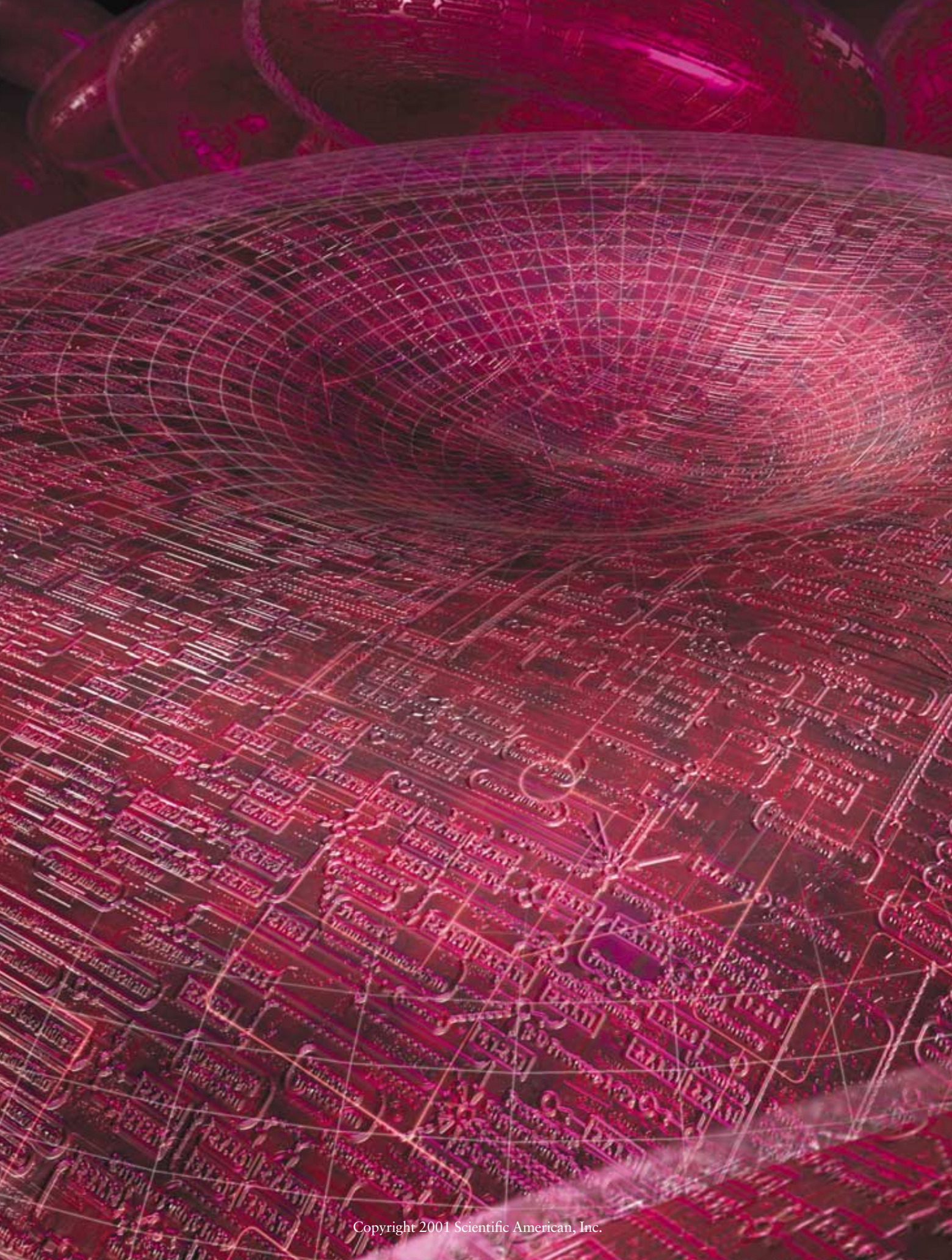


CYBERNETIC CELLS

BY W. WAYT GIBBS
Illustrations by Slim Films

THE SIMPLEST LIVING CELL IS SO COMPLEX THAT SUPERCOMPUTER MODELS MAY NEVER SIMULATE ITS BEHAVIOR PERFECTLY. BUT EVEN IMPERFECT MODELS COULD SHAKE THE FOUNDATIONS OF BIOLOGY

RED BLOOD CELLS were the first human cells to be modeled with computers.

