

# On Your Mark,

# BUILD INFRA

## The first four National Centers for Biomedical Computing take off

BY KATHARINE MILLER WITH AN INTRODUCTION BY ERIC JAKOBSSON, PhD

### WHY NATIONAL CENTERS?

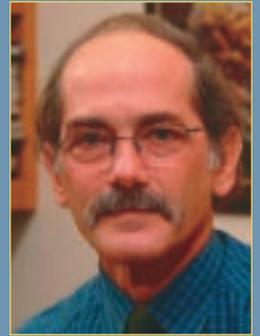
**F**our National Centers for Biomedical Computing were launched by the NIH in 2004 with \$20 million in funding for each center over five years. The reason: We need to make biomedical computing as good as desktop office computing. In particular, we must make the computational infrastructure for biomedicine more robust, efficient, easy to use, widely disseminated, and in conformity with good software engineering standards so that its components can readily be made interoperable and more versatile. In order to do this, we must build larger, more coordinated enterprises with the critical mass and appropriate mix of computer scientists and basic and clinical biomedical scientists to cover the range of expertise required for producing needed bioinformatics and computational tools.

Until now, much of the software in biomedical computing has been in the form of independent programs for particular functions such as molecular dynamics, modeling of dynamic systems, or searching for homologies in sequence databases. These programs have been developed by a scientific “cottage industry” of individuals or small groups of researchers for the purposes of their own research, and distributed, maintained, and further developed (or not) in a variety of ways.

This is similar to the old “one-program/one-function” paradigm that we used to have for our desktop office needs, in which we had separate programs for word processing, sched-

# Get Set, **STRUCTURE**





“We need  
to make  
biomedical  
computing  
as good as  
desktop office  
computing.”

—Eric Jakobsson

uling, databases, and spreadsheets. In modern desktop software all these functions are seamlessly integrated into what is effectively a single program so that all functionalities are interoperable; the results of an operation in one functionality are instantly available as input to operations in other functionalities.

Achieving desktop biomedical computation requires a focused effort. Hence the National Centers. The primary products of these centers will be the key building blocks of the ultimate national biomedical computing environment, such as: data sources, models, and model validation tools integrated within and across domains; new and enhanced algorithms; biomedical software newly engineered or re-engineered according to best software engineering practices; interfaces and portals to enhance usability for biomedical researchers and educators; and tools to integrate quantitative, symbolic, visual, and natural language representations of data. In all cases, biomedical importance and high standards of software engineering are essential.

Because the centers are aimed at creating an integrated national biomedical computing environment, coordination within and between them is crucial to their success. We’re seeing that happen already among the four centers established thus far. And in 2005, as several additional centers receive funding, I expect collaborative efforts to blossom across the country, putting an end to the cottage industry mindset, and building a burgeoning sense that biomedical computing can indeed become part of the nation’s research infrastructure.

—Introduction by Eric Jakobsson, PhD, Director of the NIGMS Center for Bioinformatics and Computational Biology at the National Institutes of Health.



Interviews with the principal investigators of

# The First Four National Centers for Biomedical Computing

## The Center for Computational Biology (CCB), at UCLA

The Center for Computational Biology (CCB), located at the Laboratory of Neuro Imaging, UCLA, focuses on the computational biology of the brain from genetics to structure and function. Its central mission: to develop the infrastructure for a “computational atlas” of the brain—a platform for addressing large-scale modeling problems that before now have been intractable. Like a road atlas, the computational atlas will help people locate a particular “address” in the three-dimensional brain. But the computational atlas will also provide many other types of information about that location—including changes over time—in order to help elucidate characteristics and relationships that would otherwise be impos-

ble to detect and measure.

We spoke with Arthur Toga, PhD, Professor of Neurology and principal investigator for CCB.

**Q:** One goal of the Center for Computational Biology is to create a “computational atlas” of the brain. What is that?

**Toga:** The normal way of thinking of an atlas is in the traditional book form. Here in Southern California we have the Thompson Guide—a book of maps in which you can look up a street to find a way to get there. The atlas tells you longitude and latitude to some named place, and it provides that information in an organized way. But a computational road atlas would also provide you with information about the location’s traffic density at given times

of the day, traffic accidents and fatalities, road closures, and nearby restaurants and hotels.

