

NewsBytes

Unraveling the Complex Functions of Proteins

At the birth of a new field, a conference can act as a midwife, making sure the infant enters the world smoothly. Such was the case for the Automated Function Prediction (AFP) conference held at the University of California, San Diego, at the end of August 2006. The field, in which researchers try to computationally determine a protein's function, appears to have arrived in good health.

"I feel we went quite far in creating an identity for this field," says **Adam Godzik, PhD**, program director in bioinformatics and systems biology for the Burnham Institute for Medical Research in La Jolla, California. The conference drew independent researchers who worked together for the first time to establish a common language, Godzik says, as well as a compilation of available methods and a way to evaluate the success of various techniques.

For decades, biochemists have tried to

solve the riddle of what individual proteins do. They've done this in the lab, painstakingly slowly. But scientists now sequence genomes far faster than they can assign functions to the corresponding proteins. When there were only a few sequenced genomes, this problem was relatively small in scale and seemed manageable. But, says **Iddo Friedberg, PhD**, the conference organizer and a postdoctoral associate in Godzik's lab, "By sheer scale the problem has changed."

Now, using computational methods, researchers can tease through to answers much more quickly. At the AFP conference, those adept at the process charted the future of this discipline.

Their approach weaves together methods biologists have used for years. They include physical analysis of a protein's structure; genomic context, in which researchers compare a protein's position in a gene to those of similar genes elsewhere with known functions; and what Friedberg

calls "guilt by association," or information gleaned in the laboratory about what the protein does when a cell undergoes a particular process. But because they're using computers and high-level algorithms, AFP researchers can now analyze more information faster and come to more robust conclusions about the workings of previously mysterious proteins. Visit <http://BioFunctionPrediction.org> for more information.

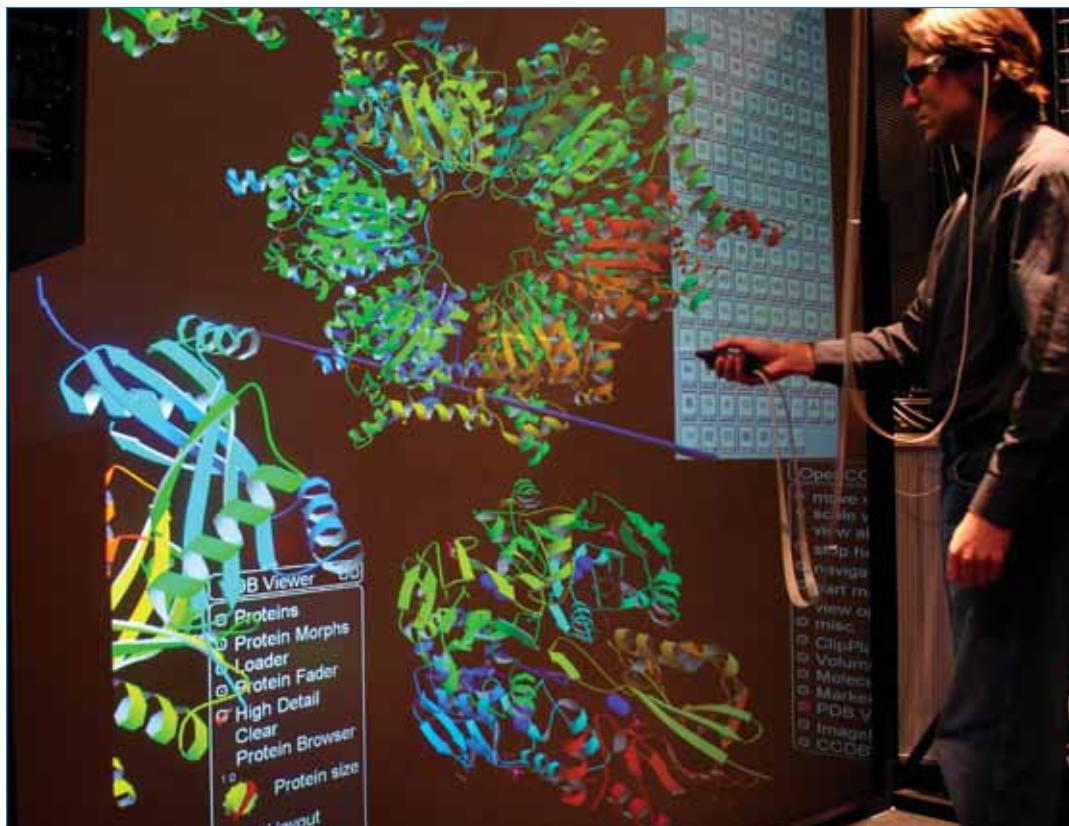
Godzik was pleased with the progress the conference made in crystallizing the field. "Until now," he says, "automated function prediction has been a black art."
—**Brittany Grayson**

