Hip Bone Is Connected to . . .

Leo P. Kadanoff

Every living cell employs a network of chemical reactions to construct and control a huge variety of chemical compounds. Biologists, physicists, and mathematical scientists have developed and applied techniques of network analysis to study how biochemical compounds interact with one another. The studies suggest that very complex biological systems may result from the interconnection of far simpler processes, and further suggest that insight into complex systems may be gained from models that are themselves conglomerates or networks of simpler models.

Here I discuss four theoretical papers, each describing how a specific network carries out a function in a particular group of organisms. Painstaking work of biologists has filled in tens of thousands of reaction processes occurring in cells and has told us what reactions are important for a given biological function. Thus the biologists traced out the interactions among the different chemical compounds that work together. Those connections have given us a partial topology of some of the most important biological networks. Further, the biologists have isolated some modules—interconnected processes that work together to produce a given biological function.

Fruit-fly embryo

One much-studied example looks at the development of an effectively one-dimensional chain of cells along the body of a developing fruit-fly embryo. Starting roughly three hours after fertilization, a simple pattern, with a spatial period of four cells, is turned into a much more complex and sharply defined pattern, with each repetition eventually becoming a segment of the adult *Drosophila* insect. A University of Washington group has modeled each cell as the interaction of the chemical products of five genes. The model involves 14 ordinary differential equations that embody the time history of each cell's concentration of its different chemical species.[1] The model, which has roughly 50 parameters, includes the interaction of a given cell with its closest neighbors. After a little adjustment of network topology, including the introduction of some new interactions that inhibit formation of two chemical species, the Washington group found a satisfactory set of parameters. The set produces a steady-state behavior that reflects well the pattern of concentrations known to exist in *Drosophila* at the termination of the developmental stage under study.

The investigators studied some 200 000 parameter sets, with each parameter varying over a wide range, usually several factors of 10. Roughly 1 in 200 of these randomly picked sets of parameter values yielded qualitatively correct behavior. The word "robust" is used to describe models or biological systems in which the proper functioning seems to persist across a wide range of parameter values or external conditions. The main result of the University of Washington work is the suggestion that this developmental module is sufficiently robust against variation of parameters that once one gets the network topology right, it is relatively easy to find parameters that give the biologically observed outcome.

Despite the fact that the authors were working with a continuum model, they described their desired outcome as a steady state in which some chemical compounds are present and others are absent. Indeed, much biological work is expressed in very qualitative form, using this kind of binary description of the overall chemical state of the system. A system is described as a network consisting of nodes and links. Each node is a particular chemical species. Each link is a line between nodes that shows how an existing chemical compound inhibits or enhances the formation of a product chemical compound. Starting from this description, it is natural to build a computer model based on a large number of statements like "If at time t, compound A is present and compound B is absent, then at time $t + 1$, compound C will be present." Putting together a large number of statements of this kind can produce a model that completely specifies the behavior of the system. Doing exactly that, Réka Albert and Hans Othmer constructed a model using the same connections as in the Washington study but built upon binary variables.[2] In every step of their model, the presence or absence of the different chemical species determined, via single or pairwise interactions, what species were available in the next step. The idea is that, given the robustness of the network topology studied by both groups, it is plausible that these simplified interactions might capture the essence of the underlying processes.

The modeling seems to be successful in that it captures some of the basic characteristics of the developmental system. Starting with any set of initial data, the model system will fall into one of ten different time-independent patterns, one of which (the "wild type") corresponds to the actual state of a viable *Drosophila* embryo. The effects of mutations that cause changes in the behavior in several of the genes have been investigated experimentally and these effects are apparently properly reflected in Albert and Othmer's simplified model. Thus the model can be used both to represent the known behavior of this biological module and to extrapolate the behavior to unnatural or new situations.

Albert and Othmer, and other authors as well, argue that their model of a developmental module has properties that are quite different from a randomly chosen network with a similar number of nodes. In general, a given complex network with binary variables will permit many different long-term behaviors involving mostly time oscillations. Instead, the model networks studied tend to have only a few different long-term behaviors, each one a steady state. In general, binary-variable networks will show vastly changed behavior after the flip of an initial value at a single node. But the models of biological networks each come to the same long-term behavior for many values of the initial state of the system. Perhaps these special features reflect the effects of evolution-

Leo P. Kadanoff *is a condensed matter theorist at the University of Chicago.*

ary selection captured in the biology of the module and reflected in its computer model. The patterns generated by this module are crucial for *Drosophila* development, and must be maintained over the lifetime of the individual insect. It would be pleasing if unusual features of network behavior closely corresponded to the network's specific biological function.

