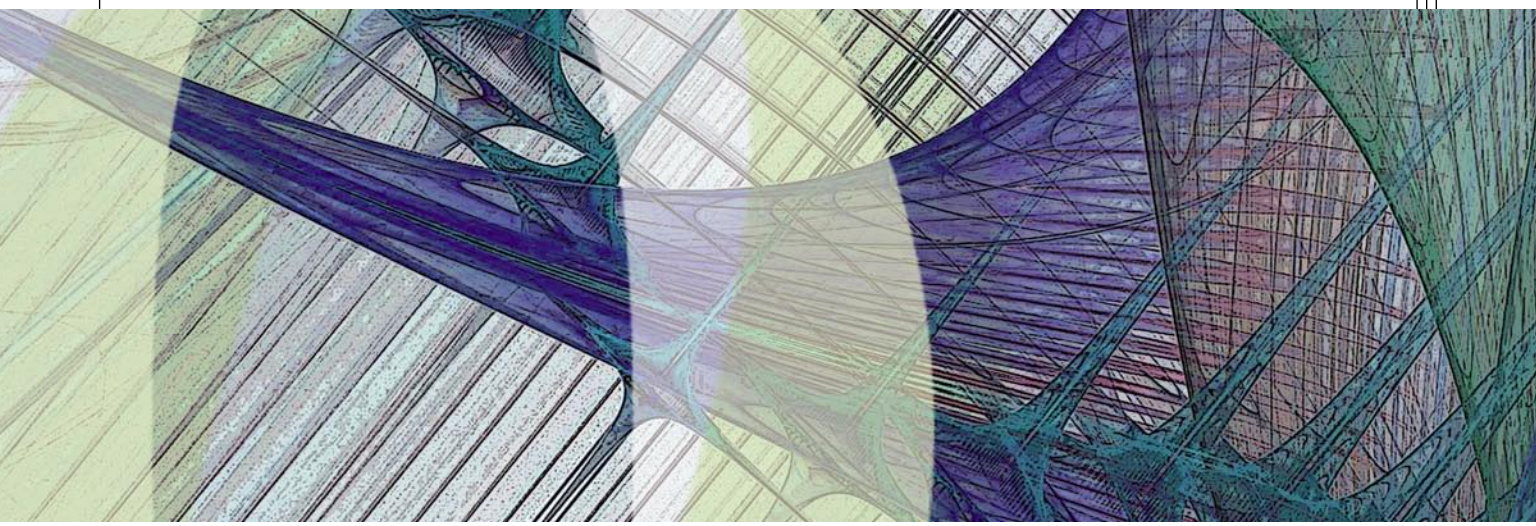


COMPUTING WITH PROTEINS

Naren Ramakrishnan, *Virginia Tech*
Upinder S. Bhalla, *National Centre for Biological Sciences, India*
John J. Tyson, *Virginia Tech*

Protein-based computers are ripe for spectacular scientific advances because we know a great deal about these circuits' molecular components. But we are only beginning to understand how they process information and make decisions.

Of the many types of information processing systems known to us, the least well understood are the complex, intracellular molecular reaction networks that control the physiology of living cells and organisms. Protein-based computers are very different from their silicon-based counterparts, and the intellectual tools necessary to understand, intervene in, and reengineer these information processing systems are likely to be quite different from the paradigms that have been so successful in electrical engineering and computer science.

Much is known about how living cells and organisms encode and process information. For example, genetic information is stored in linear sequences of nucleotides in DNA and RNA, and the information in a gene is translated into a linear chain of amino acids that folds into a protein. In the nervous system, sensory information is encoded in firing patterns of sensory neurons, and these input signals are processed by the brain, which generates appropriate responses via motor neurons and neurosecretory cells. The nervous system is dynamically programmed during use by continually revising the strengths of synaptic connections between neurons,[1] so that organisms can learn to cope with their own unique life situations. Similarly, the genome's output is dynamically programmed during development by switching on or off whole suites of genes, so that the tissues of a multicellular organism can adopt and maintain their unique roles within the body.

