NewsBytes

A Viral Closeup

The phi29 bacteriophage is an efficient infection machine—it fires its genome into a host bacterium, hijacks the host's cellular equipment, and assembles an army of new viruses for its next mission. For the first time, scientists have produced sub-nanometer resolution pictures of the virus, revealing some striking new details—including an unexpectedly tight twist of DNA suggestive of how the virus springs into action. The results appear in the June issue of *Structure*.

"We use structure as a way to try and understand how viruses function," says **Timothy Baker, PhD**, professor of chemistry/biochemistry and molecular biology at the University of California, San Diego who led the collaboration between UCSD and the University of Minnesota. "The more we can learn from structure, the better we'll understand the whole infection process and perhaps ways to circumvent it."

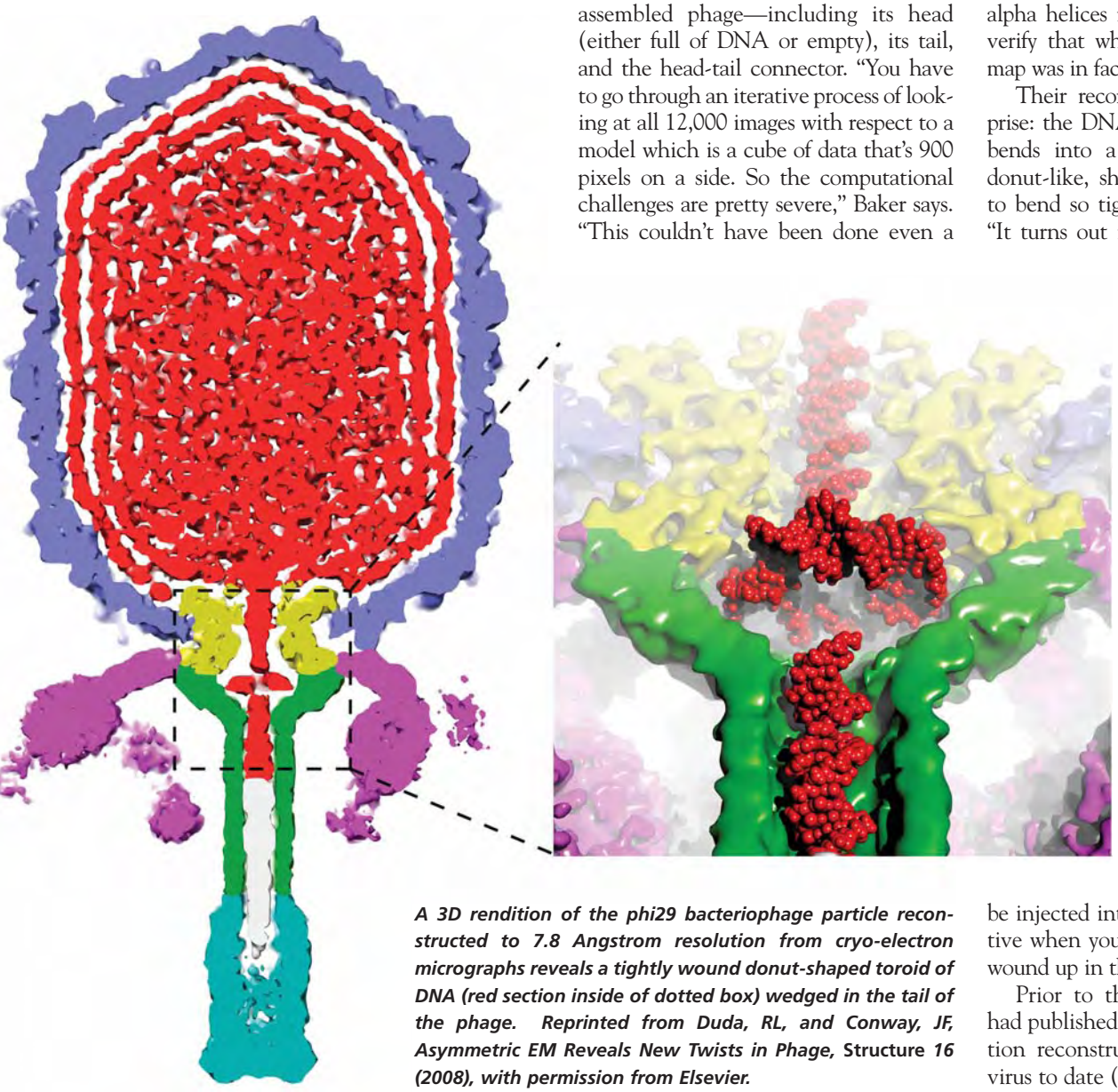
Using computer reconstruction, Baker and his colleagues aligned roughly 12,000 electron microscope images of frozen viral particles at different angles and fused them into a 3-D picture of the assembled phage—including its head (either full of DNA or empty), its tail, and the head-tail connector. "You have to go through an iterative process of looking at all 12,000 images with respect to a model which is a cube of data that's 900 pixels on a side. So the computational challenges are pretty severe," Baker says. "This couldn't have been done even a few years ago, not without really dedicated supercomputer power."

