musculoskeletal models of the lower extremity. Computer Aided Surgery,. 5: 108-119, 2000b.
4. Hoffinger SA, Rab GT, Abou-Ghaida H: Hamstrings in cerebral palsy crouch gait. J. Pediatr. Orthop. 13: 722-726, 1993.
5. Nisell R, Nemeth G, Ohlsen H: Joint forces in extension of the knee. Acta. Orthop. Scand. 57: 41-46, 1986.
6. Walker PS, Rovic JS, Robertson DD: The effects of knee brace hinge design and placement on joint mechanics. J. Biomech. 21: 965-974, 1988.

Acknowledgements

We are grateful to JoAnn Mason, Erik King, Mahi Durbhakula, Norman Fung, and Emil Davchev. Funded by NIH, NSF, and the United Cerebral Palsy Foundation.

MATHEMATICAL ANALYSIS OF ENSEMBLE DYNAMICS: MAKING SIMULATIONS OF PROTEIN FOLDING FEASIBLE

Michael Randall Shirts, Vijay S. Pande

Ensemble dynamics is a new methodology for extending simulations to very long time scales by efficiently parallelizing the calculation among many machines. Ensemble dynamics is shown to scale nearly as fast as the number of processors in many physical situations, rendering previously intractable problems within reach of large computer clusters. Interestingly, it is possible to obtain speedup greater than the number of processors under some conditions, although other systems are limited to sublinear speedup. One of the most important applications of this new work is the computer simulation of protein folding, and small peptides have very recently been folded within our group at Stanford using this method.

Ensemble dynamics is suited for problems such as molecular dynamics simulations of condensed phase systems which are characterized by infrequent crossings of energy or free-energy barriers alternating with long persistence times in energy or free-energy minima. An ensemble dynamics simulation consists of M simulations running in parallel. When one of the M simulations crosses an energy barrier, all of the M simulations are reset to (or near) the point of phase space of this barrier-crossing simulation, and the M simulations are continued from there.

We present a formalism for calculating the computational advantage (speedup using M processors) as well as interpreting the simulation data to predict rates. A heuristic argument has previously been presented demonstrating that with exponential rate processes, ensemble dynamics should give an exactly linear speedup of rates with number of processors, and that the distribution of different pathway frequencies is preserved. We find that tremendous speedups can be obtained, rendering previously intractable problems within reach.

Ensemble dynamics is a powerful technique which offers highly scaleable speedup of simulations. In many physical cases, superlinear speedup can be obtained, yielding effective efficiency greater than 100%. Deviations from linearity are representative of physical trajectories that are faster or slower than the usual trajectory, but are still physical trajectories. Under the right conditions, an ensemble dynamics simulation can ignore traps and proceed in a more direct manner along the productive reaction pathway. This method should be highly effective for the simulation of long time scales by use of hundreds or even thousands of computers, limited only by the ratio of typical simulation times to the fastest simulation times. Consider again the simulation of protein folding dynamics. While the fastest proteins fold in ~10 microseconds, a single CPU can only simulate 1 ns/day, thus requiring about three CPU years. With a 1000 processor cluster, and ensemble dynamics, once can simulate 1 microsecond/day, rendering the problem tractable.

Web Page

`http://foldingathome.stanford.edu`

SCIENTIFIC TALKS SESSION II

DISCOVERING DNA MOTIFS IN UPSTREAM REGULATORY REGIONS OF CO-EXPRESSED GENES

Xiaole Liu, Jun Liu, Doug Brutlag

The development of genome sequencing and DNA microarray analysis of gene expression gives rise to the demand for data-mining tools. BioProspector, a C program using a Gibbs sampling strategy, examines the upstream region of genes in the same gene expression pattern group and looks for regulatory sequence motifs. BioProspector uses zero to third-order Markov background models whose parameters are either given by the user or estimated from a specified sequence file. The significance of each motif found is judged based on a motif score distribution estimated by a Monte Carlo method. In addition, BioProspector modifies the motif model used in the earlier Gibbs samplers to allow for the modeling of gapped motifs and motifs with palindromic patterns. All these modifications greatly improve the performance of the program. Although testing and development are still in progress, the program has shown preliminary success in finding the binding motifs for *Saccharomyces cerevisiae* RAP1, *Bacillus subtilis* RNA polymerase, and *Escherichia coli* CRP. We are currently working on combining BioProspector with a clustering program to explore gene expression networks and regulatory mechanisms.

DEVELOPING AND EVALUATING ASSESSMENT MEASURES FROM A SIMULATION TOOL: TINKERING TOWARDS UTOPIA?

Carla Marie Pugh

Purpose

With the advent of simulation technology, the medical profession can expect significant changes in the ability to train heath care professionals. From surgical procedures to basic physical exam skills, simulation and virtual reality technology bring the promise of a new era for medical education. However, for quality assurance purposes, these new teaching tools must be evaluated. The purpose of this study is to evaluate the assessment measures developed from the pelvic exam simulator.

Materials and Methods

We have designed a new teaching tool, the pelvic exam simulator, which consists of a partial manikin - umbilicus to mid thigh - constructed in the likeness of an adult human female. The device is instrumented internally with several electronic sensors and has the ability to provide the user with immediate visual feedback regarding performance. While the user is performing an exam, sensor inputs are sampled at a rate of 30 hertz and the outputs are captured and stored in a data file. Figure 1 depicts a sample of the data generated from the simulator.

`http://www.stanford.edu/~cpugh/BCATS.html`

Because the sensor data represents information that has never been collected while performing clinical pelvic exams, we developed a method of analyzing the data. Our purpose was to extract meaningful indicators of student performance from large data files containing the electrical signals captured during the exam.

The variables developed from the sensor data include: 1) length of time required to perform a complete exam, 2) number of pressure points touched during the exam, 3) the frequency at which a given pressure point was touched, and 4) the maximum amount of pressure used while touching each pressure point. These variables were used as indicators, or measures of student performance.

To better understand the variables we created, we conducted a controlled randomized study using eighty-seven medical students. The study protocol consisted of a training phase, and an assessment phase. Only the treatment group (33 students) trained on the simulator. During the assessment phase, all students performed clinical pelvic exams on three different simulators and sensor data was collected to generate the variables discussed above.

After examining each simulator, the students filled out assessment forms regarding their exam findings. These forms were evaluated and an accuracy variable was created. As part of our data analysis, the accuracy variable was correlated with the simulator variables to determine if the variables we created were significant indicators of student performance.

Results

Pearson's correlations showed that the student accuracy variable was significantly correlated with the pressure point and maximum pressure variables, $p < .05$, establishing the validity of these two measures as indictors of student performance. The frequency variable only achieved a moderate correlation with student accuracy, $p = .056$. Time did not correlate with accuracy. The results also showed statistically significant inter-item correlations for the time, pressure point and the maximum pressure variables, $p < .01$,

further establishing the potential use of these variables as measures of student performance. The reliability coefficients for the simulator variables are as follows: Time = .7240, Pressure Points = .6329, Maximum Pressure = .7701, Frequency = .5011, and Accuracy = .6007.

Conclusion

We have developed a method of analyzing raw sensor data for the purposes of generating valid measures of student performance. Although two of the variables created seem to be valid measures of student performance, more studies need to be done.

Acknowledgements

I wish to acknowledge the following people for their guidance and support in this research project - Sakti Srivastava, M.D., Richard Shavelson, Ph.D., Decker Walker, Ph.D., Teresa Cotner, Ph.D., Beth Scarloss, MS, Merry Kuo, MS, Chantal Rawn, BS, Parvati Dev, Ph.D., Thomas H. Krummel, M.D., and Leroy H. Heinrichs, M.D., Ph.D.

A ONE-DIMENSIONAL FINITE ELEMENT METHOD FOR SIMULATION-BASED MEDICAL PLANNING FOR CARDIOVASCULAR DISEASE

Jing Wan, Brooke Steele, Thomas J.R. Hughes, Charles A. Taylor

Purpose

Current methods for vascular treatment planning rely on diagnostic and empirical data to guide the decision-making process. This approach does not enable a physician to preoperatively assess the efficacy of alternate therapies in determining the preferred treatment for an individual. We have previously described a new approach to planning treatments for cardiovascular disease, Simulation-Based Medical Planning, whereby a physician utilizes computational tools to construct and evaluate a combined anatomic/physiologic model to predict the outcome of alternative treatment plans for an individual patient. Current systems for Simulation-Based Medical Planning utilize finite element methods to solve the time-dependent, three-dimensional equations governing blood flow and provide detailed data on blood flow distribution, pressure gradients and locations of flow recirculation, low wall shear stress and high particle residence. However, these methods are computationally expensive and often require hours of time on parallel computers.

Materials and Methods

We describe, herein, a system for Simulation-Based Medical Planning based on the solution of the one-dimensional equations of blood flow using a space-time finite element method. We applied Galerkin/Least Square stabilization method in space and Discontinous Galerkin in time, which has been proven to be stable and robust.

Results

This system is applied to compute, flow rate and pressure in a single segment model, an idealized model of the abdominal aorta, in three alternate treatment plans for a case of aorto-iliac occlusive disease and in a vascular bypass graft. We demonstrate that, based on flow rate, this method can be used to rank treatments in the same order as our fully three-dimensional method.

Conclusion

Compared with three-dimensional method, one-dimensional method has the advantage of low computational cost. Although it cannot give the details, such as flow recirculation, local blood flow distribution, it can still give fairly accurate flow rate and pressure distribution along different pass. We also proved that for cases of vascular treatment planning, one-dimensional model can be used to rank treatments in the same order as our fully three-dimensional method. Further work still needs to be done to precisely calibrate the role of one-dimensional model in vascular treatment planning.

FINDING DISTINCTIVE EXPRESSION PATTERNS IN MICROARRAY DATA WITH INDEPENDENT COMPONENTS ANALYSIS

Joshua Michael Stuart, Soumya Raychaudhuri, Xiaole Liu, Russ B Altman

We introduce the application of Independent Components Analysis (ICA) for identifying distinctive expression profiles within microarray data. Instead of clustering, we use ICA to find axes of the data that distinguish a small number of genes whose expression profiles are similar to each other while dissimilar to the entire remaining set of profiles. Each collection of outlier profiles along one axis determined by ICA is naturally contrasted to the set of profiles on the opposite end of the same axis. For each outlier cluster, an "anti-cluster" is also defined. The upstream sequences belonging to a cluster of outliers can be searched for the presence of common regulatory sites. In addition, sequences from the anti-cluster can also be used as negative examples during motif identification. We provide an example where such additional information significantly increases the performance of a motif-finding algorithm. The performance of the method is demonstrated on a well-studied yeast sporulation data set. Applied to this data, ICA successfully rediscovers the MSE regulatory element known to play a role in sporulation while also increasing the number of genes known to contain such sites.

AUTOMATED QUANTIFICATION OF 4D ULTRASOUND FOR CAROTID ARTERY DISEASE

Haobo Xu, David S. Paik, Barbara Ross, Thilaka S. Sumanaweera, John Hossack, R. Brooke Jeffrey, Sandy Napel

Purpose

To develop and test the feasibility of a technique for automatic categorization carotid artery disease into those patients who are normal and those who require more definitive tests and, possibly, surgery, from a rapidly-acquired four-dimensional ultrasound examination.

Materials and Methods

3D ultrasound data (B-mode and color Doppler energy) were collected with an Acuson Sequoia 512 using a modified linear array transducer, which was translated along the elevation direction to acquire 3D ultrasonic data sets. We acquired, in addition, the electrocardiogram (ECG) of the subject and automatically annotated each acquired image with its cardiac phase. Speckle data between successive images was analyzed using a computer-based tracking technique to accurately position the successive 2D image planes in 3D space. We then used the ECG phase to parcel the images into ten separate 3D volumes, each with nearly constant ECG phase. A computer program automatically determined the medial centerline of the common/internal carotid artery from the most systolic phase volume, and the cross-sectional area of images perpendicular to this centerline were plotted. From these area vs. distance plots we determined the maximum percent stenosis for 8 assumed-normal volunteers, and 8 patients (5 M, 3 F ages 52-76; mean=66), each of whom had a comparative study (MRA:4 angio:4). We classified the results of the ultrasound and comparative studies into 4 grades (1:<30%, 2:31%-60%; 3:61%-99%; 4:occluded).

Results

4D ultrasound acquisition times averaged 12 minutes per subject. All 10 subjects with <30% stenosis (either assumed-normal volunteers or from comparative examinations) were correctly identified by ultrasound, as were two patients with complete occlusions. In addition, 2 of the 3 grade 3 patients were correctly identified by ultrasound; one was underestimated as a grade 2. The patient with a grade 2 stenosis was overestimated by ultrasound as a grade 3.

Conclusion

It is feasible to acquire and process 4D ultrasound data of the extracranial carotid artery to compute cross-sectional area stenoses. Examination time is approximately 1/4 of what is required for a conventional bilateral duplex ultrasound examination. In this preliminary study, all normal patients and those with complete occlusions were correctly identified. Although there were some misclassifications between grades 2 and 3, all patients with mild-to-significant disease were correctly separated from the normals. This rapid, operator-independent technique shows promise for identifying patients with carotid disease, though further study in asymptomatic populations is required.

DESIGNING A KNOWLEDGE BASE FOR PHARMACOGENOMICS: AN ONTOLOGY FOR GENETIC INFORMATION

Daniel Rubin, Russ Altman

With the human genome now nearly completely sequenced, attention is focusing on learning the medical significance of this genetic information. Large-scale studies in pharmacogenetics are being done to find variations in genotype in order to understand the variability in drug response among individuals. But to make sense of this information, computational tools capable of efficiently accessing and analyzing these data are needed. Genetic data are complex, and simply storing raw sequences in a relational database will be inadequate to answer the complex queries that are needed to discover the links between genotype to phenotype. We need to represent the varied features of genetic sequences and their genomic structure to allow a broad range of queries that are needed to analyze pharmacogenetics data. Ontologies specify the concepts and relationships in a given field, and they provide a means of modeling a complex domain. In this study, we developed an ontology for genetic information to represent genes, alleles, sequences, genomic structure, polymorphisms, and their relationships. We implemented this ontology in Protégé-2000, an environment for developing ontologies and knowledge bases. We tested our model by representing genetic data obtained from a research center that is actively collecting genetic data for a pharmacogenetics study. We were able to store all the data collected for a single gene, and we could reconstruct various views of the data, similar to those the study center currently constructs by hand. We believe our ontology is a rich yet flexible model of genetic information, and may be suitable for storing data and supporting queries in pharmacogenetics studies.

POSTER SESSION / SOFTWARE DEMONSTRATIONS

PROJECT MOTHRA: DESIGNING A SYSTEM FOR VIDEO-BASED, MARKERLESS, HUMAN MOTION ANALYSIS IN AN ARBITRARY ENVIRONMENT

Ajit M Chaudhari, Richard W Bragg, Eugene J Alexander, Thomas P Andriacchi

Purpose

While current approaches to motion analysis are valuable to obtain sensitive, quantitative, kinematic and kinetic measurements, most are limited to a laboratory environment in which a subject must perform activities on a force plate while wearing reflective markers fixed to relevant anatomical landmarks. These restrictions can affect a subject's comfort, which may result in data that differs from natural motion. The laboratory environment also makes it difficult to accurately reproduce sports activities. Clearly, there is a need for a method to obtain both kinematic and kinetic quantities for human motion in an arbitrary setting without markers or force plates. The purpose of Project Mothra is to achieve this goal.

Methodology

A logical approach to calculate the forces and torques associated with human motion obtained from video data is to match a subject-specific, 3-dimensional model to the recorded motion. If inertial properties are then associated with each limb segment in the model, it is possible to use inverse dynamics to calculate kinetic quantities without a force plate. Currently Project Mothra has three distinct areas of development: 1) an application for building 3-dimensional, subject-specific models, 2) a model-based visual tracking algorithm, and 3) a dynamics engine to calculate the kinematics and kinetics of the observed motion. This poster focuses on the first two areas of development, while the third is left as a long-term goal.

Model Builder

The model-building application will be implemented as a GUI that allows the user to create a 3-dimensional subject-specific model by attaching body segments to one another with idealized joints. Body segments will be chosen from a library ranging in detail from geometric primitives to realistic segments reconstructed from human subjects, then scaled and positioned to match the subject. Inertial properties will be associated with each segment, either by using the constant density assumption over the segment volume or by heuristic values published in the literature. Body joints will enforce constraints on the relative positions of the connected segments, and will range in complexity from a basic hinge joint to a joint that models the complexities of an articulating surface. Once built, the model will be used by: 1) the tracking algorithm to obtain the kinematic data, and 2) the inverse dynamics engine to calculate the kinetic data.

Tracking Algorithm

Markerless visual tracking techniques will be used to obtain kinematic data for the motion of the body segments. While most studies will only be interested in kinetic data for one or a small subset of the joints, inverse dynamics calculations necessitate tracking the motion of every segment of the body simultaneously. Therefore, a multi-mode model will be used to track the most critical segments most accurately, while saving processing time for the less critical segments. For all segments, the tracking problem will be solved by matching a model, created by the model-building application, to the images seen by multiple cameras. Subjects will initially wear colored tights, where each non-critical segment will be a unique solid color to easily identify it and determine its orientation and position. In

contrast, each critical segment will have many smaller, non-repeating patterns on it that can be tracked individually as points on the segment. Once these points are identified, the motion of the segment can be determined with much greater accuracy using an existing algorithm such as the point-cluster technique (Andriacchi *et al*, 1998). By using this approach, it should be possible to obtain enough information to calculate the kinematics of motion while minimizing processing time.

Summary

A framework is presented for a system to acquire motion data in an arbitrary setting without the use of markers or force plates. A model-building application and a tracking algorithm are proposed as key elements.

USING METACOMPUTING TOOLS TO FACILITATE LARGE-SCALE ANALYSES OF BIOLOGICAL DATABASES

Allison Waugh, Glenn A. Williams, Russ B. Altman

Given the high rate at which biological data are being collected and made public, it is essential that computational tools be developed that are capable of efficiently accessing and analyzing these data. High-performance distributed computing resources can play a key role in enabling large-scale analyses of biological databases. I will discuss using a distributed computing environment, Legion, to enable large-scale computations on the Protein Data Bank (PDB). In particular, we employed the Feature program to scan all protein structures in the PDB in search for unrecognized potential cation binding sites. I will talk about the efficiency of Legion's parallel execution capabilities and report on the initial biological implications that result from having a site annotation scan of the entire PDB. Four interesting proteins with unannotated, high-scoring candidate cation binding sites will be highlighted.

02

CLUSTER COMPARISONS

Alok Saldanha

Purpose

There are many available clustering methods, but very few ways to compare the resulting clusterings. I have implemented a visualization scheme which allows easy identification of the major differences between several clusterings.

Materials and Methods

My visualization is a web-based application which allows the user to select several clusterings and/or gene lists and displays the membership of the clusters and lists in a visual fashion. As a test case, breast cancer expression data were clustered using many methods and with multiple gene lists to determine which clusters were robust under various perturbations.

Results

Several clusters of breast cancer subtypes remained under various clusterings, and some were split. This correlated well with subtypes that were known to be heterogeneous. These heterogeous groups sometimes appear as outliers of a more stable group. Finally, the experiment clustering under different gene lists was found to be very robust, although intentional selection of a gene list which was expressed only in epithelial cells radically changed the clustering to reflect epithelial/non-epthelia origin instead of tumor subtype.

Conclusion

This method is a good way of getting a qualitative idea of how clusterings compare, and can give one a sense of which clusters are robust to different methods of clustering. However, it does not give one a quantitative measure of cluster robustness, and can be difficult to interpret.

Web Page

`http://gort.stanford.edu:8000/alok/presentations/retreat2000/`

03

PREDICTION OF NOVEL FUNCTIONAL DOMAINS USING RATES OF EVOLUTION

Alexander Simon

I have developed a method to predict putative functional regions and/or structural domains of a protein and rank their relative importance. It is based on the principle that the rate of evolution of a domain is inversely related to the strength of functional constraints on it. Thus, slowly evolving regions are more likely to be functionally important than regions that accumulate more amino acid substitutions. My technique does not require any information other than a multiple sequence alignment. Rates of evolution are calculated in a moving window along an alignment using the maximum likelihood phylogenetic tree of a gene family and plotted as a function of sequence position. Regions that are evolving more slowly are candidates for being "functional domains." The method has been validated with several functionally diverse protein families. Its predictions are accurate and consistent with known biological information.

