By Julie J. Rehmeyer, PhD

Simulated Metabolism

A First Step Toward Simulated Cells

If biologists really understood the functioning of the genome, they could in principle recreate it *in silico*. Instead of a choreographed swirl of molecules inside a living cell, electrons inside a computer would map out all those cell processes: DNA zipping and unzipping, transporters tugging molecules across cell membranes, enzymes latching on and letting go. >

It's an entrancing dream. For one thing, the process of developing such a simulated cell would help biologists find processes they were missing and show them when they finally understood the microscopic universe inside a cell. And there would be enormous practical value. Drugs could be tested on the model cell long before a single capsule touched a tongue. Your doctor could tell you which diet would be best for you, given your own personal genome. The full impact of genetic diseases could be worked out down to the cellular level.

In 2007, building on his success with *E. coli*, Palsson released a similar model of human metabolism called Recon 1. "It's a landmark study," says **Eytan Ruppin, MD, PhD**, a professor of computer science and medicine at Tel Aviv University in Israel. "I'm really excited about this, but I remind myself that we are just at the beginning."

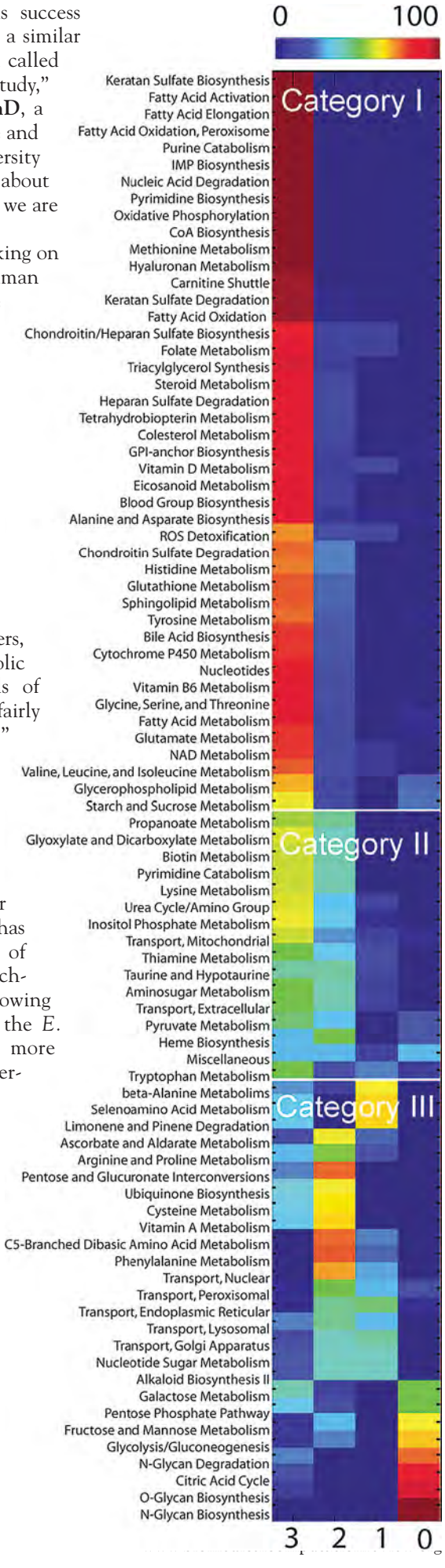
Ruppin's laboratory is working on projects using Palsson's human metabolism model. "There are many things that are really tempting to study," he says. "Degenerative disorders, cancer, a variety of metabolic genetic disorders. Hundreds of them could be studied in a fairly straightforward manner." Many other groups have also launched projects using the human model, but it's too soon for its potential to be clear.

A fully simulated cell hasn't happened yet, but one system—metabolism—has proven to be one of the simpler systems to tackle.