Similarly, a computational atlas of the brain will identify a particular location in the brain with dynamic and multifaceted information about that location. For example, for any particular site in the cerebellum, the atlas might tell you that spot's relationship to expression of particular genes, the location of specific blood vessels, and what clinical manifestations occur there at what frequency in what population. The atlas would integrate bits of information from different studies at different spatial scales, different temporal scales, and across populations—data gathered by different people in different labs. The task might seem a little daunting at first, but the field has developed a variety of coordinate systems for reconciling identical locations

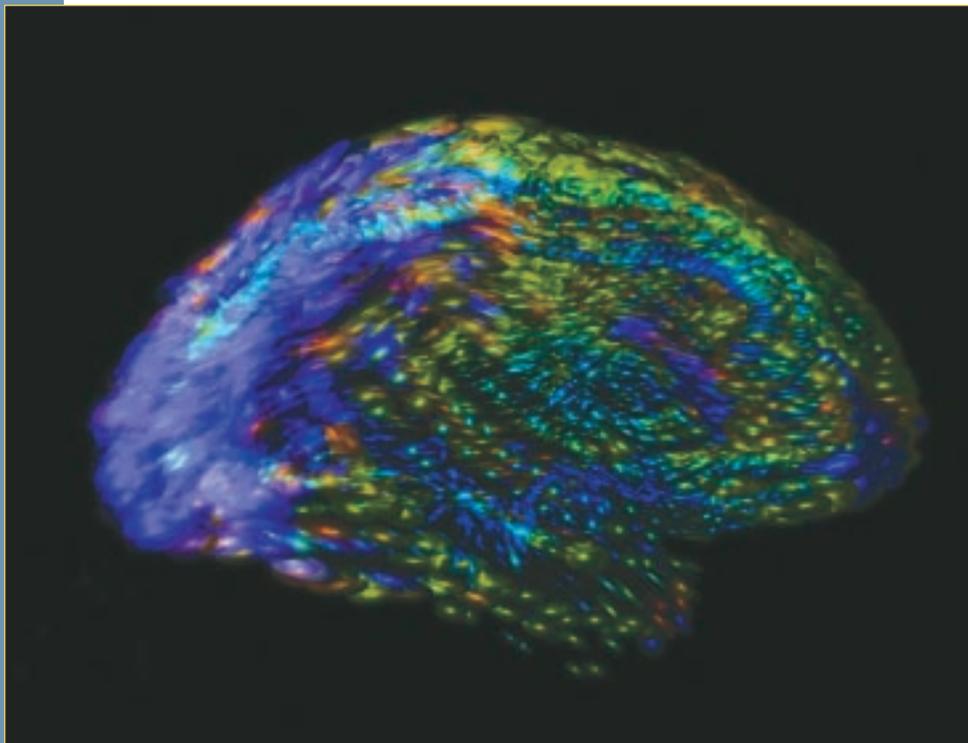
in the brains of different individuals. This advance sets the stage for adding all of this other information.

**Q:** Why would it be helpful to have such a computational atlas? Where does it get you?

**Toga:** The atlas will enable advances in both basic science and clinical medicine. For example, it can teach us about fundamental relationships between genotype and phenotype. It might be possible to determine whether expression of a particular gene at a particular location has an effect on vascular disease at that location. This is the type of basic science question that is difficult to answer now.

In clinical medicine, the atlas might map the trajectory of a disease process over time, and that trajectory could become predictive. For example, we know that relatives of Alzheimer's patients have a higher risk of the disease, but we only know if they are going to get the disease when their behavior is already starting to change. We already know that certain conformations of brain anatomy are highly correlated with Alzheimers. If we make measurements over time, to see what areas degenerate at what rate, and then calculate rate of change regionally, we might be able to project backwards in time to see what an Alzheimer's patient's brain would look like before any behavioral symptoms have arisen. That then becomes a biomarker and those patients become candidates for treatments. That's not possible now, but it might be if we develop a dynamic, computational atlas of the brain.

