

# 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.

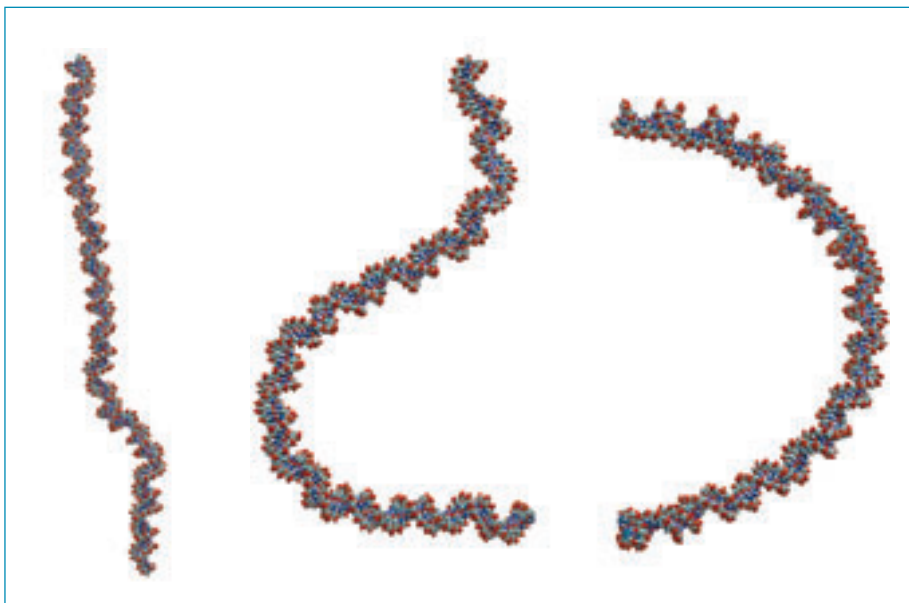
Onufriev’s results agree nicely with two independent threads of experimental evidence that have recently emerged, says **Philip Nelson, PhD**, professor of physics at the University of Pennsylvania. A 2004 paper showed that DNA of 100 base pairs spontaneously forms circles in physiological conditions; 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**

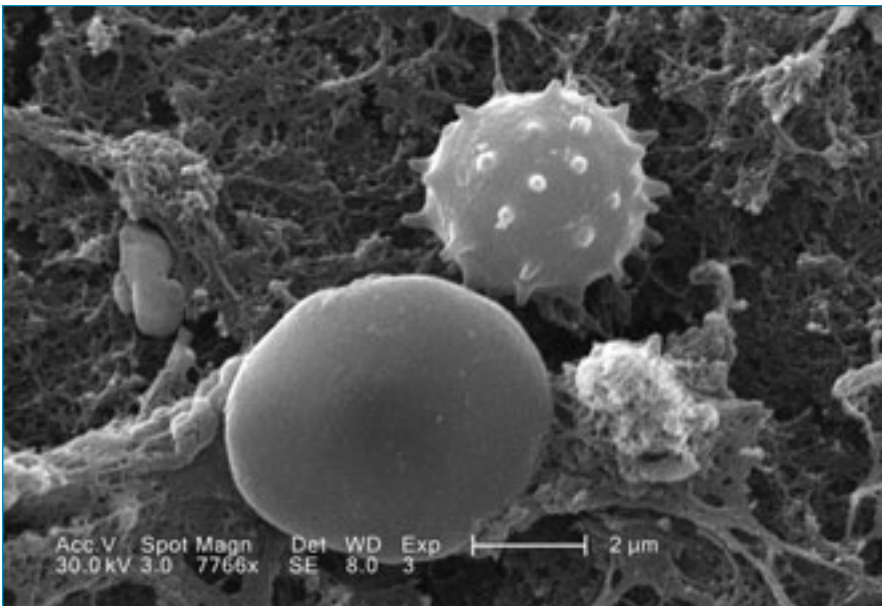
We see DNA bending quite a bit, says Alexey Onufriev. If this idea holds up, it may be a paradigm shift in how we think about protein-DNA complexes.



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 destruc-

tion 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.”

