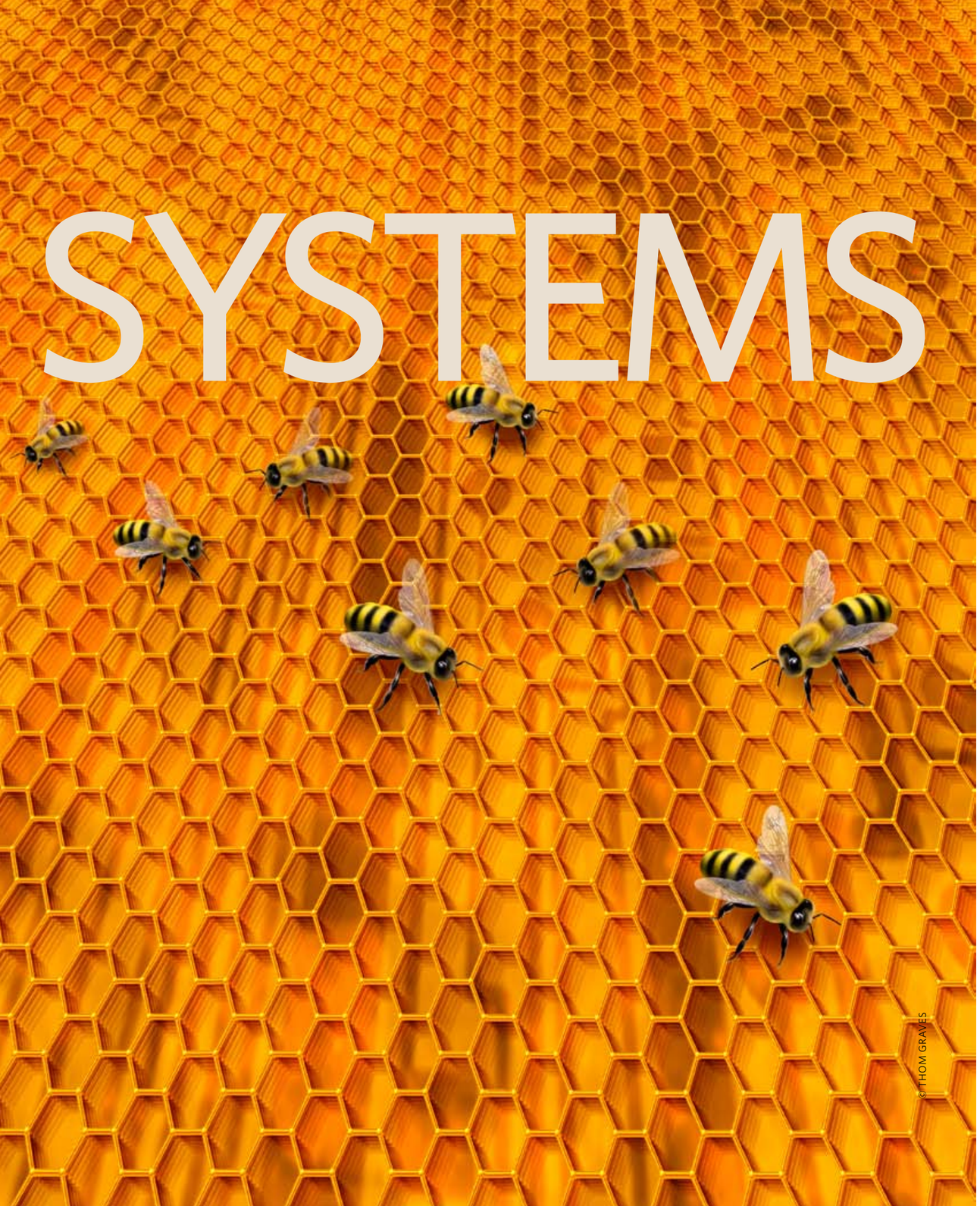


BIOLOGY

Beyond the Buzz

Lessons from EGFR research show how to kick-start a systems approach for other areas of biology

BY H. STEVEN WILEY

If you want to start an interesting debate at almost any scientific meeting, just bring up systems biology. Latched onto by the scientific and even the lay press as the "next big thing," it is clear that many scientists have misgivings about the subject. The rapidly changing landscape of biology is an exciting notion but one that can be worrisome. Regarded by some as little more than a buzzword and others as the next step in bringing biology from a descriptive to a predictive science, systems biology is host to disagreements fueled in part by a lack of a uniform definition.