**Q:** The Center for Computational Biology involves numerous partici-



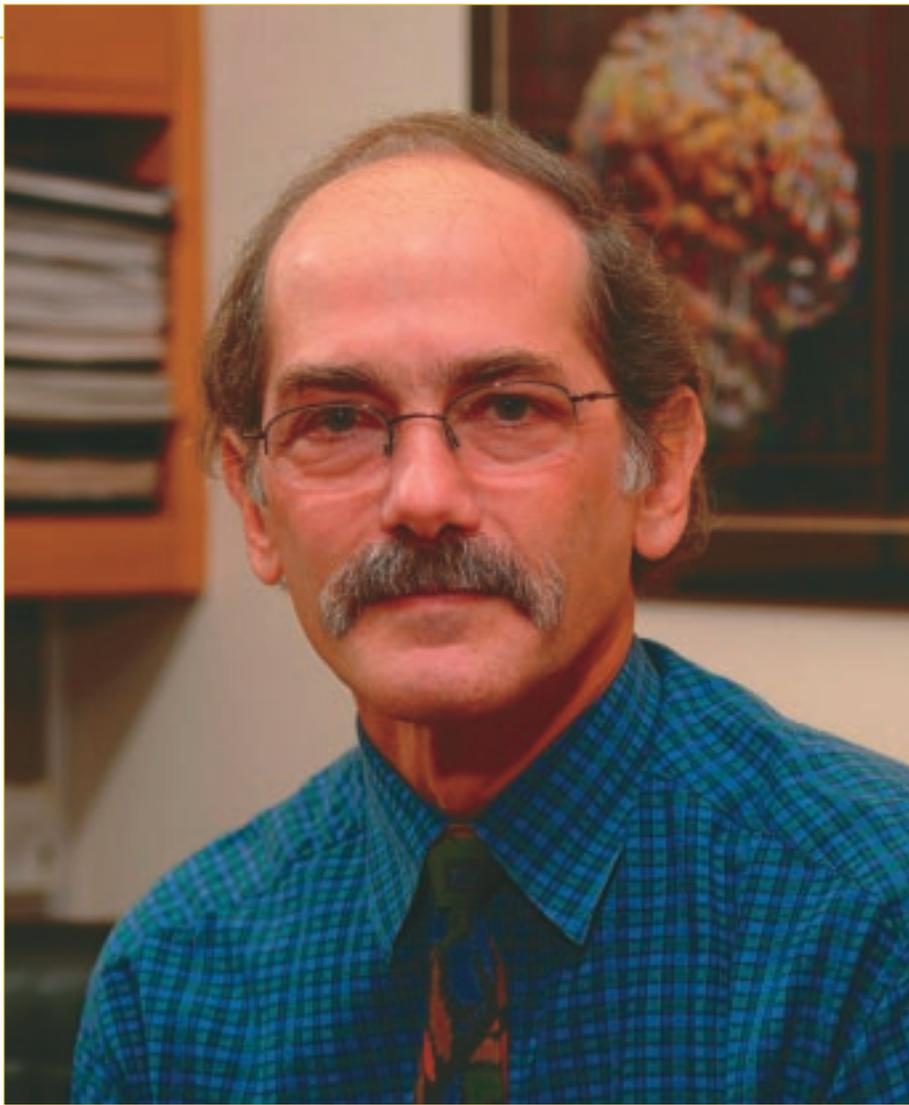
*Composite variability ellipsoids of 20 normal subjects. Purple blur represents highest variability from subject to subject in high-level regions such as the "personality centers" while blue ellipsoids represent lowest variability in low-level functions. Courtesy: UCLA Laboratory of Neuro Imaging.*

pants from a variety of disciplines including statistics, mathematics, computer science, neuro-imaging, bioinformatics, and nanotechnology. What are you doing to get the scientific cross-fertilization off the ground?

**Toga:** We've brought in researchers whose disciplines touch on our work without any intimate involvement in it. So we've been going through a tremendous learning exercise in which we identify and discuss what each of the participants does and then figure out how we can turn the algorithmic research of all these mathematicians, statisticians, and computer scientists into programs that can be distributed to a wide scientific constituency.

**Q:** In the past, computational biologists have often worked in relative isolation, creating code that doesn't get used by anyone else once their doctorate or other research project has been completed. How are you motivating people to care about creating tools for other scientists to use?

**Toga:** That's one of our grand challenges, frankly. It's really a different mentality to put our heads together and come up with a product or tool that will be used by others. Doing so requires new infrastructure and support—namely programmers who can take the research and turn it into software with documentation and support. And we're finding that there's very little common knowledge between the person who wrote the algorithm and



*Arthur Toga, PhD, is Professor of Neurology and Director of the Laboratory of Neuro Imaging at the University of California, Los Angeles.*

the person who we're now hiring to write the software. But NIH was right on the money in creating these centers and motivating them to take the work that's being done and turning it into products that can be used. It was really a very important initiative; it is very much needed, and is very hard to do. >

“A computational atlas of the brain will identify a particular location in the brain with dynamic and multifaceted information about that location.”

