NewsBytes

New Technology Reveals the Genome's 3D Shape

Try taking a human hair as long as Manhattan and cramming it—unsnarled—inside a marble. This is the challenge faced by a 2-meter-long strand of DNA as it folds into its compact array of 23 chromosomes within a cell's nucleus. Previously, scientists only theorized about how DNA squeezes inside a nucleus without becoming a hopelessly tangled mass. Now a new technique called Hi-C reveals that DNA packs knot-free into its chromosomal patterns by assuming a rare geometric shape observed in snowflakes, crystals and broccoli.

"We've developed a powerful new technique to look at chromosomes at an unprecedented resolution," says **Job Dekker, PhD**, cell biologist at the University of Massachusetts and coauthor of the study in the October 9, 2009 issue of *Science*. "What we found constitutes a breakthrough in our understanding of chromosome folding."

At the small scale, DNA wraps around proteins called histones and assumes its classical double-helix shape. At the large scale, chromosomes cluster in discrete sections within the nucleus called "territories." "Between the scale of chromosome territories and the scale of histones, effectively nothing has been known about the structure of the genome," says first author **Erez Lieberman-Aiden**, a graduate student in the lab of **Eric Lander, PhD**, professor of biology at the Broad Institute in Cambridge, Massachusetts.

Hi-C reconstructs an unbiased 3-D map of the entire genome. First, scientists soak a complete set of chromosomes in formaldehyde, which acts like glue to stick together parts of the genome that are close in 3-D space. Then they chop the DNA into a million pieces and perform massive parallel sequencing on the interacting fragments. Mapping software compares the sequences of attached fragments with a human genome reference sequence; based on the results, the scientists compute which parts of the folded DNA physically interact with each other.

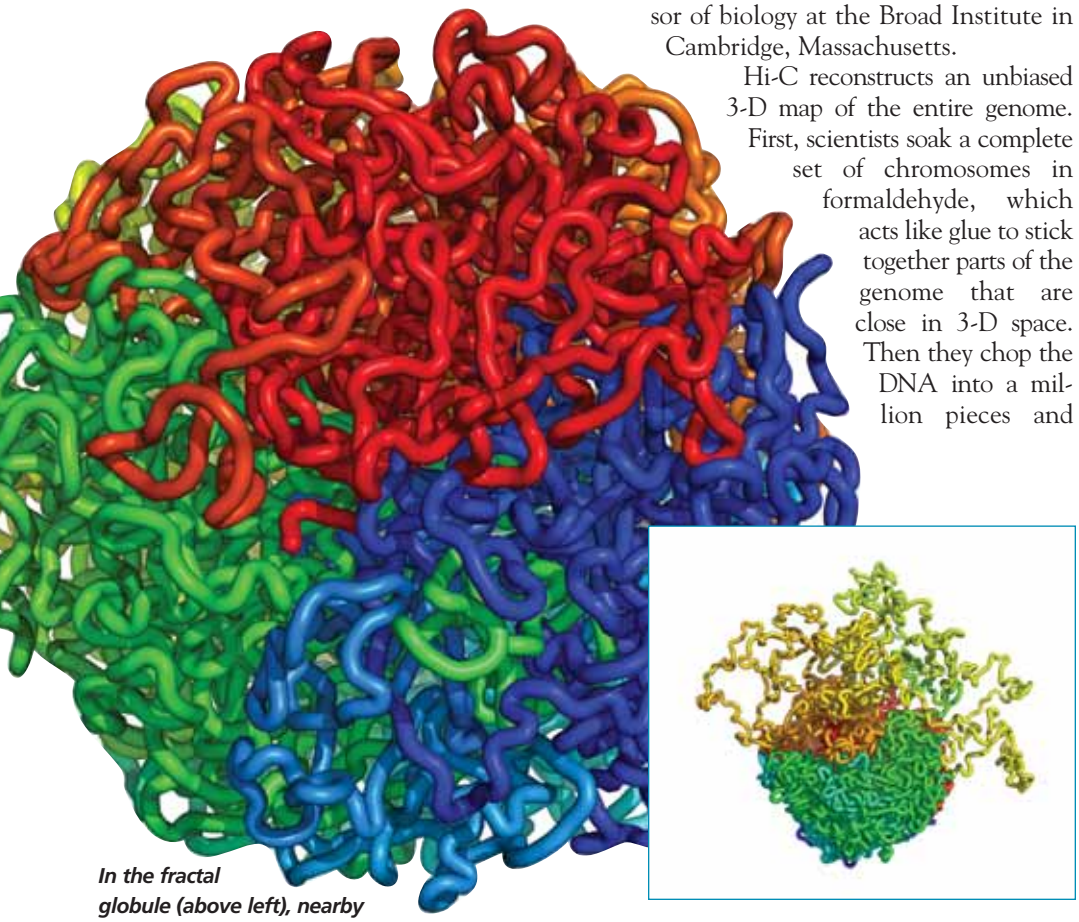
The team found that active, gene-rich and inactive, gene-poor sections cluster in separate parts of the nucleus. The active chromatin segments are like easily accessible papers spread out across a desk, whereas the inactive portions are densely packed, like folders in a file cabinet.

Simulations revealed that DNA assembles into dense fractal globules—structures that look alike at different levels of magnification, such as the intricate geometrical form of a crystal. Genes are easily accessible, but when they're not in use, the structure spontaneously collapses into a tight, knot-free bundle.

"This is the first spatial map of the genome," says **Tom Misteli, PhD**, cell biologist at the National Cancer Institute in Bethesda, Maryland. "It's a technical breakthrough that opens the doors to doing all sorts of interesting experiments."

Future experiments will investigate how the 3-D shape of DNA morphs depending on the activity of genes and disease states, like cancer. As genome sequencing becomes cheaper, Dekker says, it should be possible to obtain higher spatial resolution and even to reconstruct the shapes of individual genes.