THREE CENTURIES OF REDUCTIONISM IN BIOLOGY RECENTLY CULMINATED IN ITS ultimate triumph. Dissecting life into ever smaller pieces—organisms to organs, tissues to cells, chromosomes to DNA to genes—scientists at last hit the limit. They identified each molecular rung on the chemical ladders of the majority of the human genome. Even before the draft sequence was in hand this past February, some researchers with a philosophical bent began looking ahead to the next major phase of biology—the era of integrationism. It is clear that computer models will be the main tools with which all the biochemical pieces will be placed into a complete theory. But if the variety of "virtual cells" under development is any indication, there is no consensus yet on how best to use those tools.

"People are imagining that this is the final step," observes Drew Endy of the Molecular Sciences Institute at the University of California at Berkeley. "We have the complete parts list for a human being. Now it seems just a matter of assembling the parts in a computer and flipping the switch" to untie all the knotted mysteries of medicine. In fact, he says, "Nothing could be further from the truth."

Endy speaks as one who learned the hard way. In 1994 he and John Yin of the University of Wisconsin–Madison began programming a computer model that would incorporate virtually everything known about the way that a certain virus, T7 bacteriophage, infects *Escherichia coli* bacteria that live in the human gut. The virus looks like a lunar lander. It uses clawlike appendages to grasp the outer wall of a bacterium as the phage injects its DNA into the cell. The genetic material hijacks the cell's own reproductive apparatus, forcing it to churn out bacteriophage clones until it bursts.

Endy and Yin's model simulated mathematically how all 56 of the virus's genes were translated into 59 proteins, how those proteins subverted the host cell and even how the viruses would evolve resistance to various RNA-based drugs. That seems impressive. But peek inside the equations, Endy says, and you'll find that despite including measurements from 15 years of laborious experiments, "there are still a tremendous number of degrees of freedom." The equations can be tweaked to produce almost any behavior. "A useful model must suggest a hypothesis that forces the model builder to do an experiment," Endy says. This one didn't.

Many early attempts to re-create life in silico suffered the same problem. And so most biologists still use computers as little more than receptacles for the surge of data gushing from their robotic sequencers and gene chip analyzers. The "models" they publish in their journal articles are sketchy caricatures based on the best theory they have: the central dogma that a gene in DNA is converted to an RNA that is translated to a protein that performs a particular biochemical function.

But the past few years have seen a growing movement among mathematically minded biologists to challenge the central dogma as simplistic and to use computer simulation to search for a more

Overview/*Virtual Cells*

- Biologists have sequenced the genomes of many simple microorganisms—including germs that sicken humans. Yet they still cannot accurately predict how such cells will react to drugs or external stimuli.
- Microbiologists are now using computer models to simulate the biochemistry of cells. Some try to build models that calculate all important reactions that occur inside a bacterium. Others take an engineering approach, estimating the behavior of the cell by figuring out the basic chemical, physical and biological laws that it must obey.
- The ultimate goal is to find a way to perform virtual experiments that can speed up the discovery of new medical treatments and reduce their cost. A few companies have already begun offering such services, but the accuracy of their models has not been verified by scientific peer review.

powerful theory. "We're witnessing a grand-scale Kuhnian revolution in biology," avers Bernhard Ø. Palsson, head of the genetic circuits research group at the University of California at San Diego. Two years ago Palsson co-founded Genomatica, one of several companies that are creating computer models of cells to try to avoid some of the mistakes that make drug development so costly and slow.

Indeed, reports James E. Bailey of the Institute of Biotechnology at the Swiss Federal Institute of Technology in Zurich, "the cost to discover drugs is actually going up," despite billions of dollars invested in monoclonal antibodies, cloning, sequencing, combinatorial chemistry and robotics. One reason those technologies haven't paid off as hoped, he says, is that they are "based on the naive idea that you can redirect the cell in a way that you want it to go by sending in a drug that inhibits only one protein." The central dogma says that that should usually work. But nine times out of 10 it doesn't.

