at McGill created the Protein Subcellular Localization Tool 2, or PSLT2.

PSLT2 is composed of three modules that predict where a protein will go. The *motif* module makes predictions based on the presence of particular sequences of amino acids that suggest a protein's function—a good indication of where it belongs in the cell. The *targeting* module relies on sequences that act like a known zip code, indicating where the protein should end up—such as mitochondrial targeting peptides and transmembrane domains. And the *interaction* module concerns itself with the protein's likely comrades—the other proteins it associates with when doing its task. If protein A always interacts with protein B, and B has a known location in the cell, then A must be active in that vicinity as well.

Each module can individually predict the localization of a protein using Bayesian methods. The combination of the three modules improves the prediction when proteins lack motif and interaction data or traffic through multiple compartments.

For the entire yeast genome, the new tool predicts in which of nine compartments a protein is located with at least 72 percent accuracy. These compartments are mostly organelles but also include the cytosol and cell membrane. PSLT2 also predicts proteins' sub-compartmentalization—whether they are inside the compartment, in its membrane, or associated with its surface. The model places the proteins into 18 sub-compartments correctly 83 percent of the time.

The ability to determine sub-compartments is new to this model. "When we use classical techniques for finding the localization of a protein [in, for example, the endoplasmic reticulum (ER)], we can't use them to tell if a protein is in the ER membrane, in the cytosol, or on the periphery," Hallett says. "We need a computational method to pin down where the protein is in the organelle."

The computational model's predictions compared well with databases from two high-throughput laboratory experiments, but they didn't always agree; Hallett and colleagues suggest that the model and two databases should be used in parallel as checks on each other.

According to **Mark Gerstein, PhD**, an associate professor of biomedical informatics at Yale, the paper goes beyond what has been done before. "In particular," he says, "people have done a lot of analysis using protein subcellular localization to predict protein-protein interactions. This work turns that around to good effect."

—Linley Erin Hall

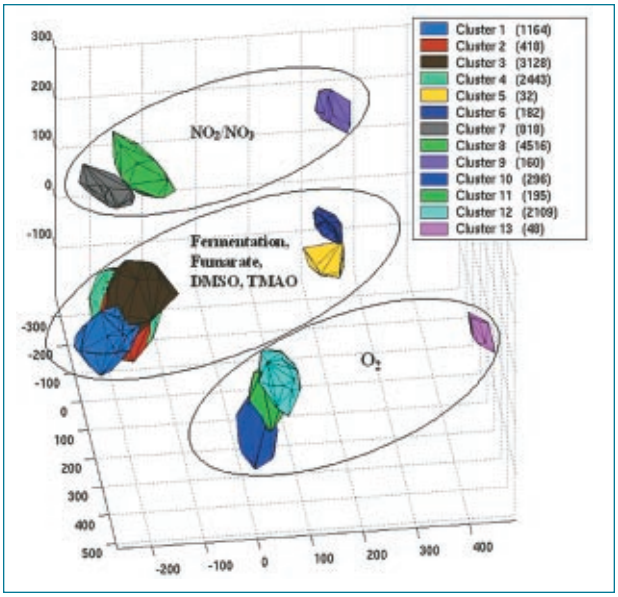
The Six Faces of *E. Coli*

Biologists' favorite bacterium grows almost anywhere—from the human gut to the pounding surf. But *E. coli*'s remarkable adaptability apparently stems from being predictable rather than accommodating. In a recent computer simulation, thousands of environments provoked only a handful of shifts in the microbe's physiology. The work was published in *Proceedings of the National Academy of Sciences* in December 2005.

"A network comprised of thousands of molecules, in response to a myriad of inputs, takes on relatively few overall responses," says Bernhard Palsson.

"A network comprised of thousands of molecules, in response to a myriad of inputs, takes on relatively few overall responses," says senior author **Bernhard Palsson, PhD**, professor of bioengineering at the University of California, San Diego. The systems biology study of *E. coli* metabolism might help scientists understand how cells function and adapt to different environments.

To simulate *E. coli*'s environment, Palsson and his colleagues first drew up a list of nutrients that could meet the



A map of possible states for* E. coli *metabolism. The axes represent the Hamming distance—a mathematical comparison of the output from different simulations. The closer two points, the more similar are those results. The ovals show the terminal electron acceptor for different types of respiration: NO2/NO3 for anaerobic; O2 for aerobic; and fumarate or DMSO or TMAO for anaerobic fermentation. Courtesy of Christian Barrett, UCSD.

microbe's needs—carbon, nitrogen, sulfur, etc. From this, they generated an exhaustive list of media that could support its growth. Then they wrote mathematical algorithms—based on 1,010 genes—for each step in *E. coli*'s well-understood metabolic process.