The resolution achieved—8 Angstroms—was two-fold higher than ever before for an asymmetric virus (where researchers cannot exploit symmetry to reduce complexity). At this resolution, individual alpha helices (in the proteins that make up the head-tail connector piece of the virus) become distinguishable as tube-like structures. Baker's team compared their picture of the viral head-tail connector with atomic-level models of this structure that were available from X-ray crystallography, and showed that the alpha helices matched up. "It helped us verify that what we were seeing in our map was in fact believable," he says.

Their reconstruction revealed a surprise: the DNA in the tail of the phage bends into a tight coil—a toroid, or donut-like, shape. DNA isn't expected to bend so tightly over short distances. "It turns out if you talk to people who know something about DNA, they say it is possible," Baker says. "They just haven't seen it in a biological system like this before."

"In terms of shock value, that was amazing," comments **John E. Johnson, PhD**, professor of molecular biology at The Scripps Research Institute who occasionally collaborates with Baker. The bacteriophage must pack its DNA into a tiny space against tremendous forces, and Johnson speculates that the toroid may act as a plug to hold the DNA inside until it's ready to be injected into the host. "It's so suggestive when you look at how this thing is wound up in this little cavity," he says.

Prior to this work, Johnson's team had published one of the highest resolution reconstructions of an asymmetric virus to date (17 Angstroms—as report-



A 3D rendition of the phi29 bacteriophage particle reconstructed to 7.8 Angstrom resolution from cryo-electron micrographs reveals a tightly wound donut-shaped toroid of DNA (red section inside of dotted box) wedged in the tail of the phage. Reprinted from Duda, RL, and Conway, JF, Asymmetric EM Reveals New Twists in Phage,* Structure *16 (2008), with permission from Elsevier.

ed in *Science* in 2006). "We saw a lot of interesting things," he says. "But this paper has pushed it to a higher level."
—By Kristin Sainani, PhD

An *In Silico* Time Machine

In biology, many exciting events happen on the millisecond timescale—proteins fold, channels open and close, and enzymes act on their substrates. Atomic-level simulations of this duration are beyond the reach of current technology, but a new specialized computer called Anton—described in the July 2008 issue of *Communications of the ACM*—may change all this. Slated to be operational by the end of the year, the machine is projected to speed up molecular dynamics simulations 100-fold.

The basic goal is to be able to visualize, at the atomic level of detail, an entire biological trajectory, such as an anti-cancer drug (like Gleevec®) inactivating its target enzyme, says **David E. Shaw, PhD**, chief scientist of D.E. Shaw Research, the independent research laboratory that is creating Anton, and a senior research fellow at the Center for Computational Biology and Bioinformatics at Columbia University. Because it provides what might be thought of as a computational microscope, Anton is named after 17th century scientist Anton van Leeuwenhoek, known as the father of microscopy.

"Our machine only does molecular dynamics. It does it blindingly fast, but it's pretty brittle and isn't designed to do anything else," Shaw explains. In molecular dynamics simulations, time is broken into discrete steps, each a few femtoseconds (10^{-15} of a second) of simulated time. At each step, the computer calculates the force exerted on each atom in the system (typically 25,000 to 100,000 atoms) and updates its position and velocity. The various time steps cannot be executed simultaneously since each is dependent on the previous, but Anton uses 512 highly specialized chips working in parallel to speed up the massive calculations within each step.