—By Janelle Weaver, PhD



In the fractal globule (above left), nearby regions on a chain of DNA—indicated using similar colors—are packed into nearby regions in 3D space. The accessible DNA chain unravels easily (above right) because the globule lacks knots. Images courtesy of Leonid A. Mirny and Maxim Imakaev, reprinted from Lieberman-Aiden, E., et al., Comprehensive Mapping of Long-Range Interactions Reveal Folding Principles of the Human Genome, Science, *326(5950): 289-293 (2009), with permission from AAAS.*

How DNA Goes A'Courtin'

Until now, scientists have known little about how complementary single strands of DNA court one another before binding to form the classical double helix. But now, molecular dynamics simulations have identified that the binding—or hybridization—mechanism depends largely on the sequence of the DNA: Ordered sequences will meet and then slither lengthwise to find the correct match; but sequences that are random will connect at key sites then rapid-

ly assemble along the molecule's length.

"One would have thought that random sequences would have more difficultly hybridizing, and that is not necessarily the case," says **Juan J. de Pablo, PhD**, professor of chemical and biological engineering at University of Wisconsin, Madison. The work was published in the October 5 issue of the *Proceedings of the National Academy of Sciences*.

Scientists have previously tried to simulate the pathways by which DNA strands combine, but the models they used included too much detail to enable sufficiently long computations, de Pablo says. So De Pablo's group developed a highly simplified model, tested on experimental data, to capture essential details of the interactions between the base pairs of complementary strands of DNA. The researchers then simulated the process by which the single strands interact using molecular dynamics and Monte Carlo simulations, taking multiple "snapshots" of the double helix as it assembled. To the team's surprise, the path to a successful union depended crucially on the sequences of the molecules.

When the sequences of both single strands are ordered or repetitive, any two sites of base pairs can come together and the two strands slowly "slither" lengthwise until complementary base pairs match along the entire chain, says de Pablo. When the sequences are random, however, single sites located toward the center of the strands unite early. "The moment they come together, then the molecule just assembles perfectly and it does so very quickly," de Pablo says.

The results could influence the design of technologies that depend on the hybridization process, such as gene chips, de Pablo says. To engineer more efficient and reliable hybridization, researchers could use random sequences, which bind more efficiently and with fewer errors.

"This is an interesting step forward," says **Nadrian Seeman, PhD**, professor of chemistry at New York University. "No one had taken the time to track the pathway previously." Seeman has used the principle of random sequencing in his own hybridization studies, and he finds it reassuring to see it vindicated by the simulation data. "It does tell people who are designing sequences to avoid repetition in the sequences," he says.

—By Jane Palmer, PhD

Modeling Bacterial Comets

Rocketing within and between human gut cells, *Listeria monocytogenes*—a motile, foodborne bacterium—leaves a comet-like tail of actin protein behind it and makes us sick. Scientists have long wondered how actin allows the bacterium to puncture through multiple cells and evade the human immune system. A new computational model shows how rapidly accumulating actin at the back of the bacterium produces that force.

"Our simulation helps us understand the basic physical properties and mechanisms by which actin can produce force," says biophysicist **Mark Dayel, PhD**, a postdoctoral researcher at the University of California, Berkeley, and lead author of the paper published in the September 2009 issue of *PLoS Biology*. "We now have an explanation of why you get a switch from the initial pulse to smooth motion."

L. monocytogenes comes from contaminated produce or milk and infects epithelial cells in the gut. Using a membrane protein called ActA, the bacterium moves by continuously building a network of actin filaments from pieces of the host's cytoskeleton. To observe this system in action, scientists have reproduced the bacterial movement *in vitro* by coating tiny beads with ActA and putting them in a cell solution. Initially, actin fibers build from the surface of the bead, pushing old actin outward and forming a shell. But when the shell gets too big, it cracks and the bead bursts out, propelled forward by