Consider, too, Bailey urges, that geneticists have engineered hundreds of "knockout" strains of bacteria and mice to disable a particular gene. And yet in many of those mutants, the broken gene causes no apparent abnormality. The central dogma also cannot readily explain how the complex behavior of myriad human cell types emerges from a mere 30,000 or so genes.

"I could draw you a map of all the components in a cell and put all the proper arrows connecting them," says Alfred G. Gilman, a Nobel Prize–winning biochemist at the University of Texas Southwestern Medical Center at Dallas. But for even the simplest single-celled microorganism, "I or anybody else would look at that map and have absolutely no ability to predict anything."

Bailey compares the confused state of microbiology with astronomy in the 16th century. "The astronomers had large archives detailing the movement and positions of celestial objects," he says. "But they couldn't predict the planetary motions with accuracy. They would never have believed that all the orbits are elliptic and described by a simple equation. Nevertheless, Kepler proved it. Now, I don't pretend there is any simple equation for the biology of a cell. But we should be looking for unifying principles that will order our facts into some understanding."

One early candidate to emerge from the more sophisticated cell simulations now under construction is the principle of robustness. Life of every kind has to cope with dramatic swings in temperature, changes in food supply, assaults by toxic chemicals, and attacks from without and within. To survive and prosper, cells must have backup systems and biological networks that tolerate interference.

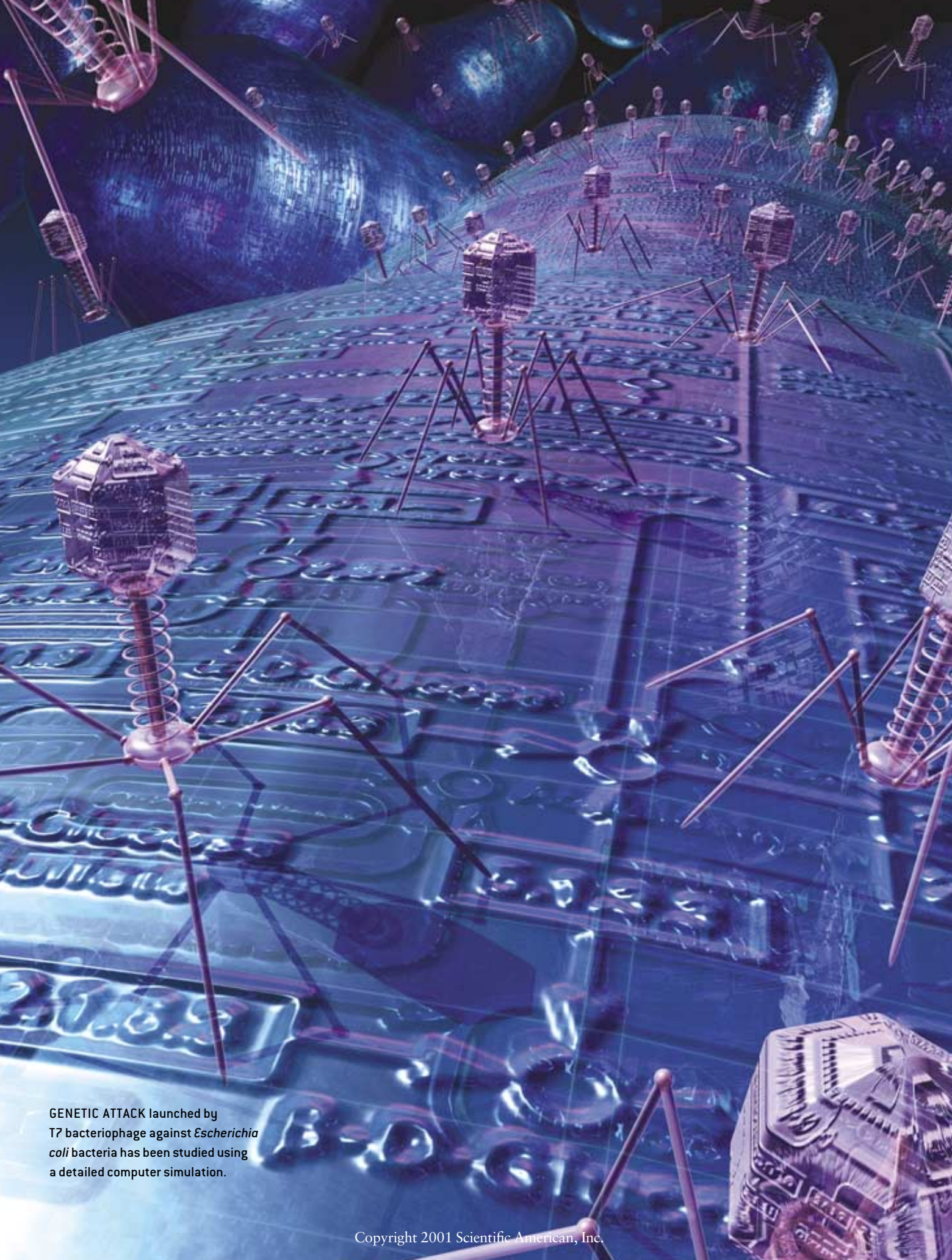
Masaru Tomita saw this property emerge in virtual experiments he ran on his E-Cell model. With teammates at the Laboratory for Bioinformatics at Keio University in Fujisawa, Japan, Tomita built the virtual cell from 127 genes, most borrowed from *Mycoplasma genitalium,* a single-celled microbe that has the smallest genome yet discovered in a self-reproducing life-form. The team's ultimate goal is to find the minimal number of genes needed to create a self-sufficient organism and then synthesize it—an eminently reductionist strategy. But Tomita was surprised when he changed by several orders of magnitude the strength at which various genes in the model were expressed: the E-Cell's behavior hardly budged at all.