Anton is designed for a specific task: molecular dynamics simulations. Here is one of the first Anton application-specific integrated circuits (ASIC), which arrived in January 2008. Reprinted from Anton, a special-purpose machine for molecular dynamics simulation, David E. Shaw, et al.,* Communications of the ACM *51:91-97 (2008) with permission from the ACM.

"They've done a beautiful job, and there are a lot of intellectually interesting aspects to the approaches they've taken," says **Vijay Pande, PhD**, associate professor of chemistry at Stanford University and director of the protein folding distributed-computing project Folding@home. Still, Pande advocates a different approach. Rather than simulating one long trajectory, which could take a million days on one general purpose computer, he simulates a large number of shorter trajectories and then merges them together with a clever algorithm. This may take just 10 days on 100,000 computers. "The approach not only gives access to long timescales, but having many trajectories allows you to do statistical testing, which you cannot do on a single trajectory," Pande says. "Most of the questions that people in the field are interested in are inherently statistical questions," he says.

Anton, slated to be operational by the end of the year, is projected to speed up molecular dynamics simulations 100-fold.

But according to Shaw, "The two approaches are very complementary and I think they may turn out to be useful for solving very different types of problems." Combining many smaller trajectories is more efficient, he says. "But there are some cases in which you'd like to have confidence that what you're seeing is one continuous, unbiased, physically realistic trajectory."

Though other groups have previous-

ly attempted to develop specialized computers for molecular dynamics simulations, most efforts have failed to stay ahead of Moore's Law, which says that the speed of general purpose computers doubles every 18 months.

"The Shaw group's effort has been one of the most exciting examples of trying to do that to date," says Pande. "Since the machine isn't out yet, it's too early to say whether they have succeeded or not. But they've got a reasonable shot."

—By Kristin Sainani, PhD

Bacteria Prepare Themselves

When we see dark clouds, we might grab an umbrella before heading outside. We've long believed that showing such foresight requires a brain and complex information-processing capability. It turns out, though, that even microbes, which do not have brains or a nervous system, can learn to use cues from their surroundings to anticipate future events, according to a new research study based on both experimental and computational techniques.

"What we have shown is that microbes too have the intrinsic capacity for predictive behavior," says **Saeed Tavazoie, PhD**, an associate professor of molecular biology at Princeton University who published the study in the June 6 issue of *Science* with co-authors and Princeton colleagues **Ilias Tagkopoulos, PhD** and **Yir-Chung Liu, PhD**. "Indeed, this may be essential for their survival." The findings could have implications for infectious disease treatment and microbial applications in industry.

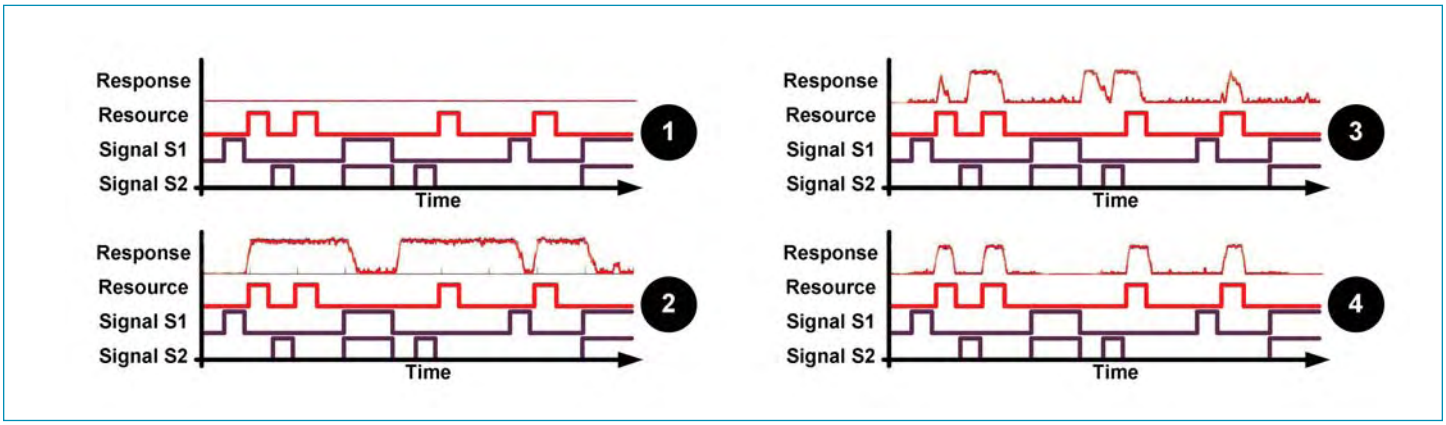
"What we have shown is that microbes too have the intrinsic capacity for predictive behavior," says Saeed Tavazoie.