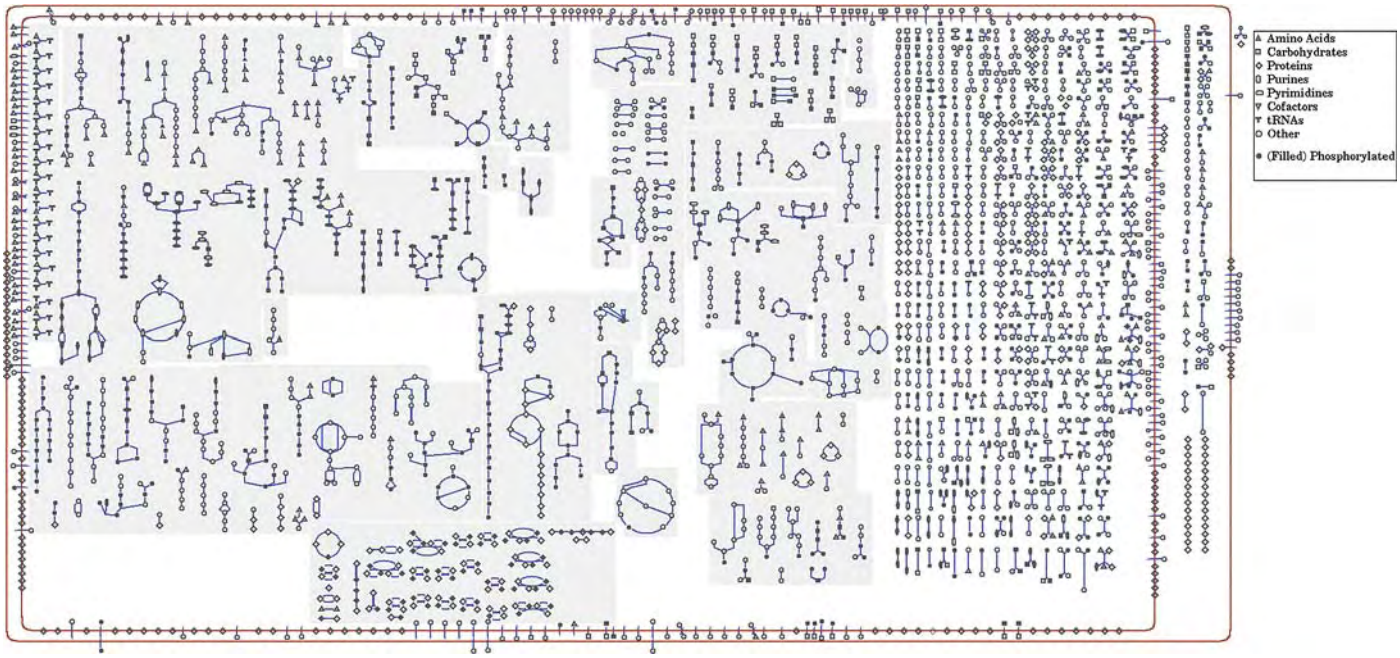
Bad news, though: "We're not even close to that," says **Joel Stiles, MD, PhD**, a computational physiologist at Carnegie Mellon University and principal co-author of MCell, a simulator of cell microphysiology. Still, it's not *just* a dream. Researchers have made great progress decoding the functioning of the genome in particular areas. Metabolism, in particular, has proven to be one of the simpler systems to tackle.

Multiple researchers have worked since the 1950s to develop an increasingly thorough and detailed understanding of metabolism in the bacterium *Escherichia coli (E. coli)*. And over the last fifteen years, **Bernhard Palsson, PhD**, a professor of bioengineering, and his colleagues at the University of California, San Diego have developed and continually improved a very successful *in silico* model incorporating all of that detail.

On the other hand, the *E. coli* model, developed over many years, has matured and has spawned whole batches of research that are already reaching fruition. In addition to showing how the genome functions, the *E. coli* model has made far more sophisticated genetic engineering possible. Researchers are turning *E. coli* into miniature factories that pump out food additives for



Modeling human metabolism revealed which aspects are well-understood (blue) and which need more research (red). All of the reactions in keratan sulfate biosynthesis (top line), for example, have direct biochemical or genetic evidence, while those involved in n-glycan biosynthesis (bottom line) haven't been evaluated at all. Inositol phosphate metabolism is in the middle, with many reactions having direct biochemical or genetic evidence, some supported by physiological data or evidence from a nonhuman mammalian cell, and a few with only modeling evidence. Reprinted from Duarte, NC, et al., Global reconstruction of the human metabolic network based on genomic and bibliomic data, **Proceedings of the National Academy of Sciences,** ***104: 1777 (Feb 6 2008).***



Each dot in this graphic is a metabolite. In the online version of this graphic (http://biocyc.org/ECOLI/new-image), clicking on a dot identifies the metabolite and the pathways it's involved in. Courtesy of Peter Karp and Suzanne Paley.

manufacturing, precursors of pharmaceuticals, and even ethanol or butanol. The development of the model has also acted as a spotlight into metabolism itself, guiding lab research.