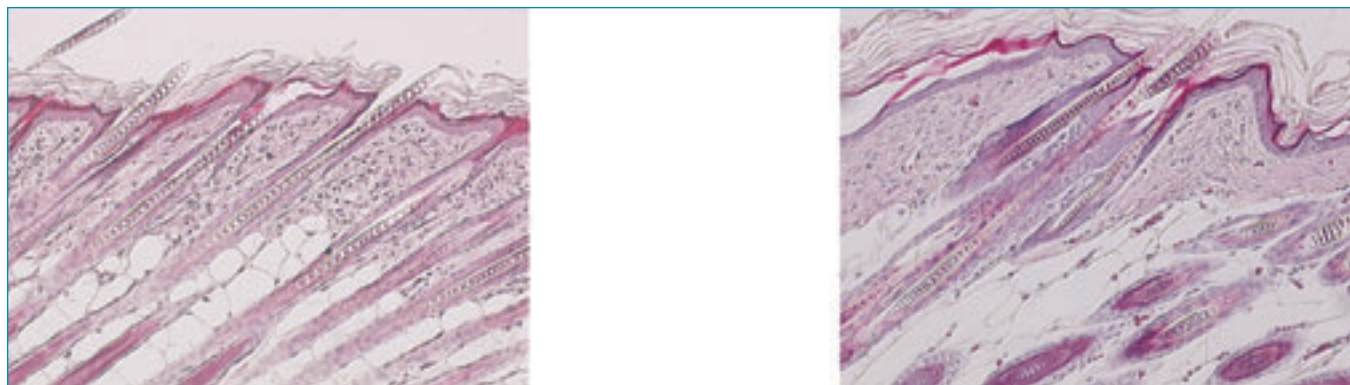
—By **Louisa Dalton**

## 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

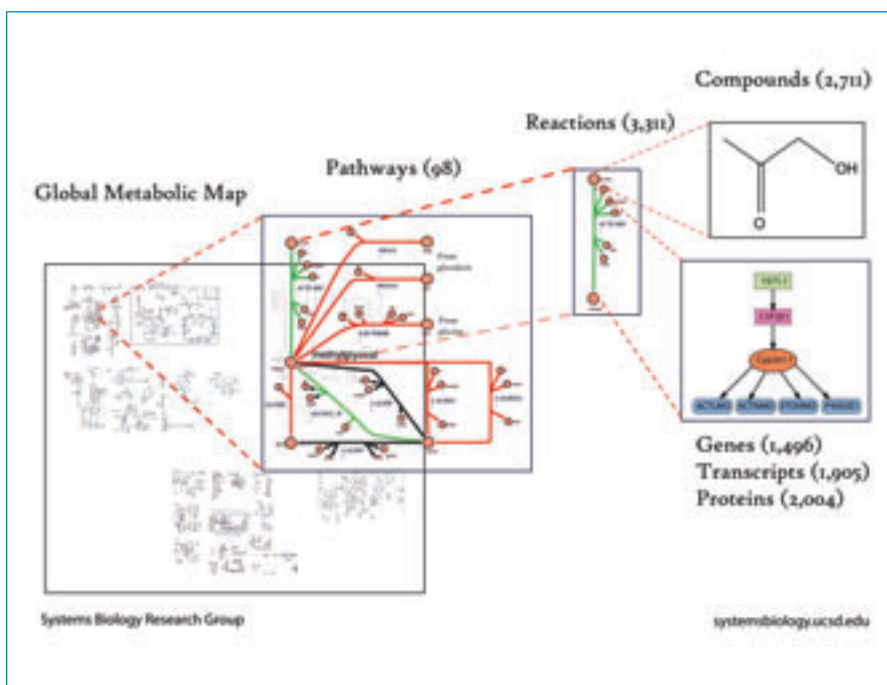
that 50 years after Turing proposed his model, biologists are just now providing detailed molecular evidence for it. "Turing," he says, "was a very smart man."  
 —By *Louisa Dalton*

## 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

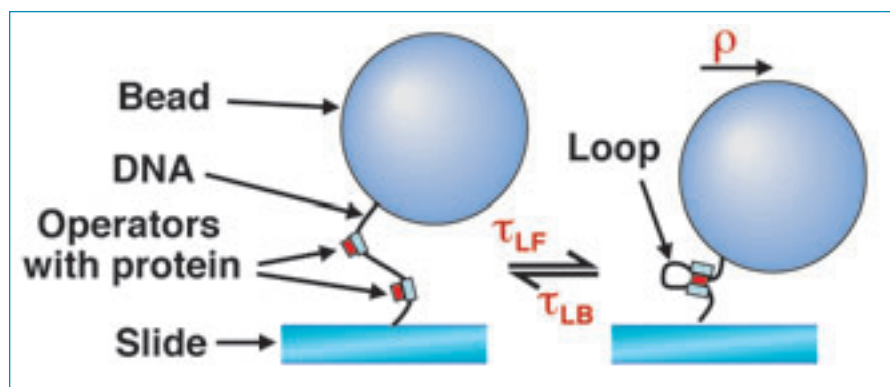
The hidden Markov model—a statistical model used for decades in fields as diverse as speech recognition and climatology—has received an update and a new application. Researchers at the University of Pennsylvania and Emory University adapted the model for tethered molecule experiments, and used it to 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 litera-

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

ture 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.

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** □