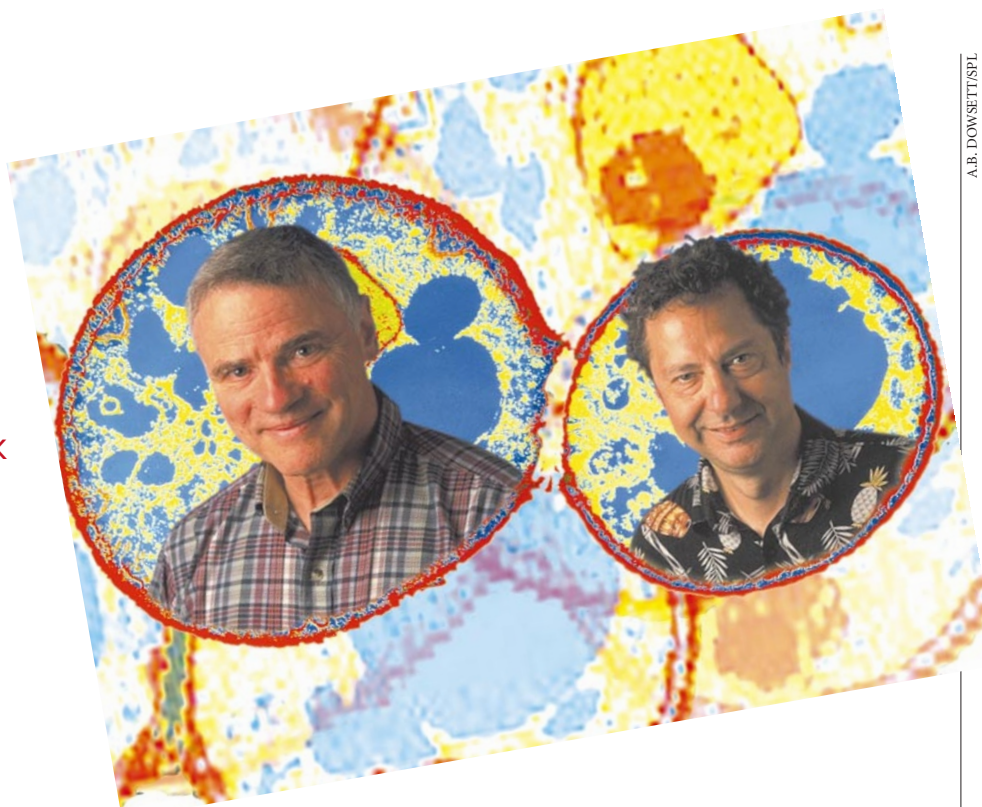


For my next trick...

Is it possible to produce a complete mathematical description of complex biological systems? Leroy Hood thinks so, as Paul Smaglik discovers.



A.B. DOWSETT/SHL

Many people in Leroy Hood's position would be resting on their laurels. In the 1980s, while at the California Institute of Technology in Pasadena, he led a team that invented the automated gene sequencer — spawning a multimillion dollar business and helping to lay the foundations for the Human Genome Project. In 1992, believing interdisciplinarity to be the key to progress in biology, Hood convinced Microsoft founder Bill Gates to establish the Department of Molecular Biotechnology at the University of Washington in Seattle. The department united biologists with mathematicians, chemists, engineers, applied physicists and computer scientists, and went from strength to strength — boasting an annual budget of US\$35 million by 1998.

But in December last year, Hood resigned from his academic post to take on his biggest challenge yet. The university, facing a long list of requests for building projects, had declined to back Hood's plan to create an institute for the emerging discipline of systems biology — so Hood decided to strike out on his own. He founded the independent Institute for Systems Biology (ISB), a non-profit research corporation housed in some 3,000 square metres of office space near the university campus. Here Hood is now assembling a multidisciplinary team to meld biology with the tools of mathematics to provide a complete description of biological processes and systems. "Biology is an information science" is his mantra.

Hood has already been joined by four of his former university's brightest stars — scientists whose expertise spans proteomics, immunology and cell biology, as well as the

application of computer algorithms to some of the thorniest problems in astrophysics. Their goal is to give molecular biology a fresh outlook. "The approach to biology for the past 30 years has been to study individual proteins and genes in isolation," says Hood. "The future will be the study of the genes and proteins of organisms in the context of their informational pathways or networks."

A model approach

Hood is not alone in embracing this systems approach. For example, a consortium called the Alliance for Cellular Signalling, which is based at the University of Texas Southwestern Medical Center at Dallas and headed by Nobel laureate Al Gilman, aims to compile a map of the interactions between 1,000 proteins in two types of mouse cell. But Hood's vision is more expansive. In addition to studying simple systems in model cells, the ISB plans to study the immune system and complex conditions such as cancer and heart disease.

To finance these projects, Hood wants to supplement grant funding by seeking endowments for the ISB worth some US\$200 million. He aims to swell the institute from its current roster of some 80 staff to around 400. An initial US\$5 million came from an anonymous donor, and in July the drugs giant Merck & Co. provided a matching sum. Negotiations are underway with further corporate donors, and other companies are providing contributions in kind. The agribiotech firm Monsanto, for instance, has let Hood keep the 30 automated DNA sequencers he used when leading the team that in April completed a draft of the rice genome.

Under the hood: astrophysicist George Lake (right) will help Leroy Hood assemble the biological equivalent of a virtual car.

But even with the finance in place, the road ahead will be difficult. Hood likens the data gathered by the Human Genome Project and other genomics efforts to an incomplete parts list for a car, in which some entries are missing and others are illegible. Although some parts have been grouped with others and seem to fit together, others remain unidentified and isolated. The ISB aims to build from this information and produce the biological equivalent of a virtual automobile.