Escherichia coli (*E. coli*) normally adjusts its breathing to match the ambient oxygen level: In the open, the bacterium breathes oxygen; inside an animal's oxygen-poor gut, it doesn't. According to prevailing notions, this switch from aerobic to anaerobic respiration is a purely reflexive response to the drop in oxygen level.

But Tavazoie and his colleagues suspected the microbes wouldn't survive if they responded only when they were already oxygen-deprived. They proposed that, instead, *E. coli* senses warmth when it enters an animal's mouth, and uses this as an early cue to switch to anaerobic breathing. In laboratory experiments, the researchers found this to be the case: When the temperature rises, *E. coli* turns off many genes needed for aerobic respiration. "By anticipating the subsequent lack of oxygen, it improves its chances of survival," says Tavazoie. "This is clearly predictive behavior." Moreover, when the researchers caused oxygen levels to *rise* shortly after an increase in temperature, *E. coli* evolved (over about 100 generations) to disregard warmth as a cue. "It rewires itself to forget the old association," says Tavazoie.

To explain how a microbe could evolve such complex behavior, the researchers devised a computational framework that mimics the essential aspects of microbe ecology. Modeled as a network of genes and proteins, a virtual bug in this virtual ecology lives and breeds when it has enough energy, or dies when it runs out of it. To gain energy, it has to be ready to eat, biochemically speaking, when "food" is available. But if it gets ready to eat and no food arrives, it wastes precious energy.

To help the virtual bugs, the researchers gave them different patterns of cues to indicate that food is coming. In one experiment, the bugs were fed



Predictive behavior of a simulated microbe species at different points along an evolutionary trajectory. The resource (food) is always given shortly after giving either, but not both, of the two signals (environmental cues). Initially (subplot 1) the response seems random relative to the food and cues. Eventually, however (subplot 4), guided by the pattern of cues, the microbe evolves its feeding response to make it synchronize with the food availability. Courtesy of Ilias Tagkopoulos. Reprinted from the supporting online material for Predictive Behavior Within Microbial Genetic Networks, Ilias Tagkopoulis, et al.,* Science *320, 1313 (2008).

shortly after they got one of two different cues—but not if they got both cues at once. "To predict mealtimes accurately in this case, the microbes would have to solve a complex logic problem," says Tagkopoulos, an electrical engineer associated with the Lewis Sigler Institute for Integrative Genomics. Sure enough, after a few thousand generations, a gastronomically savvy—and ecologically fit—strain of microbe emerged. The feeding response of such a fit bug (see figure) illustrates how interacting genes and proteins could evolve complex behavior.

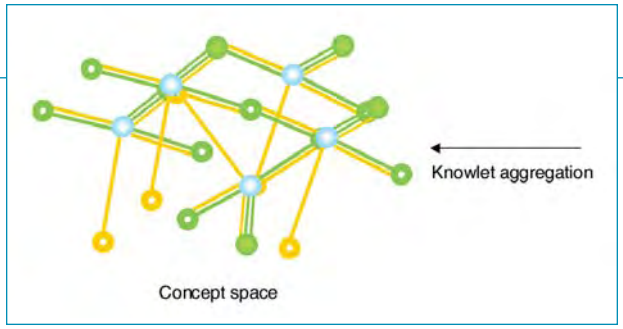
According to **David Reiss, PhD**, a computational biologist at the Institute for Systems Biology in Seattle, the researchers' computational framework is notable for incorporating more biological mechanisms than prior models did. He cautions, however, that even this model oversimplifies the behavior of real microbes. Nevertheless, Reiss says, the study is interesting and novel for showing that anticipatory behavior is not restricted to higher systems with decision-making capability.