—Arthur Toga, CCB

## I2B2: Informatics for Integrating Biology and the Bedside, at Harvard



*Isaac Kohane, MD, PhD, is Lawrence J. Henderson Associate Professor of Pediatrics and Health Sciences and Technology at Harvard Medical School, and Chair of the Informatics Program at Children's Hospital, Boston.*

The I2B2 Center at Harvard is developing an informatics framework that will build connections between clinical research data and the vast genomic data banks arising from basic science research. The ultimate goal: to better understand the genetic bases of complex diseases in order to design targeted therapies for individual patients. New bioinformatics approaches will be developed and tested in four diseases—airway disease, hypertension, diabetes mellitus, and Huntington's disease.

We spoke with Isaac Kohane, principal investigator for I2B2.

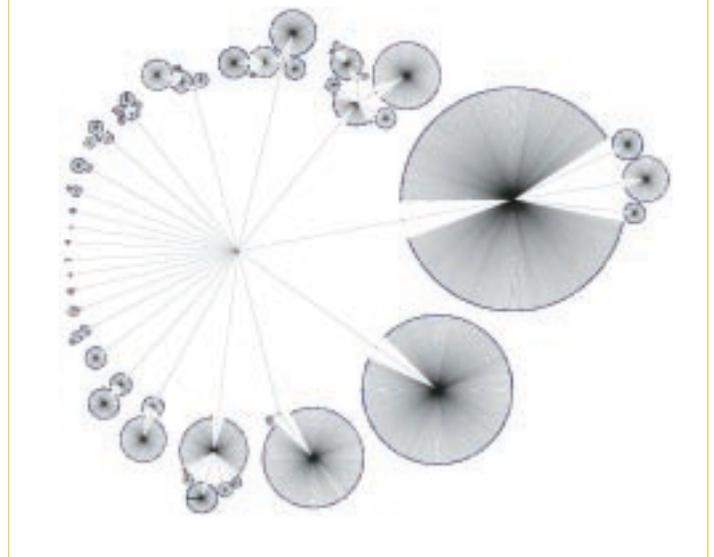
**Q:** What's the greatest challenge that I2B2 faces going forward?

**Kohane:** The overall challenge is to provide clinical researchers with the tools they need, in a format that they can understand and use in the genomic era. There has been a shift from analytic self-reliance that was the hallmark of biomedical research. In the past, a single researcher might work with a dataset that could fit onto an Excel spreadsheet and analyze it using relatively simple algorithms. But that is no longer possible. The vastly larger quantitative scale of genomic data is causing a qualitative change in the clinical research that depends on these data. When investigators start asking clinical questions involving the combined analysis of conventional clinical measures and thousands of genomic data points, the requisite computational skill set is typically beyond their training and knowledge. So, we want to provide the wherewithal in academic medical centers for clinical researchers to avail themselves of the necessary tools and data in this new time we call the genomic era.

**Q:** How does I2B2 plan to address that challenge?

**Kohane:** We've hypothesized that we can use the affiliated hospitals of

*New classification of diseases might be based on gene expression signature rather than clinical symptoms.*



Harvard Medical School as a test-bed or “living laboratory.” The Partners Healthcare System, which comprises Massachusetts General Hospital, Brigham and Women’s Hospital and a handful of other large hospitals, shares a clinical information infrastructure under the management of the information systems department. The Chief Information Officer, John Glaser, is the co-principal investigator of I2B2. So we’re very leveraged to ask questions about the system and to use the system resources to develop responses to the challenge outlined above. Ultimately, that means providing computationally-assisted design of clinical research experiments, computationally-assisted identification of large, well-defined populations, and computationally-assisted discovery of new knowledge and hypotheses for further exploration.

The Partners Healthcare System hospitals have already built a medical record system that spans the encounters with 3 million patients, including data ranging from lab studies to physician order entries. The system has succeeded in reducing clinical errors and improving quality of care, but it has not been extensively used for clinical research and particularly not for clinical research augmented by genomic data. We’re going to make that happen.