For example, in the Notch – Delta/Serrate protein families, a receptor-ligand system important in cell fate specification, my method correctly predicts 70/71 known domains. The Notch receptor has 36 tandem EGF-like repeats in its extracellular domain, which are thought to bind multiple ligands. Repeats #10-12 are required to bind the two known ligands, Delta and Serrate, however they are less evolutionarily constrained than repeat #26. The conservation of a glycosylation consensus site in this repeat that is potentially modified by Fringe, a glycosyltransferase which modulates Notch signaling, as well as Abruptex mutations which cluster near this repeat, support the prediction that it is a critical functional domain.

My analysis also reveals two uncharacterized domains present in both Delta and Serrate that are as constrained as the DSL domain, which is reported to be necessary for Notch interaction. Furthermore, these two domains are part of a large N-terminal region of Delta and Serrate that exhibits a strikingly conserved pattern of evolutionary rates. Identification of other such signatures may allow the detection of non-orthologous proteins that are functionally similar.

04

RECOGNIZING POLYPS FROM 3D CT COLON DATA

Salih Burak Gokturk, Burak Acar, David Paik, Carlo Tomasi, Christopher Beaulieu, Sandy Napel

Colon cancer is the second fatal cancer type in USA. Early diagnosis is crucial for a successful treatment. The common method is fiber-optic colonoscopy, which is invasive, time consuming and does not allow re-assessment. A non-invasive, fast and sensitive method, which is also suitable for screening, is required. It should also be able to assess the information in a high dimensional feature space. The method proposed attempts to meet these criteria for CT Colonoscopy.

Since the first medical diagnosis systems, the basic questions have been: How to (i) represent the observed data, (ii) incorporate the medical know-how, (iii) deal with the inaccuracy and uncertainty of the observed data?

The expert systems in radiology generally rely on subjective findings of a radiologist. Besides their poor reproducibility, the amount of data rapidly increasing with technological advances is not manageable with such an approach. The Bayesian approach has been the most accepted method with a sound mathematical justification. However, in this approach every single data point contributes to the decision function, irrespective of its information content. The idea of identifying and using the data points that carry the relevant information, thus focusing on the construction of the classifier itself, is utilized by Support Vector Machines (SVM).

SVM, originally proposed by Vapnik [1], constructs a classifying surface that minimizes the training error and maximizes the generalization capability of the classifier [2]. It determines the data points (Support Vectors, SV) that are closest to such a surface and defines it using the SVs only. The smaller the number of SVs, the more generalizable the classifier. A trade-off between the generalization capability and the classification performance on the training set can easily be established via a single parameter. SVM is not a clustering algorithm; it constructs the classifying surface itself, which can also be used to learn the characteristics of polyps.

Furthermore, the dimension of the classification domain can be increased indefinitely without adding much computational cost because SVM uses a kernel function in the observation space (much lower dimension) to perform inner product in the classification space. The kernel can be designed to enhance the discrimination between polyps and non-polyp structures.

In this study, the 3D CTC data is preprocessed [3] and the subvolumes at the candidate polyp locations are extracted. Each volume is sampled with 700 random slices on which four parameters are measured: (i) Distance to closest fitting circle, (ii) Distance to closest fitting line, (iii) 2nd order moment and (iv) 3rd order moment. The random slicing eliminates any possible bias due to the volume orientation and position. A histogram with 10 bins is created over all slices for each parameter. The 40 dimensional feature vector composed of these histograms for each polyp candidate is used as the input to SVM. There are 8 true polyps and 34 non-polyps in the data set. Three artificial polyp data are created by applying a small perturbation to each one of the true polyps. We use the exponential radial basis function as kernel.

The preliminary results presented demonstrate the capability of SVM in constructing classifying surfaces even in the

case of inseparable data sets and in utilizing high dimensional feature spaces. Selecting the features and the kernel is the key issue in SVM applications.

The future work includes: (i) Designing polyp descriptors and SVM kernels. ii) Using a much more significant training population. (iii) Clinical evaluation. (iv) Clinical interpretation of the classifying surface.

References

1. Vapnik V: The Nature of Statistical Learning Theory, New York, Springer-Verlag, 1995
2. http://www.support-vector.net
3. Gokturk SB, Tomasi C: A graph method for the conservative detection of polyps in the colon. 2nd Inter Symp on Virtual Colonoscopy, Boston, October 2000

FOLDING OF A COARSE-GRAINED MODEL OF THE TETRAHYMENA RIBOZYME

Bradley J. Nakatani, Vijay S. Pande

While protein folding remains one of the most intensely-studied problems in computational biology, relatively little work is being done in the related field of RNA folding. RNA folding is complicated by the large size of the polymers and the increased role of electrostatics, but at the same time simplified by the lower number of constituent monomers (4 different nucleotides in RNA compared to 20 naturally-occurring amino acids in proteins) and the primarily hierarchical folding process (secondary structure → tertiary structure). We have focused our studies on the Tetrahymena ribozyme, one of the few RNA molecules for which a three dimensional model and an abundance of experimental data exists. As a first step in the study of the folding process, we have chosen to forego a full atomic representation of the branched polymer in favor of a coarse-grained approach. In our model, we have reduced the 421-nucleotide ribozyme (8000+ atoms) to just 36 spheres, where each sphere typically represents 5-6 base-paired nucleotides in the model structure (see figure). We have simulated the folding of the ribozyme using Langevin dynamics.

In spite of the relative simplicity of our model, we have been able to reproduce many of the experimentally observed results including (1) the early collapse to a compact state, followed by a conformational search for the correct topology and (2) the existence of folding intermediates and misfolded kinetic "traps". From our simulation results, we have also been able to characterize the folding pathways and the nature of on and off-pathway intermediates. These initial findings provide an important basis for further study in this relatively unexplored field.

06

USE OF MULTIPLE CLUSTERING ALGORITHMS FOR ANALYSIS OF HUMAN LUNG CANCER GENE EXPRESSION DATA

Jessica Ross, Glenn Rosen

Purpose

The use of tools for researchers to analyze global gene expression, such as the DNA microarray, generate large amounts of biological data. Computers are necessary to analyze experimental output in order to make sense of this mass of data. Classic biological clustering methods are being assessed to determine utilities towards detecting non-random patterns from large data sets. Furthermore, modified versions of the classical methods of cluster analysis, as well as new algorithms based on mathematical techniques used for identifying patterns in complex data in a variety of other fields, are being applied to the problem of interpreting gene expression data generated from high throughput systems

Initial experiments predominantly generated data comprised from temporal expression patterns in manipulated cell lines. Algorithms used in clustering these data are based on similarities in gene expression patterns over time in response to a stimulus. Later, experiments that compare two different tissue populations in vivo were performed. Unlike time course experiments, analyses of these data is complicated by the fact that it is not possible to synchronize cells in a population, and there are multiple cells types in each tissue sample.

Analyzing data of this type is a new challenge, and there is no consensus as to the optimal method of analysis at this point in time. Furthermore, there have been no studies in the literature, which compare clusters derived from different algorithms on a given set of data. Such a study might be helpful in determining both the strengths and weaknesses of groups of algorithms with respect to these types of data.

Materials and Methods

This study will attempt to qualify groups of clustering algorithms in terms of their effectiveness in determining underlying patterns of gene expression in these type of data sets. We compare differential gene expression data from DNA microarray analysis of both tumor specimens and uninvolved lung tissue procured from patients at the time of surgery. Presently, twelve lung cancer specimens have been obtained, and we anticipate reaching a total sample size of at least one hundred. In analyzing these data we are initially focusing on algorithms that have been well reviewed in the literature with respect to microarray data, and/or used to analyze experimental microarray data. These include two types of hierarchical methods, agglomerative clustering and divisive clustering, and two types of partitioning methods, self-organizing maps and k-means clustering.

Results

Graphical representations of these data are shown in addition to a qualitative analysis.

Conclusion

Clustering algorithms have specific strengths and weaknesses when applied towards identifying non-random patterns in these data.

07

IN VIVO VALIDATION OF CARDIOVASCULAR BLOOD FLOW SIMULATIONS

Joy Ku, Gregory Wai Mong Chan, Mary Draney, Frank Arko, Chris Zarins, Charles Taylor

Purpose

We have previously proposed a new paradigm for treatment planning whereby physicians utilize computer simulations, based on patient-specific data, to evaluate the effectiveness of alternate treatment plans for individual patients. We have also described a simulation-based medical planning system for planning cardiovascular surgery (Taylor *et al*.). In order for such a system to be clinically useful, it is necessary to validate the system and demonstrate the accuracy of the simulation predictions in vivo.

Materials and Methods

We have performed in vivo validation studies by creating an aortic coarctation in pigs to simulate occlusive vascular disease and then utilizing a thoraco-thoraco aortic bypass graft to treat this condition. Anatomic data is acquired using magnetic resonance angiography (MRA) physiologic data is acquired utilizing cine phase-contrast magnetic resonance imaging (cine PC-MRI) and MR-compatible pressure catheters for both the pre-operative and post-operative states. We have utilized this data to generate computational models and perform blood flow simulations using a finite element method. Cine PC-MRI velocity measurements taken at locations downstream of the inlet to examine the accuracy of the computation simulations in vivo.

Results

Computed solutions compared very favorably with experimental data. Flow distribution between the native aorta and bypass graft was predicted within 10% as compared to experimental data.

Conclusion

Computational methods have significant application in predicting changes in blood flow due to vascular surgical intervention. In vivo validation studies are essential for the development of these methods.

08

COMPUTATIONAL ANALYSES OF THE DIFFERENCES BETWEEN DAILY AND INTERMITTENT ALENDRONATE TREATMENT

Christopher J. Hernandez, Gary S. Beaupré, Robert Marcus, Dennis R. Carter

Purpose

Alendronate has been shown to be an effective method of increasing bone mineral density (BMD) and reducing the risk of fracture in osteoporosis patients. Due to the chronic nature of osteoporosis and the special dosing requirements for alendronate, a reduction in the frequency of treatment would be an effective way to improve patient compliance (1, 2). A recent year-long clinical study has indicated that weekly and twice-weekly alendronate treatments of the same cumulative dose are therapeutically equivalent to daily treatment (2). Given these findings it is possible that even less frequent treatment could be equivalent. Computational methods are an attractive way of addressing this question because alendronate influences the bone remodeling process, a biological system that has been well-quantified in the past 30 years. In this study we use a model of bone remodeling to simulate alendronate treatment and to predict whether or not intermittent (non-daily) alendronate treatments are equivalent to daily treatment in terms of increases in BMD.

Materials and Methods

Bone remodeling is a focal process that occurs through the action of groups of cells organized into basic multicellular units (BMUs). BMUs move through bone resorbing and later reforming mineralized tissue. We use a time-dependent, non-linear feedback model of cancellous bone that quantifies BMU size, shape, origination and rate of bone resorption and formation (3). Model parameters are determined from measurements reported in the literature. Alendronate treatment is simulated by decreasing the BMU origination frequency according to the changes observed clinically (4). Intermittent treatments are simulated by allowing the origination frequency to return to pre-treatment levels between doses. This occurs at the same rate that alendronate is eliminated from the bone. Starting from an equilibrium condition corresponding to a healthy post-menopausal woman, alendronate treatment is simulated over 10 years for daily, twice-weekly, weekly, twice-monthly and monthly treatments of the same cumulative dose.

Results

The predicted increase in BMD found for intermittent treatments was less than that predicted for daily treatment (figure 1). As the frequency of dose administration decreased the percent increase in BMD was reduced. The difference in BMD increase between the daily model and the intermittent models was observed to increase with time but the efficacy of the intermittent treatments (BMD increase from intermittent treatment/BMD increase from daily treatment) was dependent only on the dose frequency. The efficacies for the different intermittent treatments were as follows: twice-weekly (96%), weekly (92%), twice-monthly (88%) and monthly (84%).

Conclusion

In this study we have used a computational model of bone remodeling to predict whether or not intermittent alendronate treatments increase BMD to the same degree as daily treatments. Decreasing the frequency of alendronate dosage resulted in reduced BMD changes as compared to daily treatment, although some treatments showed only small differences. Previous studies have used a ±1.5% BMD difference to define a range for equivalence (2). By this definition our model would predict that twice-weekly, weekly and twice-monthly treatments are all equivalent to daily

treatment over the 10 year period studied. Monthly treatments are predicted to be equivalent in short-term studies (1 year) but would not be equivalent in long-term studies (10 years). For this reason we recommend the use of efficacy for defining equivalence since the length of the study does not influence it. An efficacy range between 100 and 85% would result in similar but consistent equivalence conclusions to those mentioned above. The model is useful for comparing treatments because it is a mechanistic simulation relating cellular activity to changes in BMD. Although over long time periods other factors could influence the net change in BMD, the trends identified in this comparison model would remain the same, making this model useful during the development of clinical studies.

References

1. Bone, *et al*. (2000) Clin Ther 22: 15-28.
2. Schnitzer, *et al*. (2000) Aging (Milano) 12: 1-12.
3. Hernandez, *et al*. (2000) J Rehab Res Develop 37: 235-44.
4. Chavassieux, *et al*. (1997) J Clin Invest 100: 1475-80.

MECHANICAL INFLUENCES ON OBLIQUE PSEUDARTHROSIS FORMATION

Elizabeth G. Loboa Polefka, Gary S. Beaupré, Dennis R. Carter

Purpose

Through finite element (FE) analysis, this study analyzes the effects of stresses and strains on tissue differentiation in oblique pseudarthrosis development. A tissue differentiation hypothesis previously developed in our laboratory proposes that: 1) hydrostatic pressure directs the pluripotential mesenchymal tissue of a fracture callus down a chondrogenic pathway; 2) significant shear or tensile strain leads to fibrogenesis; and 3) given adequate vascularity, minimal levels of hydrostatic stress and shear/tensile strain allow direct intramembranous bone formation (1,3). In a low strain environment, intramembranous bone formation can be accelerated by slight hydrostatic tension (1). The objective of the present study was to test this tissue differentiation hypothesis with an FE model of an oblique fracture to determine if pseudarthrosis formation could be predicted based on stress and strain distributions within the fracture callus.

Methods

Model I: An idealized 2-D FE model of an oblique fracture was created (Fig. 1A) based upon the geometry of a typical oblique pseudarthrosis (Fig. 1B) (2). Both cortical bone and pluripotential callus were assumed to be linear elastic and isotropic materials. Cortical bone was assigned an elastic modulus (E) of 18.5 GPa and a Poisson's ratio (n) of 0.3. Pluripotential callus was assigned E = 1 MPa and n = 0.49. A compressive axial force was applied to the cortical bone ends and plane strain analysis was performed to determine patterns of hydrostatic stress and maximum tensile strain. Model II: A contact model was then developed incorporating sliding contact surfaces within the interfragmentary gap corresponding to locations of high tensile strain and regions of callus failure predicted in Model I. The same material properties and loads were assigned to Model II and patterns of maximum tensile strain and hydrostatic stress were again determined.

Results

Model I: Stress distributions showed low levels of hydrostatic tension at two periosteal corners of the fracture ends, high levels of hydrostatic pressure at the opposing periosteal corners, and intermediate levels of hydrostatic pressure throughout the interfragmentary gap (Fig. 2A). Maximum tensile strains were highest within the interfragmentary gap and lowest within the external callus (Fig. 2B). Based upon a maximum tensile strain failure criterion, callus failure is predicted within the interfragmentary gap. Bone formation is predicted at the periosteal corners exposed to low hydrostatic tension and fibrocartilage formation is predicted in the interfragmentary gap region exposed to both hydrostatic pressure and tensile strain. Model II: Hydrostatic stress distributions in the contact model were similar to those of Model I. However, maximum tensile strain distributions for Model II were quite different. Tensile strains decreased within the interfragmentary gap and increased within the external callus (Fig. 3A). These results would predict fibrocartilage maintenance within the interfragmentary gap and bone formation at the two periosteal corners experiencing low hydrostatic tension and low tensile strain (Figs 2A, 2B and 3A).

Discussion

In this study, we have predicted interfragmentary tissue failure, fibrocartilage formation, and locations of bone formation and resorption (4) consistent with initial stages of pseudarthrosis development seen in vivo (Figs. 1B and 3B) (2,3). These results suggest that pseudarthrosis formation

can be explained based upon the stresses and strains occurring within an oblique fracture.

Although pseudarthroses can arise from a multitude of factors including compromised vascularity, large fracture gaps and metabolic status, fracture geometry and load/motion at the fracture site seem to be key. To our knowledge, no prior studies have attempted to predict pseudarthrosis formation in an oblique fracture based upon the stresses and strains occurring within the fracture callus. Results from this study provide us with a better understanding of how the stress and strain distributions at a fracture site may cause delayed union, nonunion, and pseudarthrosis formation. This information may lead to improved fixation techniques and clinical outcomes for patients undergoing fracture treatment.

References

1. Carter *et al*. (1998) CORR 355S:S41;
2. McLean and Urist (1968) Bone. Chicago, Univ of Chicago Press 234
3. Pauwels (1980) Biomechanics of the Locomotor Apparatus. Berlin, Springer-Verlag 106-137;
4. Robertsson *et al*. (1997) Acta Orthop Scand 68(3):231.

Acknowledgements

Supported by VA Rehabilitation R&D grant A501-4RA.

Web Page

`http://guide.stanford.edu/People/polefka/polefka.html`

INTERNAL AND RELATIVE STRUCTURAL CONSERVATION OF DISCRETE PROTEIN SEQUENCE MOTIFS

Steven Paul Bennett, Douglas, L, Brutlag

Purpose

Protein multiple sequence alignments provide considerable information about conservation among related proteins, particularly in demonstrating which regions of a sequence are important. These alignments have allowed bioinformatics researchers to construct sensitive and highly specific motifs to describe these important subsequences. Building on the database of discrete sequence motifs previously constructed in the Brutlag bioinformatics group, this work extends the information resulting from sequence conservation, to make inferences about structural conservation as well.

Materials and Methods

Conservation of individual eMOTIFs. Accepted structures in the PDB_SELECT subset of structures were analyzed to compile a data structure relating each eMOTIF and the structures in PDB_SELECT that contain it, as well as the converse data structure relating each structure and the given eMOTIFs it contains. For each eMOTIF present in two or more of the structures in the data set, the coordinates of each residue specified in the eMOTIF were collected and rigidly aligned in pairwise fashion using a quaternian transformation algorithm. Alignments were made using alpha-carbon coordinates of the specified positions in the eMOTIF, and were scored with the resulting root mean square deviation (RMSD).

Conservation of eMOTIF pairs. As above, data structures were compiled relating eMOTIFs and PDB_SELECT structures. Here we relate pairs of eMOTIFs observed in multiple structures; for each eMOTIF-pair found in two or more structures, the alpha-carbon coordinates from each specified residue in the eMOTIF pairs were collected and rigidly aligned as described above.

Controls. Alignment results were analyzed against a background of randomly generated eMOTIFS. For each individual eMOTIF structural alignment, as well as for each eMOTIF-pair structural alignment, a population of 50 random structural alignments were generated. These random alignments were created by applying the same specification template as the eMOTIF (or eMOTIF pair) in the experiment to a randomly selected piece of sequence in a randomly chosen PDB and chain within the PDB_SELECT data set. Structural alignments were calculated using the same protocol described above for the experiments. 50 control alignments were performed in this way for each experimental alignment, in order to generate a sufficient number of samples to evaluate the experimental alignment score. Z-scores were calculated as a measure of each result's significance against a random population, and to control for eMOTIF length variation effects on the RMSD scores.

Results

We have shown that within a set of structures that are globally dissimilar in structure (having less than 25% or 30% sequence identity with one another), that our set of discrete motifs, called eMOTIFs, are observed to be highly conserved structurally, as measured by rigid structural alignment. In addition to individual eMOTIFs, we also show that eMOTIF pairs are structurally conserved with respect to one another as well.

Conclusion

These results indicate that discrete motifs such as eMOTIFs can be expected to imply structural conservation with a high degree of

confidence, both internally and with respect to one another in pairs. Implicit in the relationship between discrete sequence motifs and structural conservation is the hypothesis that multiple conserved subsequences within a protein may interact in the folded structure, even if distant in sequence. We have analyzed the structure database, and have demonstrated a significant propensity for interaction among eMOTIFs occurring in the same structure, as measured by distances between them.

CONSTRAINED GLOBAL OPTIMIZATION FOR ESTIMATING MOLECULAR STRUCTURE FROM ATOMIC DISTANCES

Glenn A. Williams, Jonathan M. Dugan, Russ B. Altman

Finding optimal three-dimensional molecular configurations based on a limited amount of experimental and/or theoretical data requires very efficient nonlinear optimization algorithms. Optimization methods must be able to find atomic configurations that are close to the absolute, or global, minimum error and also satisfy known physical constraints such as van der Waals separation distances. The most difficult obstacles in these types of problems are that using a limited amount of input data leads to many possible local optima, and that while introduction of constraints such as van der Waals helps to limit the search space, it often makes convergence to a global minimum more difficult.