Neurons Seek Their Own Solution

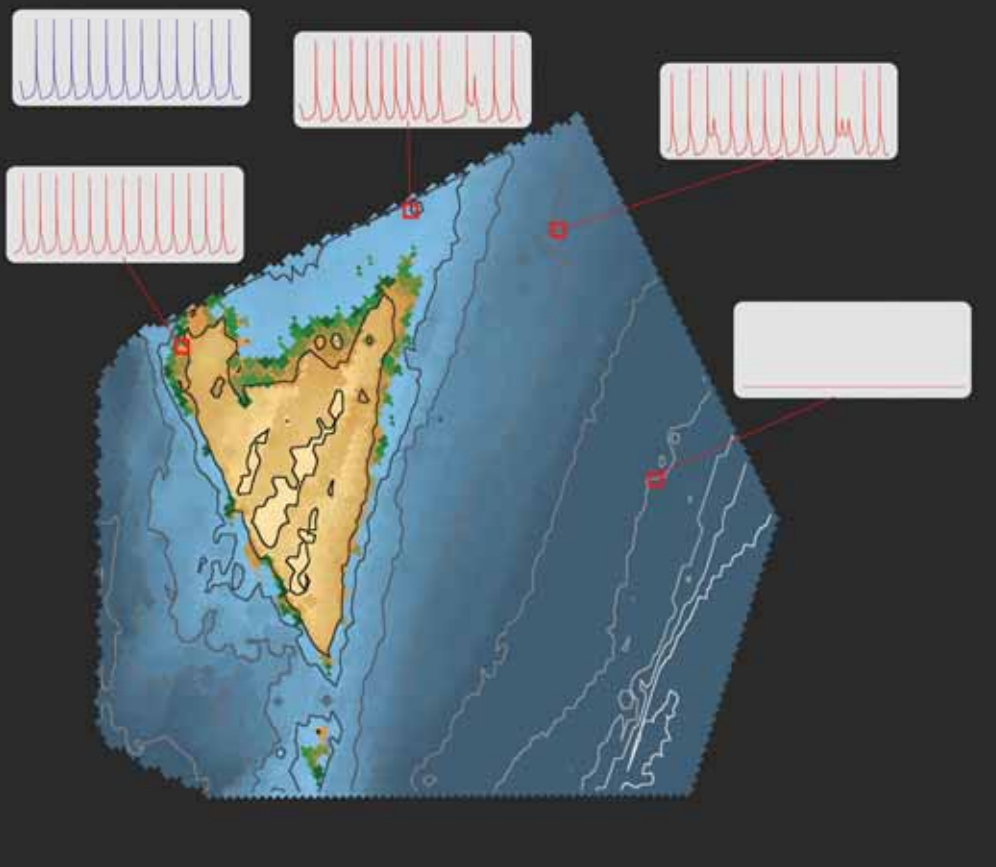
Each cell in our nervous system is an instrument in a complex symphony of electrophysiologic communication. A neuron's signaling abilities arise from its array of ion channels—tunnels within the cell's membrane that act as gatekeepers of electrical charge. But how does a cell determine the types of channels it needs and where in its membrane they should sit? The results of a new computer model suggest that even with markedly different patterns of ion channels, neurons still can come to play the same tune.

The work supports a growing paradigm shift in neurophysiology, says **Erik De Schutter, MD, PhD**, professor of neurobiology at the University of Antwerp, Belgium, lead author of the study. "We used to think of [ion channels] as LEGO blocks," he explains—with a predetermined number, type and position regulating how the neuron fires.

More recently, physiological experiments have suggested that cells with wildly different ion channel compositions could have similar firing patterns. But researchers have consistently attributed such variability to



AFP conference participant Jurgen Schulze, PhD, receiving an immersive experience of protein functional sites in the virtual reality cave at California Institute for Telecommunications and Information Technology. The software supports collaborative viewing of proteins at multiple sites on the internet. Courtesy of Jurgen P. Schulze.



One neuron firing pattern can be produced in a variety of ways. Here, typical voltage traces of different computational models (red) have to be compared with the experimental data (blue). The image shows a view of a small slice of the complex landscape for which thousands of parameter sets have been tested. Fitness has been color coded: Good models form a brown island surrounded by a blue sea of bad models. In other parts of parameter space (not shown) similar isolated islands of good models can be found.

experimental error, clinging to the notion of a “platonic ideal” of a neuron with an unchangeable firing pattern.

To look at the problem more closely, De Schutter and **Pablo Achard, PhD**, a postdoc in his lab, computationally modeled the Purkinje cell, an especially complex type of neuron best known for forming as many as 200,000 synapses. The team’s model specified 10 types of ion channels and broke the cell into four different regions that the channels could occupy. Using a mathematical approach

The work suggests, De Schutter says, that neurons are preprogrammed for a particular type of firing pattern. Each cell then decides locally how to distribute ion channels to achieve its signaling goal. How well the model predicts real biological properties is not clear and is difficult to test experimentally. Neurophysiologists can record impulses from just a few neurons at a time; the sheer quantity of recordings needed for a thorough comparison with the model doesn’t yet exist. “My estimate is you’d need about two

there was a single solution and the variance was your fault,” Marder says. “This paper is a beautiful example showing that we shouldn’t be thinking about a single solution to capture what a neuron is doing, but a family of solutions.”

—*Alla Katsnelson*

Two Ways to Merge

Cell membranes fuse with other membranes to allow material in and out. If incoming material includes invading viruses, that can be bad news for the cells. Until recently, the process of membrane fusion has been poorly understood. Now, a powerful computer model shows that neighboring membranes can merge in two distinct ways—a fact that was previously unknown. The work could help researchers clarify how viruses invade cells, and possibly lead to ways to fight viral infections.

“Ultimately we’d like to be able to control fusion in biological systems and induce or inhibit it for therapeutic purposes,” says **Peter Kasson, PhD**, a chemistry postdoctoral scholar at Stanford University and leader of the study. This paper takes steps in that direction. “Our model helps provide an explanation for how you get these two fusion processes,” he says.

Previously, scientists debated whether membranes fuse by joining directly from

“This paper is a beautiful example showing that we shouldn’t be thinking about a single solution to capture what a neuron is doing,” says Eve Marder, “but a family of solutions.”

called the phase-plane method, the model neurons were permitted to evolve their ion channel densities to produce all four firing patterns that Purkinje cells display. To the researchers’ surprise, about 20 possible combinations of ion-channel densities fit the bill. The results were published in the July 2006 issue of *PloS Computational Biology*.

people working full time for a year for that data,” De Schutter says.

