DIVERSE DISCIPLINES, ONE COMMUNITY

Biomedical Computation REVIEW

Published by Simbios, a National Center for Biomedical Computing

Modeling Cancer Biology:

How mathematical models are transforming the fight against cancer

PLUS:

BioComputation Start-ups: Where Does Value Lie?

Spring 2007

contents

ContentsSpring 2007

FEATURES

6 Biocomputation Startups: Where Does Value Lie?
BY KATHARINE MILLER

16 Modeling Cancer Biology: How mathematical models are transforming the fight against cancer
BY KRISTIN COBB, PhD

DEPARTMENTS

1 News Bytes
BY KRISTIN COBB, PhD, LOUISA DALTON, ESTHER LANDHUIS, PhD, AND REGINA NUZZO, PhD

- DNA Shows Surprising Flexibility
- The Geometry of Adhesion
- Biological Evidence for Turing Patterns
- The BiGG Picture
- Teaching an Old Model New Tricks
- Parsing PubMed

25 Simbios News: SimVascular to Simulate Cadiovascular Flow
BY KATHARINE MILLER

26 Seeing Science: Simulated Lipid Rafts
BY KATHARINE MILLER

Cover Art by Sara L. Malloure of Affiliated Design

Spring 2007
Volume 3, Issue 2
ISSN 1557-3192

Executive Editor
David Paik, PhD

Managing Editor
Katharine Miller

Science Writers
Katharine Miller
Kristin Cobb, PhD
Regina Nuzzo, PhD,
Esther Landhuis, PhD
Louisa Dalton

Layout and Design
Affiliated Design

Printing
Advanced Printing

For general inquiries, subscriptions, or letters to the editor, visit our website at
www.biomedicalcomputationreview.org

Office
Biomedical Computation Review
Stanford University
318 Campus Drive
Clark Center Room S231
Stanford, CA 94305-5444

Biomedical Computation Review is published quarterly by Simbios National Center for Biomedical Computing and supported by the National Institutes of Health through the NIH Roadmap for Medical Research Grant U54 GM072970. Information on the National Centers for Biomedical Computing can be obtained from http://nihroadmap.nih.gov/bioinformatics. The NIH program and science officers for Simbios are:

Peter Lyster, PhD (NIGMS)
Jennie Larkin, PhD (NHLBI)
Jennifer Couch, PhD (NCI)
Semahat Demir, PhD (NSF)
Charles Friedman, PhD (NLM)
Peter Highnam, PhD (NCRR)
Jerry Li, MD, PhD (NIGMS)
Wen Masters, PhD (NSF)
Richard Morris, PhD (NIAID)
Grace Peng, PhD (NIBIB)
David Thomassen, PhD (DOE)
Ronald J. White, PhD (NASA/USRA)

NewsBytes

DNA Shows Surprising Flexibility

For decades, scientists have believed that DNA of short lengths (150 base pairs or fewer) behaves as a relatively stiff rod—able to quiver a bit, but rarely forming a circle or tight angle without help from outside forces. But a new simulation, reported in the December issue of *Biophysical Journal*, puts a kink in this theory.

"We observed fairly sharp bends that are inconsistent with classical theory. We see DNA bending quite a bit," says **Alexey Onufriev, PhD**, assistant professor of computer science and physics at Virginia Tech. "If this idea holds up, it may be a paradigm shift in how we think about protein-DNA complexes."

DNA's flexibility on this length scale has implications for DNA packaging, gene transcription, and gene regulation. For example, in the nucleosome (the fundamental unit of DNA packaging), 147 base-pair segments of DNA wrap 1.65 times around a core of proteins. DNA also twists in and out of loops to turn certain genes off and on. Under the old theory, scientists had to reach for *ad hoc* explanations, such as helper proteins, to explain how unbendable DNA could manage these feats.

Onufriev and doctoral student **Jory Z. Ruscio** modeled a nucleosome worth of DNA (147 base pairs) at the atomic level. The key to their simulation was use of the "implicit solvent" method; rather than modeling every molecule of water, they modeled water as a continuous mass. This method saves enormous computing power and speeds up the simulation by about 100-fold by removing water's viscosity—the property that makes it so hard to move quickly in swimming pools, Onufriev says. "Whatever happens conformationally happens fast," he says.

At the same time, water's thermodynamic properties are perfectly preserved. "We cannot ask any questions like what are the diffusion coefficients, because those would be skewed. But we can ask thermodynamic questions—is this conformation more preferable than the other one?" Onufriev says.

This innovation plus use of Virginia Tech's super computer, System X, allowed Onufriev and Ruscio to explore DNA's range of motion on a longer length and time scale than any atomic-level simulation before them.

Their simulation showed that DNA of 147 base pairs wiggles and bends much more than traditional theory predicts—and at a much lower energy cost than expected. The bonds of the double helix remained intact in all simulations, so their results are not an artifact of the DNA simply unraveling to create soft spots.

Onufriev's results agree nicely with two independent threads of experimental evidence that have recently emerged, says **Philip Nelson, PhD**, professor of physics at the University of Pennsylvania. A 2004 paper showed that DNA of 100 base pairs spontaneously forms circles in physiological conditions; and, using atomic force microscopy, Nelson's team recently showed that DNA of this length kinks more frequently than the old theory predicts.

The emerging picture finally makes it clear how nucleosomes, DNA regulatory loops, and viral packaging are possible, Nelson says. "No *ad hoc* mechanisms for promoting tight bending are needed."

"This is one of those beautiful moments where simulation and theory and experiment all converge," he says.

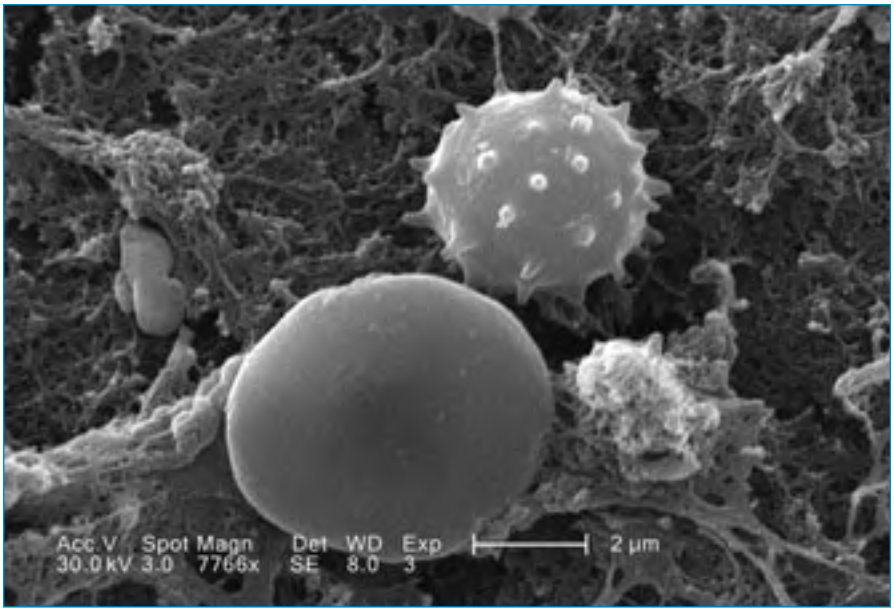
—By Kristin Cobb, PhD

> We see DNA bending quite a bit, says Alexey Onufriev. If this idea holds up, it may be a paradigm shift in how we think about protein-DNA complexes.

Three different images showing the simulation of DNA's flexibility over a length of 147 base pairs. Courtesy of Alexey Onufriev.

The Geometry of Adhesion

A single cell caught up in the flow of blood, air, or water often depends on its ability to latch onto passing surfaces—in short, its ability to stick. That's why researchers in Germany created a model



The knobby surface of a white blood cell (top) facilitates sticking, and the smooth surface of a healthy red blood cell (bottom) discourages it. Scanning electron micrograph courtesy of CDC/Janice Carr.

that addresses what geometry makes some cells stickier than others. According to their model, reported in *Physical Review Letters* in September 2006, a cell that efficiently initiates adhesion is dotted with elevated receptor patches–knobby protrusions tipped with receptor molecules. The taller the patches, the better.

"Once you start thinking about it, it's obvious." says **Christian Korn**, a PhD candidate in theoretical physics at the Max Planck Institute of Colloids and Interfaces and one of the authors. "You need these protrusions."

Cell adhesion requires two steps: encounter and docking. Korn and **Ulrich Schwarz, PhD**, a theoretical biophysicist and assistant professor at the University of Heidelberg, modeled the encounter step—to identify the cells that are best at initiating adhesion.

To create the model, the researchers simulated spheres sporting receptor patches and flowing above a flat surface with the corresponding ligands. The stickiness of cells was measured by how long it took for the first receptor-ligand encounter to occur. Korn and Schwarz then varied the number, size, and height of the receptor patches to discover the optimum receptor patch geometry. Plastering the cell with as many receptor patches as possible—akin to fully wrapping a bouncy ball in tape—is not the best strategy, they found. "The cell can have only 1% of the surface covered with receptors, and it works almost as efficiently as if it were 100% covered," Korn says. In addition, increasing the lateral size of the patches –placing bigger bits of tape on the ball—doesn't make much difference. Yet increasing the height of those receptor patches—using raised stickers instead of tape—helps the receptor patches find their target ligands sooner compared to lower receptor patches on a cell of the same size.

The researchers point to similar geometry repeated across vastly different systems in nature. Wrinkled white blood cells, which often need to dock close to an infection, place their receptor patches on the tips of finger-like microvilli. Red blood cells, in contrast, are surfboard smooth. But when a red blood cell becomes infected with malaria, it also grows knobs and new receptors on its surface to slow its progress toward destruction in the spleen. Even sticky pollen grains and wandering diatoms in the ocean, Korn says, display spiky geometry.

For experimentalists now probing such systems, says **Cheng Zhu, PhD**, a professor of biomedical engineering at Georgia Tech, the model is interesting, but only part of the equation. "Their model may explain cases where encounter is the limiting step," he says. "Without the complete equation, it's difficult to say how this might affect data interpretation in cases where docking is limiting."

Korn is now extending the model to include binding as well as encounter. He is optimistic that his model will continue to uncover general characteristics of sticky cells. "The big strength of theoretical modeling," he says, "is that you can get the big picture because you focus on a few essential aspects."

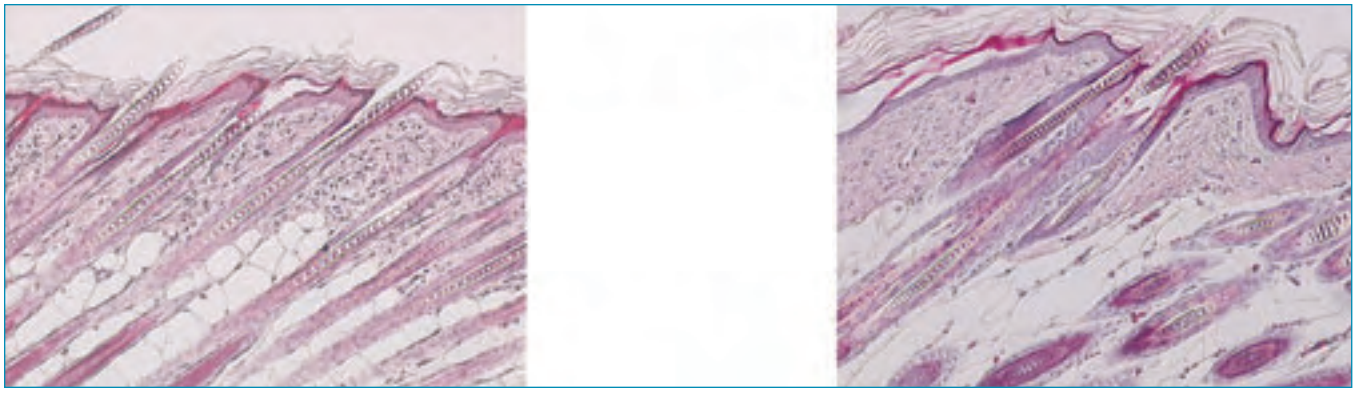
—By Louisa Dalton

Biological Evidence for Turing Patterns

In the 1950s, computer science pioneer Alan Turing suggested an elegantly simple mechanism for how biological patterns such as scales, feathers, and hair might form. Now, more than fifty years later, biologists have used a computer model and transgenic mice to confirm mathematical predictions of the Turing model of pattern formation within a specific biological system: mouse hair development.