Combining the different inputs with these mathematical algorithms, they "grew" *E. coli* in 108,728 hypothetical simulated Petri dishes, of which 15,580 nurtured bacteria growth. Each of these *in silico* cultures produced a simulated gene expression profile, which researchers visualized in 3-D using a statistical tool known as principal component analysis.

The 3-D space was mostly empty: physiological outcomes appeared as thirteen clusters organized into six groups. Cells based their metabolic decisions largely on two factors: the availability of glucose as an energy source; and the identity of the terminal electron receptor(s)—the molecules that dictate whether the cell carries out aerobic respiration, anaerobic respiration, or fermentation. These responses are reasonably similar to laboratory experiments, Palsson says, but he was surprised by the limited scope of all possible responses.

The researchers chose to study *E. coli* because it has the best-characterized DNA on the planet, but the technique could apply to other organisms. For example, ecologists might map microbial communities in an ounce of soil to see how hundreds of microbes' metabolisms interact. And engineers might use the technique to design whole bacterial ecosystems for useful tasks, such as eating toxic waste.

According to **Costas Maranas, PhD**, professor of chemical engineering at Pennsylvania State University, the study will help "to flesh out dominant organizing principles for complex systems." In addition, he says, "One could look at whether the dominant behaviors that they have elucidated will hold under different kinds of perturbations, [such as] genetic perturbations."

But the larger question of how all the complexity in the *E. coli* genome results in only a few metabolic activities, Palsson says, "is something that we still have to study, and understand."

—Hannah Hickey

A Powerful Model of Relaxation

When a heart beats, millions of muscle cells contract in unison to pump blood to the body; then they relax, allowing the heart to refill. Though scientists have carefully characterized the mechanisms that govern contraction, they are less certain about the dynamics of relaxation. But a new mathematical model of calcium ion concentration in cardiac muscle—published in March 2006 in *Biophysical Journal*—has resolved at least one controversy.

"There's been a lot of emphasis on contraction, because it's the first thing you measure experimentally," says **Nicolas Smith, PhD**, senior lecturer in the Bioengineering Institute and Department of Engineering Science at the University of Auckland in New Zealand. "But it's just as important that the heart relaxes. We wanted to be very clear that we were characterizing the relaxation properties just as well as the contraction properties in this model."

Here's what a heartbeat looks like from within a cell: An electric signal spurs the release of calcium ions, which bind to motor proteins and activate contraction; then, the calcium ions are pumped away, and the cell relaxes. The rise in calcium clearly governs contraction, but scientists still debate the key trigger for relaxation. Some have suggested that relaxation depends more heavily on mechanical factors (when the cell reaches a critical length or tension), rather than on biochemical factors (a drop in calcium levels).

Smith and his colleagues combed the literature and found decades worth of experimental data (from humans, chickens, rats, mice, ferrets, rabbits, cows and cats) on calcium concentration and binding, as well as cell velocity, length, and tension during a heart beat. They combined these diverse data into a series of mathematical equations that simulate cellular contraction and relaxation. Then they simulated the tension changes in the beat of a heart cell—and found that their predictions closely approximated tension changes measured in the lab (data that had not been used to build the model).

Their simulation also showed that cell relaxation depends predominantly on the drop in calcium levels. "In some ways this is less exciting than more esoteric ideas of length dependence and tension dependence, because it's actually quite simple. But it does clear up a lot of the debate," Smith says.

> "Often models get published that are very limited in scope, because authors are only interested in fitting their particular dataset. But these authors tried to match a diverse set of data," says John Jeremy Rice.

Smith and his colleagues are extending their model to study life-threatening biochemical changes that arise during ischemic heart disease (where oxygen is not getting to the heart). In ischemia, heart tissue becomes acidic, which wreaks havoc on calcium signaling. An unchecked overload of calcium will cause the heart to perpetually contract—a deadly deficiency of relaxation.

The model is limited to the cellular level, Smith notes, as modeling at the molecular or atomic scale would take too much computing power. But, he adds, "Because of the way we for-