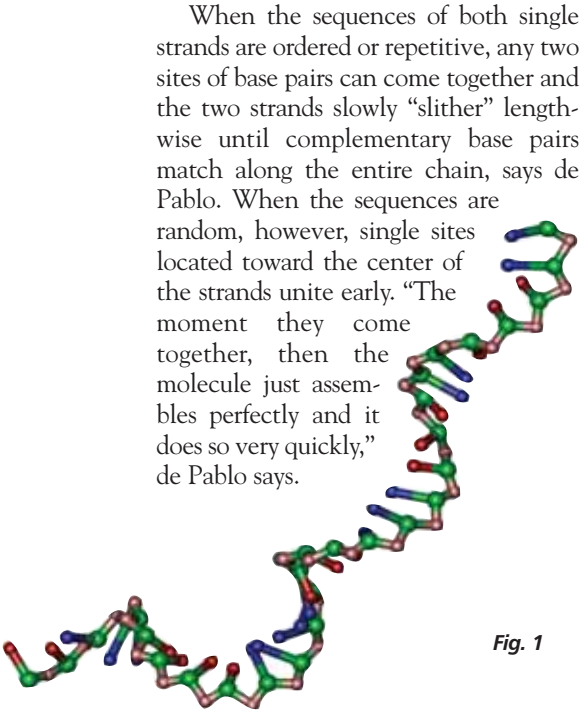


Fig. 1

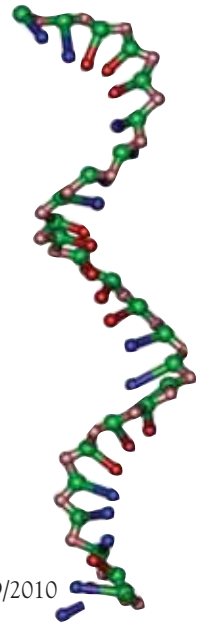


Fig. 2

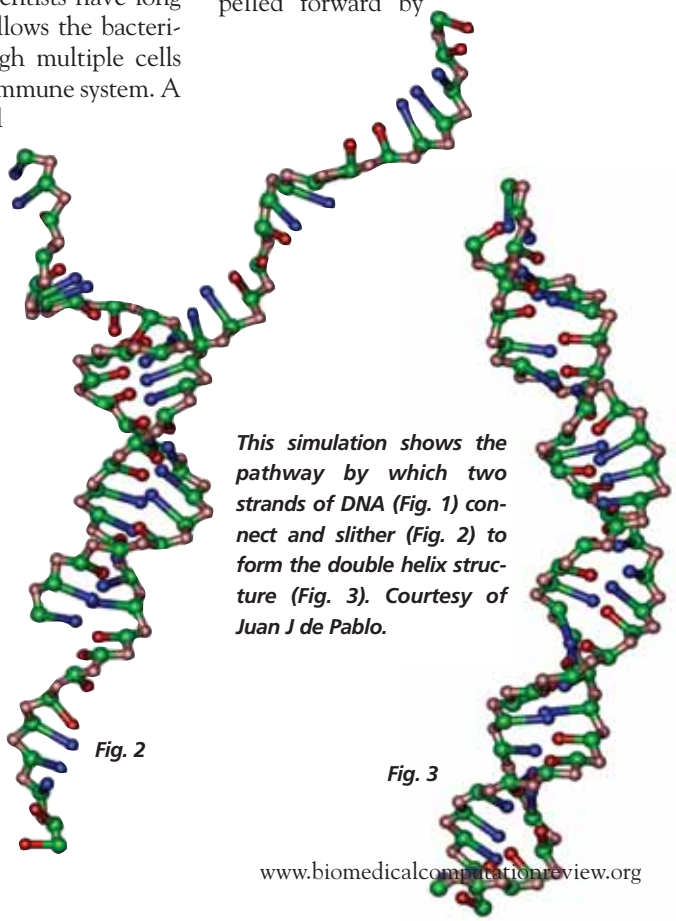
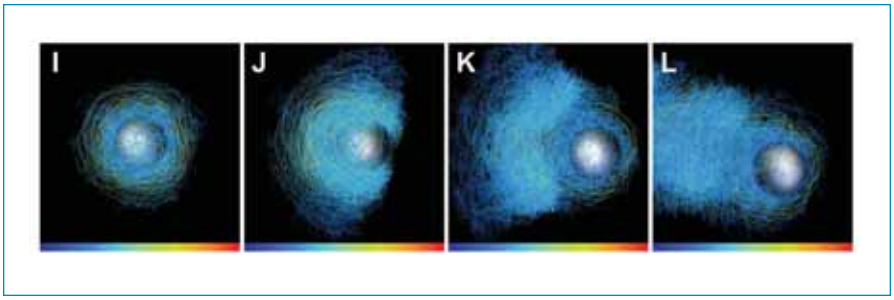


Fig. 3

This simulation shows the pathway by which two strands of DNA (Fig. 1) connect and slither (Fig. 2) to form the double helix structure (Fig. 3). Courtesy of Juan J de Pablo.



In this 3-D computer simulation time series, a bead representing the Listeria monocytogenes bacterium builds actin fibers at its surface before breaking out of its shell and moving forward, pushed by actin fibers accumulating at the back of the bead. Reprinted from Dayel et al., In Silico Reconstitution of Actin-Based Symmetry Breaking and Motility,* PLoS Biology, *7(9): e1000201 (2009), doi:10.1371/journal.pbio.1000201.

continual actin production. Until now, scientists thought that cracks in the outer shell spread inward and caused the shell to break. They also thought that the actin fibers stretched and then contracted behind the cell, squeezing it like a bar of soap.

To better understand these dynamics in detail, Dayel and his colleagues modeled the process, called "symmetry breaking." The simulation showed that the actin shell cracks from the inside, just above the surface of the bead, where tension of the actin is greatest. When the bead bursts out, surface actin accumulates against the shell left behind and pushes the bead forward, rather than squeezing as previously believed. The model then successfully predicted what would happen to the beads in novel situations, which Dayel verified *in vitro* by placing new bead shapes in different conditions. Dayel says the next step is to calibrate the model so scientists can measure forces that can't be measured *in vitro*. "We can extend its qualitative behavior to quantitative behavior, essentially allowing us to do virtual experiments," Dayel says.

"The combination of model and experiment has made a very compelling case that the mechanisms they're proposing are the real ones," says **Roger Kamm, PhD**, professor of mechanical engineering at the Massachusetts Institute of Technology. The model is "extremely simple, yet capable of capturing some fairly complex behavior," Kamm says.

—By Gwyneth Dickey

Cooking Cancer With Gold Nanoshells

Tiny gold particles that absorb laser light and convert it into heat are a promising therapy for destroying tumors. However, controlling the temperature of such gold nanoshells is crucial: The shells must get hot enough to kill tumor cells, but they must not scorch nearby healthy tissue. Now, researchers have developed a model that predicts how much these nanoshells raise the temperature of surrounding tissue.