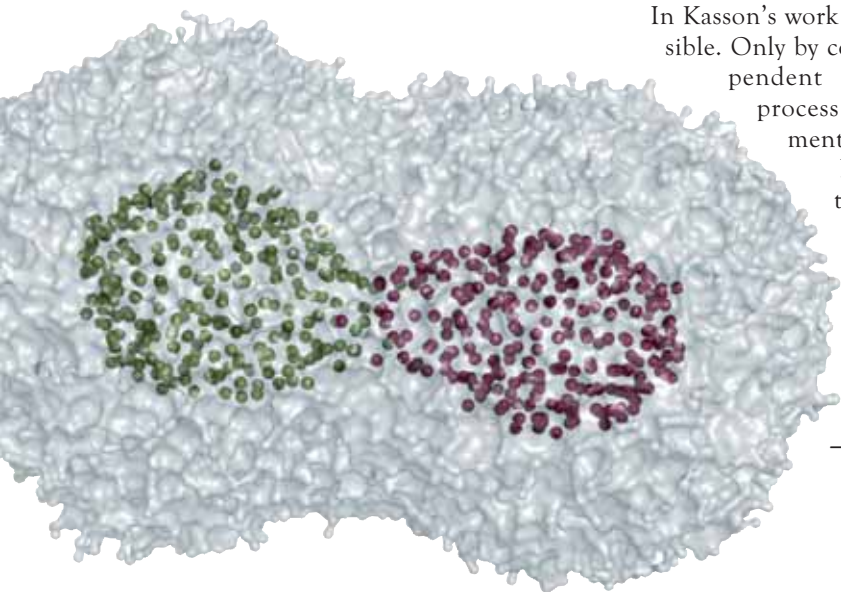
Regardless, the model provides a more nuanced version of how physiologists should conceive of their cells, says **Eve Marder, PhD**, a professor of biology at Brandeis University who studies both physiology and modeling in neurons. “The assumption has been that

their initial contact or by going through a “hemifused” halfway point, where the outside layers have merged but the insides remain separate. “We show that both could happen,” says Kasson. The work was published in the August 8, 2006, issue of the *Proceedings of the*

“Ultimately we’d like to be able to control fusion in biological systems and induce or inhibit it for therapeutic purposes,” says Peter Kasson.

National Academy of Sciences.

Observing two membranes combining in a lab is difficult because it happens so quickly—on the order of microseconds. Earlier models haven’t represented membranes in as much detail, over such long timescales, or with as many simulations as this one does, says Kasson.



As a pore forms between two vesicles, the phosphate groups from the membrane’s outer leaflet (red from one vesicle, green from the other) mingle with one another in the pore region. Courtesy of Peter Kasson.

The team ran 10,000 separate simulations of membrane fusion using a distributed computing network called Folding@Home, in which people around the world donate screensaver time to biological research. In each simulation, the fusing membranes began with different starting conditions and evolved based on laws of physics and chemistry. The result: The simulated membranes merged through either of the two routes rather than exclusively through one or the other.

Erik Lindahl, PhD, professor of bioinformatics at Stockholm University, Sweden, thinks the project sets the pace for future work in the field. “The key thing is that they’re not doing one simulation, they’re doing many,” he says. “In ten years nobody will publish a single simulation anymore.”

Siewert-Jan Marrink, PhD, head of the molecular dynamics group at the University of Groningen, the Netherlands, and creator of a previous model of membrane fusion, agrees. “I do consider this work to be a significant step forward,” he says. “In my original publication of the fusion process of the same system I was only able to look at a few events, but I could not tell how relevant these were. In Kasson’s work this has become possible. Only by comparing many independent instances of the process can global assessments be made.”

Kasson is delighted that his model explains experimental observations and can help in planning new experiments. “That’s the most exciting part,” he says, “when we can come full circle.”