Living organisms have a third, equally important, information-processing system based on complex protein signaling networks that sense a cell's chemical state and respond appropriately. These signaling networks constitute the core of what a living cell does, namely, monitor its external environment (food sources, salt balance, soluble signaling molecules coming from other cells), take into

0018-9162/09/$25.00 © 2009 IEEE Published by the IEEE Computer Society

BIOLOGY FOR COMPUTER SCIENTISTS

An organism's chromosomes contain its genetic information. Humans have 46 chromosomes, organized as 23 pairs. The nucleus of every cell in the human body contains all 46 chromosomes. Each chromosome contains a double-stranded molecule of deoxyribonucleic acid (DNA). Each strand of the DNA molecule is a linear chain of nucleotides forming a string over an alphabet of four "letters": A for adenine, C for cytosine, G for guanine, and T for thymine.

The two strands of DNA are complementary, in that A on one strand always pairs with T on the other strand, and similarly C pairs with G. Sections of the linear chain, called a *genetic locus*, contain the information necessary to construct a protein (the "open reading frame" or ORF), plus "upstream regulatory elements" that control when and where the gene will be expressed. One strand of the DNA molecule encodes the gene, and the other strand (the reverse complement) contains the information needed to make a new copy of the gene (much like the relationship between "negative" and "positive" in photography).

During a process called *transcription*, the ORF sequence is copied into a complementary strand of messenger RNA (mRNA). RNA, a single-stranded nucleic acid, consists of four bases: A, C, G, and U (for uracil). U replaces T as the base that complements A. The mRNA molecule serves as a template for constructing proteins. Compared to DNA, RNA molecules are more chemically reactive and short-lived. mRNAs are "working" copies of genes, with the "permanent" copies stored safely on the nucleus's chromosomes. Sometimes the unit of transcription is an *operon*, which is made up of several genes but which are transcribed together to form a single *polycistronic* mRNA molecule. mRNA molecules travel out of the nucleus into the cytoplasm, where proteins are manufactured. The *ribosome* is the cellular organelle where mRNA molecules are converted to proteins. Like genes, proteins are also linear chains of subunits—strings composed over an alphabet of 20 amino acids. Specifically, a codon on the mRNA (a sequence of three bases drawn from A, C, G, and U) codes for a specific amino acid in a protein's primary sequence. The association between the 64 codons and the 20 amino acids (plus "start" and "stop" codons) provides the genetic code.

For example, the codon AUG encodes the amino acid methionine (M) and also serves as the "start" codon. Translation is the process by which the ribosome interprets the genetic code of an mRNA molecule to manufacture a protein. If the mRNA molecule is polycistronic, it will code for multiple, functionally related, proteins. Once manufactured, proteins fold into characteristic three-dimensional structures that determine their functions.

Proteins serve many purposes within a cell, either singly or by binding to each other and forming protein complexes. Structural proteins give shape to cells and tissues. Enzymes catalyze the biochemical reactions of metabolism (the processing of food for energy and raw materials) and biosynthesis (the production of new cellular material).

Some proteins are transcription factors that bind to the upstream regulatory elements of a gene to either inhibit or enhance its expression. In addition, and of special relevance to this review, the primary function of some proteins is information processing. These proteins include "sensors" that bind extracellular and intracellular signal molecules, "processors" that integrate input signals and compute appropriate responses, and "executors" that carry out the responses.

account its own internal state (size, growth potential, DNA damage, attachment to neighboring cells), and compute the appropriate course of action to be taken (movement, membrane transport, gene expression, and cellular growth or death). These decisions often pose matters of life or death for the cell.