"It's the most convincing biological (as contrasted with chemical) experiment to date that claims to support the Turing mechanism," says **Irving Epstein, PhD**, a chemistry professor at Brandeis University. The work appeared online in the journal *Science* in November 2006.

Turing's 1952 proposition goes like this: Two molecules—an activator that enhances its own production, and an inhibitor that slows the production of the activator—diffuse and react. If the inhibitor diffuses sufficiently faster than the activator, repetitive patterns may spontaneously emerge.



Normal mice have well-spaced hair follicles (left). But a moderate suppression of WNT signaling changes the pattern to follicle clumps (right). Courtesy of Thomas Schlake, Max Planck Institute of Immunobiology.

Evenly spaced mouse hair is just the type of pattern that a Turing mechanism might create. That's one reason biologist **Thomas Schlake, PhD**, at the Max Planck Institute of Immunobiology started searching for key molecules involved in mouse hair follicle formation that might fit Turing's predicted pair. He found them in the signaling molecule WNT and its inhibitor DKK.

Schlake and his colleagues created a computer model describing the pair's Turing behavior and then asked the model to predict what would happen if something went wrong—if WNT or DKK appeared in too great or too small a burst. Experiments with transgenic mice verified their computational predictions. Mice that strongly overexpress DKK, suppressing WNT signaling, look like they are balding. And mice that moderately overexpress DKK form clumps of hair instead of regularly spaced follicles.

Schlake thinks it's likely that other inhibitor/activator pairs (Turing called them morphogens) form the base of other natural patterns.

Of course, stripping complex developmental pathways down to the actions of one Turing pair is a strong simplification of the real world, he adds. Mouse hair follicle placement doesn't solely depend on the behavior of two interacting molecules. Leagues of other signaling molecules stabilize and refine the process.

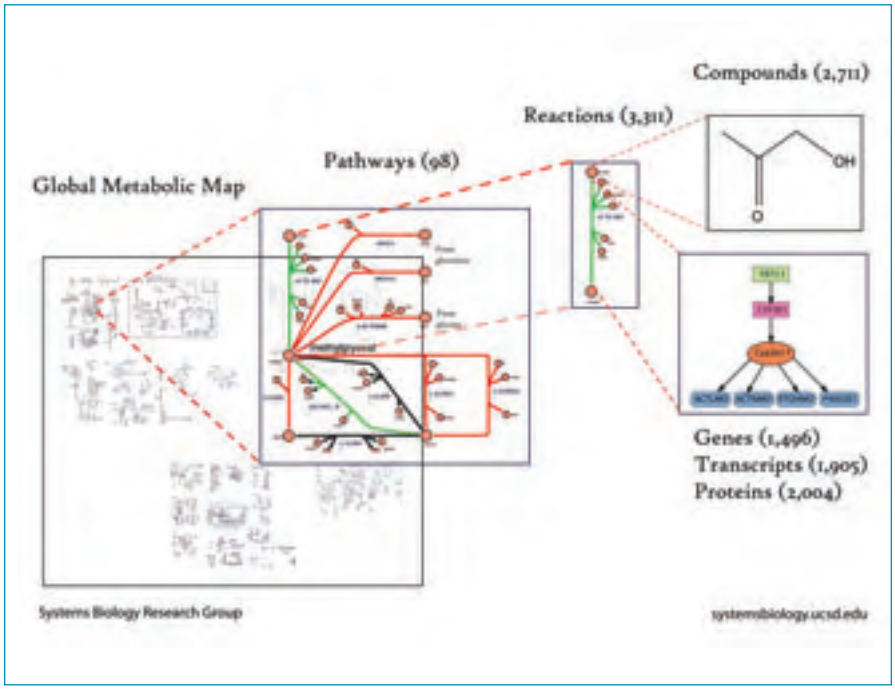
Yet it is that very power to simplify and predict outcomes from a small number of key variables that is the hallmark of a good model, Epstein says. He is not surprised that 50 years after Turing proposed his model, biologists are just now providing detailed molecular evidence for it. "Turing," he says, "was a very smart man."

—By Louisa Dalton

The BiGG Picture

It's hard to imagine a map depicting the daily flow of traffic on water, wheels and foot throughout San Diego—or any large city—over the course of a day. "That map can have many different functional states which are quite different in the middle of the night and during rush hour," says **Bernhard Palsson, PhD**, professor of bioengineering at the University of California, San Diego.

But it's even harder to imagine the map recently assembled by Palsson and his multidisciplinary research team—a virtual metabolic network representing the intracellular traffic catalyzed by more than 2,000 proteins and 3,300 bio-



Overview of the BiGG global human metabolic network. Courtesy of Bernhard Palsson and Neema Jamshidi.

chemical reactions within the human body. Construction of this first-ever genome-scale database, dubbed a BiGG (biochemically, genetically and genomically structured) reconstruction, was described in the February 6, 2007, issue of the *Proceedings of the National Academy of Sciences*.

Culled from more than a half century of published data, the computational system will allow researchers to explore hundreds of human disorders related to metabolism—the chemical processes by which the body breaks down food to build and maintain itself. For example, scientists can use mathematical optimization tools to identify sets of chemical reactions that are turned on or off together when the body makes cholesterol, explains **Neema Jamshidi**, an MD-PhD student in the Palsson lab who was a co-author on the paper. Knowing which reactions are correlated in this manner could lead researchers to alternative drug targets—components of other biochemical pathways that could be blocked to achieve the same effect as an existing cholesterol-lowering medication, Jamshidi says.

Douglas Kell, PhD, director of the Manchester Interdisciplinary Biocentre at the University of Manchester, describes another application of the BiGG database in a systems biology review published in the December 2006 issue of *Drug Discovery Today*. By computing metabolite levels under various conditions over time, he says, the network could be used to infer patterns of disease progression, providing clues as to whether a drug might reverse the degenerative process.

To give the biomedical community a shot at these lofty goals, a team of six UCSD researchers that included Palsson and Jamshidi spent 18 painstaking months gathering data to assemble the BiGG network. They combed through more than 1,500 primary literature articles, reviews and biochemical textbooks.

"What we have now is a global network," Jamshidi says. "If we found any evidence that a certain reaction occurs in a kidney cell, heart cell, whatever, we threw it in there." In the future, he says, the team will work with experts who study particular cell types—cardiac myocytes, for instance—to refine the pathways in the global system and make them more context-specific.

In the meantime, scientists such as Kell are thrilled about what the BiGG network will do for systems biology. "It is the first step on the way to a 'digital human' model," he says, "from which we can model health, disease, the metabolism of pharmaceutical drugs and so on."

—By Esther Landhuis, PhD

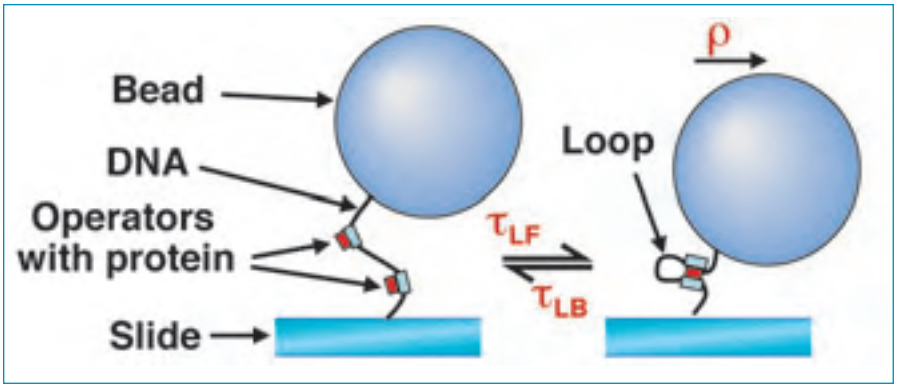
Teaching an Old Model New Tricks

The hidden Markov model—a statistical model used for decades in fields as diverse as speech recognition and climatology—has received an update and a new application. Researchers at the University of Pennsylvania and Emory University adapted the model for tethered molecule experiments, and used it to obtain the most accurate estimates to date of the kinetics of DNA looping. Their results appeared online in *Biophysical Journal* on February 2, 2007.

"Before now, no one has ever been able to measure the kinetics of DNA loop formation and breakdown in a realistically sized system," comments **Philip Nelson, PhD**, professor of physics at the University of Pennsylvania. DNA forms loops to turn certain genes off; accurate measurement of the rates of looping and unlooping are needed to build realistic models of this switching mechanism.

Researchers cannot directly see a strand of DNA in action, so they use a trick pioneered by **Laura Finzi, PhD**, (now at Emory University) and **Jeff Gelles, PhD**, (Brandeis University): they tether one end to a microscope slide and attach a visible bead to the other end. Single-particle tracking of the bead's motion is used to infer the DNA's state—when DNA is looped, the bead is pulled closer to the microscope slide and its radius of movement is more limited. Previously, researchers analyzed the data by averaging the motion of the bead within certain windows of time—called "binning

To a physicist, it's really beautiful to see the same ideas reappearing in very different contexts, says Philip Nelson.



Distance of a visible bead from its attachment point as a function of time. Sudden changes in this distance, reflecting loop formation in its DNA tether, are partially obscured by the bead's Brownian motion (diffusion). The diffusive hidden Markov model gives the most likely sequence of loop formation/breakage events. Courtesy of Philip Nelson. Reprinted with permission from: Beausang JF et al. DNA looping kinetics analyzed using diffusive hidden Markov model. Biophysical Journal, published online February 2, 2007 (Figure 1).

the data." But this method is imperfect because the results are heavily influenced by the choice of bin size. So, Nelson's team turned to hidden Markov models.

"Hidden Markov models have a long and illustrious history in the study of single ion channels, but recently they have also increasingly been the method of choice when analyzing single-molecule biophysics experiments," Nelson says. Hidden Markov models help scientists make inferences about some unobservable data (e.g., DNA states) based on a set of observable and noisy data (e.g., bead movements). The algorithm estimates the unknown rates by finding the values that make the observed pattern of data the most likely.

"For a physicist, it's really beautiful to see the same ideas getting recycled in very different contexts," Nelson says. "But we had a technical challenge, we couldn't just take it off the shelf and use it because the classic set up wasn't quite applicable." Hidden Markov modeling assumes that the noise in the observable data is purely random. However, in tethered particle analysis, this assumption is violated: the position of the bead in one moment depends on the position of the bead the instant before. So, Nelson's team made a new model—called a diffusive hidden Markov model—that accounts for this dependency.

The resulting estimates of the rates of looping formation and breakdown were robust; their rate estimates did not change when they re-analyzed the data after removing every other datapoint.

"I think their approach seems very novel and sound, and it's clear that by doing this they can obtain more accurate information about DNA looping kinetics," says **Taekjip Ha, PhD**, associate professor of physics at the University of Illinois at Urbana-Champaign. Ha has done work using hidden Markov modeling for single-molecule fluorescence studies not involving tethered molecules.

—By Kristin Cobb, PhD

Parsing PubMed

Text-mining tools such as iHOP (Information Hyperlinked Over Proteins) are doing for biological literature what hyperlinks and search engines do for the Internet: organizing interconnected information in a fast, intuitive, searchable manner. And in January 2007, the service started to provide daily updates—extending the information network by about 2,000 new papers every day.

