

# NewsBytes

## Aquaporin Simulations De-Bunk Gas Exchange Assumptions

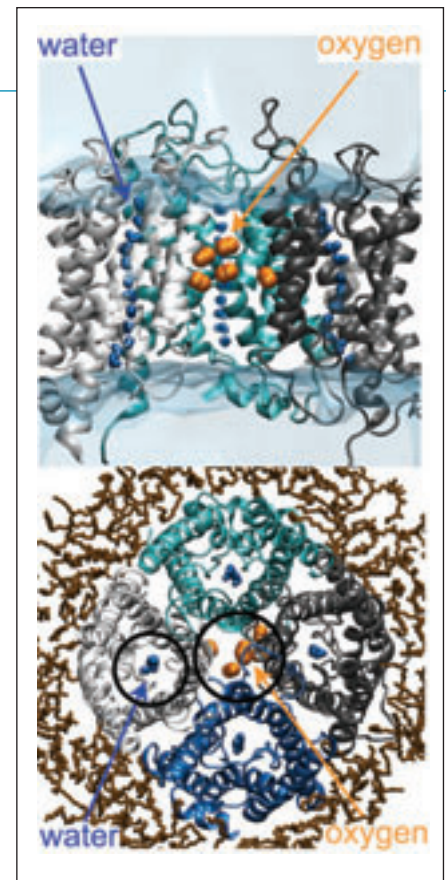
Biologists have long taken gas exchange for granted, assuming that gases simply seep through the cell's lipid membrane. Since 1998, however, evidence has been building that gases might also be exchanged through pores created by specialized proteins.

Now molecular dynamics simulations of aquaporins have weighed in on the question. The result: "It's now well established that these proteins can conduct gas molecules," says **Emad Tajkhorshid, PhD**, co-author of the work and assistant professor of biochemistry, pharmacology and biophysics at the University of Illinois at Urbana-Champaign. But, he says, some uncertainty remains: "Whether or not it's important in the human body, that's the controversial part." The work was published in the March 2007 issue of the *Journal of Structural Biology*.

Fifteen to twenty years ago, scientists believed that water permeation through lipid bilayers was enough for water transport into and out of cells. Gradually,

exchange experimentally for about ten years. To him, aquaporins are a likely suspect for gas conduction because they exist in places where oxygen must go in and carbon dioxide must come out. For example they are plentiful in cells that line the lung, in red blood cells, and in astrocytes—cells at the blood-brain barrier. But it's very hard to measure small changes in oxygen concentration at the surface of a membrane experimentally.

So Tajkhorshid's team pitched in with molecular dynamics simulations. Aquaporins occur in groups of four (tetramers), with four pores that conduct water (one through each aquaporin molecule) and one central pore where the molecules meet. The latter, until now, had no known function. When simulated using two complementary methods—explicit sampling with full gas permeation and implicit ligand sampling—the team found both oxygen and carbon dioxide were exchanged through that central pore. Carbon dioxide was also transmitted through the four water pores, while oxygen passed through those pores only rarely. The research also found, however, that a plain lipid bilayer conducts



*Simulations of the aquaporin tetramer found that carbon dioxide and oxygen are exchanged through the central pore—a site of previously unknown function. Image courtesy of Emad Tajkhorshid, a faculty associate of the NIH Resource for Macromolecular Modeling and Bioinformatics, and his UIUC colleagues Klaus Schulten, Yi Wang, and Jordi Cohen.*

**"It's now well established that [aquaporins] can conduct gas molecules," says Emad Tajkhorshid. "Whether or not it's important in the human body, that's the controversial part."**

though, researchers realized that some cells need to control water permeability, and other cells have lipid bilayers that aren't very permeable to water. Aquaporins, it turned out, carry water in and out in a controllable fashion. "I think the same might be true for gas permeability," says Tajkhorshid. "Gas permeability of a lipid bilayer is like an open free highway where everything can go through. With a protein, you can have a gating mechanism and some regulation."

One of Tajkhorshid's collaborators, **Walter Boron, MD, PhD**, professor of cellular and molecular physiology at Yale University, has been working on gas

two and a half times as much gas as one embedded with aquaporin tetramers. "The question is whether this pathway is significant and makes any difference in terms of total permeability of the membrane," says Tajkhorshid.

The researchers hypothesize that, as with water permeability, aquaporins may be physiologically relevant to gas exchange when cells have dense, rigid lipid bilayers or when aquaporins occupy a major fraction of the membrane.