**Q:** What kinds of technical problems do you face as you dive into the morass of medical records?

**Kohane:** Our computational scientists face several basic challenges up front. One is coming up with well-defined phenotypes out of the muddled and rarely structured data in medical records. It’s really a matter of very difficult linguistic parsing, and we’re working on that. A second is trying to do predictive modeling across very heterogeneous data types—clinical records, laboratory data, study-related datasets,

large volumes of genomic and proteomic data, and data from published literature. A third is figuring out which genomic information adds value. What pieces of it are really helpful diagnostically or therapeutically? Does the genomic data really add information that wasn’t already available from traditional types of data? A fourth is the question of whether we can predict disease from changes in the genome. For example, there are multiple CAG repeats in the gene for Huntington’s disease. Can we explain

“There has been a shift from analytic self-reliance that was the hallmark of biomedical research.”

—Isaac Kohane, I2B2

how the increased number of these nucleotide-triplet repeats results in worsening disease for these individuals?

**Q:** Ultimately, you hope to create something called the “clinical research chart of the future.” What is your vision for this chart?

**Kohane:** This artifact that we call the clinical research chart of the future will allow you to design and run a clinical research experiment that is targeted to the populations of interest to a particular study. The chart should allow you to “slice and dice” patient populations by phenotype without being a programmer. It should allow you to test a hypothesis without subjecting patients to imperfect and some-

times anxiety-provoking re-consents and recruitment (by ensuring that patients are properly consented or excluded from the start). It would function in a “plug and play” fashion so that a researcher can easily identify patients of interest and study them, annotate their records and combine their data with other relevant data sets. And it will need to be packaged into an interface, using bioinformatics tools that are already in existence, and novel tools that we’ll develop and that are usable by non-programmers and non-bioinformaticians.

**Q:** Why do patients opt to participate in this kind of “digital” health research?

**Kohane:** Like many patients involved in clinical research, the principal payoff for these patients in the short-run will be limited to the satisfaction that they are helping move forward the state of knowledge and hopefully the state of therapy for a particular disease, often one that they are themselves at risk for. Less likely and longer term is the possibility that particular genetic variants of these patients, given the appropriate level of consent and identification, may be found to have genomic characteristics that place them at particular risk for a disease or conversely to benefit from a specific therapy. It has been our experience that the first motivation, that of helping move forward research in a particular disease area, especially one in which the patient has a personal or family risk or experience, is powerful and energizing to many people. >

# National Alliance for Medical Image Computing (NAMIC)

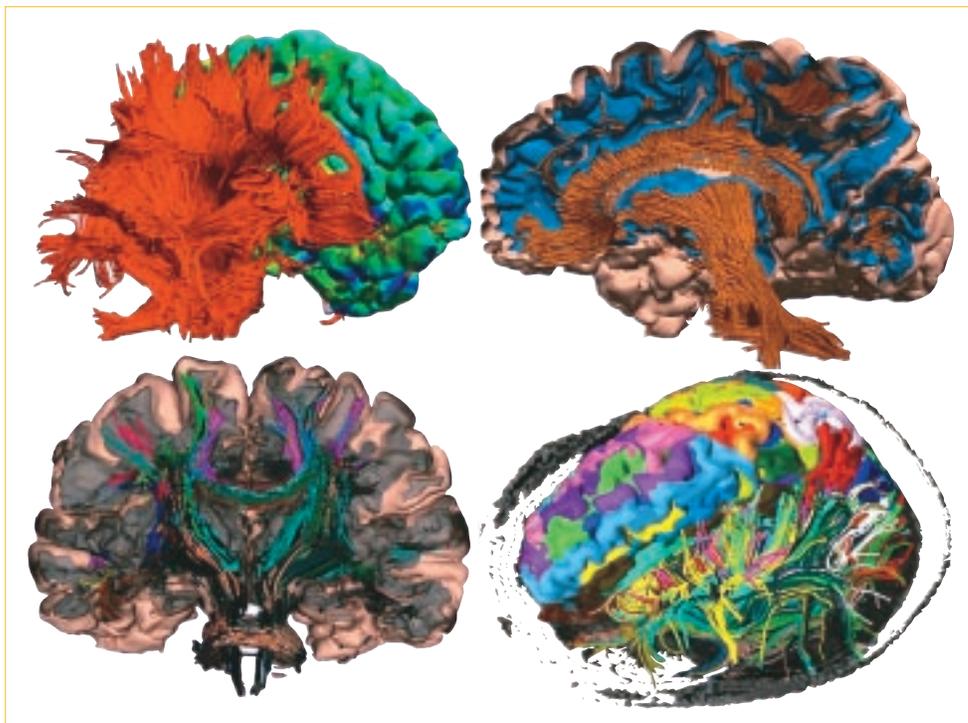
**N**AMIC's goal is to develop, integrate, and deploy computational image analysis systems that are applicable to multiple diseases and a variety of organs. The multi-institutional center, located primarily at Harvard, will integrate the efforts of leading researchers who share a vision of developing and distributing the tools required to advance the power of imaging as a methodology for quantifying and analyzing biomedical data. To provide focus for these efforts, a set of key problems in schizophrenia research was selected as the initial driving biological project for the center. NAMIC won't necessarily be developing diagnostic tools or treatments, but will instead develop concepts that can then be adopted by

commercial entities and developed into products that then go through the FDA approval process to benefit patients down the line.