—Clara Moskowitz

Cancer Proteins Show Off Their Networking Skills

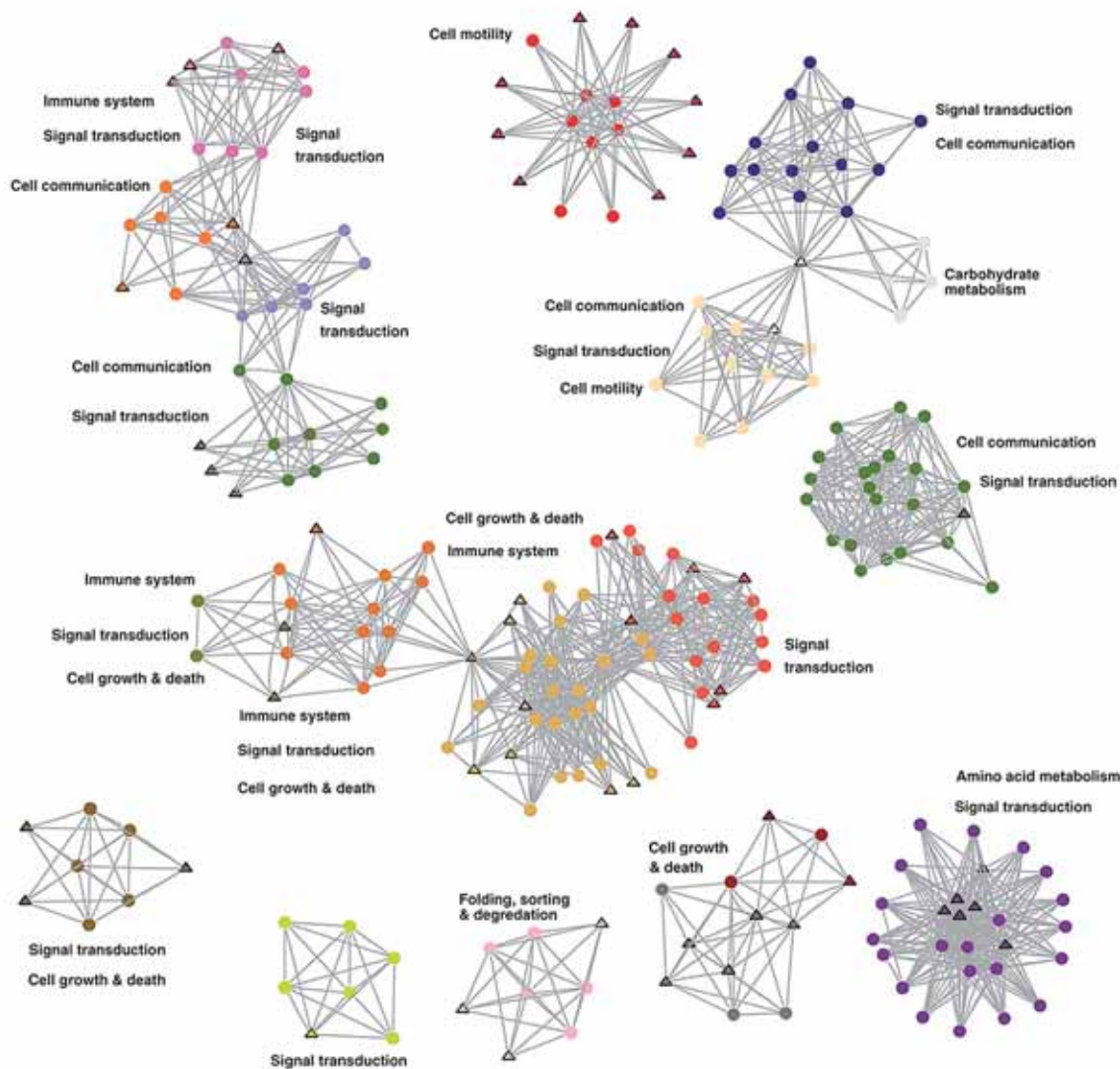
New research suggests that cancer proteins, like influential people, have the most connections. These results, from an extensive study of how human proteins interact with one another, could help explain why cancer wreaks such havoc in cells.

“We haven’t gotten to the bottom of what the increased connectivity really means, but perhaps highly connected proteins, once mutated, are more likely to cause disease,” says co-author **Paul Bates, PhD**, who heads the Biomolecular Modelling Laboratory at the Cancer Research UK London Research Institute. The research was published in the September 15, 2006, issue of *Bioinformatics*.

Bates and his graduate student **Pall Jonsson** built a model of the human proteome that contained more than 108,000 interactions using experimental information on proteins in other species, including yeast and worms. The typical protein can connect with a limited number of other proteins. The researchers scored the data and linked proteins known to interact, leading to a protein-protein interaction network, also known as the “interactome.” Then, using information from a 2004 census of 346 human genes known to mutate in cancer, Bates and Jonsson mapped 509 human cancer proteins onto their network.

The average cancer protein was linked to 23 other proteins in the network, more than twice as many as the typical protein. By analyzing protein clusters, the researchers also found the proteins from cancer genes tend to occupy intersections between protein communities that govern crucial functions such as regulating cell growth and death. This makes sense, says Bates, as proteins that are changed by cancerous mutations tend to disrupt many cellular functions. Bates adds that understanding the network properties around these proteins could help researchers identify drug targets.

Shinichiro Wachi, a doctoral candi-



A sampling of protein communities identified in the interactome. Each community is labeled by the general classes of functions in which it belongs. Cancer proteins are shown as triangles. Note that this figure shows only one community assignment per protein, although they can belong to more than one. Courtesy of Pall Jonsson and Paul Bates.

The average cancer protein is linked to 23 other proteins in the network, more than twice as many as the typical protein.

date at the University of California, Davis, says the results are hard to interpret. The mapped cancer proteins were based on genetic information rather than experimental data on how they interact, he says. “A gene may perform a function in the cell, but the mutation could either reduce the function or result in higher activity. ... It could go either way,” says Wachi, who has studied the network properties of proteins in lung cancer tissues. Wachi also cautions that because the list of cancer genes is changing dramatically, the researchers may soon need to re-examine their model.

Bates would like to do further analysis. “We’ve only got 108,000 interactions. There’s likely to be more than that—400,000, maybe 700,000,” he says. “We want to increase the map and validate it further.”

—Rachel Courtland

Watching Blood Vessels Grow and Shrink

Microscopic capillaries grow on demand, snaking toward hungry cells needing their blood supply. Understanding how to control this process could help scientists promote wound healing or halt cancer in its path. A new computer model simulates how a key molecule (VEGF, or vascular endothelial growth factor) summons vessels to sprout: It spills out of a hungry cell and travels toward a vessel, with increased concentrations in areas with few vessels. The two-dimensional model also predicts the actual number of VEGF molecules at that edge, another novel advance.

“There have been over 10,000 papers published on VEGF and not one shows a molecular-level computation,” says Aleksander Popel, PhD, professor of biomedical engineering at the Johns Hopkins

University School of Medicine. The work appeared in the September 2006 issue of *PLoS Computational Biology*.