"That was an interesting revelation for us as well," says Jeff K. Trimmer, a life scientist at Entelos. The Menlo Park,

Cybernetic Cell Projects

JAMES ARONOVSKY

Genetic Circuits Research Group, led by Bernhard Ø. Palsson (*above*) of the University of California at San Diego, is building genome-based models of *Escherichia coli, Hemophilus influenzae, Helicobacter pylori* and other bacteria that cause human illness.
E-Cell is a mathematical microbe built at the Laboratory for Bioinformatics at Keio University in Japan from the genes of *Mycoplasma genitalium.*
The Virtual Cell is a general cell-simulation package built by the National Resource for Cell Analysis and Modeling at the University of Connecticut Health Center.
MCell is a supercomputer simulation of the synapse between a nerve cell and a muscle cell developed at the Salk Institute and the Pittsburgh Supercomputing Center.
In Silico Cell, constructed by Physiome Sciences in Princeton, N.J., is written in CellML, a programming language that the company is promoting as a lingua franca through which scientists can share and combine their cell models.
Microbial Cell Project, a 10-year program sponsored by the U.S. Department of Energy, plans to spend $15 million a year analyzing single-celled organisms at the molecular level and constructing models of their biochemistry. —*W.W.G.*



GENETIC ATTACK launched by T7 bacteriophage against *Escherichia coli* bacteria has been studied using a detailed computer simulation.

Calif.–based firm has built a functional model of a human fat cell, as well as whole-body models that attempt to mimic the physiological response of obese and diabetic patients to diet and drug treatments. Pharmaceutical firms such as Eli Lilly, Bristol-Myers Squibb, and Johnson & Johnson have hired Entelos to help them prioritize their drug candidates. But when Entelos scientists adjust the virtual cell to reflect the activity of the drug, "we're often quite surprised at how little efficacy a dramatic change in cellular state has on the disease condition," Trimmer says.

Several model-building biologists suspect that what most strongly affects how a cell behaves in response to a drug or disease is not whether any particular gene is turned up or down, and not whether any single protein is blocked, but how all the genes and proteins interact dynamically. Like a connect-the-dots flip book, the story emerges from the links, which shift over time. If that is so, modelers could face a big problem: for most biochemical systems, scientists don't know what reacts with what, and when.