Tajkhorshid plans to introduce point mutations inside the central pore and manipulate the behavior of a gating loop to see how that changes the conducting

properties of the central pore. Meanwhile, Boron's group is looking for a system in which gas conduction through aquaporins is a major pathway. Says Tajkhorshid: "Even if it's 30 percent of total gas permeability, it becomes physiologically relevant because then you can control it."

According to **Nazih Nakhoul, PhD**, research associate professor in biochemistry at Tulane University, "This idea of gas transport through membrane proteins is really gaining support. It's interesting to see molecular dynamics simulations confirm some of the earliest findings."

—By **Katharine Miller**

## Parkinson's Culprit Modeled

Under a microscope, the curious protein clumps that dot the brains of Parkinson's patients stick out like the culprits they are. But no one has yet caught the protein—alpha-synuclein—in the act of causing disease. Now, investigators report in an April 2007 issue of *FEBS Journal* that they're getting closer: they've modeled alpha-synuclein's early aggregation and offered a detailed mechanism for its participation in neuron death.

"This is not just the first computational model of alpha-synuclein," says **Igor Tsigelny, PhD**, an author of the paper and a computational biologist at the San Diego Supercomputer Center. "Up to now, there was no molecular concept of the aggregation going on."

In the brain cells of Parkinson's patients, alpha-synuclein first starts to cluster as a proto-fibril. It then forms fibril chains, and finally ends up in the dense clumps of fibrils called Lewy bodies. Some researchers have suggested in the past few years that alpha-synuclein knocks off neurons right at the beginning of aggregation, long before it can be detected as a Lewy body. Biochemical and structural evidence hints that when a few alpha-synuclein molecules first self-assemble into proto-fibrils, they can form pore-like ring structures. These may interact with the cell membrane and allow ions to enter the cell. The entrance of ions such as  $\text{Ca}^{2+}$  could lead to neuron death.

The computer model created by Tsigelny and his colleagues at the University of California, San Diego, supports this theory, providing detailed dynamics of alpha-synuclein hexamers and pentamers and their interaction with the cell membrane. What's more, the model shows that another synuclein in the cell—beta-synuclein—blocks alpha-synuclein's ring-making, suggesting at least one avenue for future inhibitory drug development.

Modeling such a complex aggregation wasn't simple. Alpha-synuclein is a large protein (140 amino acids), and to model

its hexamer interacting with the cell membrane required juggling around a million atoms, Tsigelny says.

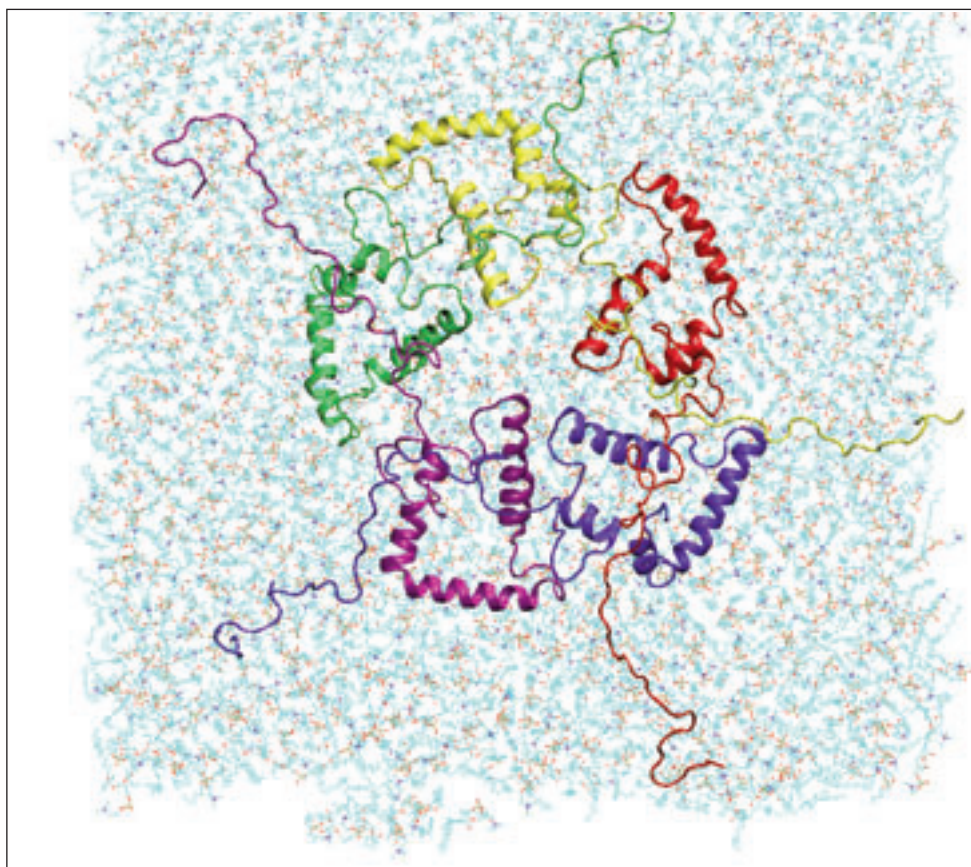
Yet more than the size of alpha-synuclein, what made it difficult to model was its lack of structure. Alpha-synuclein is an intrinsically unstructured protein—one without a distinct three-dimensional shape. Most proteins consistently fold into a favored shape to do their jobs, a form that can be crystallized, imaged, and pored over. But unstructured proteins flop this way and that, even while performing their specific tasks, making them very difficult to pin down and study.