It is a huge challenge to parse the literature on an ongoing basis, with thousands of new papers per week

With genes and proteins acting as hyperlinks between sentences and abstracts, a large part of the PubMed knowledge base becomes a giant, navigable information network, says **Robert Hoffmann, PhD**, a postdoctoral fellow at Sloan-Kettering Institute who started the iHOP project while a researcher at the Protein Design Group at the National Center for Biotechnology (CNB) in Madrid, Spain. "The new version provides current information on even more genes and chemical compounds, covering 1,500 organisms ranging from human and chimpanzee to yeast and HIV," Hoffman says. He and his colleagues also extended iHOP's results to include drug interactions, and they've provided new ways to interact with the data—such as displaying "breaking news" found in papers from the past two years.

Freely available online since 2004, iHOP parses millions of PubMed documents and selectively grabs information specific to 80,000 different biological molecules. The program displays a list of relevant sentences snagged from the parsed documents, effectively summarizing the interactions and functions of a given protein or gene. The user can also browse statistical overviews of interaction partners and associated drugs, collect interesting sentences into a logbook, and create graphical representations of the results.

The computational machinery behind iHOP has continually evolved since the program's introduction, Hoffman says.

The most important enhancement this year—daily updating—was also the most technically demanding, requiring the daily processing of about 2,000 new publications. "It is a huge challenge to parse the literature on an ongoing basis, with thousands of new papers per week," says **Chris Sander, PhD**, of the Computational Biology Center at Memorial Sloan Kettering Cancer Center. "Robert and our team can now do this as the result of new software running on a multiprocessor machine that is better suited to processing large-scale text data."

The problem, Hoffmann says, is that most parallel computing pipelines (known as Message Passing Interface frameworks) are designed for repeated number crunching, not the sort of memory-intensive, semantic database processing that text mining requires. So Hoffmann developed his own computational pipeline capable of annotating millions of documents within a few hours on an 80-node cluster, making daily iHOP updates a reality. "We're now in a good position to make the next move toward annotations of full text sources, as well as the algorithmic exploration of gene networks," Hoffmann says.

Text-mining tools such as iHOP are great for focusing on pertinent key fragments in the literature, says **Russ Altman, MD, PhD**, chair of the Department of Bioengineering at Stanford University. "There is so much published that it's hard to keep track of all the relevant information, especially in journals that end up having unexpectedly relevant material," Altman says. "iHOP is an example of an approach that helps biologists filter lots of literature."

iHOP is freely accessible at http://www.ihop-net.org/.

—By Regina Nuzzo, PhD □

Biocomputation Startups:

Where Does Value Lie?

BY KATHARINE MILLER

When discussing biocomputation startups, there's one thing people agree on: These days, they don't generate much excitement among venture capitalists.

"In the 1990s, there were a series of bioinformatics companies founded that did not succeed," says **Fred Dotzler**, managing partner of **DeNovo Ventures**, a healthcare investment firm. "Now money for these types of companies is thinner."

Why the pessimism? Simply this: There are a limited number of potential biopharma (biotech and pharmaceutical industry) customers for bioinformatics platforms, and many of those already have a suite of informatics products. Anything an outsider develops will have to be extremely promising and technically compatible with installed systems for these companies to make the change.

Nevertheless, many in the field are still trying to make a go of it by selling software platforms, tools, and services to biopharma. The pharmaceutical industry, they say, desperately needs to change how it does business. Estimates vary, but many say it costs more than a billion dollars to bring a new drug to market. Moreover, the failure rate of new drugs is extremely high, and drug safety problems are often discovered after millions have already been spent. Biocomputation, the argument goes, offers one possible way to discover new drug targets, determine drug toxicity sooner, and efficiently hasten the development of safe and effective drugs.

Colin Hill

CEO, president, chairman and co-founder of Gene Network Sciences (GNS)

Colin Hill had been doing research in theoretical physics and chaos theory in non-living systems for some time when he was drawn into applying those ideas to biological systems. He saw that the mathematical modeling of a complex many-component system wasn't really happening in the drug development world. "If one could master the source code of a living system," he says, "that would give us a huge capability to discover the underlying biological mechanisms of drug efficacy and toxicity." GNS was born of Hill's convictions.

Protein Mechanics □ Intelligenetics □ Molecular Applications Group □ Molecular Mining □ Physiome □ Rosetta Inpharmatics □ Molecular Designs □ GeneSpring □ Entelos □ Gene Network Systems □ Ingenuity □ 23AndMe □ Pathwork Diagnostics □

According to **Colin Hill**, CEO and co-founder of a systems biology company called **Gene Network Sciences (GNS)**, "At the end of the day I have to be bullish for this field, whether academic or commercial, because I don't see a way out of the pharmaceutical crisis without a better ability to discover drug mechanisms and ultimately predict efficacy and toxicity better than the industry does now." Companies such as GNS, **Entelos** and **Ingenuity,** discussed below, are betting their hopes on their ability to help the industry move forward more efficiently.

Other startup entrepreneurs, however, see greater promise in designing biocomputational products that are essential to clinical care and repeatedly needed. "I've always felt that given a chance to sell razor blades or a razor, I'd far rather sell razor blades," says **Glenda Anderson**, founder and chief technology officer at **Pathwork Diagnostics**.

A bioinformatics software platform is like a razor, she says, it's needed, but doesn't need replacing very often. Pathwork Diagnostics, by contrast, has created a razor blade—something that costs little to produce and is repeatedly needed. It's a bioinformatics tool that analyzes gene expression data to determine the likely source of cancerous tumors of unknown primary. They will charge for each biopsy result they analyze.

Anderson predicts this startup model is the wave of the future: "Biocomputational startups probably will start to look more like medical product startups with a heavy biocomputational component," she says. She points to **XDx** as another company following this model. And **23AndMe** (also discussed below) is a new company that hopes to market genetic information to the consumer—another razor blade.

The challenge for people looking to start new biocomputation companies, Anderson says, is figuring out which research ideas are commercially viable. "There's a real void here that I find quite exciting," she says. "Some of the best ideas in this field might be fantastic science, but might not translate into products that could fuel a successful startup. That's our opportunity and challenge."

Selling to Biopharma? Link to a Drug or Go without VC

According to Hill, a surge of funding for biocomputation startups in 2000 gave entrepreneurs the mistaken impression that a business could succeed just by generating data and related tools. GNS itself was funded in that wave.

"If you just want to make a lot of money, go to Wall Street. Or create the next YouTube," says Colin Hill. "The ultimate test in our field is affecting disease in a living human.... it's much more difficult than just developing technology."

"VCs want companies that are going to have a good-size market with good revenue possibilities," says Alex Bangs. "If your market is to sell to a small part of the pharmaceutical industry, then you have a small number of customers with potentially varying budgets."

"I think it gave us a somewhat slanted view of where value was going to lie for computational systems biology," says Hill. "Normally, the real currency of the industry is around a drug."

GNS survived by making sure drugs were their central focus. The company's strategy is to create models on the fly from data about specific drugs in the pharmaceutical development pipeline. They have developed a unique tool that can quickly—in hours or days—uncover how a drug is working and predict which patient populations will benefit. For example, GNS helped Johnson and Johnson discover the mechanism of action for a multi-kinase inhibitor being developed to treat cancer.

"If we're right, this will become one of the key value drivers for the whole pharmaceutical industry. If we're wrong then we'll be one more platform technology company to come and go from the stage of pharmaceutical drug discovery and development," Hill says

These days, GNS works for pharmaceutical companies on a fee-for-service basis. But, he says, "A lot of investors would say you really need to take it all the way downstream to a drug to extract maximum value from breakthrough technology."

The potential revenue from making drugs is apparent to **Entelos** as well, says **Alex Bangs**, the company's co-founder and chief technology officer. A few years ago, for example, Entelos used its bioinformatics tools to help Organon, a pharmaceutical company, identify new drug targets for rheumatoid arthritis. A co-development agreement gives Entelos a piece of those drugs going forward.

But, like GNS, the bulk of Entelos' business still revolves around fee-for-service. The company's core product is computer simulation of chronic diseases in virtual patients. They build the models and collaborate with pharmaceutical companies along all points of the drug discovery, testing, and trial process. One of their challenges as a business is to explain the full range of their tools' potential. "A tool that works in discovery and clinical is unusual," Bangs says. "It's hard for people to wrap their arms around."

Bangs says their products have proven extremely valuable in telling pharmaceutical companies when a compound is likely to fail in the marketplace. "We've had people stop those programs and spend their money in a different way," he says. "And when we get a result that suggests something's going to work, we can explain why, suggest measures to confirm what we're seeing and recommend when to take those measures to get the best result. It's very much a scientific conversation, not a black box."

Entelos is no longer a startup. The company went public last April in the

(continues on page 11)

Alex Bangs
Co-founder and chief technology officer at Entelos
While working at a management consulting company in the early 1990s, Bangs helped his cohorts develop software tools in support of modeling work for pharmaceutical companies. His contribution: the creation of a software architecture that supported the development and analysis of large scale physiology models and virtual patients. By 1996, they saw that the tools had commercial potential. That's when five partners started Entelos. In 2006, the company went public.

□ Physiome □ Rosetta Inpharmatics □ Molecular Designs □ GeneSpring □ Entelos □ Gene Network Systems □ Ingenuity □ 23AndMe □ Pathwork Diagnostics □ DeLano Scientific □ Dynameomics □ Celera □ Accelrys □ Bioimagene □ XDX □ Genomic

THE OPEN SOURCE BUSINESS MODEL

Warren DeLano, PhD
Principal, DeLano Scientific, LLC

"Scientists don't want black boxes in their software," says **Warren DeLano**, principal of **DeLano Scientific**. "They need to know how the thing works." So DeLano believes in making software that's open source rather than proprietary.

When he created PyMOL, an open source molecular visualization tool, DeLano hoped it would prove useful to many researchers and that, like other open source projects, it would benefit from an influx of good ideas, new features and code that would make it a self-sustaining project. The first part of that vision became a reality: PyMOL has proven quite useful. DeLano estimates that a quarter of all macromolecular crystal structure images published in the scientific literature today are created using PyMOL. But the software did not become independently self-sustaining. PyMOL's users aren't necessarily software engineers, says DeLano—they don't typically contribute code back to the project. Hence the need for a company—DeLano Scientific—to fulfill that role.

Four years into it, DeLano says the company can definitely support one salary, and will likely soon support two. "A business person would be trying to maximize profit by restricting the intellectual property and doling it out only for a high license fee," he says, "But because PyMOL is open source and it's important to me that it remains as such, we have to find other ways to grow revenue."

The open source nature of the software gives DeLano little leverage to charge high fees. He relies on negotiated relationships in which the clients recognize that, because they aren't contributing code back to the project, they should contribute some money instead—in the form of a subscription. For big customers, DeLano provides more individual interaction and support.

His is not the first company created to support open source software. RedHat serves that purpose for Linux users. And MySQL has a similar model. "Open source software is widely used, which translates into a very large impact on society from open source. But it's not a profit center like proprietary software has been for companies like Microsoft, Oracle, Apple and SAP."

And although he's a big proponent of open source software, he says that companies making proprietary software for biologists are starting to succeed by creating programs that act more like open source, with open architectures. "**Accelrys'** model is a very open-architecture, pluggable system," he says. "You can hook your code into these networks of capabilities which are visible and open. It's all proprietary architecture, but it's open in so many ways that a lot of flexibility is there."

As for DeLano Scientific, "On the continuum of open source to proprietary, we're somewhere in the middle," says DeLano. "You can get free versions, but you can get more value if you buy your subscription." He hopes to grow the business, but that's not the driving factor. "We're already achieving the kind of impact I wanted to achieve."

"What's unlikely to happen is to solely pursue the science and serendipitously have business opportunities and revenue drop in your lap," says Ramon Felciano.