John R. Koza, a computer scientist at Stanford University, recently conducted an experiment that may help biologists connect their genetic dots. Koza is a pioneer in genetic programming, a technique for evolving software by instructing the computer to generate random programs, mutate them repeatedly and then screen them to identify the ones that perform the desired task best. Nicely closing a circle of metaphor, Koza used genetic programming to re-create a small but complicated part of the E-Cell model, itself built from software to mimic genes.

Koza rigged his system to evolve programs that piece together known enzymes into chemical machinery that can convert fatty acid and glycerol to diacylglycerol. Each variant program was converted, for the sake of convenience, to an equivalent electrical circuit, whose behavior was calculated on a commercial circuit simulator. The biological "circuits" that most closely matched the input-output patterns of E-Cell were retained for further evolution; the rest were killed.

After a day, Koza's 1,000-processor custom-made Beowulf supercomputer [see "The Do-It-Yourself Supercomputer," by William W. Hargrove et al., on page 72] spit out a program that matched the actual reaction network. It had four enzymes, five intermediate chemicals and all the right feedback loops. It even found the correct reaction rates for each enzyme. There was a definite "right" answer; no alternative arrangements worked nearly as well.

Koza believes genetic programming can handle larger problems, perhaps one day even deducing the convoluted paths by which cells turn food into energy, growth and waste—but only in cases where biochemists have measured how cells process chemicals over time. Such data are still scarce.

The observation that many biochemical problems most likely have an optimal answer is exploited by Palsson and his colleagues in the models they have built of *E. coli, Hemophilus influenzae* and *Helicobacter pylori,* the germ found in stomach ulcers. They comb the literature to reconstruct as much of the biochemical networks as they can. "Then we subject them to constraints that they must abide," Palsson explains. Mass must be conserved, for example. Electrical charges must balance. Thermodynamics makes many reactions irreversible. "We try to home in on the range of solutions that are physically possible."

Markus W. Covert, a graduate student in Palsson's lab, says the goal is not perfect prediction but reliable approximation: "Engineers can design an airplane in a computer and test it virtually without ever building a prototype, even though they can't compute exactly how the air will flow." In February, Palsson's team reported that their simulation successfully predicted that *E. coli* is optimized for growth, not energy production.

This top-down approach to simulating cells has caught on. Gilman notes that an academic consortium called the Alliance for Cellular Signaling, which he chairs, has secured federal funding to build such models of the internal lives of heart muscle cells and B cells, key players in the immune system. He figures the effort will take a decade to complete, at $10 million a year. "But when we have these sorts of models," Gilman predicts, "it will be the most incredible drug discovery engine there ever was. You could model disease in that cell and then see what drug manipulation could do. Ultimately—though maybe not in 10 years—I have no doubt that there will be quantitative models of cell function, organ function and eventually whole-animal function."

"I would approach such a goal with a fair amount of humility," Bailey cautions. "History teaches us that simulations can help explore particular questions, but there won't be any master model that answers all questions. Eventually the models will become as complicated as the cell itself and as difficult to understand." Unless, perhaps, the next Kepler happens to be a computer wizard. SA

"When we have these sorts of models, it will be the most incredible DRUG DISCOVERY ENGINE there ever was."

—Alfred G. Gilman, University of Texas Southwestern Medical Center, Dallas

W. Wayt Gibbs is senior writer.

MORE TO EXPLORE

Modelling Cellular Behaviour. Drew Endy and Roger Brent in *Nature,* Vol. 409, pages 391–395; January 18, 2001.

Whole-Cell Simulation: A Grand Challenge of the 21st Century. Masaru Tomita in *Trends in Biotechnology,* Vol. 19, No. 6, pages 205–210; June 2001.

Details of John R. Koza's genetic programming approach can be found in the proceedings of the 2001 Pacific Symposium on Biocomputing at **psb.stanford.edu/psb-online/#PATH**