We spoke with Ron Kikinis, principal investigator for NAMIC.

**Q:** What challenges does NAMIC face as you try to get the project started?

**Kikinis:** We are in the process of standardizing a set of tools and environments that we call the NAMIC kit. It will be an infrastructure to enable the sharing of algorithmic work in the field of medical image analysis. Given that we just started this year, we're in an early phase, so we're making the kit available to our



*Upper left panel displays fiber tractography combined with cortical thickness map obtained with Free Surfer. Upper right panel demonstrates a sagittal cross-sectional view of brain parenchyma segmented into white and gray matter combined with fiber tractography map. Lower left figure shows a coronal cross-sectional view of automatic parcellation of white matter overlaid with the gray matter surface. Lower right figure shows the parcellation of gray matter surface and corresponding white matter fibers. Images courtesy of Hae-Jeong Park, PhD, and NAMIC*

NAMIC people. It makes use of existing toolkits, most developed by groups that are part of the alliance. We are leveraging existing software rather than trying to reinvent the wheel, but we want to leverage the right kind of software for the topic of NAMIC. There's lots of software around, and some of it has been developed by NAMIC participants.

As an example, ITK.org provides software that was developed by a consortium funded by the National Library of Medicine over the last five years. ITK will be one of the core components of the NAMIC kit. It exists already, so what's the value added? We will use it to address the driving biological problems of the NAMIC grant. ITK provides a very flexible framework but it's also very general. We need to do things that are more specific to the applications we're developing. ITK lets you use whatever coordinate system you want, but in working with imaging data, you have to decide on one particular one. So we are adding a layer of specifications onto the general ITK framework.

**Q:** What do you think are the grand challenges of image analysis?

**Kikinis:** Ideally, we'd want to get to the point where the algorithms used for image analysis are robust and flexible enough to handle complex data sets without a lot of supervision and without so many controlled conditions. But to reach that point, we need to deal with the field's primary challenge: complexity. Complexity is introduced by biology itself, because living things are so varied and variable. It is also introduced by the process of image acquisition, which is far from simple. To robustly and reliably extract information from images is still very difficult. You need very controlled conditions. Ten years ago, imaging data was simple. Increasingly, it's much more complicated and involves vastly more information, requiring more analysis to extract the



*Ron Kikinis, MD, is the director of the Surgical Planning Laboratory of the Department of Radiology, Brigham and Women's Hospital and Harvard Medical School, and professor of radiology at Harvard Medical School.*

**Q:** Will the NIH investment in NAMIC have a real payoff for patients down the line?

**Kikinis:** The hope is that advances in image analysis will lead to treatments and cures for schizophrenia and other medical problems. Schizophrenia is

the driving biological problem that NAMIC selected to focus on, but it's a very puzzling disease. We aren't yet sure if it is really a single disease. It might be

several diseases with a common set of symptoms. It's pretty much agreed now that there is a change in schizophrenic brains, based largely on image analysis. But the changes in the brain that we do know about are very subtle. It's not yet possible to make a diagnosis as to whether any single individual will develop this psychosis. And without a good way of diagnosing or determining whether there are different types of brain changes in people with the same symptoms, it is difficult to know which treatments are appropriate. Getting a more detailed level of

understanding will help in the development of treatments, which is what our collaborators hope to do. We are trying to help them find tools.

We believe that the software we're developing will also help patients with other kinds of problems. In fact, the same software we're using to understand schizophrenia is already being tried in clinical research for a variety of other purposes such as image-guided therapy, liver ablation, and navigation in cranial facial surgery. >

information. And so, in addition to the fundamental drive for more robust algorithms, we're facing the challenge of more and more data of a complex nature. So one of the fundamental issues is how we deal with complexity.