Car trouble

Developing the mathematical models on which this effort will depend is much more complicated than assembling a car's engine from a pile of incompletely labelled parts. For a start, biological systems encompass several different levels of information. On one level, a cell can be characterized by recording the genes that are expressed at any one time. But to understand the cell as an integrated biological system, you also need to study the structures of the proteins the genes encode, and the interactions of these proteins with each other as well as with other genes. "At the higher levels, things happen that you can't predict at the lower levels," says Hood. Consider also that each gene can code for more than one protein and that several genes can work together to produce individual proteins, and the enormity of the task facing the ISB becomes clear.

Trying to render each element's role in a process mathematically is also extremely

hard. Some systems biologists aim to characterize every aspect in a system as fully as possible, then construct the cellular equivalent of a wiring diagram — using differential equations to describe the interactions between each node. Gilman's consortium, for example, appears to be heading initially in this direction. Others believe that, because cellular systems are constantly changing, such a linear approach will fall short of providing a full description of a biological system. Instead, they are putting their faith in analysing how different components, such as messenger RNA (mRNA) levels and protein expression, relate to each other under various conditions, and then using probabilities to recreate these relationships. The idea is that this better captures the dynamic, non-linear workings of a biological system.

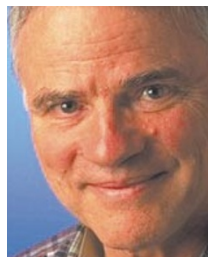
Both approaches may find their niche. But either way, Lee Hartwell, president and director of the Fred Hutchinson Cancer Research Center in Seattle, and a member of the ISB's scientific advisory board, believes that the key is to keep things simple, at least to start with. If scientists at the institute want to characterize every aspect of every system they are interested in, "then they're way too early", he says. Instead, Hartwell recommends a modular approach, concentrating on understanding each component of a system, then integrating them later.

Sweet success?

Hood seems to be heeding that advice. Returning to his automotive analogy, the institute will first produce schematics showing how parts fit together into discrete systems — the transmission, for example — and will then build mathematical models that predict what these component systems do when they work in concert. For now, the ISB is working with relatively well-understood and simple systems in model organisms — akin to understanding how the brakes work in a vintage car.

Initial studies, so far unpublished, have been on sugar metabolism in yeast. As the yeast genome has been sequenced and sugar metabolism in the organism is fairly well characterized, Hood decided to use it as a benchmark to see if systems biology could point towards the same conclusions. ISB scientists, including postdoc Vesteinn Thorsson and graduate student Trey Ideker, performed experiments on yeast fed a steady diet of either the sugar raffinose or galactose. They focused on a handful of genes known to help yeast process galactose rather than raffinose, knocking them out one at a time. After each knockout, the researchers used microarrays to measure which other genes were expressed, and used a novel mass spectrometry technique to measure the abundance of individual proteins.

Measuring both gene expression and protein production paints a clearer picture, says



The future will be the study of the genes and proteins of organisms in the context of their informational pathways or networks.

Leroy Hood



Systems analysts: in addition to Hood and Lake, the Institute for Systems Biology's staff includes (from left) proteomics expert Ruedi Aebersold, immunologist and cell biologist Alan Aderem, and Ger van den Eng, who works on high-speed cell sorting.

Ruedi Aebersold, the ISB's proteomics expert, and one of its high-profile recruits from the University of Washington. Just measuring levels of mRNA, as is typical in gene-expression studies, tells scientists that a gene has been activated, but does not detail the amount of protein it encodes, or whether that protein is functional. The automotive equivalent, explains Aebersold, is knowing that an engine contains a spark plug. The protein analysis, he says, allows you to ask: "Is this spark plug actually firing or is it just sitting in the engine doing nothing?"

The scientists detected around 1,000 genes involved with the switch to galactose metabolism. They then used a series of algorithms to group the mRNAs produced by these genes into families, based on common patterns of changes in the knockout experiments. Finally, they integrated the mRNA and protein data into a simple mathematical model that illustrates how yeast adapts to use galactose rather than other sugars.

Comparing the results with what is actually known about yeast sugar metabolism revealed that the model still needs work. "It actually does a reasonable job of predicting a lot of things, but it's way, way too simple," says Hood. The ISB's scientists will now tweak their algorithms, do more experiments, then adjust the model some more. "The whole process is incredibly iterative," says Hood.

Developing better computer algorithms will help. And the ISB has just gained a valuable asset in the form of another recruit

from the University of Washington: George Lake, an astrophysicist who is project scientist for the Earth and Space Sciences part of NASA's High Performance Computing and Communications initiative. Lake's forte is developing algorithms that, until now, have mostly been applied to problems of cosmology, astrophysics and planetary science. His algorithms, for instance, have made simulations of planetary formation run a million times faster.

Lake believes that building the models Hood needs will be tricky, but not impossible. He likens the challenge to modelling global climate, another field to which he has applied his skills. Climate models have improved through constant mathematical tinkering and by incorporating new kinds of data, says Lake, and the same will be true for simulations of biological systems. He is enthused by his crash course in biology. "It's scary and daunting," he says. "Quite frankly, it just wasn't scary and daunting to go to the office as an astrophysicist anymore."

Hood's expertise in technology development might also prove to be important when it comes to gathering the new kinds of cellular data that Lake predicts will be needed. It is still too early to tell whether systems biology's virtual automobile will ever move with the speed and grace of a sports car. But given Hood's track record, if it does, it is a fair bet that the ISB will help lead the way. ■

Paul Smaglik is *Nature's* Washington DC correspondent.

► www.systemsbiology.org