At the most basic level, systems biology seeks to understand how the molecular processes of cells are linked to higher biological functions. The relationships between low- and high-level functions must be specified to the point where one can predict how a change to a part gives rise to a change in the whole. Because of the extreme complexity of biological systems, however, this type of prediction demands sophisticated computer models. Of course, the field has not yet advanced to this point. Mostly we can show consistency between models and predictions, which leads many skeptics to conclude that systems biology is long on style and short on substance.

I have a far more optimistic impression. We have arrived at a point in biology where we will increasingly be able to understand and predict the behavior of complex biological systems. This moment is the culmination of decades of detailed studies on the molecular basis of cell function and the recent development of new analytical and computational technologies. I have seen a profound transformation in the various fields that study the epidermal growth factor receptor (EGFR) that makes me confident we will soon be able to accurately describe and predict how signal transduction pathways function in both normal and diseased cells. We are not there yet, but we are going in the right direction. And this direction will mean significant changes to the way biological research is conducted, with the potential to change the entire social fabric of biology. The EGFR system has served as a model for understanding basic receptor biochemistry for decades. It provides important lessons for our transition to systems biology as well. ▸

A MODEL SYSTEM FOR SYSTEMS BIOLOGY

We are finally starting to understand the EGFR system after more than 40 years of intense effort. EGF was discovered in the late 1950s by Stanley Cohen as a factor from submaxillary glands that induced precocious eyelid opening in newborn mice. By 1962, the protein was purified, and by 1975 its binding to specific cell-surface receptors had been characterized.[1] Stable, easy to purify, and stimulating the proliferation of many different cell types in vitro, EGF proved a wonderful protein to investigate. Because EGF is so stable, it can be labeled with numerous radioactive, fluorescent, and biochemical tags without losing biological activity. This allowed analysis of the EGFR system using approaches ranging from electron microscopy to stop-flow kinetics.

By the early 1980s, when I entered the field, the EGFR was found to possess tyrosine kinase activity and had become the focus of intense investigation by many labs. In short order, groups had cloned the receptor, mathematically modeled its dynamics, and identified some of its substrates.[2] By the late 1990s, all of its ligands had been identified, and its signaling pathways and interacting partners had been described in more than 10,000 papers.[3] The field had fragmented into multiple subfields, each investigating a part of the whole puzzle. The irony, however, was that despite all of this detailed information, we still did not know EGFR's role in normal cell physiology or how it stimulates cell division. We had essentially described all of the trees but still had no idea about the forest.

Several years ago, however, the drive towards reductionism started to reverse. Initially, each subfield in EGF research was focused on a distinct question (see "The EGFR System"), but most used the same tools of molecular biology. Eventually, data acquired in one area started to overlap with those gathered in another, and the underlying patterns started to appear. Even tangential fields of research began to see surprising connections to EGFR signaling. For example, it was known that cells that had EGFR generally make one or more EGF-like ligands. This process of self-stimulation, termed autocrine signaling, puzzled scientists until it was revealed that many other factors, such as angiotensin II and tumor-necrosis factor, activate the EGFR by stimulating the release of autocrine factors.[4,5]

This "transactivation" of the EGFR appears to be part of a process by which cells couple their specific response to their extracellular environment.[6] Stimulating a cell by an initial factor, such as tumor-necrosis factor, initiates a series of autocrine cascades involving multiple growth factors, cytokines, and EGFR ligands whose final effect depends on the extracellular environment.[7] To understand how such a complex hierarchical control system is regulated requires molecular biology as well as mathematical models.[8] Systems biology was now a requirement rather than an option.

The transition of EGFR research from reductionism to a more systems-level approach brought the field back to its beginnings. The intent of investigators entering the field 30 years ago was firmly rooted in understanding how EGF induced cell proliferation at a systems level. The most powerful tools available, however, were molecular-level tools that generated specific bits of data about small parts of the problem. It took more than 30 years of research by thousands of scientists to accumulate enough data to start seeing the underlying patterns. One of the central lessons we have learned is that systems biology is a data-driven science. Without a sufficient base of data, attempts at large-scale predictive modeling are not generally successful.