We investigate several commonly used optimization methods, and introduce a constrained global optimization algorithm that is robust and efficient in yielding optimal three-dimensional configurations that are guaranteed to satisfy known van der Waals constraints. The algorithm uses an atomic-based approach that reduces the dimensionality and allows for tractable enforcement of constraints while maintaining good global convergence properties. We evaluate the new optimization algorithm using synthetic data on the yeast phenylalanine tRNA and several proteins from the Protein Data Bank (PDB), all with known crystal structure. We compare the results to commonly used global optimization methods such as simulated annealing, continuation, and smoothing. Results show that compared to the standard optimization approaches, our algorithm is able combine sparse input data with known physical constraints in an efficient manner to yield more accurate structures in terms of RMSD.

AUTOMATED INDIVIDUALIZED DECISION SUPPORT

George Christopher Scott, Ross Shachter, Leslie Lenert

Purpose

The field of decision science has approached the problem of formally modeling decisions and analysis of decision outcomes. These methods are often applied to groups or populations to assess cost-effectiveness. However, medical decisions for individuals are often dominated by their personal preferences. By learning more about a patient's preferences and how they change over time, one should be able to determine what makes them different from the "average" patient and whether or not appropriate guidelines might need to be modified for that patient.

Obtaining estimates of how patients value different health states relative to each other is a time-consuming and, for the patient, fatiguing task. As the number of health states increases, the likelihood of mis-assignment, error or inconsistency increases. It is therefore very desirable to obtain a recommendation with certainty from a decision model with the fewest number of value assessments as possible.

Materials and Methods

We developed a method of estimating the variance contribution of each of the model parameters to the final prediction of the model. This was then used in an algorithm to reduce the number of assessments necessary by terminating a patient-computer decision analytic dialogue once one of the treatments being considered was believed to be better with 95% certainty.

A population of 100 patients was simulated for preference value assessments using the system. Each simulated patient was assigned a set of consistent preference values randomly drawn from population distributions determined by meta-analysis. The preference values of the virtual patients were then assessed by the system, applying the appropriate stopping criteria. The number of assessments and the agreement of the abbreviated assessment method with the conventional method were recorded.

Results

Using the variance reduction algorithm and 95% certainty interval criteria, we found a mean number of assessments to be 4.24 (SD = 1.97) utilities out of the seven described in the model. Further, 38% of the virtual patients required only 2 out of the 7 health states to be assessed in order to have their 95% certainty interval exclude zero. Only 47% of the simulated patients required 5 assessments or more. And 12% required all 7 assessments.

Conclusion

Although the sensitivity of each of the preferences to the model and the individual's value for each health state strongly determines the efficiency of this approach, we feel that this method is a simple and effective way to improve automated decision support systems. Due to the dependency of the algorithm's ability to reduce the mean number of assessments on external parameters, it is not possible to reliably determine this prior to its use. Any reduction in the number of assessments is desirable, and the worst case results in the current approach. In light of this, we feel that this approach is worthwhile even if its effects are not known ahead of time.

A COMPARATIVE STATISTICAL ERROR ANALYSIS OF NEURONAVIGATION SYSTEMS IN A CLINICAL SETTING

Hamid Reza Abbasi, Sanaz Hariri (CandMed), David Martin, Daniel Kim, John Adler, Gary Steinberg, Ramin Shahidi

The use of neuronavigation (NN) in neurosurgery has become ubiquitous. A growing number of neurosurgeons are utilizing NN for a wide variety of purposes, including optimizing the surgical approach (macrosurgery) and locating small areas of interest (microsurgery). Experimental and methodological assumptions (e.g. using more markers increases accuracy) have been retained in the use of NN over its development. While rapid advances in hardware and software have emerged in the last few years, there have been only few attempts at challenging the old NN tenets and applying new technology to update these systems. To identify possible areas in which new technology may improve the surgical applications of NN and to test these old NN tenets, we conducted system-independent accuracy tests of neuronavigational measurements in two currently used systems: Radionics™ and BrainLab™. An immediate goal of this project is to give surgeons information about the accuracy of NN machines; surgeons should be able to estimate the accuracy of images generated by the system to optimize their surgical accuracy.

We obtained a phantom skull to most realistically simulate the surgical setting, removed the calvaria, and installed 3 Plexiglas square rods of different heights in each of the 3 anatomical fossae (anterior, middle and posterior). We used the edges of these rods as our targets. We installed a Plexiglas ball of known diameter on the phantom's sella turcica. Replacing the calvaria, we placed a total of 12 markers bilaterally on the exterior of the skull in the following regions: 6 frontal, 2 mastoid, 2 occipital and 2 high parietal. We performed a CT of the skull in 1.25 mm slices and sent the data over the network to the two NN machines evaluated in this study. The systems utilized their respective registration and tracking systems to localize the probe's tip in 3D.

When the probe tip was placed at the edge of a rod, the NN systems visualized the probe's position on their screens in the original axial plane of the CT scans and in the sagittal and coronal planes reconstructed from the CT scans. In each of the three cross-sections, we measured how far from the actual edge of the rod (x=0, y=0) the monitor was representing the probe tip. Paging through the images on the monitor to find the largest diameter of the Plexiglas sphere, we used the known diameter of the sphere to establish a system-independent scale for measurements. These measurements were acquired for both the Radionics and BrainLab NN systems in three different marker counts. Thus, we obtained 12 series of measurements, each series consisting of 218 separate measurements, totaling 2616 discrete measurements of accuracy.

We found that, despite current NN tenets, 4 or 8, but not 6, markers yield most efficient accuracy. We are aware of the counterintuitive nature of this finding, and our lab is currently investigating this result. Additionally, the movement of skin on the skull is not included in this study and will theoretically aggravate the overall error in each setting. We also found that: placing fewer markers around the region of interest (ROI) decreases registration error at the ROI; active tracking does not necessarily increase accuracy; the spreaded marker setting increases accuracy; and accuracy of the NN machines differs both overall and in different axes. As researchers continue to apply recent developments in hardware and software technology to the NN field, an increasing number of currently held tenets will be challenged and revised, rapidly and dramatically changing the field.

NEURONAVIGATIONAL EPILEPSY FOCUS MAPPING

Hamid Reza Abbasi, Sanaz Hariri (CandMed), David Martin, Michael Risinger, Gary Heit

Epilepsy occurs with a prevalence of about 0.5 percent and a cumulative lifetime incidence of 3 percent. Approximately one quarter of these patients eventually become refractory to pharmacotherapy despite the introduction of a number of new and relatively improved drugs. For medically intractable patients, resective surgery represents the next therapeutic intervention.

In extratemporal epilepsy, an exact localization of the resection target is crucial to surgery but is often not possible with noninvasive data collection (e.g. electroencephalography) alone. Resections in extratemporal cortex require definition of a seizure focus that often lacks anatomical boundaries. In contrast, the resection site for medial temporal sclerosis is clearly demarcated as the pes hippocampi, amygala, and lateral temporal cortices. Further complicating the definition of an extratemporal resection site are structural lesions that may have a complex spatial relationship to the actual ictal focus. Additionally the potential presence of cortex involved in language, primary sensory processing, motor control or cognition can provide further constraints on the extent of tissue to extirpate.

In response to these constraints, we report on the use of a surgical navigation system (SNS) in both cases of absent structural abnormalities as well as in cases of "normal" cortex to determine the precise relation between the subdural electrodes and the underlying anatomy. This correlation is achieved by co-registering the "electrographic map" generated during sub-acute intracranial recordings to the images of the three-dimensional MRI patient gyral anatomy.

On the day of the surgery, 6 to 8 skin MR compatible registration markers are placed on the patient's head, and an MRI of the patient is obtained and sent via DICOMM transfer protocols to the SNS (Radionics Maynard MA, software OTS version 2.2). Intraoperatively, following a craniotomy and registration of the patient to the SNS using the markers, the SNS is used to center the subdural electrodes over an area of known pathology and/or to co-register the electrode to the gyral anatomy. With the SNS probe touching a representative contact on an electrode, a display screen capture is performed. This precisely localizes each electrode in the axial, sagittal, and coronal planes of the MR. The operative field is secured in the standard fashion, and the patient is taken to the telemetry unit after an appropriate post-operative recovery interval.

Post-operatively, continuous intracranial EEG and simultaneous video are recorded. Two functional maps are generated: one based on inter-ictal and ictal events recorded from specific electrode pairs and the other identifying which electrodes, if any, overlie the eloquent cortex. These maps are co-registered via common electrode contacts to the SNS maps, providing assignment of electrographic pathology and function (e.g. speech) to a specific cortical surface anatomy. Based on functional as well as anatomical criterion, this mapping permits more precise a priori surgical resection planning and better assessment of potential risk.

COMPARATIVE TRACKING ERROR ANALYSIS OF FIVE DIFFERENT OPTICAL TRACKING SYSTEMS

Jeremy Johnson, Rasool Khadem, Clement C. Yeh, Mohammad Sadeghi-Tehrani, Michael R. Bax, Jacqueline Nerney Welch, Eric P. Wilkinson, Ramin Shahidi

Purpose

The positional and angular precision of five different optical tracking system (OTS) configurations are measured. The dependence of the two precision measurements on position and within the digitizing volume and angle between the dynamic reference frame (DRF) and camera are examined. The maximum positional and angular error for all measurements and for 95% of all measurements are also presented.

Materials and Methods

Optical Tracking Systems (OTS): Four cameras from two manufacturers were tested: the FlashPoint™ (Image Guided Technology, Boulder, Colorado) and the Polaris™ (Northern Digital Inc., Ontario, Canada). Three different sizes of FlashPoint™ cameras were tested, and the Polaris™ camera was tested in both active and passive configurations.

Linear Testing Apparatus (LTA): A precision-machined assembly consisting of a movable, vertical plate with uniformly spaced holes on which the DRF was mounted.

Stepper Motor Assembly: The assembly allowed the DRF to be mounted to the LTA and be rotated about the vertical axis.

Jitter is defined as the standard deviation of a sequence of measurements about the mean of the measurements.

Positional jitter measurements were obtained at positions uniformly spaced throughout a three-dimensional volume for each OTS. The spatial x, y and z coordinates were consecutively sampled 100 times at each sensor position.

Angular jitter was measured throughout a subset of the volume, and for angles between 0° and the maximum viewable angle. The angle step size was determined by the minimum rotation of the stepper motor, 1.8°. For each position and angle, 100 angle measurements were taken and the jitter calculated.

Results

Positional Jitter for All Systems

- Dominated by the z component (camera look direction).
- Relatively constant over single z-plane (independent of x, y, and theta).
- Increases with increasing z.
- Relatively constant for varying angles up to some cutoff angle.
- Best jitter obtained with 300 mm FlashPoint™ due to proximity of digitizing volume to OTS camera.

Angular Jitter for All Systems

- Relatively constant over single z-plane (independent of x, y, and theta
- Relatively constant for a given depth up to some angle (60° for active configurations, 40° for passive).

Differences Between Systems

- For IGT systems, positional jitter increases with z; for NDI systems, it remains relatively constant over a given range of z and theta
- Both passive and active configurations of the Polaris™ camera have much larger outliers for both positional and angular measurements than do any of the FlashPoint™ systems (Figure 5).

- When considering all data, the maximum error for the NDI cameras is far larger than the error for any of the IGT configurations; when the worst 5% of outliers are ignored, the performance of the NDI configurations significantly improve and nearly reach that of the IGT systems.
- Both positional and angular jitter of the IGT systems were more predictable and well-behaved than that of either NDI configuration.
- Passive NDI behaves differently than the four active OTS configurations. Positional and angular jitters increase dramatically for orientations larger than 40°. The variation in jitter for the NDI passive configuration is also much larger than for the active configurations.

Conclusion

The precision of position and angle measurements made by five commercially available optical tracking systems has been quantified throughout a volume. The easiest way to reduce both positional and angular jitter of measurements made by an optical tracking system is to minimize the distance between the camera and the tracked instrument while staying in the camera's digitizing volume.

The method presented for jitter measurement and analysis is independent of the tracking technology, and can be used for investigating the precision of future tracking systems.

USE OF XML/RDF TO CREATE STRUCTURED METADATA FOR MEDICAL IMAGES

John J. Michon

Purpose

We have built an XML (Extensible Markup Language) schema to describe ophthalmic images. Using the knowledge-modeling tool Protégé, we have created a schema for ophthalmic images using the Resource Description Framework (RDF), an XML application for encoding and exchange of structured metadata. The schema describes the types of resources and property values allowed for a wide variety of images commonly used in clinical ophthalmology and is being used to populate an ophthalmic image database. This schema for clinical markup is expected to become incorporated into the DICOM (Digital Imaging and Communications in Medicine) standard for ophthalmology, an internationally recognized standard for the encoding and transmission of digital images.

Materials and Methods

We created the imaging schema using the knowledge-modeling tool Protégé (`http://smi.stanford.edu/projects/protege`).

Results

In our model, the class Patient has child classes MedicalClassification, Observations&Exam and Therapy. The class Eye_Image is a subclass of Observations&Exam and is the parent class for all of the imaging modalities in ophthalmology (see Fig.). The primary image classes are Angio_Image (flourescein and indocyanin green angiography), Laser_Image (optical coherence tomography), MRI_Image (MRI), Radiog_Image (CT and X-Ray), Ultrasound_Image (A scan, B scan, and biomicroscopic) and VisibleLight_Image (external, biomicroscopic, pathologic).

An RDF schema derived from Protégé describes the semantics and allowed syntax of data elements in the document that refer to it. The schema identifies the values, or value ranges, that are permitted for each property, and the types of resources that it can describe.

An RDF statement becomes part of each image document, and specifies the semantics of each property in the document. Thus, clinical meaning can be extracted from the document metadata.

Conclusion

We have built a draft schema for ophthalmic images for inclusion as a more general DICOM standard for ophthalmology. The schema defines resources and properties of the most common image types. It creates a structured data model that allows for automatic severity assessment and complex querying of image metadata.

We have begun to instantiate an image database using this schema. Further testing and validation will be performed on image sets contributed by Stanford and other institutions. Iterative development of the full schema with DICOM Working Group 9 will result in an international standard for describing ophthalmic images. We will further validate this model using another imaging domain to show its general applicability. A wide variety of medical domains can benefit from user-defined clinical criteria based on image metadata to guide diagnosis and therapy.

A long-term goal is the creation of a platform for integration of ophthalmic clinical data sets, images and genomic data for data mining, meta-analysis, and customized therapy (Fig. 5). For example, it is known that those with a first-degree

relative with open angle glaucoma are at a tenfold greater risk of the disease than the general population. There are also likely to be important genetic factors in the response to drug therapy, progression of diabetic retinopathy, and progression of age related macular degeneration from the less severe atrophic form to the blinding neovascular form. Thus the integration of genetic data with rich clinical and imaging data sets will be a powerful tool for the prevention of blindness in the future.

Web Page

`http://smi-web.stanford.edu/people/michon/APAMI-paper.htm`

A REAL-TIME FREEHAND 3D ULTRASOUND SYSTEM FOR IMAGE GUIDED SURGERY

Jacqueline Nerney Welch, Jeremy A. Johnson, Michael R. Bax, Ramin Shahidi

Current freehand 3D ultrasound techniques separate the scanning or acquisition step from the visualization step. The process leads to a single image volume dataset that can be rendered for viewing later. While satisfactory for diagnostic purposes, the method is not useful for surgical guidance where the anatomy must be visualized in real time. The Image Guidance Laboratories are currently developing a freehand 3D ultrasound system that will allow real-time updates to the scanned volume data as well as the capability to simultaneously view cross-sections through the volume as well as a volume rendered perspective view. The equipment used is not unlike other freehand 3D ultrasound systems: an optical tracking system for locating the position and orientation of the ultrasound probe, a video frame-grabber for capturing ultrasound frames, and a high-performance computer for performing real-time volume updates and volume rendering. The system incorporates novel methods for inserting new frames into, and removing expired frames from, the volume dataset in real time. The position and orientation of a surgical instrument can be tracked and used for viewing the instruments position or trajectory with respect to the imaged region, or can be used to determine the viewpoint of the perspective image. This paper reports on current work in progress, and focuses on methods unique to achieving real-time 3D visualization using freehand 3D ultrasound.

BRIDGING THE GAP: SIMULATED DYNAMICS OF LIPID BILAYERS AT BOUNDARIES

Peter M. Kasson, Vijay S. Pande

Simulation of lipid bilayers in full-atom detail is challenging because of the large number of atoms involved--on the order of 25,000 for a 60x60 angstrom patch of membrane. Furthermore, many processes of interest, such as membrane fusion, involve rearrangements of large areas of the bilayer on timescales spanning several seconds. If we assume Moore's law to hold, such calculations will become feasible for supercomputers in approximately 40 years. In this report, we investigate a much smaller-scale rearrangement of lipid bilayers. Recently developed experimental techniques allow the removal of a narrow strip from a supported bilayer, leaving a gap with bilayer on either side (1). Surfaces micropatterned in this manner can be used for the construction of biosensors or chemically defined and manipulable cell-surface interfaces. The bilayer expands only slightly but does not fill this gap, creating a situation where there are two water-bounded edges to the bilayer. No experimental approaches attempted to date can determine the structure of these edges. Molecular dynamics simulation is therefore particularly well-suited for developing a model for the behavior of such bilayer boundaries.

Using molecular dynamics, we approximated the blotting process by which such gapped bilayers are created as follows. A simulated bilayer of 128 or 256 dimyristoylphosphatidylcholine (DMPC) molecules was equilibrated in a three-dimensionally periodic box filled with water molecules for 250 ps. A 14 angstrom gap was created in the bilayer by deleting all DMPC molecules that fell within a designated region. Subsequent to gap creation, six molecular dynamics simulations were run for a minimum of 2 ns each.

Within two nanoseconds of gap creation, the simulated bilayer edges rearranged to form micelle-like structures. Although such a rearrangement is perhaps to be anticipated, it is surprising in two respects. First, the micellization occurs very quickly on the scale of normal bilayer motions-the experimentally measured mean lateral displacement of a DMPC molecule in a bilayer is less than 5 angstroms over this time period (2). Second, pure phospholipids do not normally form micelles, as this presents a packing problem for the hydrocarbon tail groups. In our model, the bilayer edges avoid this problem by bulging slightly at the end. Such a structure would solve the hydrophobicity challenge created by a new water-bilayer interface but would nevertheless be somewhat energetically strained. The DMPC molecules at the bilayer boundary are structurally and dynamically similar to molecules in an ungapped fluid-phase bilayer, suggesting that no phase transition or drastic conformational transition has occurred. This model is consistent with the experimental observation that vesicles containing labeled lipids preferentially fuse to the boundary region, suggesting that the bilayer edges are energetically disfavored even though they are kinetically stable over times greater than one week (1, 3).

In summary, our molecular dynamics data suggest a model for the structure of bilayer boundaries in which the edges of the bilayer have micellized. This micelle-like structure also resembles hemifusion intermediates postulated to occur during membrane fusion. Although a full-atom model of membrane fusion is computationally infeasible at this time, it is hoped that further experiments on the hemifusion-like structures we have generated may shed light on the larger process.

References

1. Hovis, J.S. and Boxer, S.G. Patterning Barriers to Lateral Diffusion in Supported Lipid Bilayer Membranes by Blotting and Stamping. Langmuir 2000, 16, 894-897.
2. Scandella, C.J., Devaux, P., and McConnell, H.M. Rapid lateral diffusion of phospholipids in rabbit sarcoplasmic reticulum. Proc Natl Acad Sci USA 1972 Aug;69(8):2056-60.
3. Hovis, J.S. and Boxer, S.G. Personal communication.

KB-DRIVEN MODEL BUILDING: CHALLENGES AND APPROACHES

Mike Cantor,
Peter Karp (Bioinformatics Research Group, SRI Internationa)l,
Masaru Tomita (Laboratory for Bioinformatics, Keio University, Japan)

Our research explores the following question: How can we best leverage the power of knowledge bases to help experts build simulation models.

Knowledge bases (KBs) are playing an increasingly important role in biological research. Loosely defined, a knowledge base is a collection of information that is organized into a structured representation (sometimes called an "ontology"), designed to enable the automation of complex query and reasoning tasks. KBs like Ecocyc, PubMed, and GenBank are used by researchers to solve problems as varied as predicting a metabolic pathway from a genome, finding recent references in the literature on a given disease, or searching for a homologous sequence or structure to a recently discovered gene.