(continued from page 9)

Alternative Investment Market (AIM)—a good market for small-cap technology companies that can show they have a revenue stream and growth opportunities, says Bangs. He says their success springs from three things: a technology that's proven to be of value to their customers in a short time frame; financial discipline over the last ten years; and maintaining focus on a few applications rather than expanding too rapidly.

Ingenuity Systems, Inc. is another company that strives to make drug discovery more efficient for biopharma. "It was not enough to provide a tool that would only be used by informaticians," says **Ramon Felciano, PhD**, founder and chief technology officer at InGenuity Systems, Inc. "At the end of the day, they are typically part of a larger team with a goal to discover a new drug, to understand its safety, to validate a lead. So really focusing on those more fundamental scientific and business goals I think helped us stay on track."

After eight years in business, Ingenuity's initial business idea is still working for them. They help researchers put high-throughput experimental results into the context of what is already known about a disease or cell. With the advent of high-throughput experimentation, "existing methods for understanding experimental data didn't scale well to the volume of data being generated," says Felciano. "We wanted to see if we could bridge that gap."

So Ingenuity created a large-scale platform with, at its core, a set of biomedical ontologies and a knowledge-base representing what is known about biology. "The "so what" of the data is hard to find because it's buried in research documents, figures, tables, PowerPoint slides and other non-structured repositories," says Felciano. But that's the information Ingenuity has gathered together. And they've made it available to their customers through the company's flagship product, Ingenuity Pathways Analysis: Researchers upload a dataset and run analytic algorithms to build de novo pathways linking their data and Ingenuity's knowledge base.

Industry, government and academic researchers use the knowledge base to identify or validate a new drug target or understand how a drug functions, including its toxic side effects. "We're trying to accelerate and improve the quality of scientific results that scientist users can produce," Felciano says. "We think we're doing well when they can get their work done more quickly and at a higher level of quality—better and faster."

The company now employs 85 to 90 people, not counting the part-time researchers around the world who help

Ramon Felciano, PhD
Chief technology officer & chief architect, InGenuity Systems, Inc.
Felciano and fellow Stanford students decided to start Ingenuity Systems in 1998 after they won an entrepreneurial competition held by the Business Association of Stanford Engineering Students (BASES). "Winning the competition was a threshold for us," says Felciano. Afterward, they received calls from VCs and did another round of conversations with researchers and industry people. "A lot of our initial understanding of the problem and the solution appeared to be valid. So we decided to give it a shot."

Intelligenetics □ Molecular Applications Grou
Molecular Mining □ Physiome □ Rosetta Inpharmat
Molecular Designs □ GeneSpring □ Entelos □
Network Systems □ Ingenuity □ 23AndMe □ Path
Diagnostics □ DeLano Scientific □ Dynameomics > C

Michael Sherman
Former CEO of Protein Mechanics, Now the chief software architect for Simbios, A National Center for Biomedical Computing at Stanford.
In 2000, Michael Sherman and his partner raised about four million dollars for their new company—Protein Mechanics—in one day. Their proposal: to apply their mastery of mechanical engineering and computer science to biological simulations. They started out developing software without much of a business model in mind. But, he says, the venture capitalists eventually wanted them to make drugs. Sherman saw that his company's chance of making a good drug was low. If he were to do it again, Sherman would do what he had done previously: start a sustainable business to sell useful software and make money, without relying on VCs. Although he currently works in academia, Sherman says he may eventually head back to the business world. "Academia is great because you can sit and think and do the right thing, but it has the character that basically nothing's at stake. Businesses are kind of cool because a lot of things matter, which makes it exciting. It can be a disaster too, but that's part of making it exciting."

Genomic Health □ Genstruct □ Optimata □ Cognia □ Myriad Genetics □ Protein Mechanics

"Be very cautious when taking money from venture capitalists because you might find your goals don't align well later," says Michael Sherman.

keep the knowledge base up to date. And, according to Felciano, Ingenuity's revenue and number of users are growing, as is their publication record. Over the last two to three years, the number of scientific publications citing Ingenuity's platform has grown 250-300 percent. "It's a great trajectory that we're looking at and a validation that we're doing well."

Hill, Felciano and Bangs all recognize that, nowadays, VC support for businesses like theirs is rare. As Bangs puts it, "VCs want companies that are going to have a good-size market with good revenue possibilities. If your market is to sell to a small part of the pharmaceutical industry, then you have a small number of customers with potentially varying budgets." So Bangs suggests an alternative to VC: the small business route. "Do it the old-fashioned way. Put a second mortgage on your home, get SBIR money, bank loans, and grow the business slowly. That's what you have to do in the current environment."

Michael Sherman learned that lesson the hard way. His company—**Protein Mechanics**—also saw its start in 2000 when "venture capital was flowing a little too freely." The company developed software to simulate proteins, but after a few years, their investors wanted them to make drugs. "We didn't know anything about drugs," he says, so the company sold at a fire sale in 2004.

His advice to scientists contemplating a startup: "Be very cautious when taking money from venture capitalists because you might find your goals don't align well later." And he agrees with Bangs' recommendation: go with the small business model. "The fact that Protein Mechanics isn't around anymore is a minor thing for the venture capitalists because they figure they invest in ten companies and nine might fail. But it's a big deal to me. I'd have rather built a sustainable company, but that wasn't interesting to the people who funded it."

Felciano, by contrast, has had a good experience with Ingenuity's VCs. But, he notes, "Good partners—whether VCs or others—can help you succeed, and bad ones can become major obstacles." Despite being a scientist himself, he's a realist about business. "If you want to purely drive the science there are better places to do that," he says. "What's unlikely to happen is to solely pursue the science and serendipitously have business opportunities and revenue drop in your lap."

The Razor Blade

In recent years, some biocomputation startups have drawn VC attention

not by selling software to biopharma, but by masquerading as diagnostic, proteomic, or genomics companies. To Anderson of **Pathwork Diagnostics**, this is clearly the way to go. "A well-designed biocomputation product could be the ultimate razor blade. It just needs a little invention," she says.

According to Anderson, funding for bioinformatics platforms is virtually nonexistent from this point on. It's a lesson she learned from initially naming her company Pathwork *Informatics*. No one would fund it. "There are a lot of decent platforms out there. The world doesn't need a new one," she says. "However, applications that are anchored to solving a clinical problem and that map out a path to revenue and products are something else."

Coming up with a commercial product from good science is not always easy.

"The challenges start with how to frame a problem in such a way that research can discover an answer," says Anderson, "And then how do you layer it back into something that can be produced in a reasonably cost-effective way and delivered and sold to clinicians. There are many challenges in taking an idea into the actual practice, but that's what our business is about. If it were easy it wouldn't be so much fun."

Anderson has built her career around this kind of thinking. "What I've been really fascinated with my whole life is how you craft products that are successful along with a business model that's successful as well," she says.

Now, after three years in business, Pathwork Diagnostics has its first product before the FDA. It addresses a common problem in cancer recurrence: Often, clinicians cannot determine where the cancer started in the body. So, Pathwork's Tissue of Origin Test answers this problem by analyzing gene expression data from tumor biopsies. With the analytic report in hand, a physician can better determine how to treat such cancers.

But this is just the beginning, says Anderson. "The Holy Grail in oncology diagnostics is predicting treatment response. Our real intent is to develop a platform that could answer that question." The Tissue of Origin test positions the company to do just that, she says. "This is what makes our business model compelling and what helps pull us over the hump in terms of funding."

Pathwork Diagnostics is not alone in taking this approach. XDx, founded in 2000, offers a test to determine the likelihood that a heart transplant will be rejected. "XDx's analytics are their razor

"A well-designed biocomputation product could be the ultimate razor blade. It just needs a little invention," says Glenda Anderson.

Glenda Anderson
Founder and chief technology officer, Pathwork Diagnostics
After spending nearly twenty years leading research and development organizations for healthcare manufacturers, Glenda Anderson decided in 2001 to take a closer look at bioinformatics. "It was clear that biocomputation would form the basis for new medical products over the next 20-30 years," she says. And she thought her engineering degree in computer modeling and experience taking ideas from technology or science into a clinical application and a product would be helpful in building such a business. "Biocomputation is a field that lends itself to asking questions about how to put a business model and product together to create real value." In 2003, Pathwork was born, and the company's first product is now before the FDA.

Dynameomics □ Celera Molecular Designs □ Gene
Systems □ Ingenuity □ Pathwork Diagnostics □ De
Celera □ Accelrys □ Bioimagene □ XDX □ Genomic

Valerie Daggett, PhD,
Future founder of Dynameomics
Professor of medicinal chemistry at the University of Washington
A few years ago, Valerie Daggett registered the domain name for dynameomics.com. "I've been getting ready to do this for a while," she says of the venture she has yet to launch. "This is the work we've been doing for the last 15 years or longer. I want to see it come to fruition."

GeneSpring □ Entelos □ Gene Network Systems
Ingenuity □ 23AndMe □ Pathwork Diagnostics
DeLano Scientific □ Dynameomics □ Celera □ Accelr
□ Bioimagene □ XDX □ Genomic Health □ Genstruct
Optimata □ Inforsense □ Cognia □ Myriad Genetics
Protein Mechanics □ Intelligenetics □ Molecul

blade," says Anderson. "And they can command a high price for each potential transplant recipient they evaluate because the cost of failure is so high."

Genomics startups with a heavy biocomputation focus are also gaining some momentum according to Dotzler. A number of companies offer to analyze DNA from a blood sample or cheek swab to trace a customer's genealogy or give advice about disease susceptibility. One new company, **23andMe**, launched in 2006, hopes to take that analysis further. As new gene association studies appear in the literature, they gather the results together into a database. "There's an exciting flood of information about gene/disease associations," says **Brian Naughton, PhD**, one of the company's core team. Making sense of it is not a trivial task, he says. But he and fellow team member **Serge Saxonov, PhD**, both recent graduates of Stanford's Biomedical Informatics program (BMI), say they have the skills to do it.

This year, the company hopes to release its first product, which will require a saliva sample rather than a blood sample or a cheek swab. It will analyze the customer's entire genome looking for SNPs—the single nucleotide polymorphisms that distinguish individuals from one another; some SNPs are also associated with disease. "Most people are curious about their genetics—their families, their ancestry and their health—but have nowhere to go to learn about it," says Naughton. "We want to fill that niche."

Venturing Out of Academia

Startups often blossom directly from academic research. The very first biocomputation startup—**Intelligenetics**—began that way, says **Doug Brutlag, PhD**, professor of biochemistry and medicine at Stanford. He was part of that launch, which began in the late 1970s. After sequencing methods had been developed, he says, Stanford researchers put all the accumulating data and software on a computer and made it accessible over phone lines—before the Internet really took off. "We wanted to show people how useful it was so we made it freely available to biologists and everybody started using it." But, says Brutlag, "I didn't think it was appropriate to support biologists from an academic environment. Support is better provided by commercial entities." Moreover, he says, in this area it's very important to do technology transfer. "If you just publish your work and don't provide support for it, then people

Making open source code available through an academic web site can lead to frustration, says Valerie Daggett. "People in your lab become unpaid consultants responding to every call and email that comes in about how to use the software. I don't want to do that to my research program."

SUPPORTING TECHNOLOGY TRANSFER: THE STANFORD WAY

Stanford opens up channels for getting technologies out to the real world in three important ways. First, it allows faculty to consult up to 1 day per week (20 percent of their time). Second, it has a very proactive office of technology licensing that identifies potentially useful technologies in the lab—even if faculty aren't aware of it—and then files patents and markets these to companies that may want to productize them. And third, it has a corporate affiliates program that allows industry to provide funds—e.g., fellowships—to departments. In return, the affiliates get preferential access to workshops and students (for recruitment).

This approach to commercializing technology has produced Google, Yahoo, recombinant DNA, and the Yamaha synthesizer, to name a few, and has been a model for other universities to follow.

won't use it." So Intelligenetics was born—and it lasted about 13 years.