VEGF promotes the growth of blood vessels, a process known as angiogenesis. Developing treatments to halt or promote angiogenesis is rife with complex issues: Too much VEGF can lead not only to cancer but also to abnormal

“We can model the behavior of therapeutics that you just can’t do *in vivo*,” Mac Gabhann says.

blood vessel growth. And VEGF concentration is not the only issue; to control vessel growth one needs to control the VEGF gradient across the tissue—a challenging task. Previous computational models have addressed this question *in vitro*. But this model takes that work a step further, modeling VEGF gradients in a complex *in vivo* environment.

Feilim Mac Gabhann, PhD, now a postdoctoral fellow at the University of Virginia, built the model from electron micrographs and *in vitro* data on the size and shape of vessels, muscle cells, and the organic matrix between them. He

simulated different scenarios to analyze how VEGF moved between muscle and vessel. He found that VEGF concentrations increased dramatically when the model mimicked oxygen-starved muscle. In addition, the model predicted a 12 percent change in VEGF concentration over 10 microns—a characteristic no other research had attempted to quantify. This gradient may be significant for capillary sprouting, the researchers suggest.

“This research is right on the edge,” says **Dan Beard, PhD**, professor of physiology at Medical College of Wisconsin. “They are just at the point where they’ve put everything together. They are ready for applications where there is potential for big payoffs.” But Beard warns they need to prove the model truly mimics *in vivo* angiogenesis—a task the team admits will be difficult.

Mac Gabhann and Popel next want to mimic drug activity to watch molecules interacting with VEGF. If a drug can reduce VEGF to a trickle instead of a flow so that vessels don’t grow, it might help fight cancer. If a drug increases VEGF diffusion, or guides blood vessel formation more precisely, it might help with wound healing. “We can model the behavior of therapeutics that you just can’t do *in vivo*,” Mac Gabhann says.

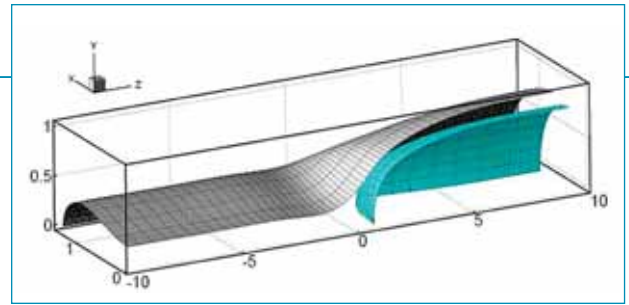
—**Megha Satyanarayana**

Opening Airways

Scientists are breathing new life into airway modeling. Using a three-dimensional mathematical model of the delicate passages in the lungs, researchers have found that strongly collapsed airways are actually easier to reopen than partially collapsed ones.

“Everybody I have mentioned it to has said, ‘No, but surely, that’s the wrong way around,’” says **Matthias**

This image shows skeletal muscle (gray circles) interspersed with capillaries color-coded to show the amount of VEGF bound to each capillary. Red capillaries have more VEGF bound and may sprout new vessels in response to the signal. Courtesy of Aleksander Popel and Feilim Mac Gabhann.



A computational model of pulmonary airway reopening. An air finger propagates (from right to left) into a strongly-collapsed fluid-filled airway. The finger reopens the airway and deposits a thin layer of fluid on its wall. The grey and cyan surfaces show the airway wall and the air-liquid interface, respectively. Courtesy of Matthias Heil and Andrew Hazel.

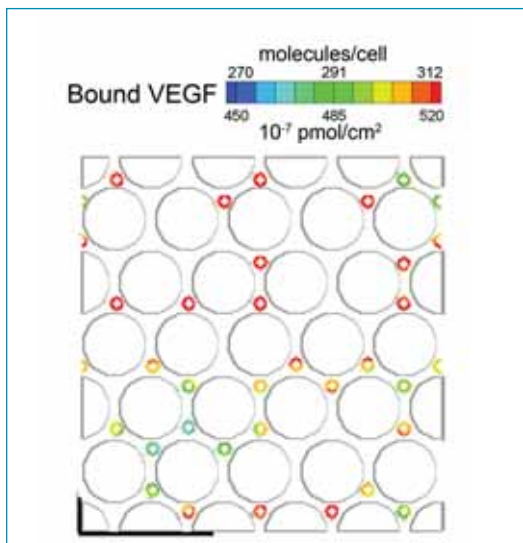
Heil, PhD, a reader in applied mathematics at the University of Manchester, United Kingdom.

Heil and **Andrew Hazel, PhD**, a mathematics lecturer at the University of Manchester, created a model of the respiratory tracts to study how a stream of air can open a closed airway. The work was published in the August 2006 issue of the *Journal of Biomechanical Engineering*.

A surfactant layer of proteins and lipids normally lines airways, reducing surface tension along the fluid-covered walls. However, many premature babies lack surfactant, and adults who have gulped air filled with smoke or noxious fumes can destroy their surfactant layers. In these cases of respiratory distress syndrome, the surface tension increases, the airways collapse, and the fluid lining becomes a blockage. Doctors treat this collapse by forcing pressurized oxygen vigorously into the airways to redistribute the fluid. They often add surfactant to avoid damaging lungs with over pressure.

“You want to reopen the lung as soon as possible,” says Heil. “The easiest way to do this is to apply an enormous amount of pressure, but then you can actually damage the lung tissue. So there’s a fine balance to be found.”

Heil and Hazel’s three-dimensional model surpassed previous two-dimensional models by taking the fluid’s viscosity, inertia, and surface tension into account. With this more realistic model, they revealed that less pressure is required to reopen airways that have collapsed more completely. The smaller cross-section of fully collapsed airways means a smaller volume of fluid to redistribute. This takes less energy, and less



air pressure, than moving around fluid that clogs a partially collapsed airway.