Protein-based computing represents and processes information in analog form. The "states" are continuous-valued: encoded in the interacting proteins' concentrations. Biochemical transformations such as protein-ligand binding, phosphorylation and dephosphorylation, and regulated protein synthesis and degradation process information. Protein-based computing spans many orders of scale, from the regulation of gene expression (when, where, and which genes to be expressed), to metabolic regulation (the flow of material and energy from food digestion to the synthesis of new macromolecules), to signaling networks (such as triggering an immune response to a pathogen attack), to entire cell-to-cell communication pipelines. Indeed, protein-based computing machines are themselves crucial in controlling how cells read out the genome during development and how synaptic plasticity remodels neuronal networks.

Protein regulatory networks in cells are quite distinct from computer architectures. For instance, their signal-response characteristics are hardwired into the molecular circuitry itself. On short timescales (seconds to minutes), the wiring is fixed; on physiological timescales (minutes to hours), the network can be rewired by changing gene expression; and on long timescales (generations), the whole logic of the network can evolve by mutation and selection. Information processing in protein regulatory networks is highly parallel (all reactions take place simultaneously in the cell) and very sloppy (molecular noise limits the accuracy of calculations to from 3 to 4 bits). For these and other reasons, traditional ways of thinking about computers, such as the hardware/software divide, are of limited utility in understanding the design and function of intracellular signal processing networks.

We focus on how the theory of biochemical reaction networks can help to unravel the complex, protein-based information processing units inside a cell. Recent advances in thinking about biochemistry in terms of computational metaphors[2] showcase some exciting new trends in modeling and analysis. To help readers understand these ideas, the "Biology for Computer Scientists" sidebar provides information about basic biological concepts.

COMPUTING USING BIOCHEMICAL REACTIONS

To illustrate how biochemical reactions constitute computational elements, we describe three basic forms of information processing: memory (information storage), oscillation (sustained cycling through a sequence of states), and pattern selection (deciding on a course of action by expressing preferences over possible inputs). Most of our understanding of biochemical computation arose through careful mathematical modeling using ordinary differential equations (ODEs). We can use numerical simulation and mathematical analysis (dynamical systems theory) to characterize the solutions of these ODEs and compare them to the physiological behaviors of cells.

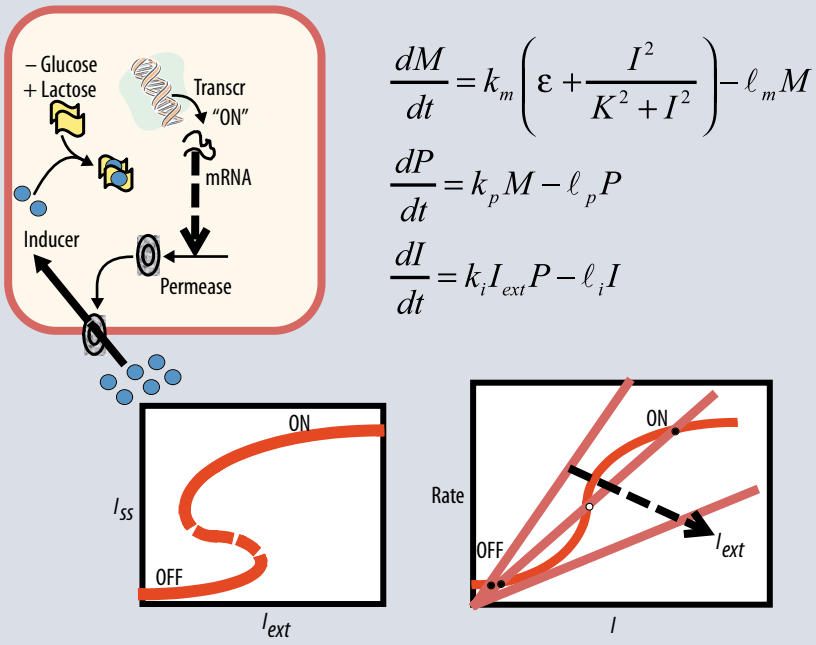


Figure 1. The *lac* operon in bacteria. A stretch of DNA (top left) encodes several proteins involved in the utilization of lactose, including a permease that lets lactose enter the cell. If glucose is available to the cell, the repressor (the yellow flags) binds to the *lac* operon and prevents expression of the genes. If glucose is absent and some lactose molecules (blue circles) reach the cell, then these molecules bind to the operon and inactivate repressors.