—By Chandra Shekhar

Molecular Biology Wikis Launched

If you build it, will they come? That's the question on everyone's mind after the launch of two pioneering initiatives in community annotation: WikiProteins and Gene Wiki, announced, respectively, in the May 28 issue of *Genome Biology* and the July 8 issue of *PLoS Biology*. The efforts create a central repository of information on genes and proteins and call on the scientific community to keep it up-to-date and accurate.

"There's no way we can handle the current growth of knowledge with central annotation only," says **Barend Mons, PhD**, who leads the WikiProteins effort. "I'm a big fan of the authoritative databases like UniProt, but we have to make them grow faster. So what we need is a shell around them of community annotation." Mons is associate professor of human genetics at the Leiden University Medical Centre and of medical informatics at Erasmus University, both in the Netherlands.



Each unique biomedical concept in WikiProteins is attached to a "knowlet" or concept cloud, illustrated here. A concept (depicted as a solid blue ball) is associated with other concepts through facts (established relationships, depicted as solid green balls), co-occurrences (co-occurrences in sentences in PubMed, depicted as green rings), or implicit associations (overlapping concepts in their Knowlets, depicted as yellow rings). Reprinted from Barend Mons, et al., Calling on a million minds for community annotation in WikiProteins, in* Genome Biology *2008, 9:R89.

"WikiProteins is more than just a Wiki; it has the whole knowledge space hovering over it," Mons says. Using text mining, WikiProteins imported structured content (adhering to computer-readable, controlled vocabularies) on 1.2 million unique biomedical concepts from existing databases, such as PubMed, Swiss-Prot, and Gene Ontology. The system also created profiles for about 1.6 million authors in PubMed, who are expected to serve as the knowledge guardians. "If you have 1.6 million people in PubMed publishing today and you have 1.2 million concepts in the Wiki, then roughly everyone could take one concept and make sure the page on that concept is correct. That's doable," Mons says.

Gene Wiki operates within Wikipedia and, in contrast to WikiProteins, emphasizes unstructured content, such as free text and images, "more akin to a review article," says **Andrew Su, PhD**, of the Genomics Institute of the Novartis Research Foundation, who leads the effort. Using data from Entrez Gene, the system added or amended about 9000 Wikipedia "stub" entries on human genes, which anyone can edit. "Being part of the larger Wikipedia community is certainly an advantage of this system. The people there are experts at welcoming newcomers, fighting vandalism, and formatting things correctly," Su says.

Su and Mons have plans to collaborate. WikiProtein and Gene Wiki entries will be linked through a common "entry page" (likely hosted in WikiProteins), making it easy to navigate between the systems. "This will allow users to take advantage of whichever system they feel comfortable with," Su says.

Getting bench scientists to participate will be a challenge, Mons says, but he believes the incentives are high. The WikiProteins system mines PubMed for new information daily, finds new explicit and implicit associations—such as predicting protein-protein interactions—and alerts scientists of all edits and updates to concepts in their purview. "I hope it becomes a daily part of their knowledge discovery process," Mons says. Since its launch, WikiProteins has also received requests to enable users to enter data as unstructured, free text, which should lower the barrier to participation.

"I'm a big fan of the authoritative databases like UniProt, but we have to make them grow faster. So what we need is a shell around them of community annotation," says Barend Mons.

Another factor that may boost participation is the development of ways to trace authorship for each entry, so that authors can get credit for their work and readers can assess the reliability of content. A recent proof of this possibility was demonstrated in

"WikiGenes," (not to be confused with the GeneWiki!) a project described in *Nature Genetics* in September 2008. WikiGenes was developed by **Robert Hoffmann, PhD**, at the Massachusetts Institute of Technology. It's part of his Mememoir project, which has, he says, "the ambitious goal to create a free collaborative knowledge base for all of science—where authorship matters."

Though the creators of the various Wikis have not yet formally quantified participation, Su says that there's been an uptick in Gene Wiki activity since the *PLoS Biology* paper came out. "It gives me hope that the system is right and that the framework is there, so if we are tapping into a desire in the community to share knowledge and harness community intelligence, then we have the structure to do it now."