FINDING MISSING PATHWAYS

To assemble the networks for both human and *E. coli* metabolism, Palsson's group began by exhaustively combing through the existing literature, gathering a list of all known metabolic reactions and their corresponding genes. The reactions form an enormous network, with 2207 reactions in *E. coli* accounting for 1260 of its 4453 genes. The human model currently contains approximately 3300 reactions, accounting for about 1500 genes, but will undoubtedly grow significantly over time.

The metabolic networks are set up like a map, with reactions forming roads connecting the metabolites. Energy flows through this network like cars do through a city. To verify this map of the "metabolic city," the researchers began a torture test. Side by side, they immersed real *E. coli* and their virtual *E. coli* in all kinds of different media. The virtual bacterium "died" if no metabolic pathway connected the medium to all the compounds the bacterium needs to live—essentially telling the researchers "you can't get there from here" on the map. If the virtual *E. coli* died when the real one survived, the researchers knew some pathways were missing in their model. Apparently, *E. coli* was capable of performing previously unknown tricks.

The researchers then evaluated the network to identify reactions that probably needed to be added to complete their map, and which genes most plausibly could make those reactions possible. Armed with these hunches, they went back to the wet lab to perform further experiments. Using this method, various teams have identified the roles of eight genes whose functions were previously unknown. They identified the gene for one reaction that had been an "orphan" for 25 years: Although researchers had known that *E. coli* could convert succinate semialdehyde to succinate, they hadn't been able to find the gene that made it possible. After adding newly identified reactions to the *in silico* model, the researchers then repeated the testing of the virtual bacterium in an effort to find more missing links.

This same iterative process was used in creating Recon 1, the human model. Because the human model is so much newer and because human metabolism is much more complex, researchers haven't yet pursued all the hints the model has provided about unknown reactions or gene functions. But the group created a "knowledge map," showing which metabolic functions are well understood and which ones clearly have lots missing. For example, cholesterol biosynthesis occurs in the endoplas-

Using a combination of literature search, lab work, and iterative computational modeling, researchers have filled in missing links in models of metabolism for both *E. coli* and humans.

mic reticulum, but the transport mechanism of its precursor from the peroxisome is still unclear, according to the reactions currently known. That's a clear indication that a transporter is missing.

The researchers also developed tools to automate some of the testing process in *E. coli*. "Can we use these systematic procedures to improve Recon 1? That is something the whole community is very interested in doing," Palsson says.