Nowadays, says Brutlag, things are different. With the development of the Internet, he does much of his technology transfer himself. "We make software tools freely available to not-for-profit institutions and allow for-profit companies to license them from Stanford University," he says.

But **Valerie Daggett, PhD**, professor of medicinal chemistry at the University of Washington, says making open source code available through a web site can lead to frustration. "People in your lab become unpaid consultants responding to every call and email that comes in about how to use the software. I don't want to do that to my research program." Moreover, she says, users of her protein dynamics simulation software really do need help. "These are very complicated simulations to setup and run, and biologists don't necessarily have the equipment or the skills to do it." The result: the academic software provider gets blamed for mistakes made by the users.

So Daggett is contemplating a new startup. She calls it **Dynameomics**. "It's very hard in academia to take your work to the next level and see it transition out of the computer lab and into an application that people will use," she says, but the technology transfer office at the University of Washington is helping out.

Daggett's lab has created the biggest database of protein simulations in the world. With upwards of 2500 protein simulations it becomes prohibitive to host the work (20 terabytes and counting) from an academic lab. This Spring, Daggett's group plans to take an early step toward commercialization when they launched dynameomics.org. The dot-com site will come later. The dot-org site will give people free access to simulations representing the top 30 folds (structures) of proteins known from the Protein Data Bank (PDB). These represent 50 percent of all known structures. From the web site, biologists will see the simulations; the structures generated by the simulations; movies of how the protein moves over time; and metadata (analyses of the simulations).

At some point, Daggett expects to take orders for simulations of specific proteins not represented in the set. Although dynameomics.com will offer much more, Daggett says, it's still unclear whether the product will be a software modeling package or drug design services or both. Several software companies and VCs have shown interest in both options, she says.

Either way, the dynamic aspect of proteins is really underutilized thus far and Daggett hopes to make it available to industry. "By basing functional analysis of proteins on static structures we're missing a good deal of the picture." And because there are loads of proteins and a lot of interesting biomedical targets that exert their actions through changing protein conformation, she thinks her protein dynamics simulation tools are a product people will need for some time to come. But until she settles on a product, it's not entirely clear whether she's commercializing a razor or a razor blade—she hopes it's the latter.

Sticking With It

Regardless of the product being sold, says Hill, biocomputation startups require perseverance. "It's not an industry that typically gives quick payoffs," he notes. "Unlike pure technology companies, we don't have to just make some widget and sell it and be a huge success. ... The ultimate test in our field is affecting disease in a living human. ... It's much more difficult than just developing technology." So, he says, you have to do it for the right reasons. "If you just want to make a lot of money, go to Wall Street. Or create the next YouTube."

Hill himself was following his scientific interest when he started GNS seven years ago: He wanted to discover how living systems work in a fundamental way. "Having that goal collide with some serious unmet and burning needs in a very practical and very lucrative industry (pharmaceutical development and biotech) is what made me do this. The fact I could do both—have my cake and eat it too—that was too good to pass up." □

Modeling Cancer Biology:

BY KRISTIN COBB, PhD

How mathematical models are transforming the fight against cancer

The most common test for prostate cancer (known as PSA screening) misses aggressively growing prostate tumors—the type typically seen in young patients. It's a fact that was accepted by the medical establishment in 2004 only after a seven-year study of 9000 men appeared in the *New England Journal of Medicine*. But **Kristin Swanson, PhD,** predicted the test's inadequacy in 2001 using a single differential equation—a "back of the napkin calculation" that "a high school student could answer." This is the type of powerful insight that mathematics can offer cancer biology, says Swanson, who is an assistant professor of pathology and applied mathematics at the University of Washington.

Unfortunately, mathematics has remained largely untapped and under-appreciated in cancer biology. Though mathematicians have been deriving formulas about cancer for decades, their work has been confined to mathematical and theoretical biology journals—a set of dense journals that the average biologist doesn't read. Biologists are also skeptical: How can cancer, which is so complex and unpredictable, be reduced to a set of neat equations?

But cancer biology may be at a turning point. Never before has there been a greater need for the field to embrace mathematics and computation. As biological data pile up at an astonishing rate, there is growing recognition that only quantitative approaches can pull it all together. As a result, quantitative cancer models are slowly making their way out of the theoretical and math journals and creeping into mainstream cancer biology. Leading biology journals like *Cell* and *Cancer Research* now contain theory sections. And, in 2003, the NIH established the Integrative Cancer Biology Program—which now funds nine inter-disciplinary centers that are applying quantitative modeling and systems biology approaches to cancer (the ICBPs).

These efforts promise enormous pay off. Modeling can streamline wet-lab experiments; give scientists deeper insight into how tumors develop, grow, and spread; and even predict a patient's prognosis and optimal treatment regimen.

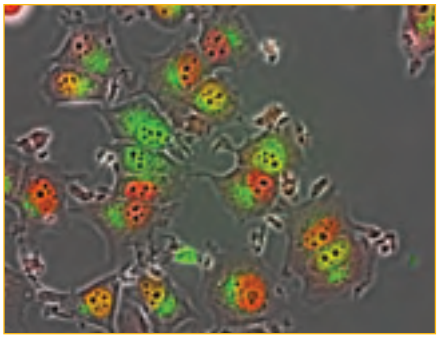
"Biologists tend to think of modeling as some sort of magical thing or black art," says professor **Philip K. Maini, PhD**, director of the Center for Mathematical Biology at Oxford University. "But we haven't done anything extra; we haven't done any jiggery-pokery or put any voodoo in there."

Mathematicians simply translate biologist's hypotheses into a formal set of testable equations, he says.

"Biologists are the first people to tell us that biology is very complicated; it's highly non-linear. Yet biologists use verbal reasoning, which is linear reasoning, which is the wrong model," he says. Mathematical models are needed to reach beyond where human intuition and linear thinking can take us, he says.

ple, **Natalia Komarova, PhD**, associate professor of mathematics at the University of California, Irvine, models the initiating event in colon cancer—the inactivation of the APC tumor suppressor gene. Normally, APC causes cells to enter apoptosis at the end of their "term" in the colon tissue, which helps prevent cancer.

Cells in the colon are constantly exposed to the elements, and thus have a high risk of mutation. Thus, it is imperative that colon cells turn over quickly. The bottom of each microscopic pit of colon tissue (called a crypt) contains adult stem cells whose job is to produce daughter cells to continually replenish the colon. These daughter cells climb up the crypt, differentiate into colon cells, and die off in about a week. It is a delicate balance, however: the quick turnover helps prevent cancer in the daughter cells, but the frequently dividing stem cells are vulnerable to accumulating mutations.



DNA-Damage Control. When cells are exposed to DNA-damaging radiation, they produce p53, an anti-cancer protein that causes damaged cells to undergo apoptosis (programmed cell death). Here, cells express fluorescently tagged p53 (green) and Mdm2 (red) following gamma irradiation. Time-lapse microscopy shows that, following DNA damage, p53 and Mdm2 levels undergo a series of pulses that vary in number from cell to cell. Courtesy of Galit Lahav's lab, department of Systems Biology, Harvard Medical School.

"Biologists tend to think of modeling as some sort of magical thing or black art," says Philip K. Maini, "But we haven't done anything extra; we haven't done any jiggery-pokery or put any voodoo in there."

What follows are some examples of how modeling is adding insight to intuition—from cancer initiation to metastasis and from the molecular to the patient level.

A Cancer Cell is Born: The Subcellular Level

Cancer arises through a series of genetic changes. Mutations in proto-oncogenes allow cells to grow and divide without the need for normal growth signals, and mutations in tumor suppressor genes allow cells to evade normal checks and balances—such as anti-growth signals and programmed cell death (apoptosis). Mutations in genes that detect and repair DNA damage facilitate the process by upping a cell's mutation rate.

Stochastic mathematical models help investigators test hypotheses about how cancer mutations accumulate. For example, **Natalia Komarova, PhD**, associate professor of mathematics at the University of California, Irvine, models the initiating event in colon cancer—the inactivation of the APC tumor suppressor gene. Normally, APC causes cells to enter apoptosis at the end of their "term" in the colon tissue, which helps prevent cancer.

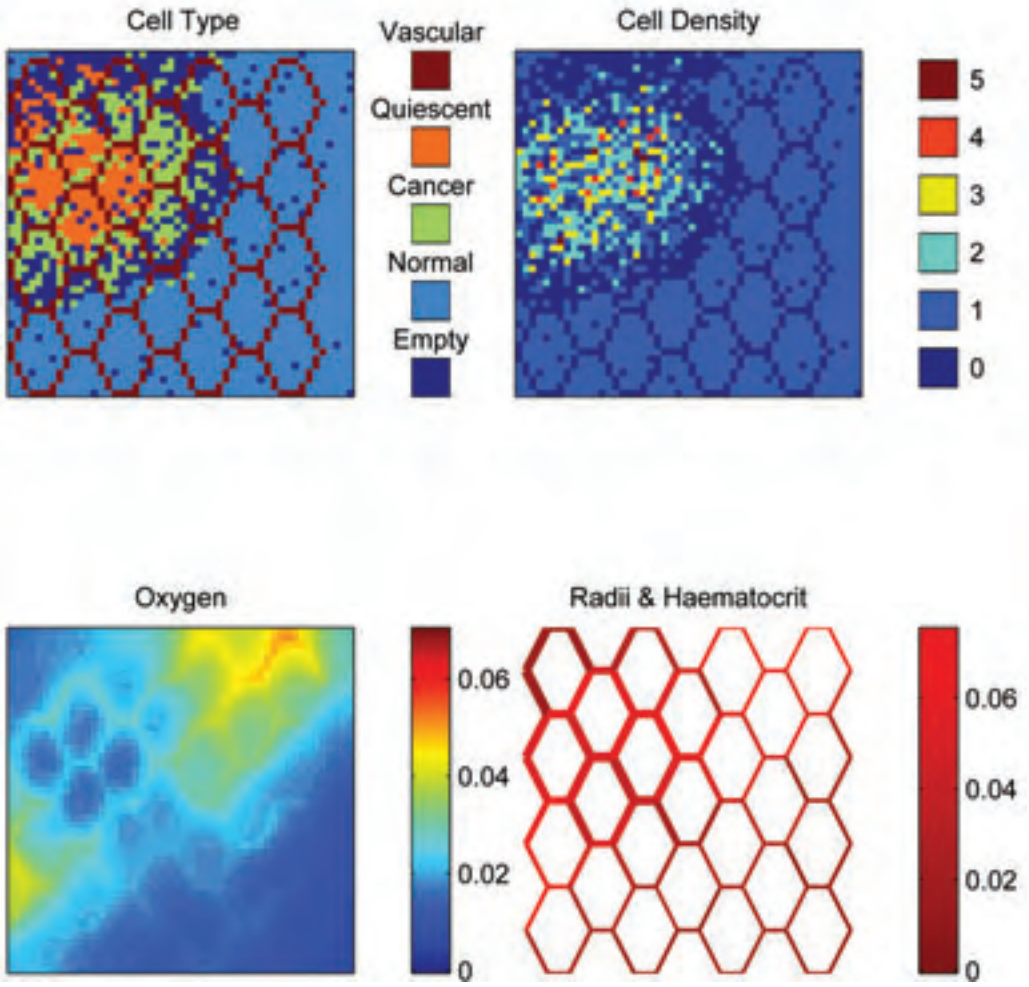
Cells in the colon are constantly exposed to the elements, and thus have a high risk of mutation. Thus, it is imperative that colon cells turn over quickly. The bottom of each microscopic pit of colon tissue (called a crypt) contains adult stem cells whose job is to produce daughter cells to continually replenish the colon. These daughter cells climb up the crypt, differentiate into colon cells, and die off in about a week. It is a delicate balance, however: the quick turnover helps prevent cancer in the daughter cells, but the frequently dividing stem cells are vulnerable to accumulating mutations.