Cell cycle of yeast

Another particularly robust network pattern may be seen in the module that controls the cell cycle in yeast. A yeast cell grows until it becomes sufficiently large and then undergoes a complex process in which it splits into two smaller cells, which then grow and split, and so on. A Peking University group has developed a binary-variable computer model to describe the module that controls the splitting.[3] The dynamics begin with the model in a stationary state representing the yeast's growth state. One of the variables in the computer model is then reset in a manner intended to mimic the "divide!" signal that can be sent to the splitting module in a real yeast cell. In response to the signal, the model changes the values of its variables and proceeds through several stages until it reaches the stationary state once again. The authors modeled this behavior using an 11-node network, gleaned from the biological literature, and they included both excitatory and inhibitory interactions. The model has 2^{11} possible states. Of these, seven were stationary states and none were cycles. Since the model is time-independent, after a finite number of steps, each one of the 2048 starting states would begin a process that would eventually fall into one of the stationary states. One such state, closely corresponding to the growth state of yeast, is the final destination of 1764 of the possible initial states. It is remarkable to see that 86% of possible starting states would thus return to the correct configuration.

Tissue differentiation

The University of Washington model mentioned earlier is robust in that its behavior does not change much as model parameters are varied. As I mentioned, its variables are not binary, but instead are numbers describing concentrations of various chemical compounds in the cell. The differential equations for concentrations contain the model's parameters. One might wish to understand in more detail the result of parameter variation. A Cornell University study with this aim looked at a signaling network involved in tissue differentiation in the development of rats and other higher animals.[4] This particular network is interesting because it shows a different response to a transient exposure to its activating chemical species (the cells proliferate) compared to a sustained exposure (the cells change into nerve cells). The model makers start by using biologist-generated knowledge about which chemical interactions are important for the biological process. To describe in mathematical terms the reactions that produce the dozen or so important chemical compounds, the model makes the chemical concentrations vary in time via a set of first-order differential equations. The equations contain 55 free parameters. For many different sets of parameter values, the model is run to generate a set of data giving the concentrations of the chemicals as a function of time. Each data set is then compared to a set of data obtained from experiments on rats. A sum of squares of differences between experimental data and model data is used as a measure of the disagreement produced by the particular parameter set. After the parameters are each varied over a very wide range, the set of parameters that produce the best fit is selected. The calculated fit matches reasonably well the experimental data used to construct it and, in addition, accurately predicts the results of additional experiments.

But that's not the point. As in the work of the University of Washington group, a wide range of parameters fit the data reasonably well, that is, produced a fit indistinguishable from that of the best fit. One conventional way of thinking is to visualize a landscape in which directions in x, y, and so on describe parameter variation while depth represents the parameter that describes the goodness of fit. With this definition, the measured goodness data might, in a simile used by James Sethna, be imagined as a river at the bottom of a canyon meandering and branching through a large territory. Some correlated changes in parameter space could produce substantially changed parameters, but an equivalent fit. These changes are, then, motion along the river. Changes in other directions in parameter space might swiftly drive an observer away from a viable solution and up the canyon walls. Underlying everything is some process, the base process of the module, produced by the topology of connections and remaining in force within some volume in parameter space. That volume is Sethna's river. The network's topology produces the branching topology of the river, as yet unknown.

Connections matter

All of these examples are hinting at the same, very exciting idea. Ten or a dozen or 20 linked elements produce a mildly complex machine, perhaps analogous to a little thermostat. We can, with study, understand its working pretty well. We can certainly simulate the machine with reasonable accuracy. Living things can be described by modules, which are thus machines performing the fundamental biological functions. These machines then work on the basis of even more fundamental processes involving chemistry and physics. Since our isolation of modules only includes a portion of the relevant network activity, we cannot expect to obtain fully quantitative descriptions of the processes in question. In particular the best-fit reaction rate parameters might be quite different from in vitro values. But since it's the connections that matter, we might nonetheless catch something of the essence of what is going on. The organism itself is much more complex, perhaps being describable as many such modules hooked together to form networks of modules, and those hooked together, and so forth. Such a global structure is too hard for us now, but students of biological dynamics are beginning to catch what is going on at the first level of interconnection.

And so by studying how "the hip bone is connected to the leg bone" we might begin to see how biological systems of unimaginable complexity can perhaps have grown from interconnections of modules of quite imaginable complexity.

This research was supported in part by the Materials Research Laboratory at the University of Chicago. I acknowledge helpful conversations with Alan Calder, Wendy Zhang, Michael Brenner, Sharon Lubkin, Chao Tang, Philippe Cluzel, Réka Albert, and James Sethna.

References

1. G. von Dassow, E. Meir, E. M. Munro, G. M. Odell, *Nature* **406**, 188 (2000).
2. R. Albert, H. G. Othmer, *Journal of Theoretical Biology* **223**, 1 (2003).
3. F. T. Li, T. Long, Y. Lu, Q. Ouyang, C. Tang, *Proc. Nat. Acad. Sci. USA* **101**, 4781 (2004).
4. K. S. Brown, C. C. Hill, G. A. Calero, C. R. Myers, K. H. Lee, J. P. Sethna, R. A. Cerione, *Phys. Biol.* **1**, 184 (2004). ■