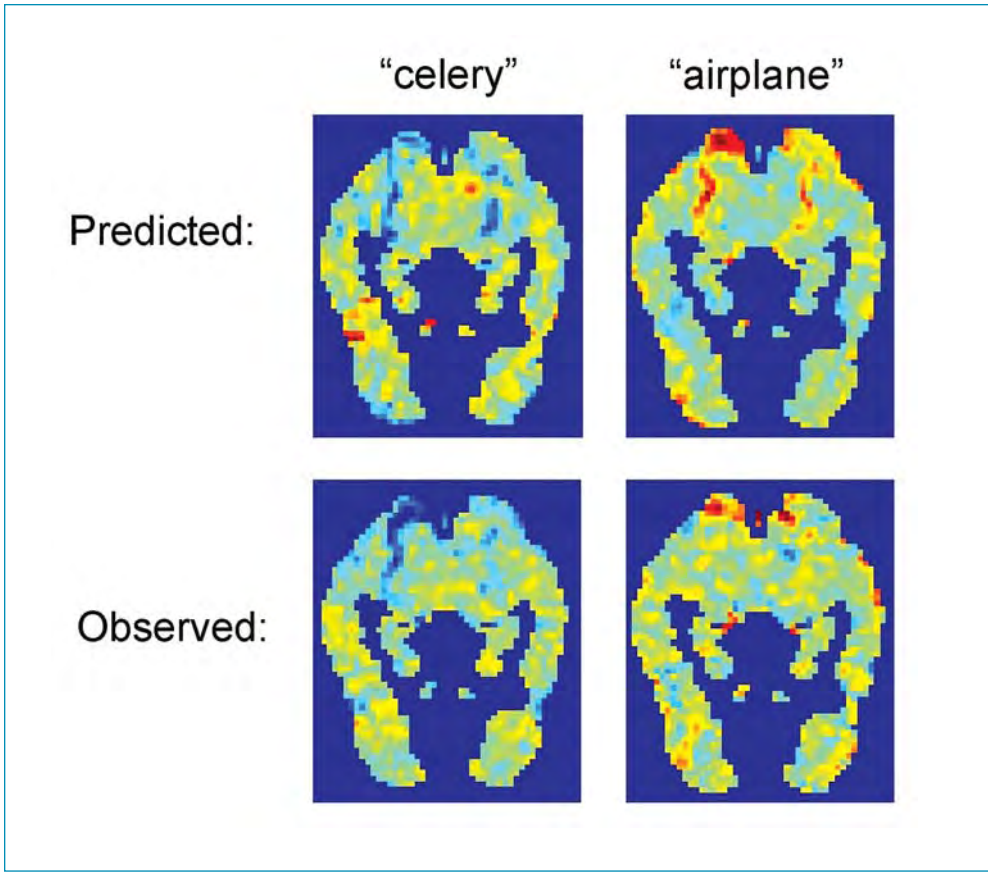
More information is available at: www.wikiprofessional.org (WikiProteins) and http://en.wikipedia.org/wiki/Portal:Gene_Wiki (Gene Wiki).

—By Kristin Sainani, PhD

Predicting Brain Response To Nouns

Thinking of a noun—a peach, train, or bird, for example—activates specific parts of the brain. Now, scientists have trained a computer to predict such activation patterns. The achievement represents a step toward understanding language processing and could one day contribute to treatments for cognitive decline.

"If we had a better model of how the brain represents language, we'd be better able to make sense of disorders like dementia," says **Tom Mitchell, PhD**, a professor of computer science at Carnegie Mellon in Pittsburgh and lead author of the research published in the May 30 issue of *Science*.



Brain activation patterns in response to nouns: The computer algorithm predicted the response to newly encountered words with 77% accuracy. Courtesy of Tom Mitchell. From Mitchell, TM, et al., Predicting Human Brain Activity Associated with the Meanings of Nouns,* Science, *320 (5880): 1191 (2008) DOI: 10.1126/science.1152876. Reprinted with permission from AAAS.

Functional magnetic resonance imaging, or fMRI, registers changes in blood flow within peoples' brains as they are asked to do a specific task—such as thinking of a specific word. Since 2000, Mitchell and **Marcel Just, PhD**, professor of psychology at Carnegie Mellon and co-director of the Pittsburgh Brain Imaging Research Center have collaborated to train a computer to produce fMRI images like those generated by humans. The training process uses two sources of data: fMRI images collected from nine people viewing 60 nouns; and a database (derived from a trillion words of text from the Internet) describing pairings of nouns and the verbs that accompany them most frequently in written text. Noun-verb pairings are the basis of language, as anyone knows who has

The computer model was able to produce a pattern of brain activity in response to words it had never before encountered with greater than 70 percent accuracy.

raised a toddler, Mitchell notes.