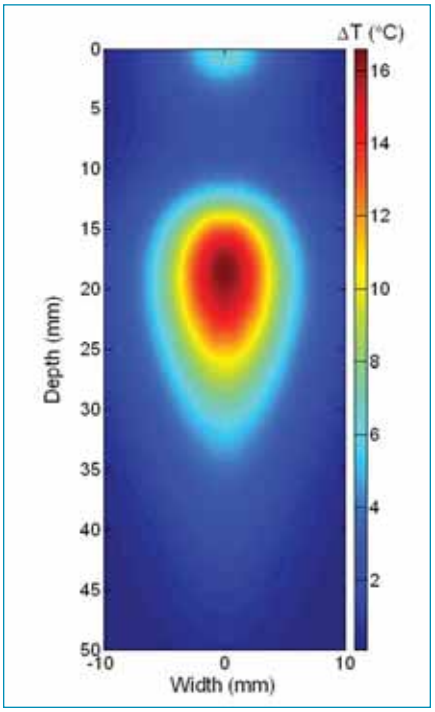
"When we tried to estimate how much heat is being generated from the process, we didn't have any good way to quantify it," says **Sang Hyun Cho, PhD**, a medical physicist at the Georgia Institute of Technology and senior author of the study in the October 2009 issue of *Medical Physics*. With the new model, researchers won't need to directly measure temperature with invasive temperature probes or magnetic resonance thermometry imaging, Cho says.

Other teams have modeled the temperatures of nanoshells in tissue. However, their models assumed that the nanoparticles spread out evenly, Cho says. "But we know that gold nanoshells are not uniformly distributed in tissue," he observes. Instead, the particles cluster tightly in some tumor regions and avoid others. That's because nanoshells travel to the growths within a tangle of misshapen blood vessels, but the vessels don't reach all parts of the tumor.

Using basic heat transfer principles, Cho's group created a computational model to calculate the heat generated by individual shells. At first, Cho assumed the nanoshells spread out evenly. But, unlike previous efforts, Cho's model is well-suited to capture the pattern of hot spots arising from a more realistic nanoshell distribution, he says.

The simulations captured the general heating profiles from past experiments but, Cho says, couldn't match the exact temperatures—probably because the team lacked good measurements of how much light is absorbed and scattered at the wavelength they used, thus affecting their calculations of the conversion to heat. His group plans experiments to pin down these values.

"They are doing very theoretically well-founded simulations," says **David Paik, PhD**, professor of radiology at Stanford University. The next impor-



Cross-sectional view of the temperature distribution in a tissue-like medium filled with gold nanoshells after three minutes of near-infrared laser treatment. Only the bottom layer of the medium (starting at a 12 mm depth) contains gold nanoshells. Reprinted with permission from* Medical Physics *36(10), 4665, 2009. doi:10.1118/1.3215536 (2009).

tant step is modeling heating in a more realistic nanoshell distribution, he says. "This is where their more computational approach would be a big win."
—By Tia Ghose

3D Angiogenesis Modeled

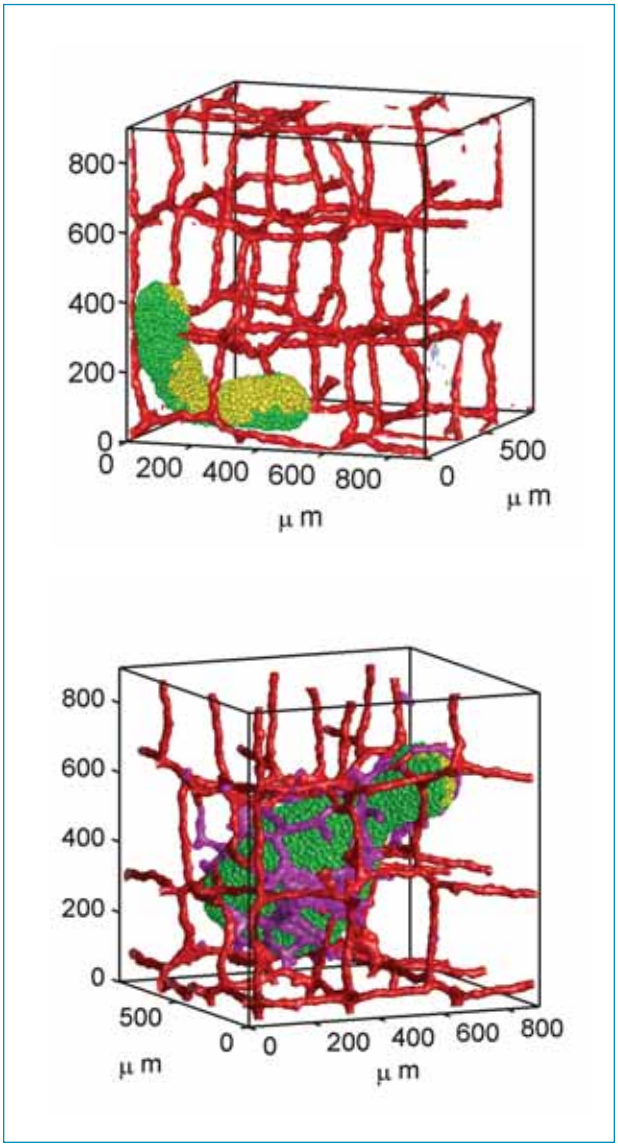
Researchers have successfully simulated how growing blood vessels affect the sizes and shapes of tumors using a 3-D model based solely on how cells behave—without reference to intracellular biochemistry. The simplified modeling system uses open-source cellular behavior "plug-ins" yet compares favorably with models laboriously coded from scratch. It also captures many essential details observed in real tumors.

"Building a computational model based on 10 to 15 behaviors is much easier than building one based on thousands of genes," says **Abbas Shirinifard**, graduate student at Indiana University's Biocomplexity Institute and lead author of the work published in the October 2009 issue of *PLoS One*.

The human body sprouts new blood vessels when they are needed. Cancer cells use this process—called angiogenesis—to their advantage. As a tumor grows, some of its interior cells become starved for oxygen and start emitting distress signals. In response, cells that create new blood vessels grow toward the distressed cells to provide them with oxygen and other nutrients. The result: a larger, actively growing tumor. Until now, researchers have modeled multi-cellular processes—such as angiogenesis—by painstakingly programming interactions among gene and protein cascades. Such models are not easily comparable between research groups, and take much longer to re-program with different conditions.