"We were not scared by an unstable protein," Tsigelny states. And he and his coworkers developed an unusual "all-dynamic" approach to modeling the protein. None of the conformations are final—they are all considered inter-

mediate and each may last only as long as half of a nanosecond. Nevertheless, Tsigelny says, even such fleeting intermediates may aggregate. The pore-like aggregates, they found, are far more stable than single molecules of alpha-synuclein.

Having this model "is one step forward," says **Hilal Lashuel, PhD**, professor at the Swiss Federal Institute of Technology in Lausanne, Switzerland. The UCSD model provides a structural basis for testing the hypothesis that alpha-synuclein forms toxic pores, he adds. But Lashuel also cautions that only biochemical and in vivo studies can prove whether alpha-synuclein pokes holes in neurons. "Isolating the toxic species is really the most difficult question we are dealing with. You have to catch it in the act."

—By **Louisa Dalton**



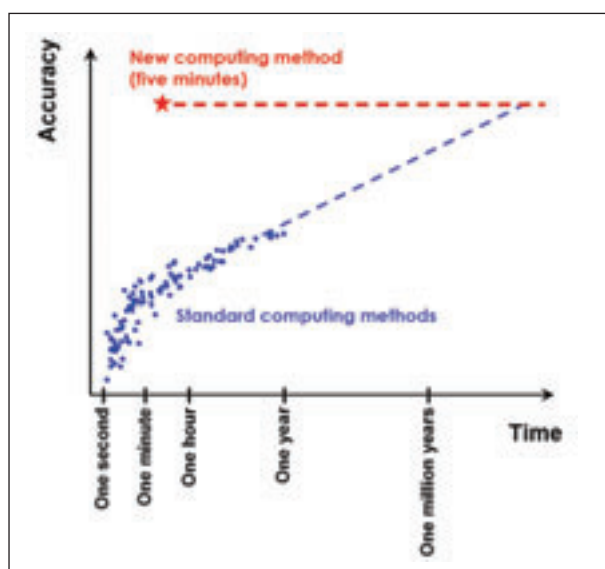
*Alpha-synuclein poses as a pentamer, pore-like, on the surface of a cell membrane. Courtesy of Igor Tsigelny*

## Clustering Without Limits

Starting in preschool we all learn how to get organized. Typically, we start with pre-determined categories (dolls, trains, blocks); pre-set ideas about what belongs in each category (Barbie: doll; Thomas the Tank Engine: train) and a fixed number of bins to put things in.

But what if you started with none of those initial limitations? Could you still group the toys? It turns out that, in a computer, such sorting is not only possible, but extremely efficient. Using a novel algorithm called affinity propagation, researchers at the University of Toronto found that they can not only cluster lots of different kinds of data appropriately, but do it better and faster than other methods. The work was published in the February 16 issue of *Science*.

“Almost all existing techniques work on a hypothesis refinement basis: they start off with a set of assumed groups and iteratively refine them,” says **Brendan Frey, PhD**, associate professor of electrical and computer engineering at the University of Toronto, co-author of the paper. “To our knowledge, ours is the first algorithm to consider all possible groupings at once.”