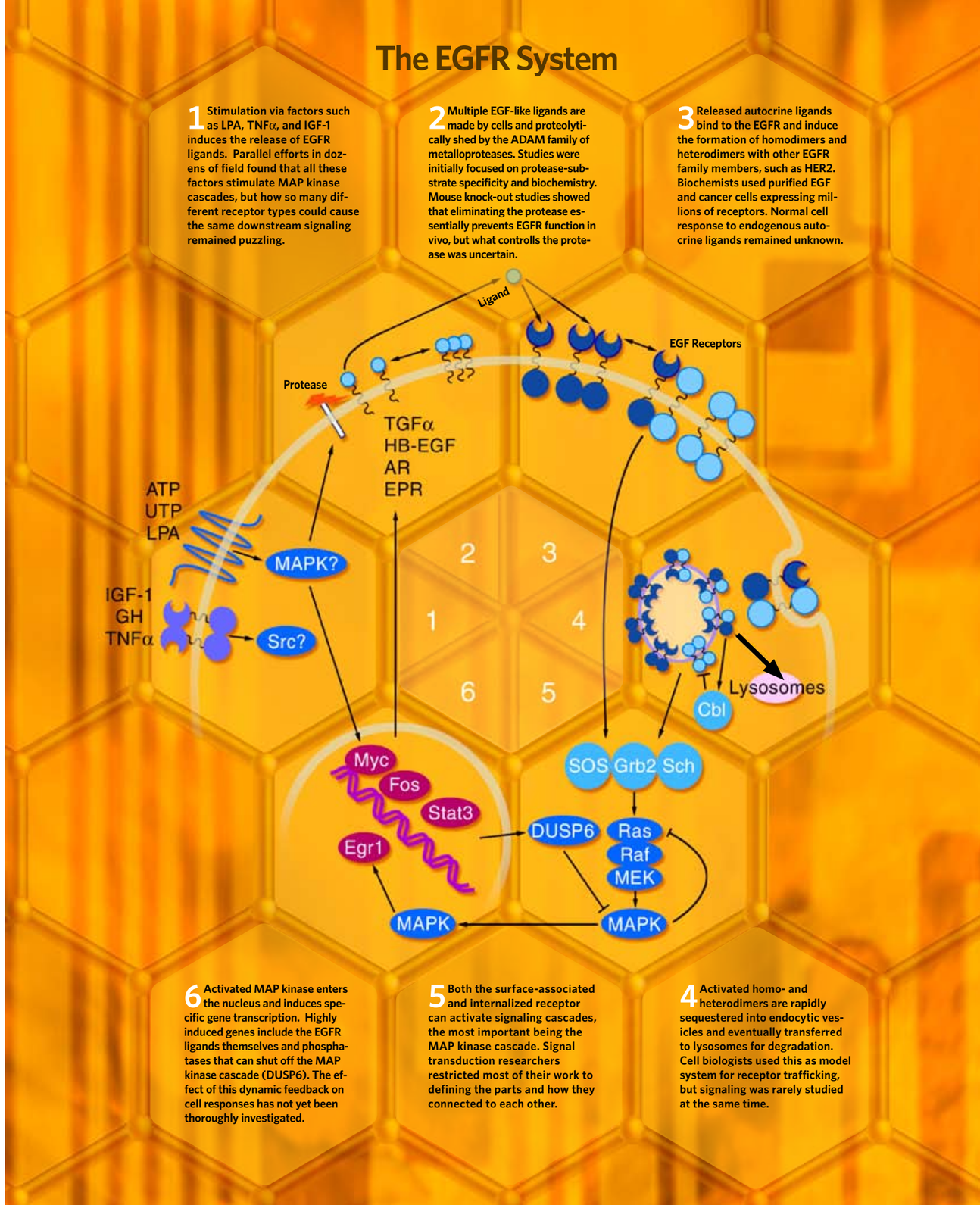
A second, equally important lesson is that you need a computational framework in which to store the data. The early success of EGFR investigators with measuring receptor dynamics gave rise to a small but active community of modelers. These investigators demanded quantitative data and this, in turn, defined the experiments that produced the most useful results. Although the initial mathematical models tended to be limited because of the lack of underlying detail, by 2003 the wealth of accumulated data allowed us to build large-scale models that incorporated details of signaling networks, receptor dynamics, and extracellular spaces.[6,9] Building on the decades of work in the EGF system, molecular detail could be connected to higher-level functions.

WHERE SYSTEMS BIOLOGY WILL WORK

The lessons learned in the EGFR field serve as an example to other areas of research, such as immunology and metabolism, which are starting to use a systems approach. Understanding fundamental regulatory mechanisms at the systems level offers a new approach for both the diagnosis and treatment of many diseases, but it requires enormous amounts of data and a sophisticated data infrastructure. The success of systems biology in understanding model organisms, such as yeast, and signaling systems, such as the EGFR, is built on decades of data.

New technologies, such as mass spectrometry-based proteomics and gene microarrays, promise to generate data at an unprecedented rate, but many different types of quantitative information are needed to understand a complex system. The information also needs to be encoded into a computational framework so that it can be analyzed. Unless a field of research meets the twin requirements of sufficient data and sufficient mathematical framework, it is essentially impossible to apply a systems approach. Parts of cancer biology are starting to reach this point, and a predictive understanding of some specific cancers is highly likely in the next decade.