**Q:** Is it really possible to build a national infrastructure for image analysis?

**Kikinis:** I believe it can be done. Key words: "I believe." It's not based on factual knowledge. I do believe there is enormous potential for this provided there's enough funding. I can't say that we will create the infrastructure by such-and-such date. The things that make science attractive are the things that you can't schedule. But, I have hopes we will be able to mesh the efforts of the various centers. It makes technical and political sense to do so.

Our mission at NAMIC is to develop the infrastructure for image analysis, and to make the platform open source so that other researchers can spend their time doing core tasks rather than building infrastructure. We have to make these tools available to other people so they can use them—and leverage our efforts.

"We have to make these tools available to other people so they can use them—and leverage our efforts."

—Ron Kikinis, NAMIC

# Simbios: Physics-Based Simulation of Biological Structures, at Stanford

**S**imbios, located at Stanford University, will develop, disseminate, and support a simulation toolkit (SimTK) that will enable biomedical scientists to develop and share accurate models and simulations of biological structures from atoms to organisms. SimTK will be an open source, extensible, object-oriented framework for manipulating data, models, and simulations.

We spoke with Russ B. Altman, MD, PhD, professor of genetics, bioengineering and medicine at Stanford University School of Medicine, and one of the principal investigators for Simbios.

**Q:** What can physics-based simulation teach us that is of value to the field of medicine?

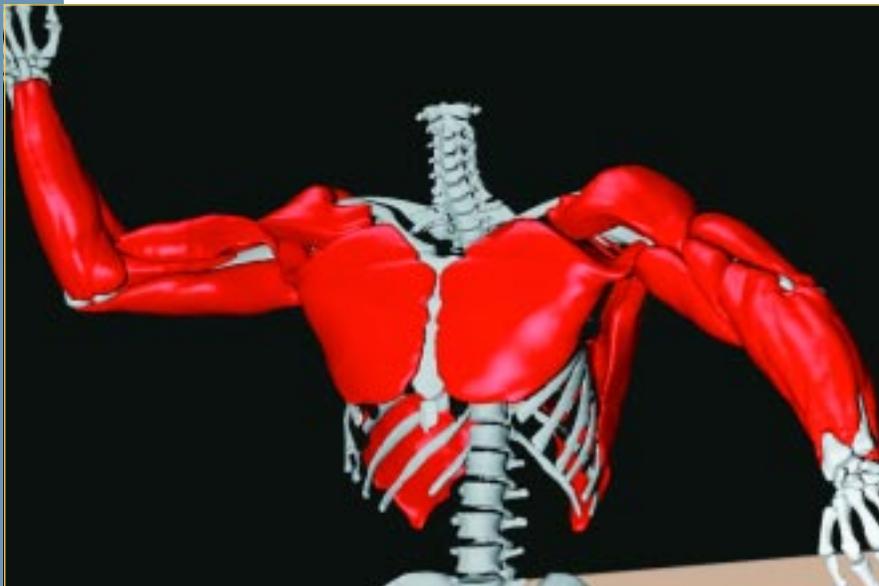
**Altman:** Physics-based simulation provides a powerful framework for understanding biological form and function. Simulations may be used by biologists to study macromolecular assemblies and by clinicians to examine disease mechanisms. Simulations help biomedical researchers understand the physical constraints on biological systems as they engineer novel drugs, drug delivery systems, synthetic tissues, medical devices, and surgical interventions.

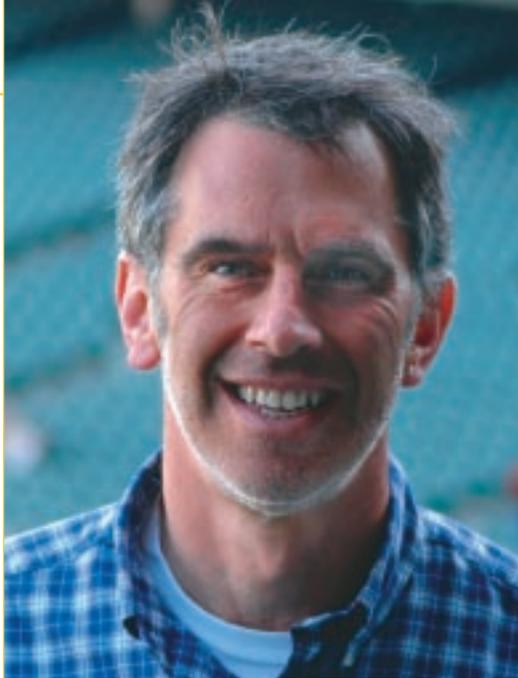
**Q:** What are the biggest challenges you face getting Simbios started?