However, the researchers acknowledge shortcomings in their model. Airways are not infinitely long tubes, as Heil and Hazel assumed. “The airway branches are relatively short before they branch again,” says Heil. “You have this tree structure, and that is something we do not take into account.”

Oliver Jensen, PhD, professor of applied mathematics at the University of Nottingham, United Kingdom, says the new model is a step forward compared to older ones. “They’ve developed an absolutely wonderful tool,” says Jensen. He notes the model also should apply to other systems with fluid-lined tubes, such as blood vessels.

For doctors who treat collapsed airways, Heil and Hazel’s work eventually could lead to fine-tuned air pressure for different patients. For now, when doctors sit down to use a ventilator, says Jensen, “It’s nice if they can at least understand what is happening in the airways.”

—*Sarah CP Williams*

New Algorithm Finds Stories in Biomedical Literature

A good story ties up all the loose ends. A new data-mining tool takes a stab at doing the same. Dubbed *storytelling*, the algorithm may make it easier to unearth unexpected connections in the avalanche of freshly published research, or among high-throughput datasets. For example, *storytelling* can sift through tens of thousands of PubMed abstracts to discover scientific links between two apparently unrelated topics; or draw connections across a knowledge structure such as the Gene Ontology.

“What we are trying to do is link data sets very far apart,” says **Richard Helm, PhD**, associate professor of biochemistry at Virginia Polytechnic Institute. “In the end, we link data set A with data set Z in the form of a story.” The work was presented at the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), in August 2006.

Researchers might not have time to find complex relationships waiting for discovery in the literature, but *storytelling* does. It finds key documents bridging from one research publication (the starting point) to another (the end point). Using System X, a Virginia Tech supercomputer, the algorithm first classifies each PubMed article’s abstract into an organized branched set of terms. It can then make thousands of comparisons and join related publications into a chain connecting start to finish.

For example, Helm and his colleagues, used *storytelling* to dig through the literature seeking ties between two remotely related papers: one on tomato genes expressed in yeast and another on how chemical stress affects yeast gene expression. The supercomputer boiled down 140,000 yeast publications to nine abstracts—stepping stones from paper one to paper two. The results included a paper that identified a novel protein, expressed only when yeast cells are exposed to cadmium, which researchers might not have immediately connected with the first two papers. Although the paper might have surfaced in an ordinary PubMed search, it would have required much sifting to find it. While not every search will yield treasures,

hopefully most results from *storytelling* will provide new insights and hypotheses that researchers can test at the bench, Helm says.

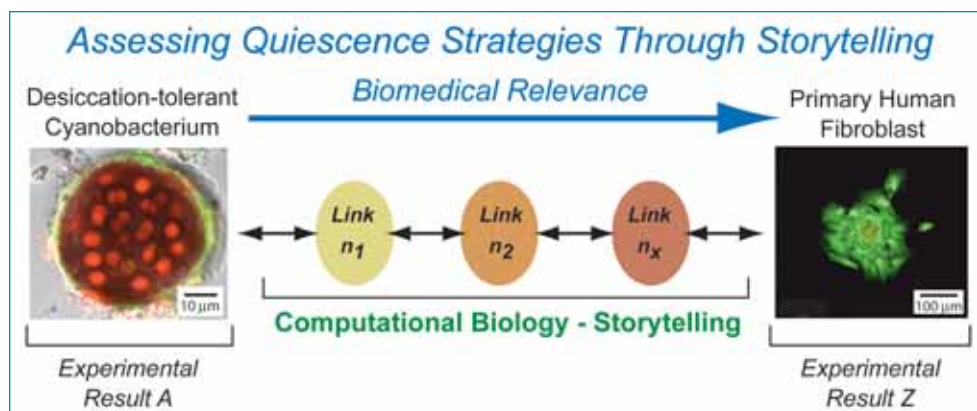
Bud Mishra, PhD, professor of computer science and cell biology at New York University, thinks *storytelling* can help biologists make new connections. “In some sense it closely resembles what biologists do, and it works in the same way that biologists think,” says Mishra.

—*Brian Lee*

A Fast Lane Through the Stomach

What goes into the stomach must come out, but perhaps not in the same order in which it entered, as gastroenterologists have long assumed. A two-dimensional computer model of human stomach digestion reveals a previously unknown narrow pathway that can funnel liquids from the top of the stomach to the intestines within 10 minutes.

“There are very few times you discover something that you weren’t expecting to find when you designed the experiment,” says **James G. Brasseur, PhD**, professor of mechanical engineering, bioengineering and mathematics at Penn State University and one of the researchers



As shown here, one might use storytelling to understand the pathways into and out of a quiescent state. Datasets evaluating desiccation-tolerant cyanobacteria (left, coccoid cells stained red and encased in an extracellular matrix, the exterior of which is stained green) could potentially be linked to studies involving the metabolic arrest and recovery of primary human fibroblasts (right, Live/Dead stain; image taken 72 hours after a two week metabolic arrest). Storytelling allows the biologist to link disparate datasets, allowing for the development of new hypotheses that can be tested at the bench and re-evaluated within the algorithm, ultimately resulting in new insights into the process of interest. Courtesy of Richard Helm.

who created the model for studying drug delivery. But this was one such time.

The curious finding could help explain the wide variability in drug activation time: A drug might get swept along in this fast digestive current one time, and not another. The work appeared online in the *Journal of Biomechanics* in August 2006.