Bistability

Cellular information processing units receive noisy, analog input signals and must often compute a digital yes/no output response. For example, a bacterium might sense the concentrations of two sugars, glucose and lactose, in its environment and decide whether to synthesize the enzymes necessary for digesting lactose. A virus might sense the level of stress in its bacterial host cell and decide whether to hide away in the bacterium's DNA (lysogenic phase) or make many copies of itself and burst out of the host (lytic phase). A dormant frog egg might sense the level of progesterone in its vicinity and decide whether to remain dormant or prepare for fertilization. An irradiated skin cell could sense the extent of DNA damage it has received and decide whether to attempt repairs or initiate its preprogrammed suicide response, called apoptosis. Molecular reaction networks often make these types of binary decisions, exhibiting alternative, stable steady states—bistability—as a function of a continuous input signal.

Having two or more stable steady states enables a biochemical network to maintain a memory. Just as a computer memory unit can store a 0 or 1 through electrical or magnetic elements, a biochemical switch can be in an ON or OFF state, where each state is modeled as concentrations of chemical species. An understanding of bistability in biochemical reaction mechanisms is thus crucial to understanding how a protein network encodes information.

Biologists have uncovered many instances of bistability in cellular networks. The bacterial *lac* operon, shown in Figure 1, provides the paradigmatic example. The proteins the bacterium needs to import and digest lactose (a type of sugar) are contained on this operon, transcription of which is controlled by a repressor protein. Lactose is used as an energy source by the cell, specifically by digesting it to form glucose.

The OFF state for this switch occurs when glucose is already available to the cell. In this state, the repressor is active, whether or not lactose is also present in the external medium. Once glucose is exhausted, if lactose is available some of it sneaks into the cell and inactivates the repressor. As a result, the genes begin to be transcribed to messenger RNA. The mRNA instructs the synthesis of some permease proteins, which help let more lactose into the cell. These lactose molecules inactivate more repressors and allow even faster production of permease, leading to the ON state. Thus the positive feedback in this circuit (Lactose ⊣ Repressor ⊣ Permease → Lactose) creates the potential for bistability as a function of the lactose concentration in the external medium. Hence, we can think of lactose concentration as the inducer signal that turns the switch from OFF to ON.

This simple regulatory circuit, shown in Figure 1, can be modeled by three ODEs, for M = [mRNA], P = [permease], and I = [inducer]. I_{ext} = concentration of inducer in external medium (a constant). The k's and l's are rate constants

for synthesis and degradation, respectively, and K is the dissociation constant for inducer binding to repressor. The steady states of this set of ODEs are given by solutions of the following algebraic equation:

$$\phi \frac{I}{I_{ext}} = \varepsilon + \frac{I^2}{K^2 + I^2} \text{ where } \phi = l_i l_p l_m / k_i k_p k_m.$$

The right-hand side of this equation is a sigmoidal function of I, shown by the red curve in Figure 1, and the left-hand side is a straight line, shown in purple, whose slope decreases as I_{ext} increases. Clearly, the operon is OFF (I small, repressor active) when I_{ext} is sufficiently small, and the operon is ON (I large, repressor inactive) when I_{ext} is sufficiently large. For intermediate values of I_{ext}, the operon can be either ON or OFF (bistable). By plotting the steady-state level of I as a function of I_{ext}, Figure 1 shows how the number of steady states depends on I_{ext}. We categorize this regulatory mechanism as a "toggle switch" that converts a continuous input signal (I_{ext}) into a discontinuous output response (I_{ss}).

Researchers have proposed several candidate bistable switches for the synapse, with direct experimental evidence for at least two.