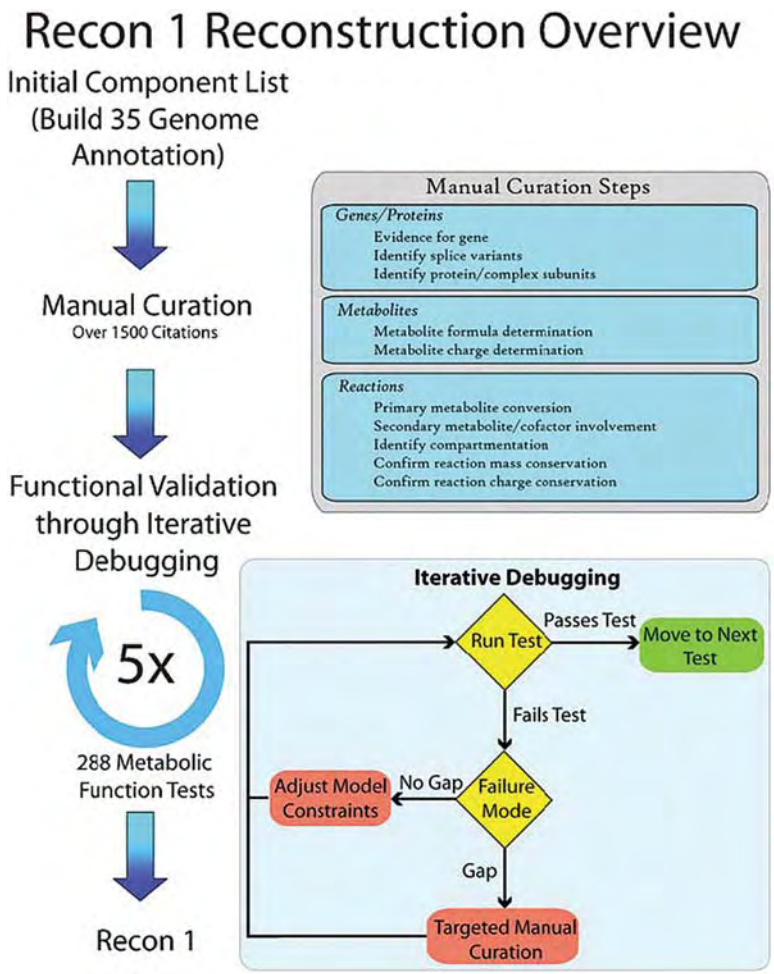
BRINGING THE NETWORK TO LIFE

Having mapped the network, the next step was to build lots of computational tools to bring the map to life. With a street map, you can figure out how to get from one spot to another, but often there are lots of different routes you might take. The same thing is true for metabolism in *E. coli*. "These models can tell you what the organism can do but not what it *will* do," says **Costas Maranas, PhD**, a professor of chemical engineering at Pennsylvania State University.

So the researchers looked for ways to deduce the pathways the bacterium is most likely to use, and to narrow the possible paths it might be taking. But they ran up against a big problem: while the reactions are known pretty well, the particular rates of the reactions aren't. Small differences in reaction rates could have a big impact on which reactions actually happen. So their map was like one that showed the layout of a city without indicating whether any particular street was a mega-highway or a dirt alley.

Masaru Tomita, PhD, a professor of bioinformatics and head of the E-Cell project at Keio University in Japan, is using high-throughput methods to identify these reaction rates and how they change in response to perturbations. This requires quantifying the rates of every different reaction in each possible circumstance—a monstrous task. Many labs have joined forces to make Tomita's project possible, and the group's work toward simulating a whole cell is ongoing.

In the meantime, and looking to simplify matters, Palsson took a different route. He created a model of how the cell functions when in a steady state. Evolution provides a big clue. "You assume *E. coli* will evolve to use the resources to grow in the most efficient way," Palsson says. "Most of the time, that's what it does." So his model calculates the pathways that most efficiently turn a particular compound into all the different compounds needed for growth. Palsson reasons that those are the ones the bacterium would most likely use.



To create the human metabolic reconstruction, the researchers first assembled a list of the components and preliminary network from the annotated human genome. They then manually reviewed more than 1500 papers to ensure that the network components and their interactions were based on direct physical evidence and reflected current knowledge. Next, they used 288 known functions of human metabolism to test the model and find missing or incorrect links. After making improvements, they repeated the tests four additional times. Reprinted from Mo, ML; Jamshidi, N and Palsson, BØ. A genome-scale, constraint-based approach to systems biology of human metabolism.* Molecular BioSystems *3: 598 (2007). Reproduced by permission of The Royal Society of Chemistry.

"These models can tell you what the organism can do but not what it *will* do," says Costas Maranas.

Experimental evidence supports this approach when the cell's circumstances aren't changing, but researchers have developed other optimization methods that the bacterium may use under dif-

ferent conditions. For example, **George Church, PhD**, a professor of genetics at Harvard Medical School, proposed that after a knockout, cells choose the metabolic pathways that will most quickly return the cell to a steady state.

Unfortunately, Palsson's observation doesn't apply so neatly to human cells, since they don't normally grow boundlessly. Nevertheless, the team developed a description of 288 known metabolic functions in humans—such as the production of the hormone melatonin—to establish restrictions for how metabolites are likely to be processed through the map. Because those restrictions only narrow the possibilities, rather than providing a unique pathway for the metabolism of a particular compound, more work remains to be done.