**Altman:** The biggest challenge is figuring out how to efficiently support extremely creative researchers who are not particularly interested in service, and have their work translated into an infrastructure that can be used by hundreds of other researchers. So it's the interface between deliverable software and talented researchers. They are two very different cultures that must be brought together in order to succeed in our mission of creating and disseminating physics-based simulation software to the general biological research community.

**Q:** What do you think will be the biggest challenges for Simbios as you move forward?

*Physics-based simulation of the biomechanics of the human upper torso. Models derived from the Visible Human dataset. Simulation of 40 muscles performed using the Invertible Finite Element Method (IFEM). Courtesy of Joseph Teran, Eftychios Sifakis, Geoffrey Irving and Ronald Fedkiw*





*Russ Altman, MD, PhD,  
is Professor of Genetics,  
Bioengineering, and  
Medicine at Stanford  
University.*

**Altman:** The really exciting challenge will be to support physics-based simulation at scales ranging from atomic to organismal. We have excellent experience and expertise at both extremes, but no one has extensive experience in the mid-range—at the level of cells and collections of cells. Our approach is to push atomic modeling up to larger scales, and pull our ability to do macroscopic simulations down to smaller scales. It will be important initially to get the atomic- and organismal-scale software under the same umbrella, so that they can be mixed and matched in our attempts to model the cellular level.

**Q:** What's an example of physics-based modeling at the cellular level?

**Altman:** A particularly challenging and exciting physical simulation that we cannot do today would be the physical interactions when sperm and ova combine, and then the following days of development where cells are formed, and begin to differentiate. Because our focus is on physics-based simulation, we would not focus on the simulation of the gene-signaling networks that are so active during development of the embryo. Instead, our focus would be on the physics of the cellular morphology, the organization of the physical components of the cell, and how they develop over time.

**Q:** Physics-based simulation is already being done at many scales in biology, but the fields have developed in isolation. What needs to happen in the field in order for the

research to become more integrated?

**Altman:** During our planning process, we realized that the basic features of physics-based simulation are the same at all scales, but the sharing of innovation is impeded by software code bases that are entirely separate and not interoperable. Simbios will create a common code base, enabling scientists to do simulations at all scales and accelerating innovation in simulation. Along the way, we need to ensure that our software architecture incorporates the innovations of colleagues at institutions other than Stanford, so that the infrastructure that we call SimTK is usable by and useful to anyone doing physics-based simulation. As an added benefit of creating a common code base, simulation technology will be easier to export to “regular biologists” who just need some help understanding and modeling the physics of the systems they study. Ultimately, our simulation toolkit should allow people to enter the universe of multi-scale simulation where simulations at different scales can be “linked” in the sense that results at the lower level are propagated up to change the parameters of the higher level, and vice versa.

**Q:** Will the Simbios project have a payoff for patients down the line?

**Altman:** Some of the Simbios application areas are immediately relevant to patients. In particular, our project on neuromuscular biomechanics aims to understand how muscle activations can lead to normal and abnormal movements. With improved understanding we could create new surgical repairs for patients with cerebral palsy, to make movements more fluid. Down the line, we could even introduce small computers to drive activation of muscles which have been denervated because of spinal injuries. In the cardiovascular fluid dynamics project, we are looking at simulations of blood flow that will allow vascular surgeons to evaluate and compare different surgical options for bypass,

in order to choose the options that deliver optimal blood flow.

Other driving problems are more at the basic level. They promise longer-term benefits by increasing our understanding of molecular processes. The human genome was completed in 2003, and we now know a little bit

“Ultimately, our simulation toolkit should allow people to enter the universe of multi-scale simulation where simulations at different scales can be ‘linked’ in the sense that results at the lower level are propagated up to change the parameters of the higher level, and vice versa.” —**Russ B. Altman, Simbios**

about the list of molecular parts that make up a human. We don't know all the three-dimensional shapes of these molecules, nor how they perform their function. By simulating molecules, we can learn about their function, and thereby learn how to modify or augment it. So these simulations are exciting because they begin to show us how the human genome project may translate into new diagnostic and therapeutic technologies. For example, I have a great interest in how drugs interact with their target molecules to perform their function. Currently, we do not fully understand how most drugs work. I hope we will be able to simulate the interaction of drugs with their targets to have a mechanistic understanding of what they do. This will allow us to design new ones, and improve the old ones. □