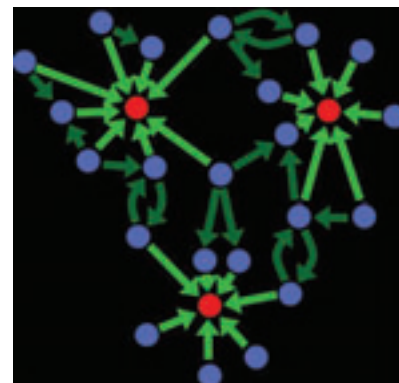
*If asked to cluster facial images, a standard clustering method (k-means clustering) would take up to a million years on a single computer to achieve the accuracy achieved by affinity propagation after five minutes.*

“Part of the attraction of the [affinity propagation] algorithm is that, although it was complicated to derive, it’s quite simple to implement and to get an intuitive feel for it,” says **Brendan Frey**.

The task sounds mind-boggling: There are a huge number of possible groupings. But affinity propagation handles that problem by sending messages between data points—pair-wise—so as to maximize

the net similarity in each group. “Each message encapsulates or summarizes a whole distribution of possible groupings for one of the data points,” says **Delbert Dueck, a PhD** candidate in Frey’s lab. “No one has done that before.”

Affinity propagation is based on an algorithm called belief propagation, which has been around in various incarnations for many years. But, say the authors, it’s an approach that has never been applied to clustering. “Certainly not to generic clustering of any type of data,”



*Frey and Dueck use affinity propagation to cluster data around “exemplars”—data points that best represent their compatriots. In this graphic, after starting with an equal chance of serving as an exemplar, candidates for that job have already emerged (red dots). Each data point sends messages to each candidate exemplar conveying how well it represents the blue point compared to other candidate exemplars. And candidate exemplars send messages conveying their availability to serve as an exemplar for particular data points.*

says Dueck. Indeed the algorithm is so generic that Frey and Dueck used it to analyze gene expression data, facial images, and airline routes, while other researchers have found applications in basketball statistics, the stock market and computer vision. And many tasks in computational biology require a computer to organize the data before using it to make predictions.

“Part of the attraction of the algorithm is that, although it was complicated to derive, it’s quite simple to implement and to get an intuitive feel for it,” says Frey. There are basically only two equations to it. “Sometimes we’ll give a talk and get emails from people who’ve implemented it the day after,” he says.

When the researchers looked at how well the algorithm performed compared to other clustering methods they found it remarkably efficient. “A problem our algorithm could solve in about five minutes on one computer would take other methods up to one million years to solve on that same computer,” says Frey.

**Tim Hughes, PhD**, of the Center for Cellular and Biomolecular Research at the University of Toronto, is considering using affinity propagation in his research. "It seems like it would do best when things really do form independent groups, and when the data are fairly sparse, so most of the correlation matrix can be dropped in early cycles," he says. "I think it will work well with exon-profiling data or genome-tiling data, where there is also a constraint that the groups have to correspond to regions near each other on the chromosome."

—By Katharine Miller

## Computer Vision that Mimics Human Vision

Our brains can recognize most of the things we pass on an evening stroll: Cars, buildings, trees, and people all register even at a great distance or from an odd angle. Now, a new computer vision program can do the same thing. It successfully rivals the human ability to rapidly recognize objects in a complex picture because it mimics how information flows during the initial stages of visual perception.

"We've built a model to be as close as possible to what is known about the human visual system," explains **Thomas Serre, PhD**, a postdoctoral associate in the Center for Biological and Computational learning at MIT and lead author of two papers recently pub-

lished out of the lab run by **Tomaso Poggio, PhD**, at MIT's McGovern Institute for Brain Research.

For decades, scientists have struggled to create computer programs that can recognize visual objects as well as humans can. Some computer systems excel at recognizing one particular object, but none are anywhere close to recognizing the wide range of objects observed by the human brain. Visual recognition is complicated by two conflicting goals: a program must be specific enough to discriminate between different objects, such as a person or a car, yet flexible enough to recognize the same type of object in different sizes, poses, and lighting.

To achieve these goals, Serre and colleagues used data recorded from real neurons in the visual system to program two fundamentally different kinds of virtual neurons called S (simple) and C (complex) units. S units recognize specific features of an image; C units monitor a range of S units in one area and allow for variation in position and size.