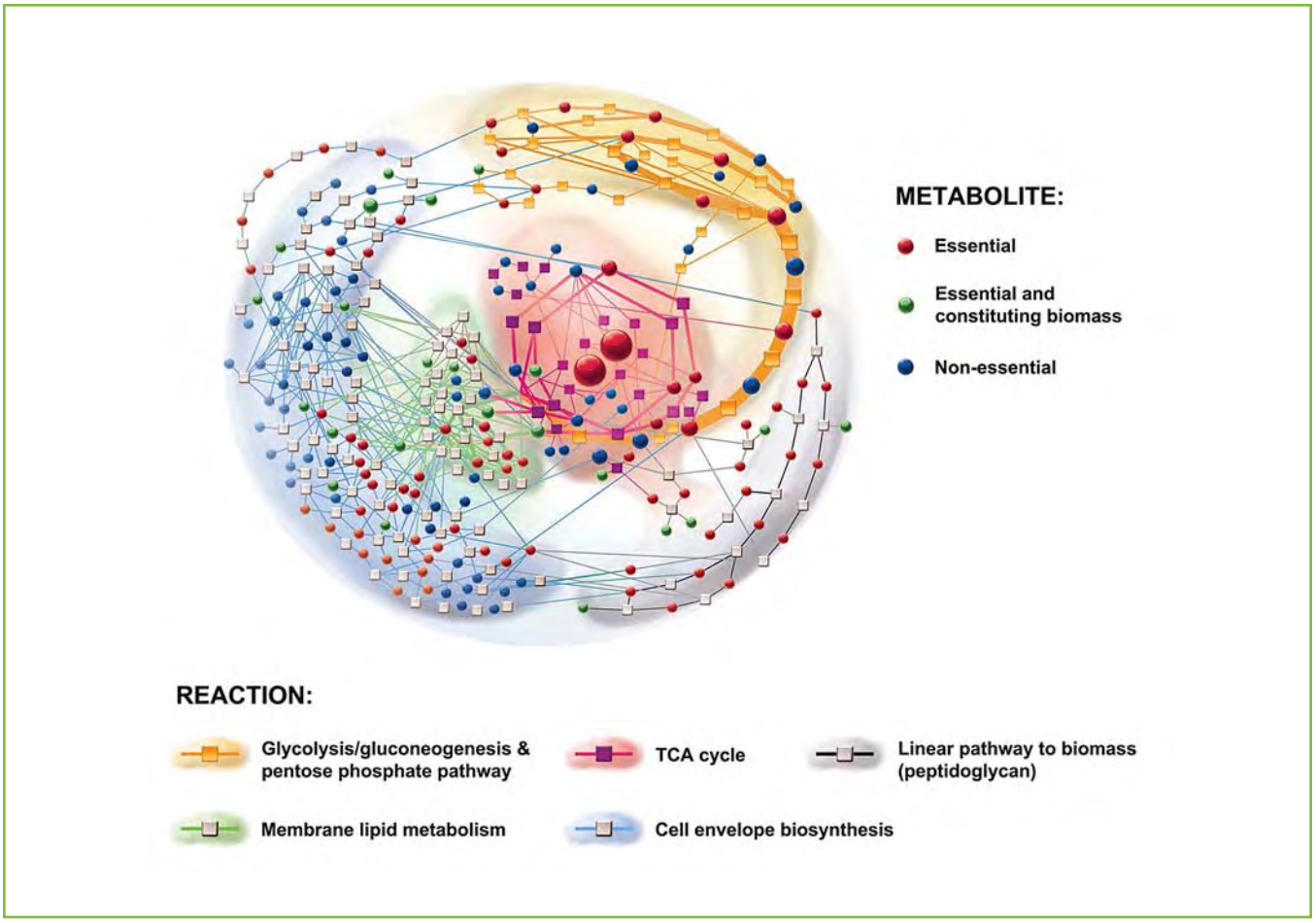
THE MODEL APPLIED: FROM BIOFUELS TO DISEASE

Palsson's group has made their models freely available, and the *E. coli* model alone has been used in more than 100 research papers by groups around the world. The most obvious way to understand the functioning of the genome is to perturb it and see what happens. This has been done systematically in the lab, but with Palsson's model, researchers can knock out a gene with a keystroke, rather than spending hours or days or weeks creating a genetically modified bacterium. The model also makes it possible to systematically study *E. coli* with multiple gene deletions, or to predict the impact of adding a gene.

"So far, the effect of the model on other peoples' research has been subtle," says **Adam Feist**, a graduate student in Palsson's lab, "but going forward, it's going to be huge. People have emailed me from around the world. They are really starting to catch on."

A big reason for the excitement is that these capabilities have made genetic engineering of *E. coli* dramatically more efficient. **Stephen Fong, PhD**, an assistant professor of chemical

...With Palsson's model, researchers can knock out a gene with a keystroke, rather than spending hours or days or weeks creating a genetically modified bacterium.