Shirinifad's team used an open-source platform called CompuCell3D (available at www.compucell3d.org) developed by **James Glazier, PhD**, and **Maciej Swat, PhD**. CompuCell3D models multi-cellular behaviors based on how each cell reacts to environmental conditions. The cells involved in tumor growth respond in defined ways, so Shirinifad and his colleagues modeled them using action-response rules, such as "If oxygen levels fall below X, send out Y signal," or "If protein X reaches concentration Y, divide." When they switched off the rule for cells to create new blood vessels in response to distress signals, the simulated tumors were small and irregular, with contours that followed the existing blood vessels. When they ran the simulation with angiogenesis "turned on," the resulting tumor grew large and rounded. These outcomes matched the appearance of such tumors in models programmed from scratch, as well as observations of real tumors treated with anti-angiogenesis compounds. In CompuCell3D, researchers can change and re-run such models in days—much quicker than if they were adapting a hand-coded algorithm, Shirinifad says.

"The exciting thing is the new technique," says **Mark Chaplain, PhD**, mathematics professor at the University of Dundee, Scotland. He notes that the current model lacks a proper simulation of blood flow; the simulated blood-vessel cells deliver oxygen itself rather than shuttling oxygen-rich blood. "If they develop this technique further by modeling blood flow, they will have a very powerful model," Chaplain says.
—By Jennifer Welsh



After 75 days of simulated growth, a tumor model looks quite different in the presence (A) and absence (B) of new blood vessel formation. Green cells in the tumor are actively dividing, while yellow cells are starved of oxygen. Red cells are blood vessel cells originally present in the model; purple cells are new blood vessel cells, present only in the model that supports angiogenesis. Axes are labeled in microns. Reprinted from Shirinifard, A, et al., 3D Multi-Cell Simulation of Tumor Growth and Angiogenesis, PLoS One, 4(10): e7190. doi:10.1371/journal.pone.0007190 (October 2009). Images provided by Abbas Shirinifard.

Improving the Sense of Touch for Surgical Robots

When a knife cuts into an organ, forces push back in ways that mechanical engineers can, to some extent, predict. But other factors are also at play: Ions shift in solution within cells, causing electromechanical changes that, researchers now say, can be predicted as

well. In a new model of soft tissue deformation, researchers for the first time include electromechanical changes as well as mechanical ones. The work could lead to better 3D surgical simulations and could ultimately provide surgeons at computer terminals with simulated feedback through surgical robot's controls.

"We want to bridge the gap between surgical simulation and surgical practice," says **Yongmin Zhong**, **PhD**, research fellow in mechanical and mechatronic engineering at the Curtin University of Technology in Perth, Australia. Zhong's novel way of modeling soft tissue deformation was outlined in the November 2009 issue of *Artificial Intelligence in Medicine*.

Robots lend a helping metal hand in surgery worldwide, cutting more precisely than trembling human fingers. But the surgeons behind the joysticks cannot feel how hard to push: slicing through fatty tissue feels the same as cutting through air. When cutting by hand, "you know how hard you're pushing, you know what damage you're doing, " says **Julian Smith**, **MD**, a heart surgeon at the Monash Medical Center in Melbourne, Australia, a co-author on the paper. "With robotic instruments, you get none of that."

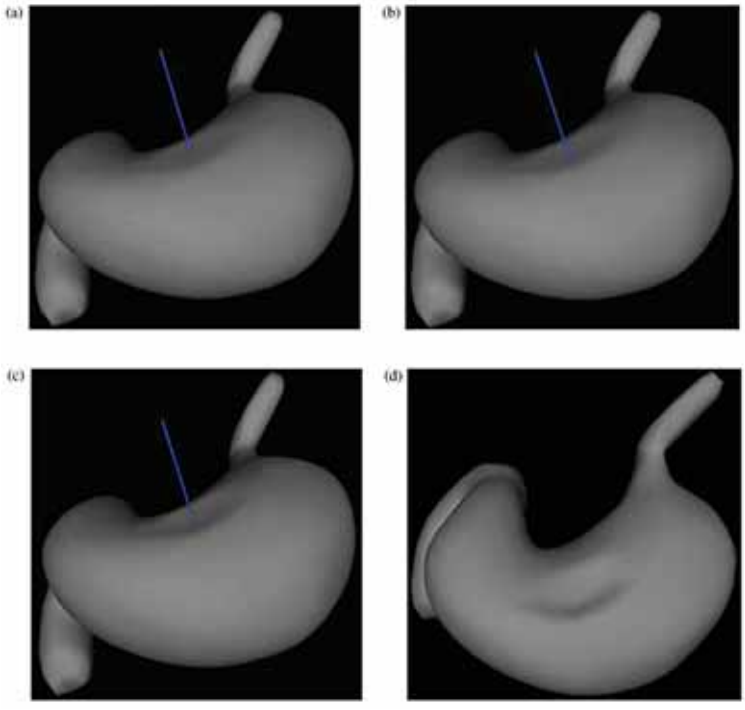
In previous attempts to provide a sense of touch in surgical simulations, researchers focused only on the mechanical force applied. While the mechanical force is important, Zhong explains, so are the electrical forces that come into play deeper within the tissue. For instance, charged particles like potassium swim in the plasma-like interstitial fluid between tissue cells, morphing the overall shape of the tissue.

By including the diffusion of charged particles in a set of sophisticated mathematical expressions, Zhong showed how prodding tissue shoves like-charged ions together, creating electrostatic repulsion. The model shows that this repulsion makes it harder to cut the soft tissue because it pushes back on the knife. Zhong tackled the equations with an artificial cell neural network, a much zippier problem-solver than numerical algorithms because the "cells" number-crunch as a team, instead of iteratively. It's the computational equivalent of six people jointly solving a jigsaw puzzle instead of taking turns. Such quick computational solutions are critical in a surgery, Zhong notes, because doctors cannot work with a time lag.