Even the EGFR field has a long way to go before systems biology becomes a mainstream approach. There is still a paucity of quanti- ▸



tative data, especially on relevant model cell types. An unfortunate reality of the EGFR field is that dozens of different cell types are studied. Combining data from all of these cells can only generate models of average cells, but average cells do not exist. Creating a model with predictive power requires collecting data on a single cell type under well-defined conditions (See "Five Simple Steps," p. 56). My colleagues and I have used a single cell type for the last decade for both modeling and experiments, but it is difficult to incorporate important data from other investigators using alternate cell types.

Although our analytical technologies have also improved over the last several years, most of them can generate only high-throughput data of low quality (i.e., microarrays), or low-throughput data of high quality (i.e., enzymatic assays). We need both high-throughput and high-quality data to create accurate cell response models. We also need technologies that simultaneously measure multiple parameters at the level of individual cells. The more information you can start with, the more likely you are to see the important stuff.

SEEING THE WHOLE PICTURE

Perhaps the most difficult changes we face in the transition to systems biology are the social ones. Today, most biological research is unapologetically reductionist. Efforts by the National Institutes of Health to encourage more ambitious, multidisciplinary research have met with resistance from study-section members who are less than enthusiastic when evaluating these proposals. We are victims, I fear, of our own success.

The most successful scientists are those who asked the "right (i.e., focused) questions" and proposed a scope of work that grant-review panels felt could reasonably be accomplished in three to five years. With the tools available over the last 20 years, this forced most of biology to become highly specialized. The more we investigated a subject area, the more complex it became, forcing us to work on increasingly narrow areas of research. Posing a big, ambitious question was probably the fastest way to have a grant application end up on the reject pile. Perhaps reductionism was needed when you had to keep all pertinent facts in your head or a lab notebook, but technology has now provided us with a way to ask bigger, more important questions. We must not allow ourselves to be ruled by yesterday's scientific approaches.

Who is going to drive systems biology to the next level? We have become comfortable with our specialized niches, and the availability of new, high-throughput technologies is unlikely to tempt most biologists to venture into a new area of research. As is usually the case with scientific revolutions, the cause will most likely be passed to the new generation. This was evident at the International Conference on Systems Biology in Boston this past October. Many senior investigators started their seminars by first apologizing that they did not do systems biology, while the poster sessions were jammed with hundreds of students and postdocs who most emphatically embraced systems biology.

Fortunately, scientific curiosity is also a strong motivator, and there is widespread appreciation in the EGFR field that systems biology can provide answers to long sought-after questions. Last year, a group of investigators formed the International Receptor Tyrosine Kinase Consortium (www.rtkconsort.org), focused on the EGFR and other related receptor systems.[10] The purpose of the group is to find ways that scientists can work together voluntarily to promote systems biology of these receptors. The effort includes defining model systems, building shared datasets, and working together on common problems. It is a framework only for large-scale systems biology, but it is a good start.

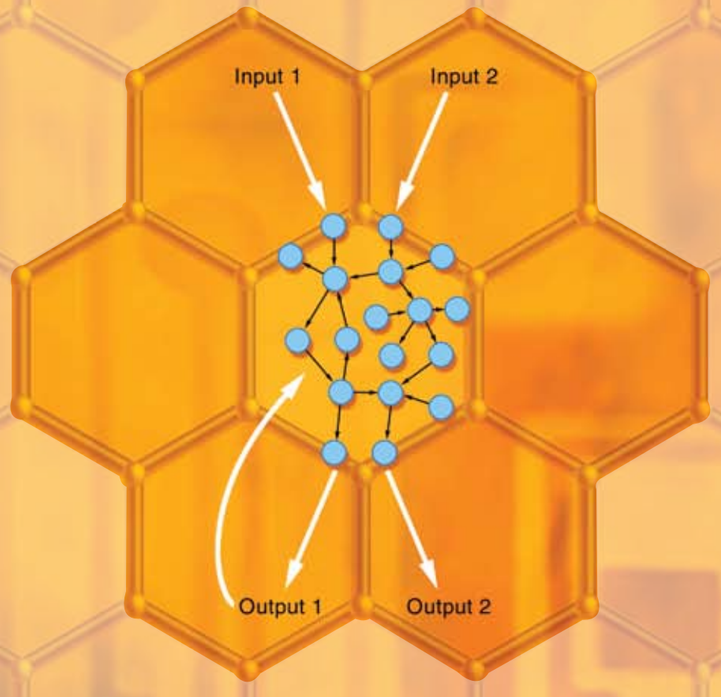
I am looking forward to the new field of systems biology and what it promises to reveal about the EGFR. Using a combination of statistics, computational approaches, novel new instruments, and molecular biology, I expect it to find ways to predict how changes in the cellular environment alter the flow of information through the multiple pathways controlled by the EGFR. A new generation of investigators will work collaboratively, using Web sites, collaborative software, and video tools to work as virtual teams. They will not ask questions about the bricks and mortar that make up cells, but about their design and architecture. Biology will move beyond the descriptive stage and join the quantitative sciences where prediction is tantamount to understanding. ■