Another method of increasing importance in molecular biology is the computer simulation of cellular processes. Simulation packages such as Ecell, MIST, Scamp, DBSolve, and Gepasi are currently used by researchers in the modeling and design (in the case of bioengineering) of metabolic, signaling, transport, and genetic regulatory pathways in a variety of organisms. Simulation tools provide sophisticated engines and interface tools for observing the behavior of mathematical models of biological processes. However, actual construction of these models remains time-consuming and labor-intensive, as the fidelity and/or scope of the model requires the synthesis of large amounts of data and information from multiple sources.

We explore the question of how this difficulty might be alleviated by taking advantage of the structured information in knowledge-bases to automate or semi-automate the generation of models for simulation engines. As our test-system, we are attempting to use EcoCyc, a metabolic and regulatory KB of *e-coli* cellular function, to generate models for E-cell, a powerful cellular simulation package. Our poster outlines some of the challenges and issues involved in attempting this connection, and the current state of our progress.

MOTIFFEATURE: AUTOMATED CONSTRUCTION OF 3D MODELS FROM SEQUENCE MOTIFS

Mike Hsin-Ping Liang, Russ B. Altman

Purpose

Identifying important physical-chemical properties around functional sites can provide information about the biochemical environment required for protein function. Machine learning methods that build 3-dimensional models from these features are powerful tools for protein function prediction. These methods promise to be useful even for detecting functional similarity between proteins of little or no sequence similarity. However, choosing a good training set for the learning task is often manually done. Manual selection of the training set is not only laborious, but is also error prone since it can lead to biases in the model due to disproportionate representation of sites. With the ever-increasing number of resolved structures, there is a growing need to quickly build accurate models of functional sites from relatively well-characterized proteins and to scan uncharacterized proteins for them. We introduce MOTIFFEATURE, an algorithm for automating the task of selecting training examples from a database of protein structures. MOTIFFEATURE also implements a weighting scheme over the selected examples to account for under- and over-representation of sites.

Materials and Methods

Given a set of sequence motifs, MOTIFFEATURE scans the Protein Data Bank for an appropriate set of sites and non-sites to include in the training set. To correct for unbalanced representation in the training set, we use an Expectation Maximization algorithm to weight the examples based on their sequence. The weighted training set can be used to build a 3D model of the functional site. The model can then be used to score previously unseen proteins for potential functional sites.

Conclusion

Automated construction of models from sequence motifs is potentially useful in detecting functional sites on proteins without detectable sequence similarity. MOTIFFEATURE eliminates the laborious manual selection of the training set as well as corrects for disproportionate representation of sites, thus providing a faster and more accurate 3D model of functional sites. Evaluation of the algorithms used to select the sites, non-sites, and weights is currently in progress.

GUIDELINE INTERCHANGE FORMAT: A REPRESENTATION FOR SHARABLE, COMPUTER-INTERPRETABLE GUIDELINES

Mor Peleg

The GuideLine Interchange Format (GLIF) is a format for sharing computer-interpretable clinical guidelines independent of platforms and systems. GLIF is based on an object-oriented logical model of concepts that can be used to model a guideline, and has an RDF-based syntax. The ability to share guidelines is central to the GLIF methodology. Sharing is supported by: (1) a multi-level representation that facilitates sharing guidelines across different institutions and software applications; (2) a consensus-based multi-institutional process for developing GLIF that involves research groups from Stanford, Harvard, and Columbia Universities; (3) an open process resulting in a product that is not proprietary; and (4) a data model that is designed to support multiple vocabularies and medical knowledge bases.

GLIF version 2 (GLIF2), published in 1998, enabled modeling of a guideline as a flowchart of structured steps, representing clinical actions and decisions. However, the attributes of structured constructs were defined as text strings that could not be parsed, and therefore such guidelines could not be used for computer-based execution that required matching of guideline criteria to patient-specific data.

GLIF3 is a developing version of GLIF, designed to support computer-based execution. GLIF3 builds upon the GLIF2 framework but augments it by introducing several new constructs and requiring a more formal definition of decision criteria, action specifications and patient data.

There are three different levels at which a GLIF3-encoded guideline may be represented. The first is the author/viewer level that models the guideline as a conceptual flowchart of temporally ordered clinical steps, which facilitates human understanding. Different guideline steps are possible. They represent clinical actions, decisions and patient states, as well as branch and synchronization steps that enable concurrency. The model is hierarchical and allows action and decision steps to contain sub-guidelines. This enables the viewer to browse the flowchart at different levels of granularity.

The second representation level is a formal representation of decision criteria and actions that can be analyzed for correctness and executed by an interpreter. In order to support a formal model, GLIF3 uses a formal expression language and a medical domain object model. The formal expression language is a superset of the Health-Level 7 (HL-7) Arden Syntax's logic grammar that is used by GLIF3 for specifying decision criteria and patient states. GLIF3's medical domain object model is being designed to enable GLIF3 steps to refer to patient data items that are defined by a controlled terminology. The controlled terminology includes standard medical vocabularies that include concept definitions and codes (e.g., the Unified Medical Language System (UMLS) of the National Library of Medicine) as well as standard data models for medical concepts and their attributes (e.g., HL-7's Unified Service Action Model, which is GLIF3's default medical domain object model).

The third representation level, which is not supported yet, will represent application-specific mappings and modifications that facilitate integration into application environments. Other features of GLIF3 include a flexible decision model, event-based control flow, and iterations in actions and decisions.

We are using Protégé, a knowledge-engineering environment developed at Stanford Medical Informatics, as a GLIF3 authoring tool. We have added to it feature that enable import and export of GLIF3-encoded guidelines that are devoid of visualization-specific details. Protégé automatically lays out guideline flowcharts that are imported. In order to validate the expressiveness of GLIF3, we are using Protégé to encode several clinical guidelines. These include: (1) Managing Cough as a Defense Mechanism and as a Symptom, a Consensus Panel Report of the American College of Chest Physicians, (2) Prevention and Control of Influenza, of the Advisory Committee on Immunization Practices, and (3) Pharmacologic Treatment of Acute Major Depression and Dysthymia of the American College of Physicians - American Society for Internal Medicine.

Web Page

`http://smi-web.stanford.edu/projects/intermed-web/`

A FINITE ELEMENT MODEL OF THE HUMAN CORNEA

Assad Anshuman Oberai, Peter M. Pinsky, Thomas A. Silvestrini

Purpose

The Intrastromal Corneal Ring (ICR), Ring Segments, and Ring Arcs are devices developed for correcting refractive defects including myopia, hyperopia and astigmatism. Simulation of the mechanics of device-cornea interaction can help in identifying parameters affecting alterations in corneal topography due to these devices, thereby providing a useful tool in designing these devices.

Materials and Methods

A finite element model of the human cornea based on a mathematical description of the ultrastructural features and material characteristics of the corneal tissue has been developed. This model incorporates sliding contact conditions to model the stroma-device interface. Shifts in corneal power are calculated from the finite element deformations using a least-squares methodology.

Results

The model has been used to simulate the instantaneous non-linear, elastic response of the cornea to implantation of Intrastromal Corneal Rings and Segments with varying design parameters. The shifts obtained from these parameters have been analyzed and explained.

Conclusion

The finite element model has provided results that correlate well with clinical measurements. Further, these results provide a valuable insight into the mechanics of tissue deformation and an explanation for the observed shifts in power. This application is indicative of the role of this technique in assisting the development of innovative ideas related fields.

MECHANICAL REGULATION OF GROWTH PLATE MORPHOLOGY

Sandra Shefelbine, Dennis R. Carter

Introduction

Long bones grow by endochondral ossification, the process in which cartilage is replaced by bone. The ossification process begins in the center of the bone and progresses toward the ends of the bone. Distinct changes occur in growth front morphology during long bone growth and development. In most mammalian long bones the growth front becomes convex when it approaches the end of the bone shaft. After the secondary ossific center appears at the end of the bone, the growth front forms a concave dip in the center.

Many have suggested that these changes in morphology are caused by mechanical stresses at the growth front in the developing cartilage. Carter and colleagues proposed that the endochondral growth and ossification process is accelerated by intermittent shear stresses and inhibited by intermittent hydrostatic pressure and that these factors influence growth plate morphogenesis. Using the theoretical framework of Carter et. al, the objectives of this study are to determine the effects of (1) material compliance of the newly formed bone behind the growth front and (2) presence of the secondary ossific center on stresses in the developing cartilage and resulting growth front progression.

Methods

An axi-symmetric finite element model was created to represent the growth front as it approaches the end of a generic long bone. The model consisted of isoparametric hybrid elements with the materials properties of cartilage (E=6 MPa, Poisson's ratio = 0.49). All material properties were assumed linear elastic, homogeneous, and isotropic. Compliance of the newly formed bone under the growth front was modeled by varying boundary conditions at the growth front. A compressive load of 0.5 MPa was placed on the joint surface to represent joint loading and muscle contractions.

The specific growth rate (de/dt) represents the rate of longitudinal growth relative to an initial length. The specific growth rate was determined by contributions from the biological growth rate, the stimulatory effects of octahedral shear stress, and the inhibitory effects of hydrostatic compression. Octahedral shear stress is always positive, thereby increasing the specific growth rate. In this model hydrostatic stress is always negative (compressive) and decreases the specific growth rate. The model was grown using orthonormal expansion of the elements in the growth region by an amount determined by the specific growth rate. A parametric study of boundary conditions was conducted to determine the effects of the compliance of the newly mineralized bone under the growth front on stresses in the cartilage and growth front morphology. In addition, a secondary ossific center composed of cancellous bone (E=500 MPa, Poisson's ratio=0.2) was introduced to determine its influence on growth front progression.

Results

With a compliant interface between the cartilage and newly formed bone at the growth front, octahedral shear promoted growth more than hydrostatic stress inhibited it at the center of the growth front. This resulted in the development of a convex growth front as the specific growth rate was higher in the center than at the periphery of the bone. This convexity was reduced and even reversed when the interface was made more rigid.

The appearance of the secondary ossific center introduced higher hydrostatic pressure and lower octahedral shear in the center of the model causing the growth front to become concave as bone growth was

inhibited more at the center than at the periphery.

Discussion

The results demonstrate that growth front and growth plate morphology can be influenced by material properties of the newly formed bone under the growth front as well as the presence of the secondary ossific center. The growth predictions in this study are consistent with clinical observations: the convex growth front occurs when the bone is growing relatively fast and the newly formed bone is relatively compliant; the growth front becomes concave as growth slows and the secondary ossific center forms. These findings indicate the important role of mechanics in skeletal morphogenesis.

Acknowledgments

This work was supported by the NSF Fellowship, the Stanford Graduate Fellowship, and the Veterans Affairs RR&D Center (Palo Alto, CA). We thank Gary Beaupré for his suggestions.

THE IMPORTANCE OF SWING PHASE INITIAL CONDITIONS IN STIFF-KNEE GAIT: A CASE STUDY

Saryn Goldberg, Steven Piazza, Scott Delp

Introduction

Persons with cerebral palsy often walk with stiff-knee gait, a condition characterized by insufficient knee flexion during the swing phase of the gait cycle. When accompanied by over-activity of the rectus femoris muscle, stiff-knee gait is commonly treated by a rectus femoris transfer. However, this surgery is sometimes unsuccessful, in part because the factors that lead to stiff-knee gait have not been adequately characterized. We believe that examining the dynamics of stiff-knee gait in individual subjects will allow us to identify these factors and aid in the determination of the most appropriate treatments.

Previously, a dynamic simulation of swing phase showed that several factors can limit knee flexion, including excessive knee extension moment, diminished hip flexion moment, and insufficient knee flexion velocity at toe-off [1]. In the present study, a dynamic simulation of a subject with stiff-knee gait revealed the influential role of all swing phase conditions at toe-off (swing phase initial conditions) in the unilateral improvement of the subject after receiving bilateral rectus femoris transfers.

Methodology

We studied an eighteen-year-old female with spastic diplegic cerebral palsy who exhibited a bilateral stiff-knee pattern and swing phase activity of the rectus femoris. Following bilateral rectus transfer (and no other surgery), gait analysis showed that swing phase knee motion improved on the left side (range of motion increased by 15°), but did not improve on the right side (Fig. 1).

A computer model of swing phase was created in which the swing leg was represented by three segments (thigh, shank, and foot) suspended from a translating hip (Fig. 2). All motion was confined to the sagittal plane. The model was used to perform inverse and forward dynamic analyses of both limbs before and after surgery. The same dynamic analyses were performed using kinematics from 13 normal subjects.

Pre- and post-operative hip, knee, and ankle kinematics recorded during gait analysis were input into the inverse dynamic model to compute pre- and post-operative muscular joint moments about each of the three joints. Combinations of these moments and swing phase initial conditions served as input into the forward dynamic simulation. Pre- and post-operative moments were paired with their respective initial conditions as input to calculate the contribution of each joint moment to the total knee angular acceleration. Preoperative moments were then combined with normal initial conditions of interest (knee angle, knee velocity, and hip velocity) as input into the forward dynamic simulation to evaluate the resulting knee kinematics.

Results

The average knee angular accelerations due to moments about the hip and knee were multiple standard deviations above normal (Fig. 3). This observation is consistent with an over-active rectus femoris muscle. However, the total knee extension acceleration was smaller than normal, suggesting that the large knee extension acceleration induced by the knee extension moment was approximately balanced by the large knee flexion acceleration induced by the hip flexion moment. Thus, these abnormal hip and knee moments were not the cause of diminished knee flexion in this subject.

Of the swing phase initial conditions, knee angle, knee velocity, and hip velocity were found to have the strongest influence on knee flexion. In both limbs, these initial values were significantly lower than normal both pre- and post-operatively (Table 1). The right limb values were further from normal than those for the left limb. When normal swing phase initial conditions of interest and the subject's abnormal preoperative moments were used as input into the forward dynamic simulation, the resulting knee flexion kinematics were near normal (Fig.4).

Discussion

This study demonstrates the importance of swing phase initial conditions in the determination of the stiff-knee gait pattern. The subject's low preoperative initial knee angle, knee velocity, and hip velocity values appear to be responsible for the diminished preoperative knee flexion and knee range of motion. When these values were raised to normal levels, the resulting knee flexion was near normal. We believe that the limited post-operative increase of these initial condition values for the right limb resulted in negligible post-operative improvement in right knee flexion, while marginal increase in these values in the left limb resulted in some improvement in post-operative left knee flexion.

The strong influence of initial swing phase conditions on swing phase knee flexion points to the importance of stance phase in the generation of the stiff-knee gait pattern, and demonstrates the need to study terminal stance phase to understand the causes of stiff-knee gait.

eferences

1. Piazza *et al*. J. Biomech. 29(6): 723-733, 1996.

Acknowledgments

Funded by NIH and a Whitaker Foundation Graduate Fellowship.

A NEW TWIST ON THE HELIX-COIL TRANSITION: A NON-BIOLOGICAL HELIX WITH PROTEIN-LIKE INTERMEDIATES

Sidney P. Elmer, Vijay S. Pande

Polyphenylacetylene, hereafter referred to as pPA, is a nonbiological polymer which has been shown experimentally to fold into a helix, with typical folding times of 10s of nanoseconds and nonexponential kinetics[1]. Nonexponential kinetics is indicative of a complex free energy landscape with intermediates, traps, and multiple pathways. Therefore, pPA demonstrates many of the same kinetic properties that proteins and other complex biological systems exhibit. Since proteins fold in the micro- to millisecond time scales, a full molecular dynamics (MD) trajectory is very difficult to obtain under physiological conditions. However, the time-scale for pPA to fold is easily attainable on modern processors, allowing us to collect 2228 all-atom MD trajectories of a 12-mer of pPA. We characterize the thermodynamic and kinetic properties of this synthetic polymer, which has relatively simple interactions, and then use these results to gain insights into the molecular details of the folding mechanism for more complex biological structures.

Our simulations result in the very interesting observation that this model of a 12-mer also folds with nonexponential folding rates in agreement with the experiments mentioned previously. The mean folding time was calculated to be 8.7 ± 3.0 ns, which is on the same order of magnitude as the experiments. Analysis of individual trajectories uncovers an intermediate state containing configurations with little conservation of local structure, suggesting there are multiple pathways to the folded state. A search for a parameter that accurately describes the progress of the folding of the polymer reveals that the length of consecutive cis dihedral angles, D, is a good reaction coordinate. In addition, a non-linear least-squares analysis of the fluctuations of the polymer reduces the complex motions of the polymer to a subspace of essential motions described by principle components of the system. The motions of the polymer projected onto the two primary components reveal a rugged free energy landscape in the folded basin, thus providing a microscopic view of the complex mechanism for folding and offering a physical interpretation of the nonexponential kinetics.

For many decades, the Helix-Coil Transition Theory has stood as a model for the formation of helical structures in Biology. Its main tenets describe the folding of helices via a rate-limiting step of nucleation of a few local residues into helical configurations. Once this event occurs, the helix formation will rapidly propagate in both directions to the ends of the polymer. The result of this theory is an exponentially distributed mean folding time for the folding of the helix, denoting a simple mechanism and a smooth free energy surface. We have shown that the Helix-Coil Theory does not hold for even simple helices, such as polyphenylacetylene. For more complex systems, such as proteins and nucleic acids, the Helix-Coil Theory clearly cannot be a reasonable model for helix formation and growth. Therefore, new theories are needed that can take into account the complex nature of the folding dynamics of helices and other secondary structures in complex biological systems.

References

1. Yang, WY; Prince, RB; Sabelko, J; Moore, JS; Gruebele M JACS, 2000, 122, 3248-49.

SEQUENCE ANALYSIS AND STRUCTURE COMPARISON OF THE SH3 DOMAIN FAMILY

Stefan M. Larson, Alan R. Davidson

As biological research rushes through the genomic and proteomic era, the resulting accumulation of protein sequences and structures is creating huge demand for efficient analytical methods. For example, the complexity of the protein-folding problem requires computational analysis of protein sequences and structures to provide insights into their relationships and to aid in experimental interpretation and design. In this study, we developed and applied a rigorous set of computational and statistical analyses to a single protein family, the SH3 domain. The SH3 domain was chosen as a model system due to its biological importance as a ubiquitous mediator of protein-protein interactions, its relatively small size and simple fold, and its well-behaved experimental nature. These factors have also led to it being well characterized experimentally in our lab and many others, allowing for direct comparison of theoretical and experimental results.

To start, a non-redundant alignment of 266 SH3 domains was carefully assembled. Henikoff weighting was used to reduce sequence bias and Shannon entropy was calculated at each position as a measure of residue conservation. Eighteen SH3 domain structures were aligned and a number of all-vs-all comparisons were performed to quantify structural variation in the domain. No direct correlations between sequence identity or positional entroy and RMSD between structures were observed. However, conserved residues were found to consistently play important structural and/or functional roles in the SH3 domain. It was found that residues playing consistent structural roles in ligand-binding were much more highly conserved than those which contact the ligand differently in different structures. This points to the less conserved residues as being responsible for the exquisite binding specificity of SH3 domains.

To further understand residue interactions in the SH3 domain, an algorithm for covariation analysis was developed. Because of its potential to aid in structure prediction, a focus of this work was on elimination of artifactual covariations and accurate prediction of residue contacts. The vast majority of covariations in the SH3 domain involved residues in the hydrophobic core and in the ligand-binding pocket. Several networks of three covarying residues were also identified. Two of these triplets have been dramatically confirmed experimentally, by combining three destabilizing mutations into a triplet mutant of near wild-type (Fyn SH3) stability. Contact prediction was successful: 84% of the highly covarying residue pairs are within 8 angstroms in at least one of the eighteen SH3 structures studied. Fifteen additional domain alignments were analyzed using the covariation algorithm. Six of these produced significantly accurate contact predictions.

Sequence alignment analysis and structure comparison of the SH3 domain produced useful data not previously available through structural or experimental studies. Much of this data has already been incorporated into other studies to interpret results and design new experiments. New work in the Pande group at Stanford aims at large-scale sequence design of SH3 domains (among others). By reconciling the results of computational sequence design with detailed analysis of naturally occurring sequences and structures, we hope to more rigorously define what features of a protein sequence are necessary to define its fold.