In 2004, Brasseur and research associate **Anupam Pal, PhD**, had published a

The researchers dubbed their discovery the “Magenstrasse,” which means “stomach road” in German. It is a sort of passing lane, where digested matter reaches the intestines more quickly than food and liquids outside the road. “We do not know if the Magenstrasse has a physiological function or if it is only a byproduct of the contraction waves in the distant stomach,” says Brasseur. The team showed that when the simulation excluded contractions known to exist at the bottom of the stomach, the Magenstrasse didn’t form.

“Brasseur’s modeling has revealed a potential new mechanism of liquid emptying that can explain why some patients take a very short time to absorb drugs in their bloodstream,” says **Michael Fried, MD**, director of the division of gastroenterology and hepatology at the University Hospital Zurich, Switzerland. He says pharmaceutical companies who wish to achieve constant and reproducible patterns of absorption of drugs must take this study into account. However, Fried notes, the model “has to be validated in animals or in humans.”

Brasseur agrees more research is necessary. “We would need further studies that checked how the path works with nutrients and drugs of different densities,” he says. He believes the Magenstrasse would still exist if the stomach contents are thicker than the viscous liquid used in his team’s simulations, but it may assume a different shape.

—**Maria José Viñas**

Stem Cells’ Existential Crisis Explained

To differentiate or not to differentiate? That is the question constantly faced by embryonic stem cells. And they seem to answer it decisively at the behest of a molecular trio of transcription factors. A new computational model shows how it is possible for three proteins to control the switch in the observed, clear-cut manner. The model also gives researchers a hypothesis they can test in the lab.

“We’ve shown that these three players are able to define the embryonic stem

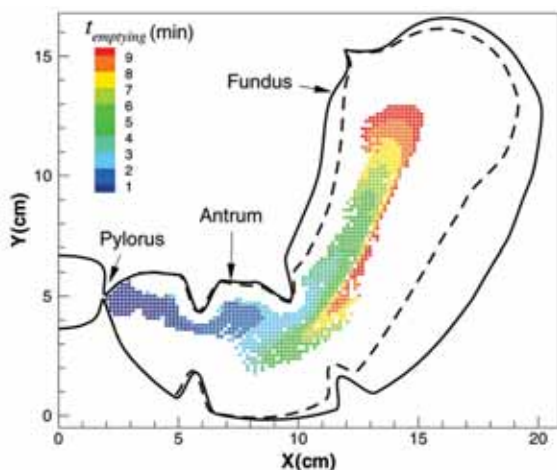
cell switch,” says **Carsten Peterson, PhD**, professor of computational biology at Lund University in Sweden and lead author of the study. By explaining this either/or molecular circuitry, the model could eventually help scientists harness stem cells for treatments. The work appears in the September 2006 issue of *PLoS Computational Biology*.

Embryonic stem cells have two defining traits: They can divide forever to remain stem cells, and they are pluripotent, meaning they have the potential to become any type of somatic cell in the body—gut, muscle, skin, blood or nerve. But whether in a Petri dish or the body, stem cells receive a barrage of molecular signals that they must interpret and respond to with a binary decision. “Either you are a stem cell or you commit yourself,” says Peterson.

Over the last few years, biologists have discovered that a handful of proteins determine an embryonic stem cell’s fate. Three in particular—named OCT4, SOX2 and NANOG—seem to coordinate the decision by cueing the actions of hundreds of target genes.

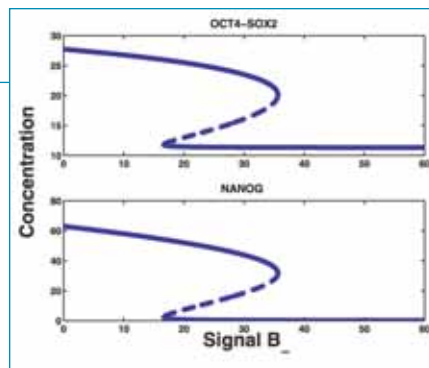
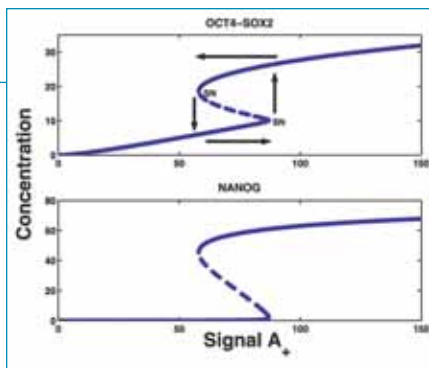
Peterson’s team derived mathematical equations governing the rate at which these three transcription factors bind and unbind to DNA, thereby regulating the expression of differentiation genes and stem cell genes. The team then used the Systems Biology Workbench to simulate how this trio controls other genes to engineer opposing outcomes: self-renewal or differentiation. They infer that the proteins reinforce one another’s actions through a positive feedback mechanism, creating a bistable switch: either on or off, with no middle ground. When all three proteins are active, stem cells remain stem cells; when the trio is inactive, the cells differentiate, with no middle ground. Peterson and others hypothesize that stem cells receive external signals to control the switch. Because of the positive cascade of interactions, the cells effectively ignore slight changes in those signals and respond with a single outcome every time.

The model was based on previous work by **Laurie Boyer, PhD**, a postdoctoral fellow at the Whitehead Institute in



This image of Brasseur’s two-dimensional stomach model shows the initial locations of all particles that left the stomach over the course of 10 minutes. The pattern of gastric emptying suggests the existence of a Magenstrasse in the stomach. The dashed line shows the shape of the stomach after 10 minutes. Courtesy of James Brasseur.

computer simulation of gastric flow and mixing. In the current work, they used the same basic simulation, but studied gastric emptying. They gave unique numbers to thousands of points (fluid particles) uniformly distributed around the stomach. They then watched the order in which these particles left the stomach during a 10 minute simulation, while keeping track of each particle’s position at each time-step. When they then ran the simulation in reverse, they could watch the particles’ changing positions. After color-coding the points by time of leaving, they observed a ribbon-like path of gastric emptying that originates in the top of the stomach and passes along its side of least curvature. The entire animated sequence can be seen at <http://mne.psu.edu/Brasseurlab/gastric/>.