The researchers were surprised to find that a simple system, consisting of four alternating layers of S and C units,

was able to classify pictures of a busy street scene as well as other leading mathematics-based computer vision systems, as described in the March 2007 issue of *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Serre's team then built a more complex system, consisting of many S and C layers designed to closely match the flow of information in a human brain during the first 100-200 milliseconds of perception. This enhanced system performed as well as humans on a rapid object recognition task: distinguishing animals from non-animals when images were flashed in front of humans and computers. The work appeared in the April 2007 issue of the *Proceedings of*

*the National Academy of Sciences*. The computer system even made errors similar to the errors made by humans, suggesting that the model recapitulates the early processes of the human visual system.

The model will be used as a tool by neuroscientists to better understand the human visual system, and also has practical applications for surveillance, driving assistance, and autonomous robotics. According to Poggio, the team's next

"We've built a model to be as close as possible to what is known about the human visual system," says Thomas Serre.

*When presented with a real-world street scene (left), Serre's computer vision system successfully recognized pedestrians, cars, buildings, trees, sky, and the street (right). Although not pictured, the model also successfully identified bicycles. Note the error in this example: the model mistakenly classified a street sign as a pedestrian. Graphic courtesy of Stanley Bileschi, PhD, McGovern Institute for Brain Research at MIT.*



goal is to extend the model to include the “back projections” from other parts of the brain that allow feedback processing of visual information after 200 milliseconds.

“This is the first demonstration that a purely bottom up approach to visual object recognition, inspired by recordings from the neurons in the brain, is effective as a practical computer vision system,” says **Terry Sejnowski, PhD**, head of the Computational Neurobiology Lab at the Salk Institute. “There is much more work to do, both to improve its performance, and also to use it to better understand how our own visual system works.”

—By **Matthew Busse, PhD**

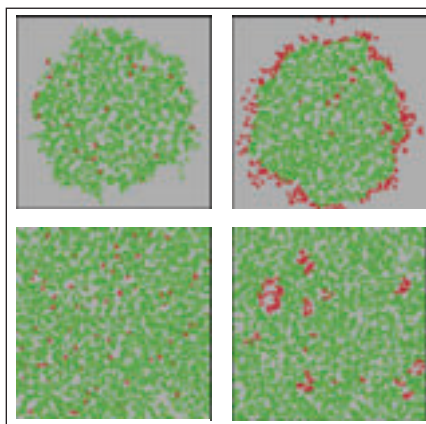
## Nature Versus Nurture *In Silico*

Every generation, a few nonconformists crop up in tissue cultures of genetically identical cells. The question is: are the wayward simply born that way, or did something in the environment affect them? “You have these two possibilities—*intrinsic* or *extrinsic*, *nature* or *nurture*,” says **Andras Paldi, PhD**, a biologist at Genethon in France.

Now, Paldi and his colleagues have modeled such cultured cells to determine whether *extrinsic* or *intrinsic* influences play a key role in the spontaneous emergence of phenotypic variation. It turns out that for spatial patterns beyond randomness to arise, there has to be some effect of sensing neighboring cells—i.e., *extrinsic* factors must play a role. And the *extrinsic* model resembles results seen in real cells. The work appears in April in *PLoS One*.

Paldi’s work was motivated in part by the open question among stem cell biologists of what triggers a stem cell to differentiate. Why, in the same warm spot, getting the same rich media, do some cells differentiate and others stay stem cells? It is commonly assumed that this is because the decision to differentiate is *intrinsic*—that is, purely random.

To test that assumption, Paldi’s group started by designing two simple, multi-



*Agent-based computer models predict the pattern (left) produced when genetically identical cells have an inherent probability of changing (from green to red and vice versa), and the pattern (right) produced when cells are triggered to change by an extrinsic factor, such as cell density. Top images represent exponential growth; bottom are at equilibrium. Courtesy of Andras Paldi.*

agent based models of a tissue culture plate. In each model, all cells act independently and can switch between two cell types: A or B. In the “*extrinsic*” model, A cells turn into B cells when it gets crowded, and back to A cells when they have more space. In the “*intrinsic*” model, each cell has fixed probabilities of switching from A to B and back again.

When the scientists ran the models, they found each produces a stable, heterogeneous population, yet they differ in the cell patterns. The *intrinsic* model predicts lone A cells distributed evenly throughout a largely B population. *Extrinsic* predicts that the A cells will cluster. The result held even though the cells were allowed to migrate.

This pattern difference allowed the researchers to compare their computational simulation with real cells. Using a muscle cell line that can switch between two distinct phenotypes, a stem-cell like progenitor state and a differentiated state, they found that the cell pattern mostly resembles that of the *extrinsic* model. Many of the rare, stem-cell like cells cluster; a few are solitary.