REPRESENTING CONTEXTUALLY CHANGING DECISION MAKING BEHAVIOR IN MEDICAL ORGANIZATIONS

Carol HF Cheng, Raymond E Levitt

In medical practice and scientific research, many members of a team work together to synthesize data about a patient or problem. As problems become more complex, the communication and coordination tasks among team members becomes non-trivial. Individuals participating in this process are ill equipped to understand their role in it, and efforts to improve process frequently focus on the elimination of local errors. A recent Institute of Medicine report [1] argues that local errors may be caused by systematic factors, and an understanding of the global work processes may assist in diagnosing probable areas of error commission [2]. Such diagnostic tools are especially lacking in medicine, a highly flexible service industry with varied roles and many concurrent processes.

We have created a simulation environment, the Virtual Design Team (VDT) [3, 4], to represent the many professions who coordinate in the care of a case-mix of patients. This discrete-event simulation tool provides a virtual test-bed for designers of clinical protocols to assess their impact on the workflow of the organization. VDT allows the description of the actors in the organization, their roles, skills, and experience levels, as well as the activities of the organization. We make explicit the responsibilities of each actor, the relationships between the activities, and the communication and coordination requirements of each activity. Using an information-processing framework [5], we assume that each activity can be represented by the amount of time it requires in direct work, coordination, and rework. Protocol designers can thus describe their protocols succinctly and create computer simulation models of organizational behavior, a controlled, cheap, and quick alternative to experimental studies.

However, in industries as diverse as call centers, banks, and healthcare, managers have already attempted to institute quality standards through standardized protocols, best practice guidelines, and workflow management systems [6]. These devices outline the ideal process, but inadequately anticipate the contexts in which decision makers systematically deviate from the ideal. These contexts include the time of day, the workload of the decision maker, the workload of collaborators, and the schedule status of an process. In these contexts, decision makers' behavior deviates in predictable ways from the ideal protocol, in order to fulfill objectives not described by the protocol. In general, failure to consider context-dependent changes in decision making behavior can lead to unanticipated results in the load on individual workers, the communication and coordination requirements of activities, and the amount of error and rework necessary. The alternatives workers develop to the desired workflow thus impact service quality and efficiency.

In keeping with information-processing theory, we model contexts induced by the activities and environment of the worker, but not directly related to the content of the work. Our goals for incorporating contextual effects are two-fold: to describe the heuristics used by decision makers to respond to recurring contexts, and to measure the effects of such local behaviors on the organization-level process performance. We do not pretend to describe contexts exhaustively, but focus on those which have had a documented effect on decision making behavior in the medical domain. We model the effects of delaying work, allowing decisions to be made by lower-skilled workers, and hurrying through

tasks. In our initial scenarios, we find that although such contextual behavior may decrease the time spent on direct work for activities, it may lower quality through fewer attempts at coordination, and increase the total time of the project because of necessary rework. Thus, the aggregation of small, isolated decisions to optimize individual work can lead to significant changes in the "macro-behavior" of the organization.

Contextual changes in decision making is an understudied phenomenon in workflow analysis with potentially large effects on process execution. We have defined a small subset of contexts and their effects on decision-making behavior. We represent these in a framework, which highlights the communication requirements of coordinated work processes. We plan to evaluate the representation of contexts by modeling a medical clinic, and evaluate its generality by modeling an airline service organization. We hope that the investigation of contextual responses will inform protocol design, prevent implementation failures, and lead to more flexible decision-making capabilities in existing workflow management software.

References

1. Kohn, L., J. Corrigan, and M. Donaldson, eds. To Err is Human: Building a Safer Health System. , ed. I.o. Medicine. 1999, National Academy Press: Washington, DC.
2. Chen, R. and R. Altman, Automated diagnosis of data-model conflicts using metadata. JAMIA, 1999. 6(5): p. 374-392.
3. Fridsma, D. Representing the Work of Medical Protocols for Organizational Simulation. in AMIA Annual Symposium. 1998. Orlando, FL.
4. Kunz, J., *et al*., The Virtual Design Team. Communications of the ACM, 1998. 41(11): p. 84-91.
5. Galbraith, J., Designing complex organizations. 1973, Reading, MA: Addison-Wesley.
6. Massaro, T., Introducing physician order entry at a major academic medical center. Acad Med, 1993. 68(1): p. 25-30.

Web Page

`http://www.stanford.edu/group/CIFE/VDT/index.html`

MEDLINE QUERY-BY-EXAMPLE

Elmer Bernstam, Olga Troyanskaya, Jeff Chang

Purpose

Medline is a database of over 10,000,000 citations to the world's biomedical literature and is growing at a rate of over 30,000 new citations per month. Although Medline has proven extremely valuable, retrieval from Medline is difficult. Recall and precision rates are quite variable, but 25 - 60% are typical rates for both parameters. The goal of the MedlineQBE project is to facilitate the use of Medline by: (1) allowing non-expert users to query Medline by giving examples of what they want, rather than by specifying a query using traditional query languages and (2) to create a flexible, extendible framework that allows developers to easily create modules implementing novel search strategies.

The general paradigm for MedlineQBE use is: (1) perform initial search (2) select relevant articles from the retrieved set, (3) repeat as necessary.

Materials and Methods

We implemented MedlineQBE using industry standard technologies. The system is currently running on a Linux based Tomcat WWW Server on a Dell Inspiron 7000 (Pentium II, 300 MHz) notebook computer.

The user interface is written in Java. A Java Servlet controls the display of multiple Java Server Pages (JSPs). JSPs are responsible for interacting with modules implementing specific search strategies.

There are two general classes of modules: (1) modules to perform the initial search and (2) modules to refine searches (i.e., input includes a list of relevant articles as selected by the user). Modules can be written in any compiled or interpreted language, which can execute from the Linux command-line, but the example modules are implemented in Perl and Python.

We implemented modules to perform "Power Search", where the user's input string is sent directly to PubMed, "Simple Search" where the user is able to fill out a form to issue a structured boolean query to PubMed and "Related-by-MeSH" a module that allows search refinement using the combination of MeSH terms of user-selected articles. The base module, which handles all interaction with PubMed is written in Python.

Results

The system, as described above has been successfully implemented. As of September 2000, it has not been made available to the public, though there are plans to do so.

Modules written in Python and Perl have been successfully integrated into MedlineQBE. Given example modules, the developer only has to write enough code to construct a valid PubMed query given user input.

Conclusion

We have created a flexible, extendible framework for a Query-by-Example interface to Medline. Evaluation of usability and performance is planned.

OFFLINE TESTING OF A COMPUTERIZED DECISION SUPPORT SYSTEM FOR MANAGEMENT OF HYPERTENSION

Susana Martins, MK Goldstein, BB Hoffman, RW Coleman, SW Tu, R Shankar, M O'Connor, MA Musen, SB Martins, N Hastings
VA Palo Alto Health Care System, Palo Alto, CA
Stanford University, Stanford, CA.

Complex decision support systems (DSS) require evaluation before they can be safely deployed for clinical uses, either prospectively to make recommendations or retrospectively for quality review.

Methods

We developed ATHENA, a DSS implementing the JNC 6 guidelines for hypertension. A physician (MD) previously unrelated to the project developed with us by consensus a written document (RULES) detailing our operationalization of the JNC 6 rules. We selected from electronic medical records a random sample of 100 hypertensive cases, stratified by comorbid disease. The MD reviewed the same case material available to ATHENA and made recommendations based on the RULES. A physician and a pharmacist compared MD recommendations with those from ATHENA with another physician adjudicating disagreements. After identifying and correcting problems in the DSS, physician and pharmacist carried out a second review of all cases.

Results

In the 87 cases that met inclusion criteria for ATHENA review, 224 drug recommendations were made by MD and/or ATHENA. Of these, there were 87 disagreements. In 25 of the 81 disagreements ATHENA (12) and the MD (13) deviated from the RULES prescribed recommendation. 11/12 ATHENA deviations were due to incorrect coding of a drug dosage form and 1 case was due to a wrong entry in the pharmacy database. No errors in program logic were observed. In 10 of 13 MD disagreements the MD did not note all technically possible drug recommendations. These omissions were without clinical significance (e.g., MD noted "substitute X for Y, or A for B" but did not also note "substitute X for B, or Y for A."). In 32 disagreements, MD applied different interpretation (30) or used additional medical knowledge (2) to make a clinically appropriate recommendation. For example in 14 heart failure cases MD recommended a beta blocker while the RULES stated that this issue was beyond the scope of ATHENA recommendations and in 4 cases ATHENA recommended either DHP or NDHP calcium channel blockers while MD decided on only one of these subclasses. After corrections were made to the DSS a second review revealed that disagreements previously noted due to error in the knowledge base or drug tables were corrected.

Conclusions

A complex hypertension DSS can work remarkably well. As expected, the DSS was more complete in listing all possible combinations of recommendations. It is interesting to note that the MD who participated in development of the RULES document consciously deviated from it in many cases, suggesting that the MD's overall impression of the best therapeutic decision overrode the rules. The evaluation of a DSS before implementation in clinical practice is imperative to detect errors that could affect clinicians' confidence in using information from the DSS.

COMPARISON OF RIBOSOMAL MODELS TO EXPERIMENTAL DATA WITH THE RIBOWEB SYSTEM

Michelle Whirl Carrillo, Russ B. Altman

The RiboWeb system was designed to provide a web-based computational environment facilitating ribosomal modeling and evaluation. It links a knowledge base of experimental structural data regarding the ribosome to molecular modeling programs and other computational tools. One available tool supports the comparison of molecular ribosomal models with experimental data in the knowledge base. This type of comparison has important implications for modeling. Determining which data is consistently compatible, or not, with models can be a clue to understanding the reliability of the data for model building. Trends in data satisfaction may show data that are consistently incompatible with other data, or suggest that certain data types are not being interpreted accurately. This information is valuable for future model building.

We compared five widely accepted models of the 30S ribosomal subunit to all of the footprinting, crosslinking and cleavage experimental data in our knowledge base. We saw trends in the overall satisfaction of the data by the models. We were also able to rank models by overall data satisfaction, and view the "problem areas" in each model, according to the data comparison.

THE MOUSE SNP DATABASE: MAPPING QTLS IN SILICO

Jonathan Usuka

Understanding the underlying genetics of human diseases is the focus of many current research projects. Because of the experimental limitations in human genetics, mouse intercross models exhibiting phenotypes observed in human disease are analyzed instead. Genes that are identified in mouse experiments often belong to the same pathways involved in the human disease and therefore yield a better understanding of the human disease process. Two of the slower steps in genetic analysis involve determining the appropriate mouse cross and the subsequent genotyping of the intercross generation. In order to accelerate these steps we developed two computational tools: a searchable mouse SNP database with allele information for the 13 most commonly used inbred mouse strains, and a SNP based linkage prediction program that predicts the most likely QTL's from quantitative phenotype data across three or more mouse strains. The computational QTL prediction method correctly predicted the experimentally identified QTL's from six published mouse models with various phenotypes.

A NEW METHOD FOR DETERMINING PROTEIN FUNCTION SIMILARITY BASED ON KEYWORDS AND GENE ONTOLOGY

Yueyi Liu, Russ, Altman

Sequence homology search programs such as BLAST are very useful in getting some idea of the function of a gene or protein when nothing except the sequence is known. They have also been used for clustering genes or proteins based on function, since similarity in sequence tends to lead to similarity in function. For genes or proteins that have been studied experimentally, text documents are a useful source in determining their function. Natural language processing (NLP) is one way of clustering based on documents, but it is usually computationally intensive. We propose to use Swiss-Prot keywords in comparing protein functions and in clustering based on function. Swiss-Prot is a protein sequence database that provides keywords for the function of a protein. The keywords of proteins with similar function are more closely related then keywords of proteins with completely different functions. The relatedness of two keywords is captured by their mappings on Gene Ontology, which consists of three distinct ontologies for three areas: molecular function, biological process and cellular component. Since Gene Ontology is hierarchical, we can measure the pair-wise distance for the mappings of two keywords. The keyword distance is defined as the minimum number of edges from the two mappings to their first common root. The closer their mappings are, the smaller the keyword distance. The pair-wise distances between all pairs of mapped keywords are then calculated. We downloaded the keywords for all proteins of five genomes from Swiss-Prot and calculated the pair-wise distance between each pair of proteins. The distance between the keywords for a particular protein and some other protein is defined as the sum of minimum keyword distances between each keyword for this protein with every keyword for the other protein, divided by its total number of keywords. We found that this distance is sensitive enough to pick up proteins with similar function. We hope this method will compliment the sequence comparison methods in clustering proteins with similar function, since not every protein involved in similar function has similar sequence.

OPTIMIZING KNOWLEDGE-BASED ENERGY FUNCTIONS: FROM LATTICE STUDY TO REAL PROTEINS

Yu Xia, Michael Levitt

Given the increasing number of known protein structures and the limited success of physical potentials in discriminating native structures against misfolded structures, knowledge-based energy functions extracted from a database of known protein native structures have been widely used in protein structure prediction.

We propose a general framework for extracting knowledge-based energy functions. We assume that the total energy for a protein structure is a linear combination of certain basis functions, and a set of native protein structures with corresponding libraries of decoy structures are known. In our scheme, the energy function is optimal when there is least chance that a random structure has a lower energy than the corresponding native structure. The optimal energy parameters depend on the distribution of decoy structures in the structure space. Subject to certain approximations of this distribution, most current database-derived energy functions fall within this framework, including mean-field potentials, Z-score optimization, and constraint satisfaction methods.

We propose a fast and effective method for energy function parameterization based on this framework. We go on to compare our method to other methods using a simple lattice model in the context of three different energy function scenarios. We show that our method, which is based on the most stringent criteria, performs best in all cases. Z-score optimization also performs well.

We go on to derive energy parameters for real proteins with optimal performance. We choose residue-residue contact potential as an example. We select a representative set of protein sequences with experimentally determined structures from the Protein Data Bank. For these sequences, we use fast Monte Carlo methods and off-lattice models to generate over forty million randomly sampled misfolded conformations that have protein-like features such as self-avoidance, compactness and preferred bond length, angle and dihedral angle values. We then compare these misfolded conformations to their corresponding native conformations, and optimal energy parameters are derived from these data.

Our method is optimal in that given a specific energy function representation and a large set of randomly sampled misfolded structures, this approach will find the energy function parameters that give the best discriminating power. We apply our optimal energy function parameters to discriminatory tests and compare its performance to other energy functions.

MONTE CARLO SIMULATIONS OF FOLDING OF SIMPLE ALPHA HELICES

Bojan Zagrovic, Jessica Shapiro, Vijay Pande

Purpose

Alpha helices are basic elements of protein structure, but the manner in which they fold is still not nearly fully understood. In this study, Monte Carlo (MC) simulations of capped 21-residue peptides, (Ala)21 and (Ala)5-(AlaAlaAlaArgAla)3-Ala, were conducted to analyze the preferred location of helix nucleation sites, speed and direction of helical propagation, and the influence of bulky arginine side-chains on these attributes of folding. In addition, the results were compared with the results of molecular dynamics (MD) simulations of the same systems.

Materials and Methods

The simulations involved standard Metropolis Monte Carlo using OPLS force field and Tinker software for energy evaluation. The simulations were performed in implicit solvent starting from an antiparallel beta sheet configuration with no pre-equilibration. Temperature was set at 300K. The simulations involved 3 kinds of Monte Carlo moves: a) major backbone phi/psi moves (dihedral angles locked to values characteristic of alpha helices, 3/10 helices, parallel beta sheets or antiparallel beta sheets); b) minor backbone phi/psi moves; c) rotamer moves for arginines. Data analysis was performed using Mathematica software.

Results and Conclusions

Polyalanine peptides exhibit a preference for nucleation at the C-terminus and, concomitantly, tend to extend in the C to N direction. Their folding times are roughly exponentially distributed with the mean of 20,000 MC steps. Finally, the residues at both ends of polyalanine helices adopt helical conformation more quickly compared with the residues in the middle of the helix. The dynamics of folding of the arginine containing peptides depends strongly on the characteristics of the allowed moves for arginines. “Slow arginine” peptides (1 backbone angle move per 81 rotamer moves) exhibit no preferred nucleation sites, fold in ~120,000 MC steps, and tend to get trapped in collapsed states. “Fast arginine” peptides (1 backbone move for each rotamer move) fold in a manner that is indistinguishable from the polyalanine helices.

AUTOMATIC DETECTION AND QUANTIFICATION OF ABDOMINAL AORTIC THROMBUS IN CT ANGIOGRAMS BASED ON CLUSTERING AND GLOBAL GEOMETRIC INFORMATION

Feng Zhuge, Sandy Napel, David Paik, Geoffrey D. Rubin

Purpose

Detection of thrombus based on CT attenuation alone will unavoidably generate errors in certain local regions with the relatively low contrast to noise ratios. This problem is aggravated by the occurrence of adjacent tissues such as bowel loops, with similar attenuation that should be excluded from the detection result. The purpose of this research is to develop an algorithm that detects aortic thrombus edge in the presence of noise and other interfering structures.

Materials and Methods

Our method use a classical edge detector to find all possible edges based on gray level information only. Edge candidates are organized by the distance and angle to a given center point. Our segmentation model assumes that the real thrombus edge should not contain high frequency components; the variation of these distances with angle is therefore restricted. Then, edge candidates are clustered according to the distance and angle to the center point. Edges caused by noise and other structures are determined to be in a different cluster than the true edge because of sharp changes of distance in a small angle range. The surfaces comprised of these false edges are assumed to have smaller area than the true edge surface. Thus we reject clusters corresponding to small surface areas. Interpolation is applied where edge candidates are judged to be false.

To evaluate our algorithm, we used a helical CT simulation program to simulate a human abdomen including a thrombosed aortic aneurysm with adjacent vena cava, bowel loops, and spine at various locations. We performed 4 simulations: one without noise and 3 with realistic CT noise.

Results

We evaluated detection results by two metrics: (1) False negative volume fraction (FNVF = undetected thrombus volume / true thrombus volume) and (2) False positive volume fraction (FPVF = falsely included thrombus volume / true thrombus volume).

For the noiseless case, PNVF = 0.06% and FPVF = 5.22%. For the noisy cases: FNVF = 0.14% ± 0.02%, and FPVF = 5.56% ± 0.05%. Residual error is due to noise and interpolation; all false edges were rejected by our algorithm.

Conclusion

Adding global geometry constraints to gray level information improves detection and quantification of aortic thrombus in a phantom model. Accurate delineation of abdominal aortic thrombus will allow accurate and reproducible quantification of aortic aneurysm size and growth, which has become a critical issue in the era of stent-graft therapy.

QUANTIFICATION OF THE HYDROPHOBIC INTERACTION BY SIMULATIONS OF THE AGGREGATION OF SMALL, HYDROPHOBIC SOLUTES IN WATER

Tanya M. Raschke, Jerry Tsai, Michael Levitt

We have used molecular dynamics (MD) simulations to investigate the aggregation of a series of small, hydrocarbon molecules in water. MD simulations were performed on systems containing increasing numbers of solute molecules in water-filled boxes of different sizes, sampling a hundred-fold range of solute concentrations. Throughout these simulations, the formation and disruption of solute clusters was observed. By treating the data from the trajectories as a series of equilibrium measurements, we directly measured the free energy of adding a single solute molecule to a cluster. This derived free energy is proportional to the loss in exposed molecular surface area that occurs when a solute molecule joins a pre-existing cluster. Furthermore, the constant of proportionality (45 cal/Å2) is in complete agreement with experimental measurements of the hydrophobic effect. This is the first direct calculation of the hydrophobic interaction from MD simulations; the excellent agreement with experiment indicates that force fields with van der Waals interactions and atomic point-charge electrostatics account for the most important driving force in biology.

VIEWFEATURE: INTEGRATED FEATURE ANALYSIS AND VISUALIZATION

D. Rey Banatao, Conrad C. Huang, Patricia C. Babbitt, Russ B. Altman, Teri E. Klein

Visualization interfaces for high performance computing systems pose special problems due to the complexity and volume of data these systems manipulate. In the post-genomic era, scientists must be able to quickly gain insight into structure-function problems, and require flexible computing environments to quickly create interfaces that link the relevant tools. Feature, a program for analyzing protein sites, takes a set of 3-dimensional structures and creates statistical models of sites of structural or functional significance. Until now, Feature has provided no support for visualization, which can make understanding its results difficult. We have developed an extension to the molecular visualization program Chimera that integrates Feature's statistical models and site predictions with 3-dimensional structures viewed in Chimera. We call this extension ViewFeature, and it is designed to help users understand the structural Features that define a site of interest. We applied ViewFeature in an analysis of the enolase superfamily; a functionally distinct class of proteins that share a common fold, the a/b barrel, in order to gain a more complete understanding of the conserved physical properties of this superfamily. In particular, we wanted to define the structural determinants that distinguish the enolase superfamily active site scaffold from other a/b barrel superfamilies and particularly from other metal-binding a/b barrel proteins. Through the use of ViewFeature, we have found that the C-terminal domain of the enolase superfamily does not differ at the scaffold level from metal-binding a/b barrels. We are, however, able to differentiate between the metal-binding sites of a/b barrels and those of other metal-binding proteins. We describe the overall architectural Features of enolases in a radius of 10 Angstroms around the active site.