A map of* E. coli *metabolism, based on Palsson's models. Metabolites marked in red or green are essential, those in green constitute biomass, and those in blue aren't essential. Reprinted from Kim, PJ, et al., Metabolite essentiality elucidates robustness of* Escherichia coli *metabolism.* Proceedings of the National Academy of Sciences *104: 13638-13642 (Aug 21 2007).

and life science engineering at Virginia Commonwealth University, says that before these models existed, his bioengineering work was vastly more time-consuming. He would make one modification based on his best guess of what would work, and then test the result. Based on what he found out, he'd make another, and another and another. "Each cycle takes several months," he says. "Simulations literally take less than a second to do." That has revolutionized the process, he says. "You have a way of screening through all the things that seem like they have the highest probability for success before you do any experiments at all."

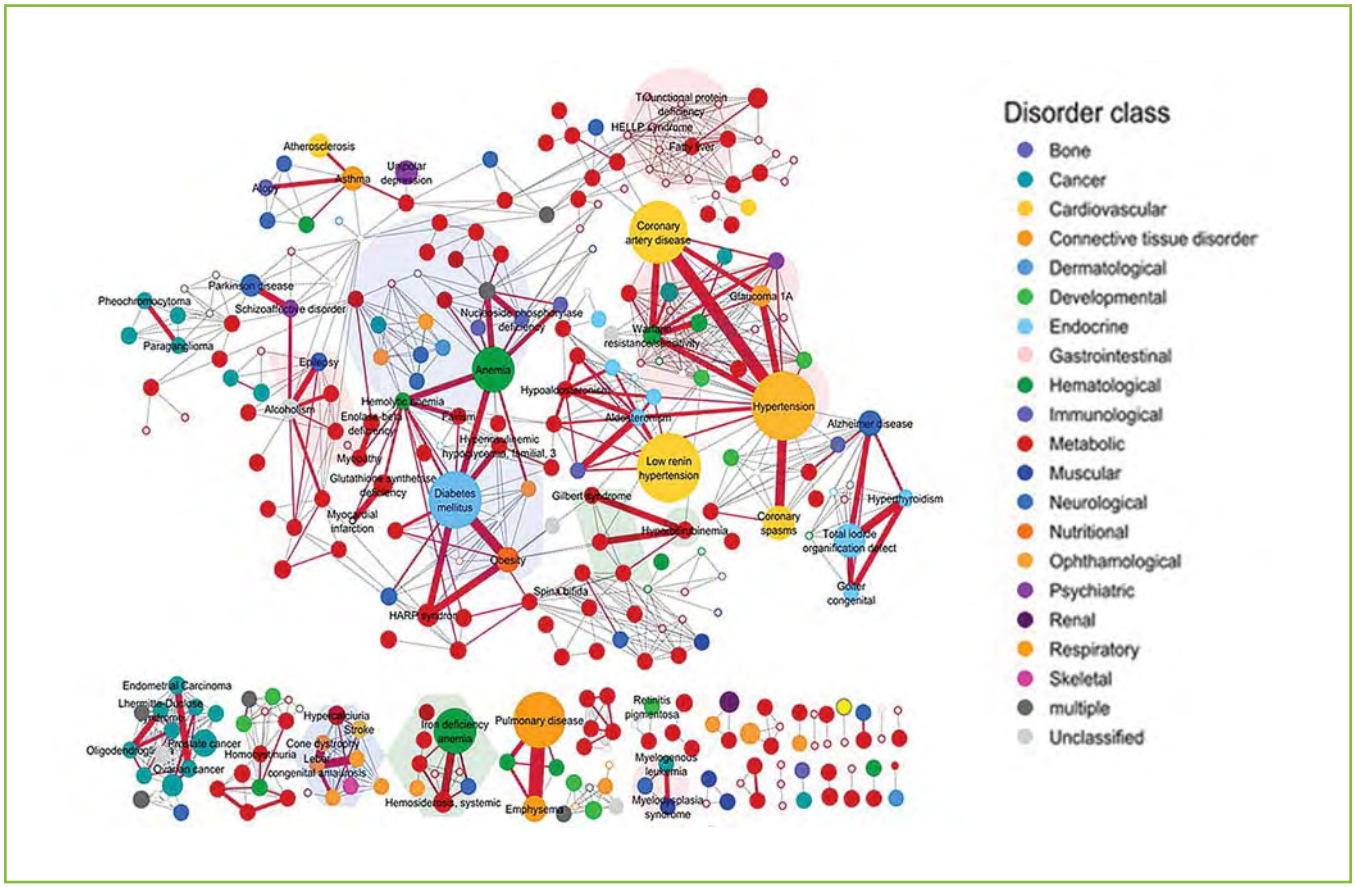
In 2005, Fong pioneered the use of Palsson's model in genetic engineering, creating *E. coli* that produce lactic acid, which is used as a food additive and to create scaffolding for tissue implants. His strategy was to essentially cripple the bacterium by eliminating genes so that its metabolism would be less efficient at turning its food into the compounds it needs to grow. Instead, the bacterium would convert some of its food into a waste product—lactic acid. Fong used Palsson's model to identify the most promising genes to knock out to achieve this, and then he experimentally modified the bacterium to confirm the model's predictions.