What’s important here, Paldi says, is that they find environment playing a role—a significant one. In the case of stem (progenitor) cells, it means neighbor cells

can affect the differentiation process. “The stem cell nature is not an intrinsic property of the cell,” he says. “It is a property of the whole cell population.” Paldi further believes the work supports the effort to find a way of converting adult, differentiated cells into stem cells (and avoid the need for harvesting embryonic stem cells)—a possibility that has not just

scientific, but social and political implications as well.

**Christa Muller-Sieburg, PhD**, however, disputes that scientific

Why, in the same warm spot, getting the same rich media, do some cells differentiate and others stay stem cells?

conclusion. “The idea that mature cells can turn into stem cells is very attractive to many modelers but has little support through experimental data,” says the professor at the Sidney Kimmel Cancer Center.

**Sui Huang, MD, PhD**, at Children’s Hospital Boston, would have liked to see Paldi’s group perturb the cell line or the culture to confirm their model. But both he and Muller-Sieburg believe the study addressed an important question, that of heterogeneity of a genetically identical population of cells. And, says Huang, it certainly “contributes to the discussion in the community.”

—By **Louisa Dalton**

## Simulating Populations with Complex Diseases

Diabetes, breast cancer, multiple sclerosis, Alzheimer's disease. All are associated with several genes' alleles interacting in complex ways with one another and the environment. Now, using a computationally intensive method known as forward-time simulation of human populations, researchers are hoping to gain a better understanding of how such complex diseases become established.

"In a real population you just see people with the disease," says **Marek Kimmel, PhD**, professor of statistics at Rice University and co-author of the work. "You don't see who in the population has the disease genes because people carrying these genes do not necessarily become diseased." But in the model population, he says, "you see both." And the researchers' approach allows them to simulate a very complicated scenario—including changes in types of selection pressure.

"This lets us evaluate how well statistical genetics tests determine what genes are responsible for the symptoms of a disease and how frequently those genes appear in the population." That's a non-trivial exercise, he says, because it has been impossible, until now, to compare the many existing gene-mapping methods head-to-head. The work was published in *PLoS Genetics* in March 2007.

Before now, the most commonly used approach to simulating diseases in human populations—called the "coalescent" method—worked by coalescing backward in time to a most-recent common ancestor. But it's extremely difficult to take selection into account using the coalescent method, says co-author **Bo Peng, PhD**, a postdoctoral fellow at the University of Texas MD Anderson Cancer Center. Moreover, that approach gets too complicated if more than one disease gene is involved. So Peng and his colleagues turned to forward-time simulation, an approach that's been around for about one hundred years.

But that technique is not without its problems. When a population evolves forward in time, there are simply too many possible outcomes. Most notably, when you introduce a disease allele, it can rapidly be eliminated and replaced with new alleles. So Peng came up with a trick: He pre-sets desired disease allele frequencies in

based on Python. The software is freely available at <http://simupop.sourceforge.net>, under a GPL license.

When Peng and his colleagues used their method to compare several gene mapping techniques they found that certain methods worked better for loci that were located distantly from one another; and other methods were more effective when loci were close together. Overall, though, says Kimmel, "We're mildly pessimistic" about current gene mapping approaches. "When the number of loci involved in complex disease is greater than two, the methods rapidly lose their power." Until recently, gene mapping for complex diseases has been disappointing, he says. Loci identified in such efforts have later turned out to be statistical artifacts. "Our modeling could figure out if this is inevitable," he says—and help guide people toward more effective approaches.

**David Balding, PhD**, a professor of statistical genetics at Imperial College in London, does similar work using forward-time simulations of large genomic regions. He has become pessimistic

about the method's usefulness for understanding complex diseases because no one really knows what kind of selection is going on. Nevertheless, he says, this work can be useful for studying selection itself. "People tend to look at selection one allele at a time," he says, "But forward-time simulation lets us do it with complex interactions."

—By *Katharine Miller* □



"In a real population, you just see people with the disease," says Marek Kimmel. "You don't see who in the population has the disease genes..."

the current generation, extrapolates them backward, and starts the simulation from there. As Kimmel puts it, "We are restricting potential variability in one aspect of the present in order to produce a simulation that resembles something close to the actual variability that exists now."

The simulation uses a scripting language called simuPOP, a general-purpose forward-time simulation environment