One question that cannot be reliably answered experimentally is how many stem cells are in each crypt. Komarova tries to answer this question mathematically—by calculating the optimal number to minimize a person's chance of getting mutations in the APC gene.

"A situation like this is perfect for the application of modeling because in the model we can assume that there is one stem cell or that 50 percent of them are stem cells and we can see what happens," Komarova says. It turns out that, for young people, having many stem cells minimizes the chance of cancer. But for older individuals, having a few stem cells is the best strategy. Likely, evolution favored the optimal strategy for young people, since evolution acts on those of reproductive age, she says.

Besides probabilistic models of mutation "hits," other researchers model the signaling pathways involved in growth, anti-growth, and cell death. Typically, these models consist of systems of ordinary differential equations. Each equation describes the rate of change in the concentration of a particular enzyme, substrate, receptor, or signaling molecule as a function of its production, degradation, and reaction with other network players.

For example, **Galit Lahav, PhD**, assistant professor of systems biology at Harvard Medical School, models the p53 signaling network. p53 is a tumor suppressor gene that plays a crucial role in apoptosis, among other important anti-cancer functions. If specialized sensors in the cell detect DNA damage (or other dangers, such as oncogene over-



Virtual Angiogenesis. In these snapshots from a computer simulation of tumor growth and angiogenesis, the top panels show the presence and density of tumor cells at time=5; cells in the core of the tumor become quiescent because oxygen cannot reach them. As the tumor grows, tumor cells secrete angiogenesis factors that cause new blood vessels to grow and supply extra oxygen and red blood cells to the tumor (bottom panels). Courtesy of Philip K. Maini.

expression), they trigger p53 to initiate a cascade of events leading to the cell's death. More than half of all human cancers contain a mutation in p53, making it the most common cancer mutation.

Lahav studies the p53 network both experimentally and theoretically. "We go back and forth from the bench to the computer," she says.

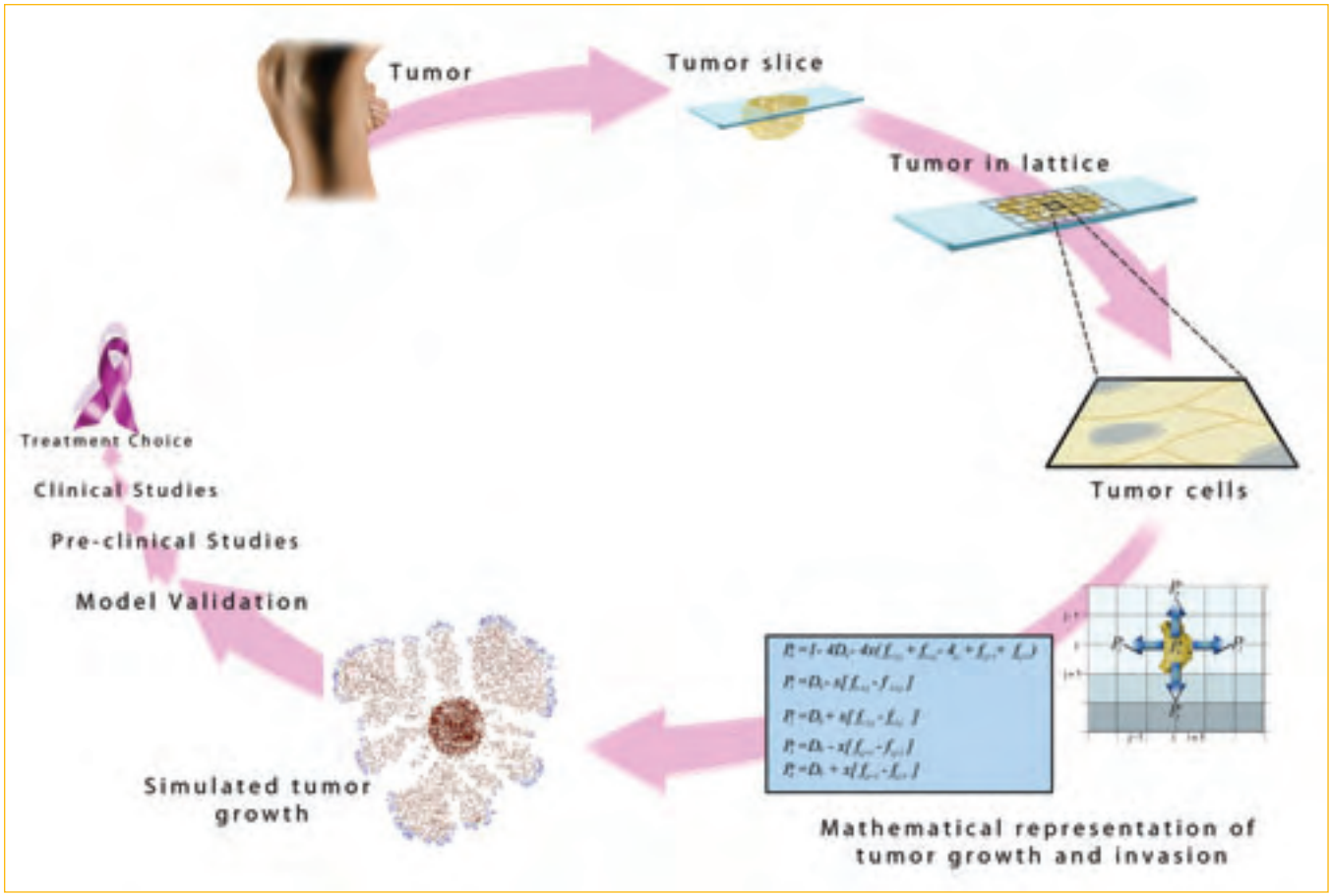
In the lab, Lahav uses fluorescence microscopy to measure the changing levels of p53 and other proteins of interest (all tagged with fluorescent markers) after a cell is exposed to DNA-damaging gamma radiation. On the theoretical side, she uses a series of ordinary differential equations to predict the changing levels of p53 and related proteins, such as Mdm2, which is involved in a negative feedback loop that regulates p53.

"The idea of the models is to help us predict how the network will behave in response to different treatments and to suggest new experiments," she says.

For example, she discovered that levels of p53 oscillate following gamma irradiation, and she is using modeling to help understand these oscillations. If they are important for apoptosis, then some cancer drugs may work better if delivered in pulses rather than continuously, she says.

The Growing Tumor: The Cellular Level

Once tumor cells have acquired the ability to propagate unchecked, they grow into a small ball of cells—which mathematicians model as a growing spheroid. Initially, the tumor feeds on oxygen and nutrients that diffuse to its surface. But these supplies cannot penetrate deep into the tumor, so cells in the core become dormant or die of starvation. The limited nutrient supply curbs the tumor's growth to about half a millimeter in diameter—and if the story ended here, the tumor would be harmless.



Forecasting Invasion. This graphic depiction of a mathematical model developed by Vito Quaranta and Alexander Anderson predicts whether a tumor will become invasive. The tumor is represented on a two-dimensional grid. Each virtual cell is accounted for on the grid and its behavior (e.g., growth, movement, death) is tracked based on mathematical functions and partial differential equations. Solving these equations in sequential time-steps generates a computer simulation of tumor growth and invasion. This approach has the potential to predict disease outcome based on precise quantities measured in the tumor of a specific patient. The model was described in: Anderson et al.* Cell. *2006 Dec 1;127(5):905-15. Courtesy of the journal* Cell. *Graphic by Dominic Doyle.

Unfortunately, as cells in the center become starved of oxygen (hypoxic), they release chemicals that stimulate angiogenesis—the growth of new blood vessels. These chemicals encourage blood vessel cells (endothelial cells) to migrate toward the core of the tumor and supply it with blood. Now the hungry tumor can feed unhindered. At the same time, the tumor gains a connection to vessels throughout the body, giving it an escape route for metastasis.

One strategy for modeling angiogenesis is to set up systems of partial differential equations that describe how the tumor and vasculature are changing in both time and space (how their shapes are changing). For example, **Zvia Agur, PhD**, President of the Institute for Medical BioMathematics in Israel, has modeled angiogenesis using three interconnected modules of partial differential equations. Her equations describe: the changing volume of tumor cells (which depends on factors such as oxygen concentration); the changing volume of immature blood vessels (which depends on how quickly tumor cells release VEGF, a potent angiogenesis factor); and the changing volume of mature blood vessels (which depends on molecular signals that promote maturation). "The simplest model we could make was quite complex," Agur says.

She also set up an experimental system to validate her model. Her team implanted small balls of ovary cancer cells into mice and measured changes in the size and shape of the tumors and the blood vessels using MRI. For each individual tumor, Agur simulated its expected growth in the computer and then compared the simulation results to the actual results from the lab—and the prediction was quite good, she says.

She then simulated what would happen if tumors were treated with anti-angiogenesis drugs, and got a surprising result: The model showed that treatment with a single anti-angiogenesis drug is not sufficient to eliminate a tumor; rather, combinations of anti-angiogenesis drugs are needed.

"At the time, the anti-angiogenesis drug Avastin was very much in the news, and people thought that it could be used on its own," Agur says. "Genentech was doing extensive clinical trials using Avastin monotherapy, and it took them another year or so to realize that we were right."

Invasion and Metastasis: The Tissue Level

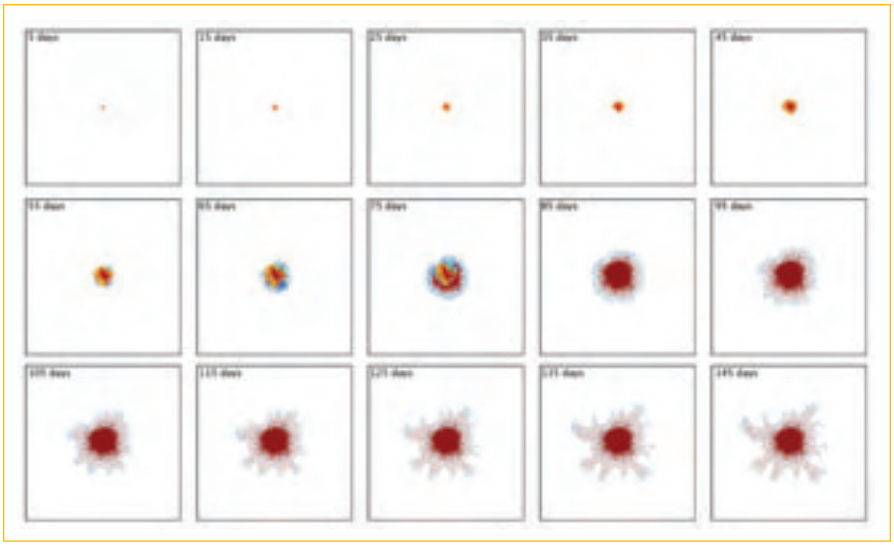
For a while, the tumor continues to grow as a cohesive ball of cells with smooth edges. At this point, the tumor is still often curable, as a surgeon can just scoop it out, says **Vito Quaranta, MD**, professor of cancer biology at Vanderbilt University and also principal investigator of the Vanderbilt Integrative Cancer Biology Program (one of the nine ICBPs).

But, eventually, some rogue cells break away from the growing tumor and invade the local tissue. To become invasive, tumor cells have to pick up certain abilities—they must escape cell-to-cell adhesion, migrate along the extracellular matrix (the surrounding connective tissue), and secrete enzymes that digest the extracellular matrix.