"They did a very good job and it's closer to what we can get in the real world, but it doesn't mean the problem is solved," says **Xiaobu Yuan**, **PhD**, associate professor in computer science at the University of Windsor. For example, poking the stomach causes it to shrink because it's connected to the nervous system, but the new model doesn't take that into account.

Smith plans to test if the model matches reality by putting animals under the knife. "The model is yet to be applied," says Smith, but it has "outstanding potential."

—By Marissa Cevallos



Researchers at the Curtin University of Technology used a cellular artificial neural network to simulate how soft tissue deforms under pressure, say, from a surgical knife—as shown here in blue deforming a virtual human stomach. Reprinted from Zhong, Y, et al., An electro-mechanical based deformable model for soft tissue simulation,* Artificial Intelligence in Medicine*, 47(3):275-288 (2009), with permission from Elsevier and Dr. Yongmin Zhong.

Conducting Medical Research from Electronic Health Records

To discover links between genes and disease, researchers typically recruit individual patients with and without the disease of interest; have them sign consent forms; take their medical histories; and analyze their blood samples. As well as being time-consuming and expensive, it can be hard to get a large enough sample of patients. But now researchers have shown there might be another way—using electronic medical records to identify patients with the

desired phenotypes and then obtaining their anonymized leftover blood samples to test for genetic information.

"We showed that we can actually conduct full-blown association studies to find the right patients with the right phenotypes and connect them to the right samples," says **Isaac Kohane, MD, PhD**, professor at Harvard Medical School and director of i2b2 (Informatics for Integrating Biology and the Bedside), the National Center for Biomedical Computing that conducted the study published in the September 2009 issue of *Genome Research*. "It's soup to nuts work."

With the help of natural language processing (NLP), the i2b2 researchers set out to use a large, available, cheap data pool: the electronic medical record archives for 2.6 million patients at Partners Healthcare System in Massachusetts. Although doctor's notes are notoriously unstandardized, NLP tools can break them into their smallest components, analyzing parts of speech and how words are joined. The i2b2 team sought to identify pools of patients with rheumatoid arthritis, asthma, secondary illnesses and risk factors for asthma (for example, smoking history). Along the way, clinical experts gauged the accuracy of the process and helped refine search terms. "It takes three to four months of iteration with expert clinicians until we get it just right," Kohane says. In addition, the researchers developed a system to access anonymously saved leftover blood samples from the identified populations to use for future studies requiring genetic data.

And the NLP tools did a pretty good job: Of about 98,000 patients identified as having asthma, 82 percent of the time the experts reviewing the files concurred in that diagnosis; 90 percent of the patients identified with a history of smoking had such a history; and of the 4,618 NLP-identified rheumatoid arthritis sufferers, 92 percent had definite arthritis (according to expert review) while 98 percent probably did. By studying these electronic patients, the researchers successfully reproduced several results from past clinical research. And while the clinical studies had paid an average of $650 to characterize and obtain blood samples from each patient, i2b2 spent $20 to $100.

Medical Record Snippet	*Smoking History*
SOCIAL HISTORY: The patient is married with four grown daughters, uses tobacco, has wine with dinner.	Positive
SOCIAL HISTORY: The patient is a nonsmoker. No alcohol.	Negative
SOCIAL HISTORY: Negative for tobacco, alcohol, and IV drug abuse.	Negative
BRIEF RESUME OF HOSPITAL COURSE: 63 yo woman with COPD, 50 pack-yr tobacco (quit 3 wks ago), spinal stenosis, ...	Positive
SOCIAL HISTORY: The patient lives in rehab, married. Unclear smoking history from the admission note...	Insufficient data
HOSPITAL COURSE: ... It was recommended that she receive ...We also added Lactinax, oral form of Lactobacillus acidophilus to attempt a repopulation of her gut.	Insufficient data
SH: widow,lives alone,2 children,no tob/alcohol.	Insufficient data

After lengthy training, i2b2's natural language processing software scans clinical histories, tagging words and phrases that describe smoking history and making a diagnosis (right-hand column). With training, the NLP tools were able to equate "smoking history" with "smokes often," distinguishing both from "non-smoker." Clinical experts also reviewed random results and computer scientists refined the search terms to clarify ambiguities like "tob." Reprinted from S. Murphy, et. al, Instrumenting the health care enterprise for discovery research in the genomic era,* Genome Research *, 19(9): 1675–1681 (2009).

"This paper represents very encouraging results using free open-source software," says **Chunhua Weng, PhD**, assistant professor of biomedical informatics at Columbia University. She says the next step is to include information such as how long an individual smoked or when symptoms began in patient descriptions. Kohane agrees, noting that researchers are working to include time-varying data in i2b2's model.

—By Daniel Strain

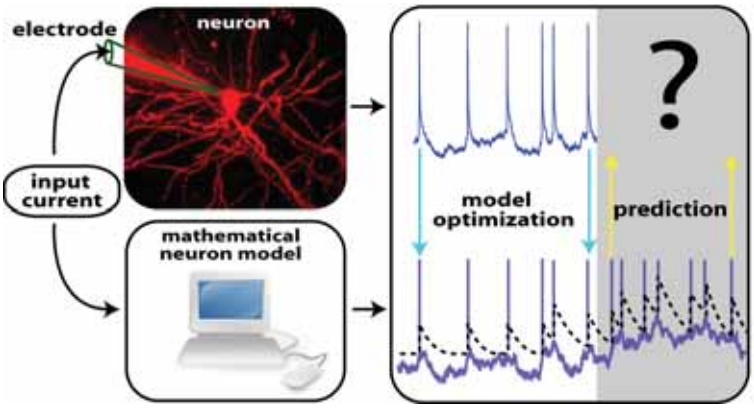
Neuron Models: Simpler Is Better

During the summer of 2009, the International Neuroinformatics Coordinating Facility in Stockholm dangled a nearly $10,000 cash prize in front of neuron modelers and challenged them to do better. And they did. The winners of the competition, which was described in the October 16, 2009 issue of *Science*, produced a neuron model that became more accurate as they stripped away pieces of a much more complex starting model.