USING HUMAN LANGUAGE ABILITY TO LEARN AND RECOGNIZE PROTEIN FOLDS

Neil F. Abernethy

Purpose

Although a wealth of protein sequence and structural data is now widely available, this information remains difficult to digest for typical biologists. Even displays of protein structures often seem complex and difficult to recognize, understand, and remember.

The question this theoretical research seeks to address is whether innate human language skills can be used as a mechanism to help people learn to recognize protein folds. Our existing pattern-recognition skills may be useful for both raw amino-acid sequences and the two-dimensional projection of three-dimensional folds. For example, the 1996 CASP2 competition in protein-fold recognition was won by Alexey G. Murzin, a protein structure expert who scored higher than any of the competing bioinformatics applications. Is it possible that all biologists could become "experts" by using their natural language skills?

Methods

As a first test, a simplified set of known secondary and tertiary structural rules corresponding to particular amino-acid residues and sequences will be selected. These rules will be taught to non-biologists who will then be tested on their ability to recognize the simplified patterns in amino acid sequences. For instance, one rule could be that a string of characters consisting of (A,D,E,F,H,I,L,K,M,Q,W,V) would be classified as "spiral" (alpha-helix-forming), since these are the amino acids with alpha-helix-forming tendencies. Use of non-biologists will control for background experience and encourage a natural approach to the pattern-recognition problem.

A second test will examine a subjects' ability to recognize either the two-dimensional icons representing the topology of proteins or the actual projection of a three-dimensional structure. In both cases the subjects would attempt to identify the fold class of a protein. Prior work in the field has generated hand-drawn iconographs representing protein secondary structure and basic fold.

Applications

If successful, this research could lead to better education for molecular biologists, enabling them to visually recognize protein motifs. This could greatly deepen their understanding of the sequence/structure/function relationship in proteins, and help them better utilize the structure images generated by display software. Such skills would prove valuable in the coming era of therapeutic protein design.

A "legible" protein iconography would be extremely useful to scientists attempting to visually understand gene/protein networks on a page. Finally, it may motivate earlier training of structural rules as a part of the rapid language acquisition of childhood years.

Future work

If non-biologists show an ability to recognize these patterns, we may suspect that they are using linguistic or spatial-cognitive abilities. To further refine our understanding of what cognitive abilities are under use, this ability could be explored with functional neuroimaging. This technique pinpoints the areas of the brain being used to process information or perform tasks. The results could then be compared and contrasted to existing data from various language and spatial information processing tasks.

This work is in a largely theoretical stage - interested potential collaborators are encouraged to contact the author.

STRUCTURE AND STABILITY OF COLLAGEN

Sean Mooney, Teri Klein

Collagen is the most abundant protein in mammals, comprising over 28% the total dry protein weight. Unlike globular proteins, collagen is a fibril protein identified by the presence of a triple helical domain that has a regular X-Y-Gly repeating amino acid sequence. Interruptions in the X-Y-Gly repeat in collagen can cause diseases such as Osteogenesis Imperfecta and disorders such as Ehrlos Danos Syndrome. An understanding of the structure and the factors that contribute to the stability of collagen will lead us to a better understanding of how mutations in collagen causes disease.

Because of its regular structure, the triple helix of collagen can be modeled using short, idealized collagen-like peptides. We are using clinical, experimental and theoretical results to guide molecular mechanics studies of collagen-like peptides, with the goal of building a model that can predict structural and thermodynamic changes that occur when mutations are introduced into the triple helix of collagen.

We have built models of several collagen-like peptides to address these questions. Our models quantitatively reproduce the thermodynamics of introducing mutations into the glycine position of the repeating X-Y-Gly triplet motif (Klein, *et al*. Biopolymers, 1999 and Mooney, *et al*. Biopolymers, In Press). We are currently using our models to better understand how hydroxyproline stabilizes the triple helix and to better understand the structural changes that occur when mutations that cause lethal Osteogenesis Imperfecta are present.

COMBINING KINETIC INFERENCE WITH A PREDICTOR-CORRECTOR METHOD TO MODEL GENETIC REGULATORY CIRCUITS THAT ARE CONSISTENT WITH HETEROGENEOUS EXPERIMENTAL DATA

Nizar Batada, Mike Laub, Harley McAdams

Purpose

To infer kinetic parameters regarding gene transcription and translation and to investigate regulatory circuitry by correlating and checking the consistency of heterogeneous experimental data (genomics, proteomics) using mathematical modeling and simulations.

Materials and Methods

Microarray experiments were done with mRNA samples taken from synchronized Caulobacter cells taken at ten time points with 15-minute intervals over the 150 minute cell cycle to obtain gene expression data on genes involved in flagellar biosynthesis (Laub, M *et al*, Science, 2000, in press). Results from this time series expression profiles of genes involved in flagellar biosynthesis is compared to delay differential equation model of flagellar gene regulatory cascade and simulated using Matlab/Simulink software. The simulation model includes synthesis and degradation kinetics of proteins and mRNA as well as estimated time delays associated with transcription initiation and protein folding events.

Results

We have developed a functional simulation model that implements the differential equation model of a three gene class regulatory cascade involved in flagellar biosynthesis, and have extending it to incorporate several regulatory feedback mechanisms that regulate mRNA stability as variant regulatory architectures. We have demonstrated the utility of the predictor-corrector method where results predicted from the model can be compared to the time series to check for consistency and to propose and test new regulatory circuits.

Conclusion

We have demonstrated the utility of taking a systems-level perspective, by combining forward modeling with the reverse problem of network inference from the rich data generated from genomics and proteomics research. The potential for this predictor-corrector approach which takes into consideration heterogeneous data is enormous and enables identification of genes subject to postranscriptional regulation. Current work is focused on investigating whether it is possible to distinguish between autoregulated and non-autoregulated genes by identifying distinctive “signature” time series profiles.

AN INTERACTIVE BIOMECHANICAL MODEL OF THE HUMAN HAND

Robert Pao-Feng Cheng, Jean Heegaard, Parvati Dev, Sakti Srivastava, Leroy Heinrichs, Tonia Sengelin

Purpose

The human hand is controlled by a complex interaction of muscles, tendons, and other soft tissues. The manner in which the tendons act on the bones is not always obvious. In order to understand hand function, medical students often require the opportunity to interact with cadaver specimens. Due to the limited availability of specimens, students typically rely on static anatomic images to gain insight on the function. With current graphics hardware and software, the ability to create a virtual model of the hand is possible. A computational model also allows for repeated simulations of various tendon lesions, while the specimen can only be used once. In this project, we are developing a software application for the interactive manipulation of a 3D hand model. The model is designed to behave with the appropriate biomechanical constraints to produce realistic motions.

Materials and Methods

A full model of the human hand has been obtained from Primal Pictures (London, UK). The model contains detailed geometric representations of the bones, cartilage, tendons, ligaments, muscles, and nerves. These models were placed into a 3D environment, using the CosmoWorlds software (SGI, Mountain View, CA), where the joint axes could be defined. The models were also decimated (up to 40%) to increase the responsiveness of the model during interaction. Only the bone and cartilage models are used because we are mainly concerned with the motion of rigid bodies in the system. Soft tissue deformation algorithms are required if the tendons, muscles, and skin are to be included.

We are using an open source graphics software, the visualization toolkit (vtk), for development of the application and interface. The flow of data in the application goes from an input motion or force applied to the model to a dynamics solver, which computes the new position of each of the objects in the system. The updated positions are then relayed back to the visualization software for display. The dynamics equation solver is implemented in C++ to ensure adequate update rates for the graphics display.

Results

A set of animations demonstrating the function of normal fingers, as well as fingers with tendon lesions, has been created. The series of motions include: flexion at the MCP (metacarpophalangeal) joint, flexion of the DIP (distal interphalangeal) and PIP (proximal interphalangeal) joints, thumb abduction/adduction, thumb flexion/extension, and thumb opposition. These 3D animations can be viewed with any VRML (Virtual Reality Modeling Language) enabled web browser. An image from the animation is shown in Figure 1. The function of the flexor digitorum profundus, flexor digitorum superficialis, extensor digitorum, and intrinsics are included.

A prototype of the application interface has been developed using the python and Tkinter programming languages with vtk. The prototype uses slide bars to control the joint angles of the finger (shown in Figure 2). The limits of the joint rotations are constrained to lie within physical limits.

Conclusion

While the primary work accomplished has been based on preset animations, continued work with vtk is aimed at permitting interaction with the model. The inclusion of soft tissue behavior will be required in the

computational model to deliver physiologic motion. Though the prototype includes the behavior of a single finger, the goal is to have a model of the complete hand. In the next step, the dynamics model will be interfaced with a haptics device to provide force feedback while interacting with the hand.

Web Page

`http://www.stanford.edu/~alief/bcats.html`

IMPLEMENTATION OF A RADIO-FREQUENCY INTRAVASCULAR ULTRASOUND SYSTEM FOR QUANTITATIVE TISSUE CHARACTERIZATION IN CORONARY ARTERIES

Brian Courtney, Abel L. Robertson, Paul G. Yock, Peter J. Fitzgerald

Introduction

A method to perform in vivo characterization of tissues within coronary arteries would have several important applications. Clinically, knowledge of the structure and composition of an atherosclerotic lesion provides valuable information with respect to the likelihood of the onset of acute myocardial infarction. Tools to assist in the identification of vulnerable plaques in vivo would therefore assist greatly in the clinical management of coronary artery disease. Similarly, research of coronary artery disease can benefit from in vivo methods to monitor changes in the composition of coronary arteries over time. Such methods would provide new insights into the progression of disease and facilitate studies involving different modalities of intervention, such as angioplasty, atherectomy, brachytherapy, stenting, angioplasty and pharmacological agents.

Intravascular ultrasound (IVUS) is a method that produces two-dimensional cross-sectional images of arteries and is currently used for assessing coronary lesions in many clinical and research centers. Such assessments generally consist of qualitative descriptions of real-time video images and geometric measurements within the images.

In order to enable more quantitative measurements of intravascular ultrasound, and to extract different measurements from the ultrasound signals used to produce the image, a radio-frequency ultrasound data acquisition and analysis system has been developed. The system enables quantitative measurements through an ultrasound image-based interface. The hardware and software acquire high frequency (500 MHz) digitally sampled records of unprocessed ultrasound signals that may contain more information about the tissue structure than the envelope of highly processed ultrasound signals used to produce traditional IVUS images.

Method and Equipment

The system consists of a personal computer equipped with a 500 MHz 8-bit analog to digital converter, controlled by custom software and connected to a radio-frequency output connector of an IVUS console. The software has several additional functions, including digital filtering, image reconstruction, transducer calibration and the interactive measurement of several ultrasound parameters. Measurable parameters include backscatter intensity, statistics regarding the envelope of the signal, attenuation, geometric information of the vessel components and frequency content. Video reconstruction, high volume data management and data exportation make the system flexible for several research purposes.

Uses and Future Directions

This system has been and continues to be used in several in vivo and in vitro models and has provided insights into the interactions of ultrasound with different tissue types under various conditions. New measurements and signal processing techniques will continue to be added to the software. Ultimately, it would be desirable to include an inference engine to the software so that tissue components and important geometric features could be algorithmically detected based on a collection of several measurements within segments of the vessel cross-section.

Although greater processing power, data storage mechanisms and higher resolution of the digitized ultrasound signals will eventually be incorporated to facilitate

widespread adoption of radio-frequency IVUS, the current system is easy to use and provides important information into the development of minimally-invasive tissue characterization methods based on IVUS.

TWO SIDED CLUSTERING FOR YEAST GENE EXPRESSION USING PROBABILISTIC RELATIONAL MODELS

Eran Segal, Ben Taskar, Daphne Koller

DNA microarray technology is currently producing a wealth of gene expression data on genome-wide scale. Much work has focused on clustering genes and experiments with similar expression level. However, most methods perform clustering of genes and experiments separately and then combine the results. Furthermore, these methods ignore significant information about both genes and experiments that can aid in discovering more accurate and significant clusters. For example, cellular role, biochemical function, and localization may be known for some genes, and type of tissue and conditions are often known about the experiments. We present a novel approach that incorporates information about genes and experiments in a unified probabilistic model and allows to cluster both genes and experiments simultaneously.

Our methods are based on probabilistic relational models (PRMs), which extend the standard attribute-based Bayesian network representation to incorporate a rich relational structure. A PRM specifies a template for a probability distribution over a set of (complex) objects. It specifies, for each type of entity in the domain, a dependency model for each attribute of that entity. This model encodes the dependence of the attribute of an object on other attributes of this and related objects.

Our PRM schema consists of an object for each gene and an object for each experiment, with a many-to-many relation between them containing the expression level measured for the gene in the experiment. Each gene and each experiment have a hidden attribute corresponding to the cluster, both influencing the expression level of the gene for that experiment. Thus, the expression level depends only on the cluster assignments of the gene and experiment. We can also consider richer structures where both the experiment object and the gene object have additional attributes that can influence the expression level. Thus, the cluster attribute would capture the residual dependence not explained by the observed attributes (e.g., the experiment type).

Given the measured expression levels for all genes, we learn the model parameters using EM. The EM procedure requires that we compute, in each iteration, the distribution over the hidden attributes. For gene expression data, the probabilistic model resulting from our two-sided clustering schema consists of tens of thousands of highly dependent objects, making the inference task computationally intractable. We therefore use an approximate belief propagation algorithm due to Pearl. Several groups have recently reported excellent experimental results by using this approximation scheme.

We ran our experiments on yeast gene expression data (`http://rana/clustering`). Each column represents an experiment, and each row a probe on the microarray designed to detect the expression level of a particular gene. We clustered this data using our two-sided clustering algorithm, and compared to a standard clustering approach (EM on a Naïve Bayes model) on genes and experiments separately. The results, see `http://robotics/~erans/twosided.html`, show that the clusters obtained by our approach are substantially more coherent.

PRMs provide a flexible general-purpose framework for representing models of complex biological processes such as gene expression. They can represent additional attributes, time series for the experiments, and gene expression pathways. We show that we can effectively learn even these very complex models from data.

44D

WEB APPLICATIONS FOR MICROARRAY DATA ANALYSIS AND PRESENTATION

Christian A. Rees, Charles M. Perou, Douglas T. Ross, Jonathan R. Pollack, J. Michael Cherry, Patrick O. Brown, David Botstein

Microarray experiments generate large datasets that require computational tools for analysis and visualisation. Making these tools available as web applications allows researchers simple access with a web browser. We have developed two web applications for visualisation, analysis and publication of microarray data.

GeneXplorer uses a web browser as its interface to visualize large matrices of microarray gene expression datasets. A dataset matrix can contain thousands of genes and hundreds of experiments. The matrix is displayed as a color coded image. This makes it easy to detect patterns of similarly expressed genes if the matrix has previously been ordered by a clustering algorithm.

GeneXplorer provides different ways of analyzing the data: 1) The user can visually browse by clicking on the matrix image in an overview frame. A zoom of the selected region is then displayed in a separate frame together with gene names and hyperlinks to more detailed gene information. 2) A keyword search on the gene descriptions can be performed. All genes with a matching keyword in their description are displayed in the zoom frame. For different organisms gene information is easily configurable. For human genes the following fields are searchable: names, gene symbols, cDNA clone identifier, UniGene cluster identifier, GenBank accession numbers, chromosome and cytoband. The display images and hyperlinks of this information can be configured with a stylesheet. 3) To find the neighbors of a gene of interest, the user can click on the image representation of an individual gene. This will display the most similar genes in order of similarity together with the correlation value.

CaryoScope is a visualisation and analysis tool for microarray based Comparative Genome Hybridisation - aCGH (Pollack *et al.*, 1999). While determining genomic copy number changes with CGH has a resolution at the megabase level, array CGH can increase the resolution by orders of magnitude. Copy number changes can be mapped at the single gene level. Taking the CGH technique to the microarray platform requires a new approach to visualisation and analysis. The genomic location of cDNA clones is obtained by comparing their sequence - usually ESTs - to the human genome draft assembly (data provided by Jim Kent, UCSC). The ratios of normal DNA vs. cancer DNA are drawn by the CaryoScope web application as barcharts representing the chromosomes. Genetic markers are included as reference landmarks in the human genome. This way, chromosomal regions of amplification or deletion in human cancer cells can be determined. The graphical output of CaryoScope is provided in GIF, PostScript and Portable Document Format (PDF). Hyperlinks from each clone to additional gene information are added to the PDF output.

Web Page

`http://genome.stanford.edu/~rees`

45D

IRACS: A LITERATURE MINING TOOL FOR FAST INTERPRETATION OF MICROARRAY DATA

Sep Kamvar, Eldar Giladi, Jeanne Loring, Mike Walker

Purpose

The advent of the microarray has made large-scale gene expression studies thousands of times faster than previously possible. However, meaningful interpretation of this expression data is still a bottleneck to the discovery process, often taking several weeks or months of literature review. We seek to automate this interpretation procedure by developing a literature-mining tool to organize and summarize the literature in a manner helpful to biologists analyzing large-scale expression data.

Materials and Methods

We aim to (a) retrieve all MedLine articles pertaining to coexpressed or differentially expressed genes in a microarray experiment, and (b) find clusters of closely related articles within this document collection and summarize these clusters.

For document retrieval, we use standard term-matching methods. Clustering is achieved using Principal Direction Divisive Partitioning (Boley, 1998), a clustering algorithm based on generalizations of graph partitioning.

Results

This tool has proven to be highly effective in trials. In all trials, the analysis of large-scale gene expression data took under an hour, while the analysis of these same results took several weeks using traditional methods. In addition, this tool has led to novel insights in Alzheimer's disease and osteoporosis.

Conclusion

We present a new tool to aid biologists in interpreting gene expression data. This tool has been shown to be highly effective in organizing and summarizing the relevant literature in far less time than traditional methods used by biologists. Currently, gene expression data can be produced much faster than it can be interpreted, and we suggest that this tool can be significant in widening the bottleneck that slows the discovery process in functional genomics.