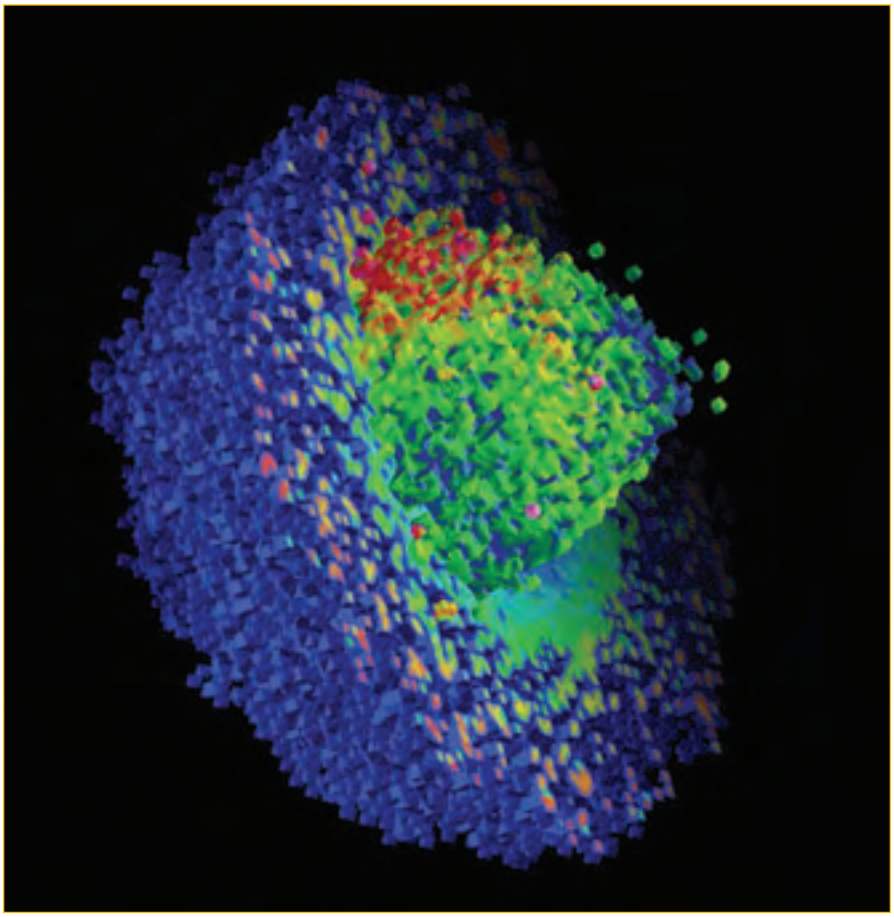
Eventually, these invading cells burrow their way into the blood or lymph systems and spread (metastasize) to distant sites, where they seed new tumors. Now it is impossible to just reach in and scoop out the tumor—and the cancer is much more deadly.

Quaranta, who is an experimentalist, collaborates with mathematician **Alexander Anderson, PhD**, senior lecturer of mathematics at the University of Dundee in Scotland, to model the process of invasion. They use a "hybrid discrete-continuum" model, which means molecules and proteins—such as oxygen and matrix-degrading enzymes—are modeled as continuous densities, but cells are modeled as individual, discrete entities that make autonomous decisions. Such agent-based models are computationally intensive, so simulations are limited to about five million cells (in contrast, a tumor may have a few billion cells).

Cells move on a two-dimensional grid that represents the changing micro-environment—including the concentrations of nutrients, enzymes, and extracellular matrix proteins. Cells have a certain probability of moving to each adjacent point on the grid (called a biased random walk). For example, cells are more likely to move to regions where oxygen levels are high. Cells are also allowed to adhere to each other, migrate, degrade their surrounding tissue, divide, even die, according to cer-



Cancer Invasion. Starting with only 50 cancerous cells, this mathematical simulation shows how a tumor grows first into a smooth ball of non-invasive cells and then—under the right conditions—into an invasive mass that fingers into the surrounding environment. Blue cells are highly aggressive; orange cells are less aggressive, and brown cells are dead. Courtesy of Alexander Anderson



Virtual Tumor. A simulation of one half of the whole living tumor cell population (outer half sphere) and the complete necrotic (dead) tumor cell population (inner sphere). Coloration relates to cell-adhesion value—cells on the outer surface of the tumor all have zero cell-to-cell adhesion. Courtesy of Alexander Anderson

THE INTEGRATIVE CANCER BIOLOGY PROGRAM

Established by the National Cancer Institute in 2003, the Integrative Cancer Biology Program (ICBP) funds efforts in computational modeling and systems biology approaches to cancer. "It's difficult to do this type of research because you have to do both experimental biology and sophisticated computational approaches. Pulling those kinds of groups together really requires a structure like a center," says **Jennifer Couch, PhD**, IT/Computational Biology Coordinator for the ICBPs. "Our vision is always that these centers will sort of form the locus for the development of a community focused on integrative cancer biology." Currently, the ICBP funds nine centers:

Todd Golub, M.D., Dana-Farber Cancer Institute, Boston, Mass.
Identifying protein kinase signatures in cancer.

Joe W. Gray, Ph.D., Lawrence Berkeley National Laboratory, Berkeley, Cailf.
Modeling signaling networks to identify patients for targeted therapeutics.

Tim H-M Huang, Ph.D., Ohio State University, Columbus, Ohio.
Epigenetic changes in cancer genomes.

Timothy Kinsella, M.D., University Hospital of Cleveland, Cleveland, Ohio.
Modeling mismatch repair defective malignancies.

Sylvia Plevritis, Ph.D., Stanford University School of Medicine, Stanford, Cailf.
Regulatory and signaling pathways in neoplastic transformation.

Joseph Nevins, Ph.D., Duke University, Durham, N.C.
Cell signaling pathways in cell proliferation and oncogenesis.

Thomas Deisboeck, M.D., Massachusetts General Hospital, Boston, Mass.
Model and simulation of multicellular patterns in cancer.

Richard Hynes, Ph.D., Massachusetts Institute of Technology, Boston, Mass.
Modeling cancer progression.

Vito Quaranta, M.D., Vanderbilt University Medical Center, Nashville, Tenn.
Model and simulation of cancer invasion.

tain parameters—which Quaranta measures experimentally—such as speed of migration and the rate of cell division. Moreover, as cells divide, they acquire mutations that make them more aggressive and invasive (better able to proliferate, migrate, and enter the surrounding tissue).

The resulting computer simulation—which shows a slice of a growing tumor—looks a bit like a weather forecasting map, Quaranta says. Virtual cells divide, move, and change colors to represent their changing phenotypes—for example, blue for highly aggressive, orange for less aggressive, and brown for dead. Depending on the conditions, tumors will either grow with smooth margins (remain non-invasive) or will finger out into the surrounding tissue (become invasive).

When they ran their model, they got a surprising result: "We found that if the surrounding environment is a smooth, easy environment, then the cells tend to be non-invasive. But if you put pressure on the cells, say by reducing oxygen or making the landscape very hard to deal with, then the tumors become invasive," Quaranta says. In gentle conditions, many different tumor cell phenotypes co-exist, but when the conditions become harsh one or two super-aggressive phenotypes prevail.

Anti-angiogenisis drugs, inflammation, even chemotherapy and radiation therapy might create conditions for aggressive phenotypes to become dominant, Quaranta says.

Their findings were published in the December 1 issue of *Cell*, a leading biology journal. Anderson says that before his collaboration with Quaranta he would never have dreamed of submitting a paper to *Cell*.

"There was a bit of a wrestling match over the exact wording. But that ultimately paid off because it produced a paper that was really aimed at their audience, and that they could understand," Anderson recalls.

"Ultimately I'm hoping this is going to be good for the math biology community, because if I can get a paper published in *Cell*, then why can't somebody else?" he adds.

Brain Tumor Revealed. Only 10 percent of glioma cells are visible on MRI (the bright white area above); a computer simulation superimposed over the MRI helps doctors visualize the rest. The yellow/pink/red areas show that glioma cells may have diffused way beyond the borders of the mass seen on MRI. Courtesy of Kristin Swanson

"It's a nice change," Alex Anderson says. "To have mathematics driving experimentation, instead of us just always playing catch up with the biology."

Quaranta says the partnership has changed his biology as well. "Our experiments now are actually driven by mathematics. So we're entering an era of mathematics-driven experimental biology that is going to be interesting to see."

"It's a nice change," Anderson says. "To have mathematics driving experimentation, instead of us just always playing catch up with the biology."

Gliomas: The Patient Level

Kristin Swanson (of the University of Washington) also works on modeling tumor invasion, but in glioma—a specific type of brain tumor that is particularly invasive and deadly. By the time a glioma mass is detectable on MRI, invasive glioma cells have already wandered far into the brain. Swanson compares it to an iceberg: the mass you can see represents only about 10 percent of the total tumor cells in the brain; the rest are undetectable, making it impossible to remove them.

Swanson's model consists of a series of partial differential equations that describe how the mass of glioma cells spreads within a virtual brain—a three-dimensional lattice complete with areas of white and grey matter (glioma cells migrate at different rates in these different tissues). Her computer simulations show the changing density of glioma cells along sections of the virtual brain—for example, red where the tumor density is high and blue where density is low.

A glioma patient's MRI reveals only the detectable part of the tumor, so Swanson uses her simulations to visualize the undetectable portion and predict how the tumor will spread.

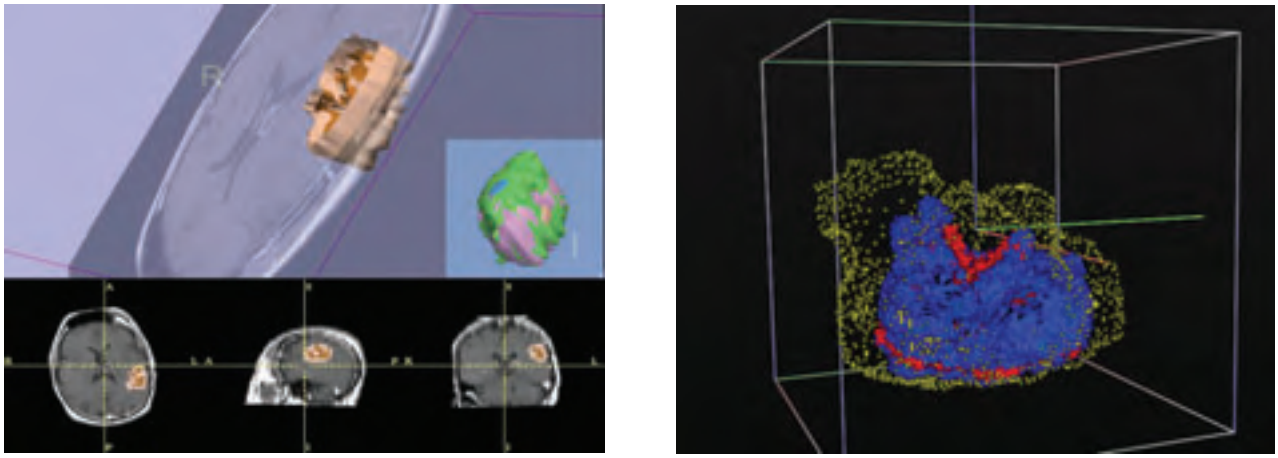
"Just using diagnostic MRI and this mathematical model, you can predict survival with very reasonable accuracy for an individual patient," she says.

Her model can also be used to run *in silico* clinical trials. "It's hard to test therapies for gliomas because patients don't live long and you can't see what's happening with most of the tumor," Swanson says. "But if you have a model for the expected behavior of an individual patient's tumor, then you can assess the success of therapy relative to the expected behavior."

Another investigator working on gliomas is **Thomas S. Deisboeck, MD**, who is assistant professor of radiology at Massachusetts General Hospital and Harvard Medical School, as well as principal investigator of the Center for the Development of a Virtual Tumor (CViT), one of the nine ICBPs. Deisboeck uses a discrete, cell-based approach, rather than a continuous approach, to predict how cells will spread through a three-dimensional virtual brain. This allows him to connect what is happening at the subcellular to the cellular and tissue levels. "Our main interest is multi-scale, multi-grid, multi-resolution modeling," he says.

His virtual cells can proliferate, migrate, die, and respond to the environment and each other. They also contain a nucleus, cytoplasm, membrane and even working biochemical pathways. The actions of particular biochemical pathway components can influence the behavior of the cells and the spread of the tumor. For example, Deisboeck is modeling how the EGFR (epidermal growth factor receptor) pathway acts as part of a molecular switch that turns glioma cells from proliferative (dividing) to migratory (invading local tissue).