"It was amazing for us physicists to see the description become simpler as we tried to make the performance better," says **Shigeru Shinomoto, PhD**, a physicist at Kyoto University in Japan who, along with two of his former students, snagged the grand prize.

Modeling the electrical behavior of individual neurons is crucial to understanding how thought and other cognitive functions arise in complex neuronal networks. Current neuron models can predict some neuron behavior, but with limited accuracy and at high computational cost.

The international competition has grown from eight entrants in 2007 to 33 this year and included teams around the world. "We had different people from different backgrounds using methods we would never have thought of," says **Wulfram Gerstner, PhD**, a computational neuroscientist at the Ecole Polytechnique Federale in Lausanne, Switzerland who co-authored the *Science* paper.



To set up one of the challenges for the neuron modeling competition, an artificial current was injected into a live neuron (upper left) and the resulting electrical activity was recorded for 60 seconds (blue trace, top right). Competitors used data from the first 38 seconds of the recording to fine-tune the parameters of a mathematical neuron model receiving an identical current injection (purple trace, lower right). Model performance was measured by the percentage of spikes correctly predicted in the final 22 seconds of the recording. Graphic courtesy of Richard Naud.

Contestants had to predict the precise timing of electrical spikes in individual neurons from different parts of the brain. Since different neurons can respond differently to the same signal, competitors used the first 38 seconds of data from a neuron to adjust their model parameters to better fit that neuron. They used the freshly tuned model to predict spikes in the subsequent 22 seconds of data. Shinomoto's winning model predicted 59.6 percent and 81.6 percent, respectively, of the spikes from two different neurons.

Electrical activity in a real neuron spikes when its membrane potential passes a set threshold value. Shinomoto's model neuron has an adapting threshold that increases immediately after a spike and decays exponentially to its initial value. The decay is modulated by two time constants of 10 ms and 200 ms, chosen to reflect the timing of ion currents in the neuron membrane.

The competition will evolve with the field, Gerstner says. Computational neuroscientists will soon draw on an emerging body of molecular knowledge to improve their models, says **Erik De Schutter, PhD**, a professor of computational neuroscience at the Okinawa Institute of Science and Technology. Advanced molecular techniques should reveal the physical structures and electrical properties of neurons in much greater detail than is currently known. These data may help modelers account for the effects of variations in temperature and chemical conditions and in the physical structures of the neurons.

"Neuron modeling is still a work in progress," De Schutter says. "It's much more difficult than we thought."

—By Sandra M. Chung

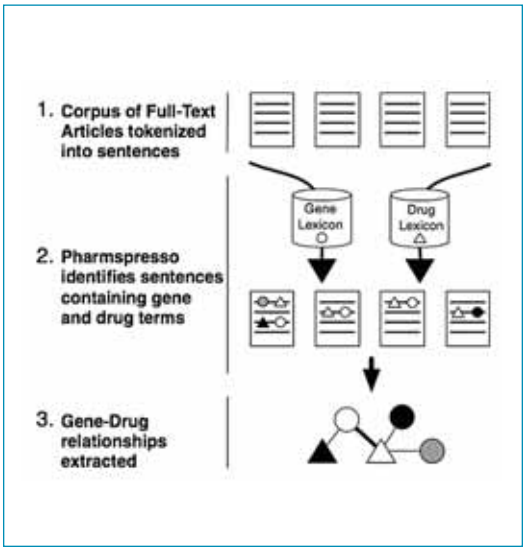
Trawling for Drug-Gene Relationships

When a drug saves one person but makes another ill, a bitter lesson in genetic differences often follows. With many such lessons already under our collective belts, researchers are using existing knowledge to predict additional drug-gene relationships as a way to forestall future calamities. A new software program can trawl published papers for gene-drug relationships, plug those relationships into known genetic networks, and predict which genes are likely to affect a patient's response to a drug.

"Our contribution is using text mining and taking decades of research and folding that in to inform the prediction," says **Yael Garten**, biomedical informatics PhD candidate in the lab of **Russ B. Altman, MD, PhD**, at Stanford University and a lead author of the work. "We showed that this is as good as and sometimes even better than manual curation," in which scientists painstakingly enter published drug-gene interactions into a database. Garten will present the team's research January 2010 at the Pacific Symposium on Biocomputing in Hawaii.

The previous version of the algorithm, designed by Altman and others, relied more heavily on manual labor. Called PGxPipeline, it employed a database of gene-drug relationships manually compiled from scientific articles by a team of scientists at Stanford Medical School. PGxPipeline wove these relationships into an orderly web, along with a database of gene-gene interactions and other data, to predict how strongly each of 12,460 genes affects response to a specific drug.

The team has now cut PGxPipeline loose from the manually created drug-gene database, automatically mining the information from published papers. This faster, cheaper method will inform the drug-gene rankings with constant updates from new literature. The manual-curation- and text-mining-based



The text-mining-based version of PGxPipeline automatically dissects journal articles into component sentences and marks where a drug or a gene is mentioned. Reading the sentence syntax and vocabulary, it tracks the interactions between drugs and genes. A network/web of interactions is established (bottom), in which the thickness of each edge corresponds to the number of articles that support the interaction. The web of relationships is later enhanced using a database of gene-gene interactions and other information. Image reprinted from Garten, Y., Tatonetti, N., & Altman, R., Improving the prediction of pharmacogenes using text-derived drug-gene relationships, Pacific Symposium on Biocomputing, Hawaii, January 2010.

versions of PGxPipeline predicted with similar accuracy a test set of 682 drug-gene interactions. And the text-mining-based version was slightly better at identifying genes that play the largest roles in response to a specific drug.