Now, he and many other groups are after a more exciting quarry than lactic acid. They want to revolutionize the creation of biofuels using the same process. *E. coli* has been engineered to produce ethanol, as well as fuels like butanol or alkanes that are better substitutes for gasoline.

Church is associated with four different companies that are using these approaches to develop biofuels, and he says the only challenge is to scale up the manufacturing process to create the fuels inexpensively enough. "Inevitably as it scales up, it'll be able to beat petroleum out of the ground," he says.

Church's dreams extend beyond bio-

"Simulations literally take less than a second to do." That has revolutionized the process, Fong says. "You have a way of screening through all the things that seem like they have the highest probability for success before you do any experiments at all."



Barabasi and his colleagues analyzed a Medicare dataset to create this map showing the relationship between diseases. Diseases in the network are connected if mutated genes associated with them catalyze metabolic reactions that are closely related. Diseases that occur more frequently are depicted with larger dots, and two diseases that tend to occur in the same person are connected with a heavier line. Reprinted with modifications from Less, DS, et al., The implications of human metabolic network topology for disease comorbidity.* Proceedings of the National Academy of Sciences *105: 9880-9885 (July 22 2008).


E. coli

fuel as well. "This kind of biochemical engineering is pleasantly easy these days," he says, because of the rise of these computational methods. "It's really interesting and fun to use these cellular models to think of all the different products you can make that are currently fairly expensive."

One of his ideas is to create *E. coli* that he can feed off agricultural waste to produce non-biodegradable precursors for plastics. "If you pull carbon dioxide out of the air and make a wax or a plastic, and don't burn it and don't let it degrade, then you've had a net loss of carbon dioxide from the atmosphere. Rather than sequestering carbon at the bottom of the ocean, why not sequester it into roads and schools?"

Palsson's lab is working on biofuels as well, but they're also pushing to make further improvements in the *E. coli* and human models. They're inching the *E. coli* model closer to the vision of a fully functioning cell inside a computer by integrating the metabolism model with models of gene regulation and transcription. At the same time, they're applying the knowledge they've gained from *E. coli* to human pathogens like salmonella.

E. coli has been engineered to produce ethanol, as well as fuels like butanol or alkanes that are better substitutes for gasoline.

The human cell model is developing quickly, both because of Palsson's work and that of others. Currently, the model includes all metabolic reactions known to happen in any human cell. But only a portion of those reactions occurs in a specific cell type, say, a liver cell or a brain cell or a heart cell. A model identifying which reactions occur in which types of cell will allow for the study of specific cell types, and that is expected to come out soon.

Then there are applications of Recon 1. "There are so many different ones that it's hard to choose," Palsson says. "It's become clear in recent years that metabolism is involved in all of the major human diseases, either as a consequence or a cause."

Already, the model is beginning to be used. A team led by **Albert-László Barabási, PhD**, of University of Notre Dame, USA and **Zoltán Oltvai, MD**, of the University of Pittsburgh School of Medicine, used the Recon 1 metabolic network to discover relationships among various metabolic diseases in a Medicare dataset of 13 million patients and 30 million hospital visits. They found that if two genetic diseases were caused by mutated genes whose associated enzymes were close to one another in Palsson's network, then odds were increased that someone with one of the diseases would have the other as well.

"My expectation is that the human applications will develop much faster than in *E. coli* because of the interest," Palsson says. "The momentum is enormous." □