Once trained, the computer model was able to produce a pattern of brain activity in response to words it had never before encountered with greater than 70 percent accuracy. "We now have a model that is capable of extrapolating beyond the data on which it was trained," Mitchell says. For example, after training, the model could predict that a food noun would provoke activity in the area of the brain mediating eating sensations, the so called gustatory cortex: "peach," for example, frequently occurs in English paired with the verb "eat." Similarly, a noun will activate motor areas of the brain to the degree that it co-occurs with the verb, "push," or cortical regions related to body motion to the degree that it co-occurs with "run."

Harvard cognitive psychologist **Alfonso Caramazza, PhD**, cautions that the model may be imperfect. He says it fails to capture an area of the brain that is damaged in semantic dementia, one form of brain damage in which people cannot understand the meaning of words. "Our understanding of concepts, and representation of this information in the brain, is not only sensory-motor," Caramazza says. Evolution likely has sculpted our brains to react appropriately to inanimate things that may be either potentially dangerous or pleasurable. Emotional areas of the brain respond differently to a hammer than to a dog, he points out.

"These are deep questions to which no one has the answers, so one should be cautious," Caramazza says, adding, "I think (the Pittsburgh team) would agree, these tools are in their infancy and we are only beginning to know how to use them."

—By Roberta Friedman, PhD

A Finer Fat Model

When it comes to heart disease risk, "bad" and "good" cholesterol—also known as low density lipoproteins [LDL] and high density lipoproteins [HDL]—do not tell the whole story. These particles that carry fat through the blood can be broadly classified based on their density, but they actually vary widely in their composition and clinical risk. A new computational model, described in the May issue of *PLoS Computational Biology*, allows scientists to see this diversity for the first time, providing additional information to aid in diagnoses and treatment planning.

"We look at lipoprotein profiles in greater detail in order to find possibly relevant abnormalities in the lipid values that would remain undetected by looking only at LDL or HDL," says lead author **Katrin Hübner, PhD**, a postdoctoral research fellow at the University of Heidelberg who completed much of the work while a PhD student at the Charité University hospital in Berlin. The model has several potential clinical applications.

Unlike previous models of blood lipid metabolism, Hübner and colleagues modeled the whole spectrum of individual lipoproteins.

Unlike previous models of blood lipid metabolism, which considered just four lipoprotein density classes (very low, low, intermediate, and high), Hübner and colleagues modeled the whole spectrum of individual lipoproteins—by combining any of three proteins (apoB, apoA, and other) and three fat molecules (cholesterol, triglycerides, and phospolipids) in varying amounts. The particles undergo 20 reactions, including particle birth from the liver, particle death from cell uptake, and transfer of fats between particles.

In initial simulations, Hübner and colleagues generated virtual blood lipoprotein profiles that closely matched experimental values from healthy individuals. Then they tweaked the parameters in their model to mimic three known lipid disorders. For example, to simulate familial hypercholesterolemia, which involves a malfunctioning LDL receptor, they decreased the rate of cellular uptake of apoB-containing particles (which are recognized by the receptor) by 75 percent. The simulations accurately reproduced the characteristic lipid profiles of the three diseases.

The model could help pinpoint the underlying molecular defect in patients with abnormal lipid profiles of unknown origin, Hübner says. It could also be used to predict the impact of specific treatments, such as drugs or lifestyle changes, on a patient's lipid profile.

"This work addresses an important issue in modeling lipoprotein metabolism, which is the heterogeneity of lipoproteins," says **Brendan O'Malley, PhD**, Project Leader of Systems Biology of Lipid Metabolism at Unilever Corporate Research in the United Kingdom, who also works on lipoprotein modeling (using a different approach).

"This is one of the first works in this area, so there's still quite a lot of work to be done," he says. For example, the model needs to be further validated with high quality patient data. But, in the future, it could lead to improved diagnostics and personalized treatments for cardiovascular disease, he adds.

"It's not ready for the clinic yet," Hübner agrees. "But we've made a promising first step."

—By Kristin Sainani, PhD □