Garten hopes to use the revised PGxPipeline to parse all relevant scientific literature for drug-gene relationships. Better predictions will save researchers time in deciding which of the possible interactions to test in the lab and eventually influence how doctors prescribe drugs, she maintains.

"There is an emerging trend in bioinformatics to combine information from curated databases with information extracted from text," says **Tom Rindflesch, PhD**, principal investigator for the semantic knowledge representation project at the National Institutes of Health in Bethesda, Maryland. "This is an excellent example."

—By Olga Kuchment, PhD

Scientific Discovery Through Video Games

When it comes to folding proteins, even modern supercomputers don't always get things exactly right. Enter FoldIt, an online video game that harnesses the human brain's natural pattern-recognition abilities to tweak computer oversights. Since its release in May 2008, FoldIt has captivated a core group of several thousand dedicated players. Contestants manipulate three-dimensional protein chains into the best configuration they can find, exposing effective and previously unknown algorithms. In recent months, the puzzles have focused on medical applications. For example, a puzzle released in October called "Finding Home" asks players to bind a potential gene therapy tool—a homing endonuclease—to DNA. In another, called "Pack the Holes and Fight Cancer," gamers will help design a protein that could activate a new kind of cancer drug.

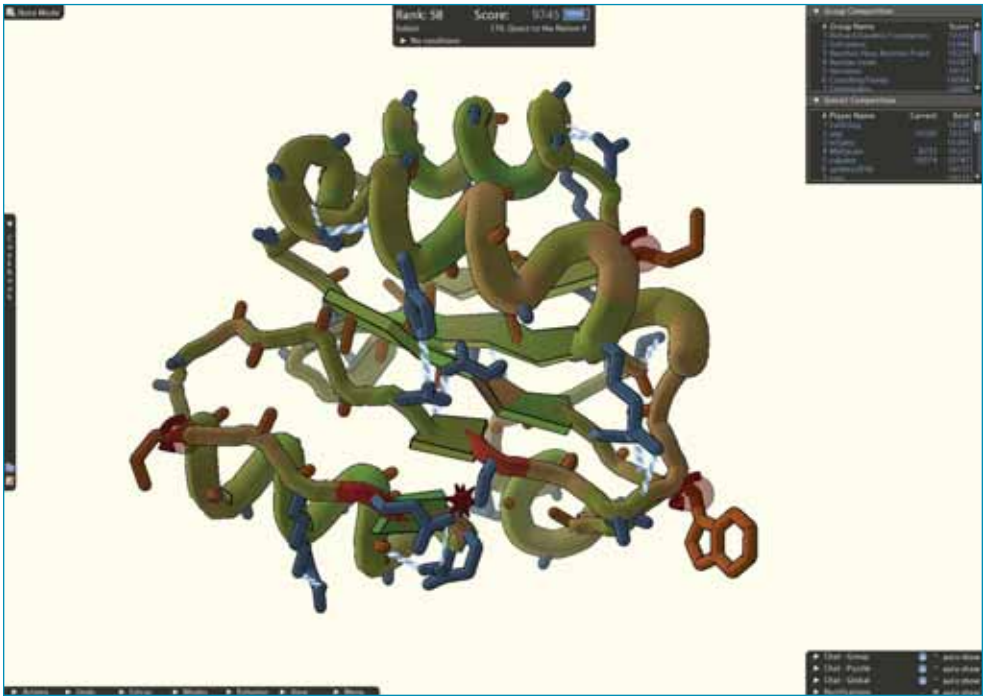
"The players, most of whom are non-experts, have sort of become protein scientists," says **Adrien Treuille, PhD**, assistant professor of computer science at Carnegie Mellon University. Treuille helped create FoldIt with a team at the University of Washington led by graduate student **Seth Cooper**, computer scientist **Zoran Popović, PhD**, and biochemist **David Baker, PhD**.

Researchers often must correct obvious errors in computer-folded proteins. FoldIt was developed to allow amateurs to spot and fix these computer inaccuracies. Players rack up points by pulling, wiggling, and tweaking a polypeptide sequence into the most chemically and physically accurate orientation. Most gameplay has concentrated on uncovering new folding algorithms, but FoldIt's current focus is producing player-designed proteins that can interact with particular biological targets, such as a small DNA strand. The game's creators recently released a puzzle asking players to generate a better design for human fibronectin, a protein used to mimic antibodies. One player modified fibronectin's peptide chain in a way that may turn out to be more stable than the original. Chemists at the University of Southern California are currently fabricating the novel structure for testing.

"FoldIt is a seminal and important project," says **David P. Anderson, PhD**, research scientist at the University of California, Berkeley Space Sciences Laboratory who created an online astronomy volunteer project called Stardust@home. But he encourages the team to focus more on hard scientific data in the future. "I hope they are able to quantify what they've actually done," he says.

Despite such concerns, Treuille thinks other researchers might imitate FoldIt's approach to computational analysis. "Everywhere you look in science there's labor that could use many people," he says. Treuille believes that similar projects could draw on the power of crowds while entertaining and educating the public.

—By Adam Mann □



A team at the University of Washington designed the online game FoldIt to improve protein-folding algorithms. Players maneuver polypeptide chains, such as this 2HSH sequence, into their lowest energy configuration to get the highest score. Image courtesy of Seth Cooper at the University of Washington.