46D

SYMPOSIUM PARTICIPANT LIST

Neil Abernethy
Biomedical Informatics
nfa@smi.stanford.edu

Rami Aburomia
Genetics
aburomia@leland.stanford.edu

Burak Acar
Dept. of Radiology, School of Medicine
bacar@stanford.edu

Annette Adler
Agilent Technologies
annette_adler@agilent.com

Aneel Advani
Medical Informatics
advani@smi.stanford.edu

Susanne Elizabeth Ahmari
Molecular and Cellular Physiology
sahmari@stanford.edu

Ahmed Murad Akhter
Computer Science
murad@cs.stanford.edu

Gene Alexander, PhD
Mechanical Engineering
genealex@leland.stanford.edu

Russ Biagio Altman
Medicine
Russ.Altman@stanford.edu

Dong Anton An
Computer Science
antonan@stanford.edu

Kirk Anders
Genetics
anders@genome.stanford.edu

Clay Anderson, Ph.D.
Mechanical Engineering
fca@stanford.edu

Nancy Anderson
Undergraduate Advising Center
stfelix@stanford.edu

Kok Long Ang
Gyn & Ob
K.L.Ang@stanford.edu

Mehmet Serkan Apaydin
Electrical Engineering
mapaydin@stanford.edu

Lauren Marie Aquino
Mechanical Engineering
lauren.aquino@stanford.edu

Allison Arnold
BME Division, ME Department
asarnold@stanford.edu

Blake Ashby
Mechanical Engineering
bmashby@stanford.edu

Srinivasan B
Asia Pacific Research Centre, Inst of International Studies
bsrini@stanford.edu

Brian Babcock
Computer Science
babcock@stanford.edu

Virginia Bachrach
Pediatrics
Virginia.Bachrach@stanford.edu

Pierre Barbero
Biochemistry
pbarbero@cmgm.stanford.edu

Leah Ortiz-Luis Barrera
Math and Computational Science
leahob@stanford.edu

Dr. John Bashkin
SRI International
john.bashkin@sri.com

Sanmit Basu
Mechanical Engineering, Division of Biomechanical Engineering
sbasu1@leland.stanford.edu

Nizar Batada
Developmental Biology
nbatada@stanford.edu

Gary Beaupré
Biomechanical Engineering Division
beaupre@bones.stanford.edu

Donia Larissa Bencke
Biology
dbencke@stanford.edu

Sandra Elizabeth Bendeck
medicine
sbendeck@stanford.edu

Steve Bennett
Biochemistry
encino@stanford.edu

Aviv Bergman
Center for Computational Genetics and Biological Modeling
Aviv@Stanford.edu

Elmer Victor Bernstam
General Internal Medicine: Stanford Medical Informatics
elmer.bernstam@stanford.edu

Gail Binkley
Department of Genetics
gail@genome.stanford.edu

Jon Binkley
Genetics
binkley@genome.stanford.edu

Terrence F Blaschke
Medicine/Clinical Pharmacology
blaschke@stanford.edu

Silvia Salinas Blemker
Biomechanical Engineering Division, Mechanical Engineering Department
ssblemker@stanford.edu

Jason Bock
Molecular and Cellular Physiology
jbbock@leland.stanford.edu

Roy Bohenzky
Roche Diagnostics
roy.bohenzky@roche.com

David Botstein
Genetics
Botstein@Genome.Stanford.Edu

Leah Bowser
biology
lbowser@stanford.edu

Edward Stuart Boyden
Neurosciences
boyden@stanford.edu

Richard William Bragg
Mechanical Engineering
rwbragg@stanford.edu

Dena Bravata, M.D., M.S.
Primary Care & Outcomes Research
bravata@healthpolicy.stanford.edu

Andrew Broderick
Stanford Research Institute
abroderick@sric.sri.com

Igor Eric Brodsky
Microbiology and Immunology
ibrodsky@leland.stanford.edu

Michael Brudno
Computer Science
brudno@cs.stanford.edu

Anh Bui
Biological Sciences
abui@stanford.edu

Elaine Carlson
Buck Institute
ecarlson@buckinstitute.org

Michelle Whirl Carrillo
Biophysics
mikki@leland.stanford.edu

Dennis R. Carter
Mechanical Engineering
dcarter@stanford.edu

John Cavallaro
Management Science and Engineering
jcavalla@stanford.edu

Urszula Chajewska
Computer Science
urszula@cs.stanford.edu

Dr. Albert S Chan
Department of Family Medicine
albert.chan@alumni.stanford.org

Lap Fung Chan
Electrical Engineering
berchan@stanford.edu

Ravi A. Chandrasekaran
Biological Sciences and Chemistry
ravic@stanford.edu

Jeffrey Chang
Medical Informatics
Jeffrey.Chang@stanford.edu

Doug N. Chang
cs
dougc@leland.stanford.edu

Chi Chang
Mechanical Engineering: Biomechanics Division
chichang@leland.stanford.edu

Celene Chang
Business/Engineering
chang_celene@gsb.stanford.edu

Austin Che
Computer Science
austin@stanford.edu

David Yu Chen
CS
dychen@stanford.edu

Kenneth Chen
Statistics
kenchen@stanford.edu

Mingying Chen
Computer Science
myc@stanford.edu

Christopher Paochung Cheng
Mechanical Engineering
Christopher.Cheng@stanford.edu

Carol Hsen-Fae Cheng
Biomedical Informatics
cheng@smi.stanford.edu

Jason Cheng
Biology and Computer Science
jjcheng@stanford.edu

Phillip Ming-Da Cheng
BMI
pmcheng@stanford.edu

Robert Cheng
Mechanical Engineering
alief@stanford.edu

Christine Cheng
Computer Science
shiangc@stanford.edu

Yu-Che Eddie Cheng
EE
chengye@stanford.edu

Wendy Cheng
Mechanical Engineering
wcheng@alum.mit.edu

Mike Cherry
Genetics
cherry@stanford.edu

Steve Chervitz
Neomorphic, Inc. (Affymetrix)
sac@neomorphic.com

Ming Chiang
ISIS Pharmaceuticals
mchiang@isisph.com

Kyeongjae Cho
Mechanical Engineering
kjcho@leland.stanford.edu

Edward Choice
Pediatrics
ed_choice@hotmail.com

Evan Chou
Computer Science
echou@cs.stanford.edu

Grace Chou
SRI International
grace.chou@sri.com

Douglas Chow
Graduate School of Business
Chow_Douglas@gsb.stanford.edu

Andrzej Chruscinski
Medicine
andrzej@cmgm.stanford.edu

Su Chung
San Diego Supercomputer Center
suchung@ispchannel.com

Janice Yu-Hsin Chyou
Undeclared
janice.chyou@stanford.edu

Erin Cline
Molecular and Cellular Physiology
erin.cline@stanford.edu

Brian Courtney
School of Medicine
bcourtney@computer.org

Craig Anthony Cummings
Microbiology and Immunology
cummings@cmgm.stanford.edu

Dr. Ronald L. Dalman
Surgery
rld@stanford.edu

Oranee Daniels, MD.
Division of Clinical Pharmacology
oranee@stanford.edu

Mindy Davis
Chemistry
midavis@stanford.edu

Jerel Clayton Davis
Biological Sciences
jerel@stanford.edu

Ed Davis
SRI International
edward.davis@sri.com

Ulrike DeMarco
Psychology
ulrike.demarco@stanford.edu

Peter Dehlinger
Iota Pi Law Group
peter@iotapi.com

Scott L Delp
Mechanical Engineering
delp@stanford.edu

Corrie Detweiler
Stanford Department of Microbiology
corried@leland.stanford.edu

Parvati Dev
SUMMIT, School of Medicine
parvati.dev@stanford.edu

Erlind Nasufi Dine
Graduate School of Business
dine_erlind@gsb.stanford.edu

Kara Dolinski
Genetics
kara@genome.stanford.edu

Magdalena Dorywalska
Structural Biology
magg@stanford.edu

Mary Therese Draney
Mechanical Engineering
draney@stanford.edu

Katerina Athena Drouvalakis
Medicine
drouvaka@stanford.edu

Chenggang Duan
Computer Science
duancg@stanford.edu

Chris Duffield
Stanford Materials Science & Engineering Dept.
chris@iptq.com

Jonathan Dugan
BMI
dugan@stanford.edu

Maitreya Dunham
Genetics
maitreya@stanford.edu

Phillip Marks Ecker
med
pme@stanford.edu

David Elgart
Genencor
davidelgart@hotmail.com

Sid Elmer
Chemistry
sidnasty@stanford.edu

Gerald Engel
Mechanical Engineering
gerengel@stanford.edu

Carol C. Epstein, Ph.D.
BioInfoStrategies
carol@bioinfostrategies.com

Christian Eversull
Medicine
eversull@stanford.edu

Rob Ewing
Carnegie Inst
ewing@genome.stanford.edu

Larry Fagan
Stanford University
fagan@smi.stanford.edu

Zhenbin Fan
Urology
Zhenbin.Fan@stanford.edu

Daryl Faulds
Berlex Biosciences
daryl_faulds@berlex.com

Chris Feezor
Guidant Corporation
cfeezor@guidant.com

Gonzalo Raul Feijoo
Mechanical Engineering
grfeij@stanford.edu

Yanan Feng
Genetics
feng@stanford.edu

Tracy Ferea, Ph.D.
Applied Biosystems
fereatl@appliedbiosystems.com

Dr. Michael John Fero
School of Medicine
mfero@genome.stanford.edu

Patrick Alexander Fleisch
Engineering
pfleisch@leland.stanford.edu

Kelly Frazer
Affymetrix, Inc.
kelly_frazer@affymetrix.com

Robert French
robert.french@mciworld.com

Jie Gao
Computer Science
jgao@leland.stanford.edu

MArgarita Garcia
TAIR/Carnegie Institution of Washington
garcia@acoma.stanford.edu

Audrey Gasch
Biochemistry
agasch@leland.stanford.edu

Lise Carol Getoor
Computer Science
getoor@cs.stanford.edu

Dr. Gary Gilbert
Telemedicine and Advance Technology Research Center
Gilbert@TATRC.ORG

David McPherson Goehring
Biological Sciences
goehring@stanford.edu

Salih Burak Gokturk
Electrical Engineering
gokturkb@stanford.edu

Saryn Goldberg
Mechanical Engineering, Biomechancis Division
saryn@stanford.edu

Seri Gomberg
iKnowMed
seri@wcf.com

Matthew Gonzales
Div. of Infectious Diseases
lgonz@stanford.edu

Justin Graham
Stanford Medical Informatics
jus10@stanford.edu

Randy Grow
Applied Physics
rgrow@stanford.edu

Ming Gu
Computer Science
minggu@cs.stanford.edu

Kyle Alan Gurley
Developmental Biology
kylegurley@stanford.edu

Katsuyuki Hoshina
Vascular Surgery
khoshina@leland.stanford.edu

Scot Lee Haire
Mechanical Engineering Dept, Flow Physics and Computation Div
slh@stanford.edu

Joan Hebert
Genetics
jhebert@genome.stanford.edu

Adolff Theodorus van Der Heide
Mechanical Engineering (Division of Biomechanics)
dolf@stanford.edu

W LeRoy Heinrichs
Gynecology & Obstetrics/SUMMIT
leroy.heinrichs@stanford.edu

Christopher John Hernandez
Mechanical Engineering
chernand@stanford.edu

Catherine Hettinger
eng
chetting@stanford.edu

Carol A. Hill
CliniCon
CliniCon@aol.com

Tad Hogg
Xerox PARC
hogg@parc.xerox.com

Karin Hollerbach, Ph.D.
karin.hollerbach@alum.mit.edu

Bret Alan Holley
Biological Sciences
bholley@stanford.edu

Bret Holley
Biological Sciences
bholley@stanford.edu

Lena Hong
lyhnahong@yahoo.com

Zachary Dolph Hornby
Biological Sciences
zhornby@stanford.edu

Brita Hornung
Anesthesia
brita.hornung@stanford.edu

Gabriel Howles
Biology
ghowles@stanford.edu

Allison Yen-Ling Hsieh
alihsieh@stanford.edu

Jerry Yungchi Hsu
Cancer Biology
jhsu@stanford.edu

Kurt Huang
Biomedical Informatics
khuang@smi.stanford.edu

Suttiporn Janenawasin
psychiatry
jsutti@stanford.edu

Guha Jayachandran
Computer Science
guha@stanford.edu

Michael Christopher Jewett
Chemical Engineering
mjewett@leland.stanford.edu

Xuhuai Ji
Department of Medicine/Division of Gastroenterology
xuhuai.ji@stanford.edu

Audrey Jia
Protein Design Labs, Inc
ajia@pdl.com

Rong Jiang
computer science
kjiang@stanford.edu

Jeremy Aaron Johnson
Electrical Engineering
jeremyj@stanford.edu

Betsy Johnston
Infectious Diseases
betsyj@leland.stanford.edu

Keith Joho
Abgenix, Inc.
joho_k@abgenix.com

Sunghae Joo
Nanogen, Inc.
sjoo@nanogen.com

Andy Kacsmar
Computer Science
andy@db.stanford.edu

Herbert Kaizer
Medicine
kaizer@smi.stanford.edu

Oliver Kaljuvee
Computer Science
oliver.kaljuvee@stanford.edu

Sep Kamvar
SCCM
sdkamvar@stanford.edu

Kenneth Sye-young Kang
Computer Science
ksykang@leland.stanford.edu

John Kang
Alpha Innotech Corporation
jkang@aicemail.com

Rami Kantor
Division of Infectious Diseases
rkantor@stanford.edu

Fiona Kaper
Radiation Oncology
fkaper@stanford.edu

Peter Kasson
Biophysics Program
kasson@stanford.edu

Brett Taketsugu Kawakami
Civil and Environmental Engineering
brettk@stanford.edu

David Kiang
Medicine/Gastroenterology
dkiang@leland.stanford.edu

Charlie Kim
Microbiology and Immunology
cckim@stanford.edu

Dong-Hyun Kim
Electrical Engineering
dhkim@stanford.edu

Teri E. Klein
Stanford Medical Informatics
klein@smi.stanford.edu

Uwe Klein
Advanced Medicine, Inc
uklein@advmedicine.com

Tod Klingler
Prospect Genomics
tod@prospectgenomics.com

Pete Klosterman
UC Berkeley / Lawrence Berkeley Lab
pete@compbio.berkeley.edu

Alex Kobler
GSB
alex.kobler@stanford.edu

Robert Kohlenberger
Applied Biosystems
kohlenrw@appliedbiosystems.com

Daphne Koller
Computer Science
koller@cs.stanford.edu

Charlene Kon
Developmental Biology
charlene_kon@stanford.edu

Christian Johannes Korth
MS and E
korth@stanford.edu

Kalpagam Kowsik
Chabot College
kaku25@hotmail.com

John R Koza
Stanford Medical Informatics
koza@stanford.edu

Lee G. Kozar
Bioinformatics Resource
kozar@stanford.edu

Aruna V. Krishnan
Medicine/Endocrinology
Aruna.Krishnan@Forsythe.stanford.edu

Tom Krummel
Surgery
tkrummel@stanford.edu

Christopher Kueny
Lawrence Livermore National Laboratory
kueny1@llnl.gov

David Kulp
Neomorphic, Inc

Eric J. Kunkel
Pathology
ejkunkel@cmgm.stanford.edu

Chuck Pui Lam
Electrical Engineering
chucklam@stanford.edu

Stefan Larson
Biophysics
smlarson@stanford.edu

Ivy Ann Lee
Biology
ivylee@leland.stanford.edu

Choonghyun Lee
Computer Science
chlee@cs.stanford.edu

Ann Lee-Karlon
Business
AnnL@stanford.edu

Michael Levitt
Structural Biology
michael.levitt@stanford.edu

Kaijun Li
Pathology
kli@stanford.edu

Wenchuan Liang
Biochemistry
wliangy@yahoo.com

Mike Hsin-Ping Liang
Stanford Medical Informatics
mliang@stanford.edu

DeYong Liang
Neurobiology
dyliang@stanford.edu

Yung S. Lie
Biological Sciences
yung@stanford.edu

Jason Lih
Genetics
jasonlih@stanford.edu

Min Chin Lim
Biological Sciences
mclim@stanford.edu

Connie Lin
Biology
cslin@stanford.edu

Zhen Lin
Medicine, Stanford Medical Informatics
Zhen.Lin@stanford.edu

Richard Lin
Computer Science
richlin1@leland.stanford.edu

Yuhong Liu
geological and environmental science
yuhong@leland.stanford.edu

Michael Ming-Cheng Liu
Civil and Environmental Engineering
mmliu@leland.stanford.edu

Yueyi (Irene) Liu
Biomedical Informatics
yliu@stanford.edu

May Liu
Biomechanics Division, Mechanical Engineering
mql@stanford.edu

Xiaole Liu
Stanford Medical Informatics
xliu@stanford.edu

Shuo Liu
Biomedical Informatics
shuo.liu@stanford.edu

Michael Liu
Biology/Computer Science
michaelliu@stanford.edu

H.F. Machiel Van der Loos, PhD
Functional Restoration (Consulting Assistant Professor)
vdl@stanford.edu

Rita Lopatin
Cygnus, Inc.
mlopatin@cygn.com

Ann Loraine
Neomorphic
loraine@neomorphic.com

Hui-Ling Lu
EE
vickylu@leland.stanford.edu

Charity Yueh-chwen Lu
Computer Science
clu@cs.stanford.edu

Walter Jaren Luh
Computer Science
wluh@leland.stanford.edu

Terry S. Desser M.D.
Radiology
desser@stanford.edu

Jiong Ma
Biology
jiongm@stanford.edu

Gregory Marsden
Computer Science
gmarsden@stanford.edu

Susana Martins
VA Palo Alto Health Care System
susanamartins@writeme.com

Mary Mata
Nanogen, Inc.
mmata@nanogen.com

John Charles Matese
Genetics
jcmatese@genome.stanford.edu

Alexander F. Mayer
Affymetrix
amayer@neomorphic.com

Frederic Mazzella
National Biocomputation Center
fred@biocomp.stanford.edu

Scott E Mcphillips
Stanford Synchrotron Radiation Laboratory
scott.mcphillips@stanford.edu

Craig Meyer
Electrical Engineering
cmeyer@stanford.edu

John Joseph Michon
Biomedical Informatics
michon@smi.stanford.edu

Nesanet Senaite Mitiku
Medicine, Genetics
nesanet@leland.stanford.edu

Subhasish Mitra
Electrical Engineering
smitra@CRC.stanford.edu

Shannon Elizabeth Moffett
medicine
smoffett@stanford.edu

Joshua Irving Molho
Mechanical Engineering
josh.molho@mems.stanford.edu

Ja Moon
Cooley Godward LLP
jmoon@cooley.com

Sean Mooney
Stanford Medical Informatics
mooney@smi.stanford.edu

Edward William Moore
Human Biology
tedmoore@stanford.edu

Katherine Moore
Las Positas College
kitty_m@pacbell.net

Dan Morris
Computer Science
dmorris@tiqit.com

Joseph M. Morris
Affymetrix/Neomorphic
joe@neomorphic.com

Willy Moss
Microbiology and Immunology
wmoss@cmgm.stanford.edu

Yannick Moy
CS
ymoy@stanford.edu

Lukas A Mueller
Carnegie Institution of Washington
Lukas.Mueller@stanford.edu

Mark A Musen
Medicine (Medical Informatics)
MUSEN@smi.stanford.edu

Ankur Nagaraja
Biology
nagaraja@stanford.edu

Rob Nail
Velocity11
rob@velocity11.com

Brad Nakatani
Chemistry
bjn@stanford.edu

Girish Narayan
Cardiology
gnarayan@stanford.edu

Rosa Ines Navarro
Biological Sciences
rnavarro@stanford.edu

Krishna S. Nayak
Electrical Engineering
nayak@lad.stanford.edu

David Neale
Applied Biosystems
dpneale@earthlink.net

Kelvin Ming-Wei Neu
Immunology
neu@stanford.edu

Lan T. Nguyen
Graduate School of Business
nguyen_lan@gsb.stanford.edu

Rakesh Nigam
Mathematics
rakesh@quake.stanford.edu

Dave Nix, Ph.D.
Medicine
dnix@stanford.edu

Nassim Nouri
Affymetrix
nassim_nouri@affymetrix.com

Patrick John O'Brien
Biochemistry
pobrien@leland.stanford.edu

Mary O'Connell
Biomechanical Engineering
oconne1l@leland.stanford.edu

Brian O'Connor
iScribe
brian.oconnor@stanfordalumni.org

Assad Anshuman Oberai
Mechanical Engineering
oberai@stanford.edu

Christine Olsson
Deltagen, Inc.
colsson@deltagen.com

John G. Olyarchuk, MD
Medicine
jgo@smi.stanford.edu

Jessica Hammond Owens
Cancer Bio
jesshs@stanford.edu

Ramesh Padala
CS
rpadala@stanford.edu

Rasmus Pagh
Computer Science
pagh@theory.stanford.edu

Jodi Paik
HRP
jle@stanford.edu

David Paik
Stanford Medical Informatics
paik@smi.stanford.edu

Chana Palmer
Genetics
cpalmer@stanford.edu

Shyam N. Panchal
Cardiovascular Medicine
snpanchal@cvmed.stanford.edu

Hoi-Cheung Pang

hcpang@stanford.edu

Hyunsun Park
Iconix Pharmaceuticals, Inc.
hpark@iconixpharm.com

Aarati Parmar
Computer Science
aarati@cs.stanford.edu

Phil Payne
Protein Design Labs
PPAYNE@PDL.COM

Boris Peker
Biophysics
boris1@stanford.edu

Mor Peleg
Medicine
mor.peleg@stanford.edu

Kent Peterson
SRI International
kent.peterson@sri.com

William Petitt
Biomedical Informatics
bill@smi.stanford.edu

Nicolas Peyret
Applied Biosystems
peyretnn@appliedbiosystems.com

Hamid R Abbasi MD PhD
Neurosurgery
hamid@igl.stanford.edu

Jan Benjamin Pietzsch
Management Science and Engineering
pietzsch@stanford.edu

Zachary Pincus
Biological Sciences
zpincus@stanford.edu

Elizabeth G. Loboa Polefka
Mechanical Engineering
egloboa@stanford.edu

Murali Prakriya
Molecular and Cellular Physiology
prakriya@stanford.edu

Carla Marie Pugh
School of Education
cpugh@stanford.edu

Prasanth Pulavarthi
Computer Science
prasanth@stanford.edu

You-Wen Qian
Medicine
ywqian@leland.stanford.edu

Attila Racz
UCSF
ati@itsa.ucsf.edu

Tanya M. Raschke
Structural Biology
tanya.raschke@stanford.edu

Rosalind M Ravasio
Medicine (Medical Informatics)
ravasio@smi.stanford.edu

Soumya Raychaudhuri
Medecine
tumpa@stanford.edu

Paul Reicherter, MD
Dermatology
drpderm@aol.com

Leonore Reiser
The Arabidopsis Information Resource/Carnegie Institution of Washington
lreiser@acoma.stanford.edu