Though he eventually hopes to use his models to improve patient treat-



From Patients to Molecules and Back. MRI images from a brain tumor patient (left) are used to build a 3-D* in silico *model of the growing tumor (right). Each cancer cell is represented as an autonomous agent that can move in space and change phenotypes (proliferation* = blue; *migration* = red; *quiescence* = green*). Each cell's behavior is determined by equations that represent the cell's intracellular networks, cell-to-cell interactions, and cell-microenvironment interactions. Images Courtesy of Thomas S. Deisboeck. The underlying multi-scale model was described in Zhang et al. J. Theor Biol. 244(1): 96-107,2007.

ment, his first goal is more modest—to improve diagnostics and patients' quality of life.

"What would be already a very significant achievement is if you could argue that instead of taking three MRI images say over six months, the combination of *in silico* modeling with two images would be just as informative," he says.

As principal investigator of CViT, Deisboeck's broader vision is to build an online community of cancer modelers and a toolkit for multi-scale *in silico* cancer research. CViT is creating new infrastructure, including a digital model repository that will allow people to share and combine models (www.cvit.org).

Bridging the Divide

The above examples share a common theme—a tight link between the lab or clinic and the computer. But these examples are still the exception rather than the rule. The major obstacle in bringing modeling to cancer biology remains the lack of communication between modelers and experimentalists.

On the one hand, biologists and clinicians tend to be mathematically illiterate and fearful of mathematics, says **Robert A. Gatenby, MD**, professor of radiology and applied mathematics at the University of Arizona.

On the other hand, mathematicians tend to neglect the biology, he says. "Mathematicians will set up equations and then they'll do uniqueness theorems and things like that, which are very mathematical approaches but utterly meaningless biologically. This just reinforces the biologists' opinion that this is meaningless and can't be even remotely helpful to them."

Getting these two groups to speak a common language and embrace a common objective is a major challenge. But efforts like the Integrative Cancer Biology Program are helping to bridge this divide and to train a new generation of scientists who are eager to cross disciplines.

"A lot of the students nowadays don't want to get locked into just one field; they are looking for these multi-connections between a lot of disciplines. They may be engineering majors, but they want to know something about biology," says **Daniel Gallahan, PhD**, Project Director of the Integrative Cancer Biology Program at the NIH. "That's been a pleasant surprise to me and it's something I see as a critical component for the future of this effort."

"Maybe in one or two generations, we'll have experimental biologists who are fluent in the language of mathematics," agrees Vito Quaranta of Vanderbilt University.

The Cutting Edge

Quaranta believes that a new era of cancer biology is fast approaching. "The way we do experimental oncology is going to change dramatically as these mathematics-driven simulations become more and more common place," he says.

As quantitative modeling moves from the margins of cancer biology to the mainstream, it is also presenting cutting-edge challenges for modelers.

"It's raising issues that mathematicians and modelers have never had to face before," says Philip Maini of Oxford University. For example, how do you model the mechanics of a growing tissue? How do you build multi-scale models that are accurate across different biological and time scales? How simple or complex is the optimal model?

"It's a very interesting time for graduate students and post-docs to be involved, because it's an area that's now really beginning to take off," Maini says. "Yet it isn't so far developed that you can't immediately start making inroads." □

SimbiosNews

BY KATHARINE MILLER

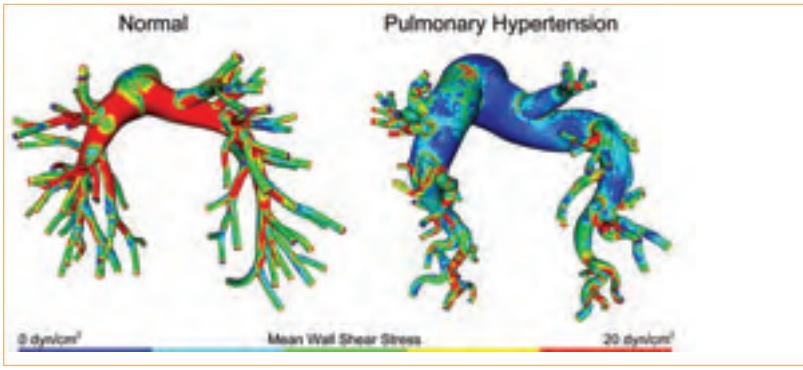
SimVascular to Simulate Cardiovascular Flow

On the computer screen, vessels throb realistically with each pump of the heart while the river of blood swirls and pools at curves and intersections. This is a simulation built with SimVascular—an integrated software system for doing cardiovascular modeling. Starting this summer, it will be available for testing by beta-users.

"Most people doing simulations of blood flow focus on several inches of the vascular system and treat it as a rigid tube," says **Charles Taylor, PhD**, associate professor of bioengineering at Stanford University and PI for the cardiovascular dynamics project within Simbios. "This software allows you to do things that are much more extensive."

SimVascular creates geometric models from medical image data; converts those into finite element meshes; models blood flow through these different geometries; solves governing equations of blood flow, wall dynamics and pressure; and then visualizes the results. It is uniquely able to simultaneously model blood flow and muscle wall dynamics; handle patient-specific geometric modeling; and take into account the part of the cardiovascular system beyond the limits of the medical image data.

"Where SimVascular really shines is in handling complex modeling," says **Bill Katz, MD, PhD**, senior scientist for Simbios. At Stanford, over the last few years, the software has been used to help understand and quantify the relationship between blood flow and cardiovascular diseases such as atherosclerosis and aneurysms. And Taylor is working with clinicians to assess its ability to plan and predict the outcome of interventions for adults with cardiovascular disease as well as for children with congenital heart defects.



Here, SimVascular simulates mean wall shear stress in the pulmonary arteries of patients with (right) and without (left) pulmonary hypertension. The disease has altered the arterial geometry in the lungs.

WANT TO BE A BETA-TESTER?

Taylor is currently planning a SimVascular user-training course for July or August of 2007. Contact Bill Katz, william.katz@stanford.edu, for more information.

Because SimVascular includes commercial components, its release to the scientific community as an open source project has posed some challenges, says Katz. "It required a good degree of encapsulation so that we can eventually allow open source alternatives to the commercial components." At the same time, he says, the various commercial entities they've dealt with have been very cooperative—UGS gave starter grants to alpha users for their solid modeling software, and companies have pre-negotiated the terms of their relationships with future users.

For Taylor, the public release of the software feels like letting go of his baby. He conceived of the technology just over 11 years ago and has been nurturing it ever since. But, Taylor says, it's time to let others use it as well. "There are so many applications for this technology to different manifestations of congenital and acquired cardiovascular disease. We won't be able to do all the work here at Stanford."

WHAT IS SimVascular?

SimVascular is a software application for patient-specific cardiovascular modeling and simulation. It integrates best-in-class commercial components and custom open-source code, including an integrated flow solver with outflow boundary conditions and fluid-structure interaction for cardiovascular problems. SimVascular includes:

- Image processing and visualization using VTK and ITK from KitWare, Inc.;
- Patient-specific geometric modeling using the Parasolid® solid modeling kernel from UGS;
- Automatic mesh generation using MeshSim from Simmetrix, Inc.;
- Parallel finite element flow solver, developed jointly by RPI and Stanford, which incorporates an iterative solver library (LesLib) from AcuSim, Inc..

Simbios is a National Center for Biomedical Computing located at Stanford University.

DIVERSE DISCIPLINES, ONE COMMUNITY
BiomedicalComputation REVIEW
Published by **Simbios**, a National Center for Biomedical Computing

Nonprofit Org.
U.S. Postage Paid
Permit No. 28
Palo Alto, CA

Biomedical Computation Review
Simbios A National Center for Biomedical Computing
Stanford University
318 Campus Drive
Clark Center Room S231
Stanford, CA 94305-5444

seeing science

SeeingScience

BY KATHARINE MILLER

Simulated Lipid Rafts

According to traditional theory, lipid membranes consist of a "fluid-mosaic" in which molecular components, including membrane proteins, are randomly distributed and move freely against a fluid background.

In recent years, however, this idea has been challenged by experimental results suggesting membranes may contain nano-sized rigid patches known as lipid rafts. Some have suggested that these rafts are involved in membrane trafficking, signal transduction, and regulation of membrane proteins.

But it's nearly impossible to observe rafts in action. So **Perttu Niemala**, a graduate student at Helsinki Institute of Technology, and his colleagues decided to simulate lipid rafts on a computer. They compared the structures of three different mixtures of a background lipid, cholesterol and sphingomyelin—with ratios of 1:1:1 (top); 2:1:1 (middle); 62:1:1 (bottom). Cholesterol and sphingomyelin are believed to induce raft formation.

The researchers found that at higher cholesterol and sphingomyelin concentrations, the membrane becomes thicker (as shown here) as well as more rigid. Moreover, the measured lateral pressure profiles within the studied membranes were considerably different, which is suggested to be an important factor for regulating the action of membrane proteins.

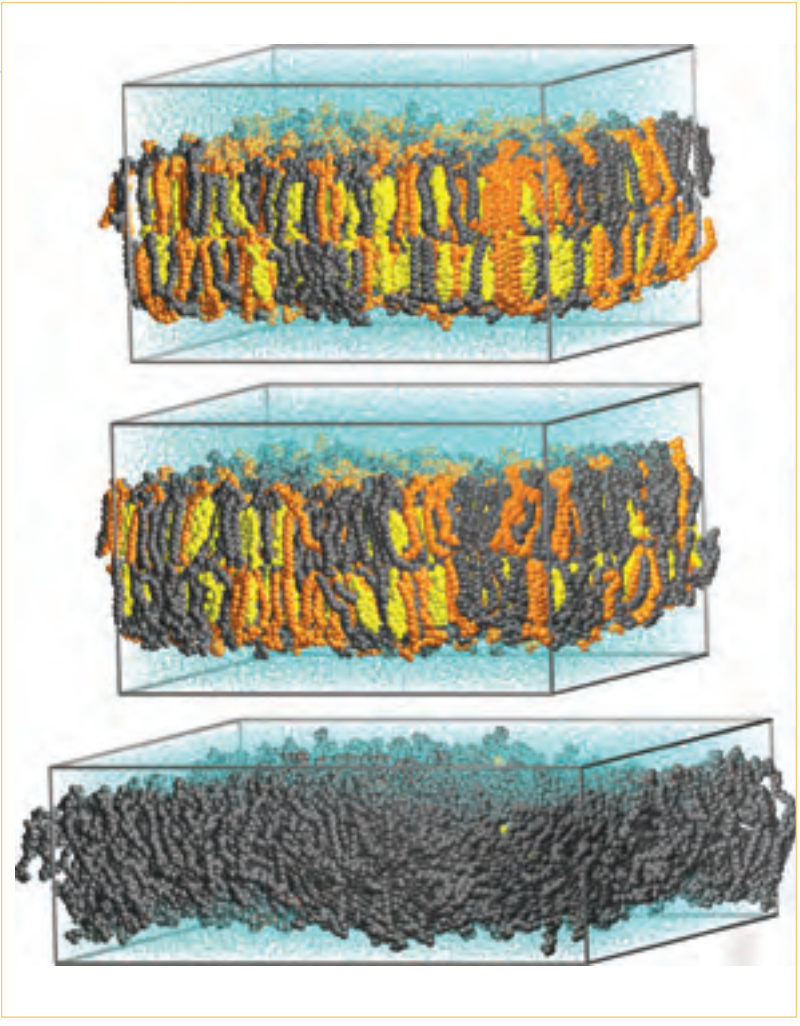


Image provided courtesy of Perttu Niemala. It was published as part of a paper titled "Assessing the Nature of Lipid Raft Membranes" in PLOS Computational Biology in February 2007.