A+ represents hypothetical factors that activate OCT4 and SOX2. In response to increasing that signal, the OCT4-SOX2 dimer and NANOG switch from all off to all on. **B-** represents factors that repress NANOG, which has the effect of turning the switch off. Courtesy of Carsten Peterson.

Cambridge, Massachusetts. She thinks other components must function with the trio of proteins. However, she says, the work “provides a testable model to explain how OCT4, SOX2 and NANOG may contribute to these seemingly opposing activities.”

The heart of the model—explaining how stem cells reconcile their dual identities—is vital, says Boyer. “If you are ever going to realize the therapeutic potential of these cells, you really need to find the key for understanding how embryonic stem cells balance their ability to self-renew or differentiate.”

—Ewen M. Callaway

Connecting the (Microarray) Dots from Drug to Disease

Normal cells, diseased cells and cells on drugs share a common language: They all produce their own patterns of gene expression. And the patterns can be compared in useful ways—given a disease in which a certain set of genes are up- and down-regulated, one would like to find a drug that specifically counteracts those changes. A powerful new web-based tool allows researchers to draw just such connections. Researchers from all around the world can use the tool, called the Connectivity Map, to identify potential new drugs for a variety of diseases, and potential new uses for existing drugs.

“The objective is to connect diseases with the genes that underlie them and the drugs that treat them,” says **Justin Lamb, PhD**, a senior scientist at the Broad Institute of Harvard and MIT. He and his colleagues described their work in the September 29, 2006, issue of *Science*.

The advent of DNA microarrays more than a decade ago made the

Connectivity Map possible by allowing scientists to study thousands of genes all at once. Lamb and his colleagues used microarrays to determine which genes were turned up or down in cells treated with 164 different biologically active compounds (mostly drugs). They also gathered together previously published gene expression data for various diseases, including obesity, Alzheimer’s Disease, and various types of cancer. Using pattern-recognition software the team developed, any researcher around the world can query the database at <http://www.broad.mit.edu/cmap> to identify drugs or diseases with patterns that are either the same as or opposite to a particular gene signature of interest.

The most exciting use of this approach is in suggesting potential drug therapies for real diseases, Lamb says. In one of the team’s initial tests, gene signa-

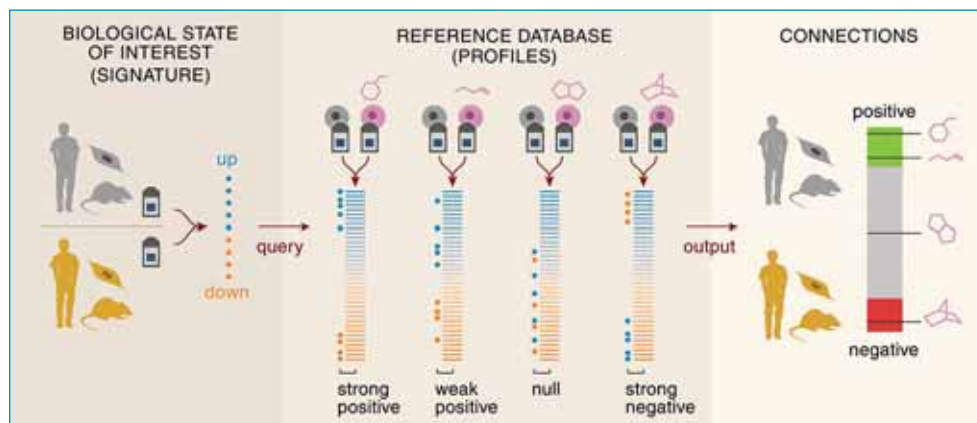
tures exposed a possible new treatment for children with drug-resistant acute lymphoblastic leukemia. The researchers also have used the database to seek novel treatments for obesity and Alzheimer’s disease.

The Connectivity Map also can unveil how drugs work on a molecular scale, says Lamb. For instance, similar genomic patterns in the database pointed to a connection between a compound called gedunin and a class of well-understood drugs. This implies gedunin might operate in a similar fashion to disrupt hormone signals in cancerous cells.

Having completed the Connectivity Map’s pilot stage, the team intends to expand to cover all drugs approved by the Food and Drug Administration. One limitation is that it is restricted to specific cells in a petri dish, which neglects other interactions that occur in a human body, says **Eric Schadt, PhD**, senior scientific director of Rosetta Inpharmatics in Seattle. Still, Schadt says, “It’s a great piece of work. It’s a great tool, and it will be well-used.”

“The whole field is on fire with these large-scale experiments,” Schadt adds. “It’s the only way you can uncover these networks that drive diseases. I think it’s right on the money.”

—Marcus Woo □



*The Connectivity Map lets users compare the gene expression signature of any cell state of interest (left) with a reference database of profiles from cultured human cells treated with various drugs. Pattern-matching algorithms score each reference profile for the direction and strength of enrichment with the query signature to produce a “connectivity score.” Here the columns for the first and second drug treatments show a positive connection to the state of interest; the third has no relationship to it; and the fourth has a negative connection. From: *The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease*, by Lamb, et al., *Science* 2006 Sep 29;313(5795):1929-35. Reprinted with permission from AAAS. <http://www.sciencemag.org/cgi/content/full/313/5795/1929>*