Martin Axel Reznek
Surgery
mreznek@hotmail.com

Wito Richter
GYN/OB
witorichter@web.de

Gabriel del Rio
The Buck Institute
gdelrio@buckinstitute.org

Adam Josef Rodriguez
rodriguez_adam@gsb.stanford.edu

Rob Rogers
Business
rrogers@gsb.stanford.edu

Jessica Ross
BMI
ccross@leland.stanford.edu

Michael Ross
Computer Science
micross@stanford.edu

Daniel Rubin, MD
Stanford Medical Informatics
rubin@smi.stanford.edu

Daniel B. Russakoff
Computer Science
dbrussak@stanford.edu

Michael G. Shulman
Biomedical Consulting
mshulman@pacbell.net

Pinkesh Sachdev
Electrical Engg
pinks@stanford.edu

Bauback Safa
School of Medicine
bsafa@stanford.edu

Khaled Nabil Salama
EE
knsalama@leland.stanford.edu

Alok Jerome Saldanha
Genetics
alok.saldanha@stanford.edu

Peter Salzman
MS&E
peter.salzman@stanford.edu

Ram Samudrala
Structural Biology
ram.samudrala@stanford.edu

Brynnen Noelle Sandoval
Biology
Brynnen.Sandoval@stanford.edu

Kavita Yang Sarin
Medicine
ksarin@stanford.edu

Serge Saxonov
Biomedical Informatics
saxonov@stanford.edu

Peter Leif Schilling

peter.schilling@stanford.edu

Thomas Michael Schmid
Business
thomas.schmid@stanford.edu

George Christopher Scott
Biomedical Informatics
gcscott@stanford.edu

Eran Azriel Segal
Computer Science
eran@cs.stanford.edu

Adam Seiver
Surgery
adam.seiver@forsythe.stanford.edu

Peter K. Seperack
Skjerven Morrill MacPherson, LLP
pseperack@skjerven.com

Anand Sethuraman
Biochemistry
asethura@cmgm.stanford.edu

Ross D Shachter
Management Science and Engineering
shachter@stanford.edu

Maulik Kamlesh Shah
Computer Science
Maulik@stanford.edu

John Sheehan
Affymetrix, Inc.
john_sheehan@affymetrix.com

Sandra Joan Shefelbine
Mechanical Engineering - Biomechanics
sandras@stanford.edu

Earl R. Shelton
Kowa Research Institute
eshelton@kowa.com

Smadar Shiffman
Psychiatry
SHIFFMAN@STANFORD.EDU

Hidetoshi Shimodaira
Department of Statistics
shimo@stat.stanford.edu

Michael Randall Shirts
Chemistry Department
mrshirts@stanford.edu

Eiketsu Sho
Vascular surgery
jsho@stanford.edu

John Shon
Internal Medicine
john@smi.stanford.edu

Jennifer Ann Shumilla
Pediatrics
jennifer.shumilla@stanford.edu

Arend Sidow
Pathology and Genetics
arend@stanford.edu

Mark Siegal
Biological Sciences
Mark.Siegal@stanford.edu

Natalie Simmons
natalie.simmons@stanford.edu

Alexander Simon
Pathology and Genetics, Program in Cancer Biology
asimon@stanford.edu

Nita Singh
EE
nita1@stanford.edu

Rohit Singh
Computer Sciene
rohitsi@cs.stanford.edu

Amit P. Singh
Biomedical Informatics
apsingh@cmgm.stanford.edu

Sheela Singla
medicine
ssingla@stanford.edu

Katharine Elise Skillern
Medicine
meow@leland.stanford.edu

David Alan Socks
GSB
dsocks@windamerevp.com

Manoon Somrantin
Cardiovascular Medicine
manoon.somrantin@stanford.edu

Ruchira Sood
Biochemistry
rsood@stanford.edu

Alexis Sowa
Human Biology
asowa@stanford.edu

Kunju Joshi Sridhar
Hematology
hf.kun@forsythe.stanford.edu

Brooke Noelani Steele
Mechanical Enginering
bnsteel@stanford.edu

Carl Steeves
Agilent Technologies
carl_steeves@agilent.com

Daniel Steines
Radiology
steines@stanford.edu

Fredrik Sterky
Department of Plant Biology, Carnegie Institution of Washington
sterky@genome.stanford.edu

Dr. Veronika Stoka
Buck Institute
vstoka@buckinstitute.org

Renee Patricia Stokowski
Genetics
labrat@stanford.edu

Derek Stonich
derekstonich@internetconnect.net

John David Storey
Statistics Department
jstorey@stanford.edu

Joshua Michael Stuart
Biomedical Informatics
stuart@smi.stanford.edu

Ted Su
Chemistry, Economics
teemu292@stanford.edu

Cenk Sumen
Microbiology & Immunology
csumen@stanford.edu

Ray-Hon Sun
SCCM
rsun@cs.stanford.edu

Patrick David Sutphin
Radiation Oncology
psutphin@stanford.edu

Alrik Suvari
Genentech, Inc.
suvari@gene.com

Srilatha R. Swami
Medicine/Endocrinology
sswami@cmgm.stanford.edu

Jim Swartz
Chemical Engineering
jim.swartz@stanford.edu

Michael Sykes
Biophysics
sykes@stanford.edu

Yuichiro Takagi
structural biology
ytakagi@stanford.edu

Mary Tang
Electrical Engineering
mtang@snf.stanford.edu

Hua Tang
Statistics
huatang@leland.stanford.edu

Ashish Tara
Graduate School of Business
tara_ashish@gsb.stanford.edu

Benjamin M Taskar
Computer Science
btaskar@stanford.edu

Charles Anthony Taylor
Surgery
taylorca@stanford.edu

Kavitha Thangavelu
INCYTE GENOMICS
ktkavi@yahoo.com

Yvonne Thorstenson
Stanford Genome Technology Center
yvonne@sequence.stanford.edu

Rabin Tirouvanziam
Psychology
rabin@psych.stanford.edu

Carlo Tomasi
Computer Science
tomasi@cs.stanford.edu

Simon Tong
Computer Science
simon.tong@stanford.edu

Thodoros Topaloglou
Gene Logic Inc
thodoros@stanford.edu

Lorenzo Torresani
Computer Science
ltorresa@cs.stanford.edu

Kristina Nikolova Toutanova
Computer Science
kristina@cs.stanford.edu

Joseph D. Towles
Mechanical Engineering
towles@stanford.edu

Olga G Troyanskaya
Biomedical Informatics
olgat@stanford.edu

Matthew Tsang
Biological Science
matthewtsang@stanford.edu

James Turner
Molecular Dynamics
jim.turner@am.apbiotech.com

Jonathan Andrew Usuka
chemistry
usuka@Stanford.edu

Priya Venkatesan
Symbolic Systems and Biology
priyav@stanford.edu

Hugo O Villar
Telik, Inc.
hugo@telik.com

Jing Wan
Petroleum Engineering
jingw@pangea.stanford.edu

Justin Wan
SCCM Program
wan@sccm.stanford.edu

Ping Wang
CS/EE/Bio
pingman@stanford.edu

Alfred Yu-Leen Wang
Molecular Pharmacology
aywang@stanford.edu

James Warren
Scientific Computing and
Computational Mathematics
warren@sccm.stanford.edu

Allison Waugh
Computer Science
alli@cs.stanford.edu

Thomas Scott Wehrman
Molecular Pharmacology
wehrman@leland.stanford.edu

Silvia Weinberger
San Jose State University
silviaolah@hotmail.com

Jacqueline Nerney Welch
Medicine, Mechanical Engineering
welch@stanford.edu

Peizhong Wen
Cardiovascular medicine
vpwen@stanford.edu

Yi-shin Weng
health research and policy
yweng@stanford.edu

Jason Brian Whitt
Business School
whitt_jason@gsb.stanford.edu

Sutanto Widjaja
WineShopper.com
sutanto_widjaja@hotmail.com

Gio Wiederhold
CSD and Medicine
GIO@CS.stanford.edu

Eric P. Wilkinson
Image Guidance Laboratory,
Department of Neurosurgery
epw@stanford.edu

Marna Williams
Pathology
mwilliam@cmgm.stanford.edu

Kim Williams
Biological Science
kew@leland.stanford.edu

Glenn Williams
Medical Informatics
gaw@smi.stanford.edu

Cyrus A. Wilson
Biochemistry
cyrus@stanford.edu

Lisa Wong
Biophysics
ljwong@stanford.edu

Stacey Woo
Human Biology
staceywoo@stanford.edu

Jim Wood
Crosby, Heafey, Roach & May
jwood@chrm.com

Kim Woodrow
Chemical Engineering
kwoodrow@stanford.edu

Kristina Nicole Woods
biphysics
kwoods@stanford.edu

Dr. John Wooley
University of California San Diego
jwooley@ucsd.edu

Shu-Hsing Wu
Carnegie Institution
shu@andrew2.stanford.edu

Jenny Wu
Ciphergen Biosystems, Inc
jennywu@ciphergen.com

Yu Xia
Department of Structural Biology
yuxia@csb.stanford.edu

Qunong Xiao
Computer Science
qxiao@ucla.edu

Wenzhong Xiao
biochemistry
wzxiao@stanford.edu

Yu Katherine Xu
Developmental Biology
ykxu@leland.stanford.edu

Chengpei Xu, MD, PhD
Surgery
cx11@stanford.edu

Haobo Xu
Appled Physics
HaoboXu@stanford.edu

Sanae Yamada
Graduate School of Business
Yamada_Sanae@gsb.stanford.edu

Jian Yang
Iconix Pharmaceuticals, Inc
jiyang@iconixpharm.com

Iwei Yeh
Biomedical Informatics
iyeh1@leland.stanford.edu

Krishna C Yeshwant
Computer Science
kcy@cs.stanford.edu

Golan Yona
Dept. of Structural Biology
golan@gimmel.stanford.edu

Elizabeth M Yu
eyu@stanford.edu

Ron Yu
SCCM
ronyu@sccm.stanford.edu

Xiang Yu
psychiatry
yuxiang@stanford.edu

Bojan Zagrovic
Biophysics
zagrovic@stanford.edu

James Francis Zawada
Chemical Engineering
jimmyz@stanford.edu

Shuli Zhang
Medicine
szhang@leland.stanford.edu

Hong Zhang
Genetics
zhanghon@leland.stanford.edu

Lu Zhang
Department of Medicine, Division of Hematology
ymysun@yahoo.com

Jian Zhang
Pherin Pharmaceuticals
pherin@earthlink.net

Kemin Zhou
Neomorphic
kzhou@neomorphic.com

Ji Zhu
Statistics
jzhu@stanford.edu

Feng Zhuge
Electrical Engineering
zhugef@stanford.edu

Jenny H. Zou
medicine
zou@stanford.edu

SPONSOR PARTICIPANTS / CONTACTS

DoubleTwist
Chris Campbell
chrisc@doubletwist.com

Andrew Kasarkis

Informax
Dennis Bittner
dbittner@informaxinc.com

Andrew Cogill
acogill@informaxinc.com

Tim O'Brien
tobrien@informaxinc.com

Jim Dickey
jdickey@informaxinc.com

Joel Haaf
jhaaf@informaxinc.com

Incyte Genomics
Timothy Nelson
tnelson@incyte.com

Tofoi Yandal-Moore
Cindy Georgette
Karen Wood
Rick Silvers

GeneLogic
Thodoros Topaloglou
thodoros@genelogic.com

Madhavan Ganesh
mganesh@genelogic.com

Kevin McLoughlin
kmcloughlin@genelogic.com

Krishna Papaniapan
krishna@genelogic.com

SGI
Jeffrey Hausch
jjh@sgi.com

Sun Microsystems
Jon Arikata
jon.arikata@eng.sun.com

Guidant
James Hong
jhong@guidant.com

Christopher Feezor
cfeezor@guidant.com

Reid Hayashi
rhayashi@guidant.com

Northern California Pharmeceutical Discussion Group
Eric Schuur
eschuur@pacbell.net

Jeffrey Flatgaard
ncpdg@best.com

Genencor International
Molly B. Schmid
mschmid@genencor.com

Donald Naki
dnaki@genencor.com

Jian Yao
jyao@genencor.com

Skjervan, Morril, MacPherson, LLP,
Peter Seperack
pseperack@skjerven.com

Christopher Allenby
callenby@skjerven.com

Gregory Powell
gpowell@skjerven.com

Signe Holmbeck
sholmbeck@skjerven.com

SYMPOSIUM SPONSORS

Full Sponsors

- **DoubleTwist**
- **Incyte**
- **Informax**
- **GeneLogic**
- **Guidant**
- **SGI**
- **Sun Microsystems**

Half Sponsors

- **Genencor International**
- **Northern California Pharmeceutical Discussion Group**
- **Skjervan, Morril, MacPherson, LLP,**

Please refer to the end of the Participant List for individual contact information.

DoubleTwist.com (privately held)

Vertical ASP

www.doubletwist.com *Oakland, CA*

DoubleTwist is an application service provider (ASP) devoted to empowering life scientists. The company provides research environments that leverage information technology and the World Wide Web to simplify and accelerate genomic research.

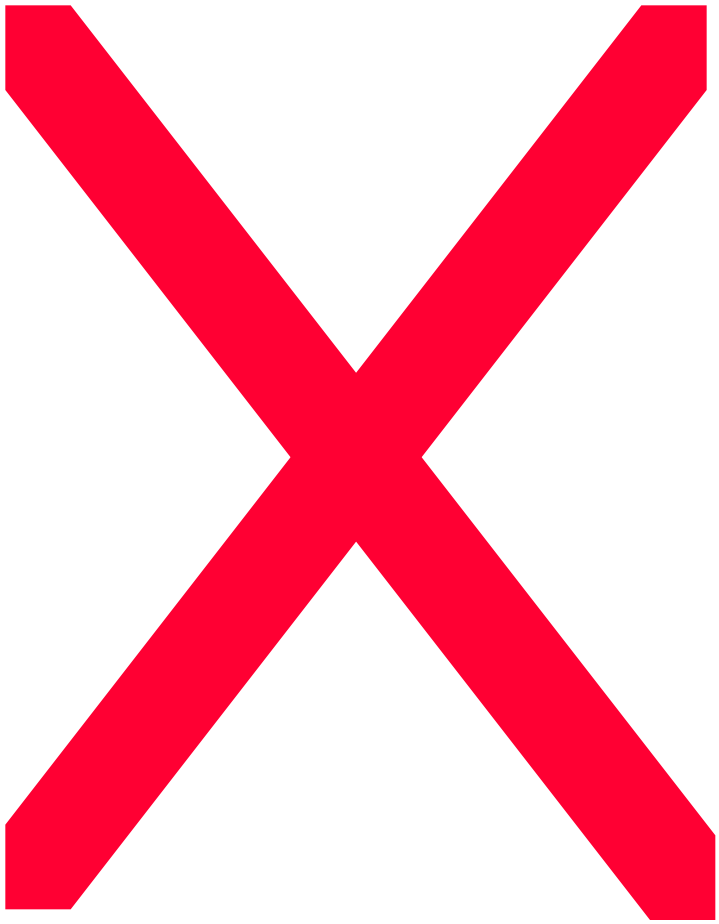
The company's leading product, DoubleTwist.com ™ is a secure and comprehensive online research environment that enables life scientists to perform sophisticated genomic analysis without requiring bioinformatics expertise. Subscribers to DoubleTwist.com receive access to intelligent and automated analysis tools and advanced, interactive software for the visualization of their research results. In addition, DoubleTwist.com provides a number of resources that support life science research, as well as e-commerce functionality and value-added content of relevance to life scientists.

The technology platform underlying DoubleTwist.com integrates more than 25 disparate genomic databases. These databases include public databases, databases licensed from third parties and strategic partners and the DoubleTwist, Inc. proprietary databases, such as the Annotated Human Genome Database and the Annotated Human Gene Index, which are created by processing and annotating the public genomic data. DoubleTwist has established several strategic relationships as a means of integrating additional features and content into DoubleTwist.com. Included in theses strategic relationships are Derwent Information Ltd., Myriad Genetics, Inc., Molecular Simulations, Inc., Chemdex, BioTools Inc., and Eragen Biosciences, Inc.

Launched in January 2000, DoubleTwist.com is located in Oakland, California, with additional offices in Germany and Switzerland. Current DoubleTwist.com customers include Affymetrix, Inc., Bristol-Myers Squibb Company, Chiron Corporation, E.I. du Pont de Nemours and Company, Elan Pharmaceuticals, Hitachi Ltd., Merck 7 Co., Inc., Millenium Pharmaceuticals Inc. and Monsanto Company.

DoubleTwist is currently hiring for a variety of positions at our Oakland offices. Within a burgeoning new scientific discipline at the intersection of computer and life sciences, DoubleTwist is looking for people with experience in areas such as bioinformatics, molecular biology, chemistry, computer science, sales, and customer support. We offer the opportunity to join a leading-edge company in a field that is only beginning to take off.

For a complete listing of open positions, please log onto our website at: www.doubletwist.com. You may also e-mail your resume directly to DoubleTwist at: hr@doubletwist.com.



aaps

American Association of Pharmaceutical Scientists

Northern California Pharmaceutical Discussion Group

Since 1983 the NCPDG has provided a forum for the Bay Area pharmaceutical/biotechnology industry for development of the community and discussion of topics important to our industry. The NCPDG holds monthly dinner meetings that are attended by individuals representing every aspect of industry life and from nearly every pharmaceutical/biotechnology company in the Bay Area to hear talks on subjects ranging from genomics to contract manufacturing to financing company operations. Our unique combination of fellowship and education provides several material benefits to our members.

- Opportunities for effective networking
- Increased understanding of industry issues
- Expanded knowledge of various pharmaceutical/biotechnology businesses
- Self-Improvement and education

Your company can also take advantage of NCPDG involvement. Our membership spans the width and breadth of the pharmaceutical/biotechnology industry in the Bay Area and we regularly have presentors and attendees from as far away as Europe and Japan. By becoming a NCPDG Sponsor your company can gain increased visibility in the Bay Area and beyond while helping to support individual professional development. Benefits of sponsorship include:

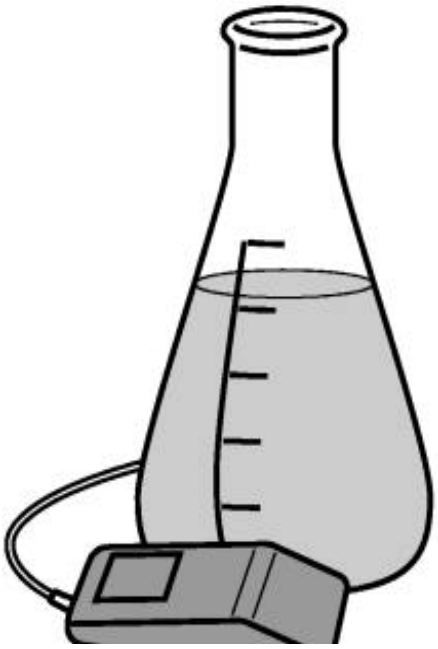
- Name exposure on our printed materials, web site and e-mail distributions
- Tailored sponsorship opportunities
- Distribution of job announcements and other materials at NCPDG meetings
- Another way to help your employees develop professional and social skills

For more information on the NCPDG visit our web site or contact Ben Borson at (415) 362-3800, Eric Schuur at (650) 224-4178, or Helen Wang at (415) 922-3868.

WWW.NCPDG.ORG

www.skjerven.com

An Interdisciplinary Firm



For an Interdisciplinary Field

SKJERVEN MORRILL MACPHERSON LLP

A law firm serving high technology clients from offices in San Francisco, San Jose, Newport Beach, and Austin is pleased to sponsor

BCATS - Biomedical Computation @ Stanford 2000

Genencor International, Inc. is proud to have provided sponsorship for the **BCATS 2000 Symposium**

Genencor International, Inc.®

Genencor is a diversified biotechnology company that develops and delivers products into the health care, agriculture and industrial chemicals markets. Using an integrated set of technology platforms, our products deliver innovative and sustainable solutions to many of the problems of everyday life.

Find more information about us at: www.genencor.com