Five Simple Steps To Systems Biology

While not the only way to do systems biology, these steps have proven effective and reflect the current abilities and limitations of the approach. In the case of the EGFR system, we used a cell type (human mammary epithelial cells) that displays a strong and reproducible response to EGF. We made many engineered variants of these cells that serve as the basis for mathematical models of receptor trafficking and signal transduction. Although the approach may seem theoretical, it has been used successfully to understand the complex relationship between multiple growth factors and cell responses, such as apoptosis, as discussed elsewhere in this article.

1 Have a clearly defined set of inputs and outputs. The processes between them define the system. If input and output are too closely linked, say between receptor occupancy and receptor activation, the system becomes trivial. Too far apart, say between receptor mutations and tumor formation, and the system becomes indeterminate with current technology. A process such as apoptosis or cell migration that takes place over several hours is a good compromise.

2 Define the relevant parts of your system. This requires knowing many of the relevant changes in gene expression, protein phosphorylation, enzymatic activities, and redistribution of proteins that occur during the time between application of the inputs and measurement of the outputs. Global microarrays, proteomics, phosphoproteomics, and high-throughput enzyme assays are ideal for initial studies, but once the relevant subsystems are discovered, a smaller set of assays can usually be used to define the "state" of the system. This state is typically represented in terms of a network.



3 Next, move the system through many different states and then quantify the relationship between the state and the output. Using multiple inputs is essential here and can be achieved by adding more than one extracellular factor or changing the levels of a critical component by siRNA or gene overexpression. The more kinds of output measurements you can make using multiple altered system states, the easier it will be to establish functional correlations.

4 Relate the changes in system state to output using one of a variety of different mathematical tools, such as principal component analysis or self-organizing maps. This provides a quantitative foundation for structuring the system's components into modules that can provide a mechanistic explanation for the observed effects. This work can be greatly facilitated by detailed kinetic studies, as this helps to order the processes into a sequence of events.

5 Modify the original network map to be consistent with the results of the previous experiments. This is where computational models can be particularly helpful in that they can generate quantitative outputs from any given set of inputs. One can then test the hypothesized system structure by introducing additional perturbations. Repeat the cycle of experiment-analyze-model until the model and results converge.

H. Steven Wiley is director of Pacific Northwest National Laboratory's Biomolecular Systems Initiative, a multidisciplinary program designed to understand complex biological systems.
hswiley@the-scientist.com

References

1. G. Carpenter et al., "Characterization of the binding of 125^{II}I-labeled epidermal growth factor to human fibroblasts," *J Biol Chem*, 250:4297–304, 1975.
2. G. Carpenter, S. Cohen, "Epidermal growth factor," *J Biol Chem*, 265:7709–12, 1990.
3. Y. Yarden, M.X. Sliwkowski, "Untangling the ErbB signalling network," *Nat Rev Mol Cell Biol*, 2:127–37, 2001.
4. M. Asakura et al., "Cardiac hypertrophy is inhibited by antagonism of ADAM12 processing of HB-EGF: metalloproteinase inhibitors as a new therapy," *Nat Med*, 8:35–40, 2002.
5. W.N. Chen et al., "Induced autocrine signaling through the epidermal growth factor receptor contributes to the response of mammary epithelial cells to tumor necrosis factor alpha," *J Biol Chem*, 279:18488–96, 2004.
6. H.S. Wiley et al., "Computational modeling of the EGF-receptor system: a paradigm for systems biology," *Trends Cell Biol*, 13:43–50, 2003.
7. K.A. Janes et al., "The response of human epithelial cells to TNF involves an inducible autocrine cascade," *Cell*, 124:1225–39, 2006.
8. K.A. Janes et al., "A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis," *Science*, 310:1646–53, 2005.
9. H. Resat et al., "An integrated model of epidermal growth factor receptor trafficking and signal transduction," *Biophys J*, 85:730–43, 2003.
10. Y. Sakaki et al., "The International Receptor Tyrosine Kinase Networks Consortium," *The Scientist*, 19(8):8, April 25, 2005.