The cell suicide response apoptosis works similarly. Every mammalian cell contains death-dealing enzymes called caspases whose activities are controlled by suites of proapoptotic (P) and antiapoptotic (A) proteins. P and A proteins create a mutually antagonistic feedback loop that creates two stable steady states: "A ascendant" or "P ascendant." Normally, the A proteins are ascendant, and the caspases are kept inactive. But in response to a sustained level of DNA damage, the switch flips irreversibly to the P-ON state, the caspases are activated, and the cell disassembles itself by degrading its own proteins, nucleic acids, and organelles.

Synaptic memory

Neurons exchange information through tiny contacts called synapses. A typical neuron has about 10,000 synapses, and the human brain hosts about 10^{15} synapses. Each synapse is a complex molecular machine packed into a structure about half a micron in diameter. Current experimental studies and theoretical models suggest that modulation of synapses' connection strengths forms the cellular basis for memory. Bistable switches seem to be obvious candidates for controlling synaptic strength.

The requirements for synaptic switches are quite stringent, however, and determining how molecular switches can fit all the known conditions presents a continuing challenge. First, electrical stimuli as short as one second can trigger changes that last for days, and some synaptic changes presumably sustain the information for a lifetime. Paradoxically, the lifetime of individual synaptic molecules spans from minutes to, at most, days. Further, a free molecule can diffuse away from the synapse in just a few seconds. Finally, the synaptic volume—0.1 femtoliters— is so small that there are perhaps only five free calcium ions and a few tens to hundreds of other key signaling molecules. With molecules in such small numbers, random fluctuations might cause the switch to flip spontaneously and thus not retain its memory.

Given all these engineering difficulties, a synaptic molecular switch seems an extremely unpromising substrate for stable memories. Nevertheless, researchers have proposed several candidate bistable switches for the synapse, with direct experimental evidence for at least two of them.

The first putative switch involves the protein Calcium/Calmodulin-dependent protein kinase II (or CaMKII for short), present at very high levels in the synapse. Its baroque chemical regulatory mechanisms include the capacity for self-activation that, in some models, leads to bistable behavior. Further, the switch responds rapidly to strong synaptic input. Its long-term state stability in the presence of signaling noise raises more debate, but some models predict that its spontaneous switching time could exceed a century.[3]

Another experimentally supported synaptic switch, based on the mitogen-activated protein kinase (or MAPK for short) signaling network shown in Figure 2, also involves a positive feedback loop.[4,5] This switch can also occur in a completely different context: the decision about whether or not a cell should grow and divide. Making this decision incorrectly is a key step in the formation of cancer cells.

Which mechanism really supports synaptic memory? Quite likely all of them. We know, for example, that different molecules are critical at different stages of memory consolidation following an initial trigger. Based on such studies, we might conclude that CaMKII is needed for short-term storage, MAPK for longer-term memories, and very likely yet more stages remain to be worked out.

Oscillators

In some cases, a cell's proper response is not a simple yes or no answer but an oscillatory output. For example, hormone-secreting cells often release their hormones in periodic pulses. Newly fertilized eggs undergo a rapid, periodic alternation of DNA replication and cell division. Bread mold mycelia generate a 24-hour rhythm of gene-regulatory proteins that trigger the production of

fruiting bodies at the optimal time of day. Vertebrae and musculature are laid down in a periodic traveling wave during the development of a fish embryo.

The paradigmatic oscillatory response in cellular systems is the circadian (~ 24 hour) rhythm in fruit flies, as shown in Figure 3. A genetic locus (per) encodes mRNA (M) for the PER protein (P). PER must combine with a partner protein, CRY, to form a heterodimer (Q), which moves into the nucleus and there serves as a repressor (R) for transcribing the per gene. The negative feedback in this loop ($M \rightarrow P \rightarrow Q \rightarrow R \dashv M$) creates the potential for sustained oscillations, as Figure 3 shows. For intermediate values of l_q, the steady state is unstable and surrounded by a stable "limit cycle" oscillation. The dark black line in the figure indicates the maximum and minimum values attained by $[PER]_{tot}$ on the periodic solution. The vertical red line indicates the value of l_q.

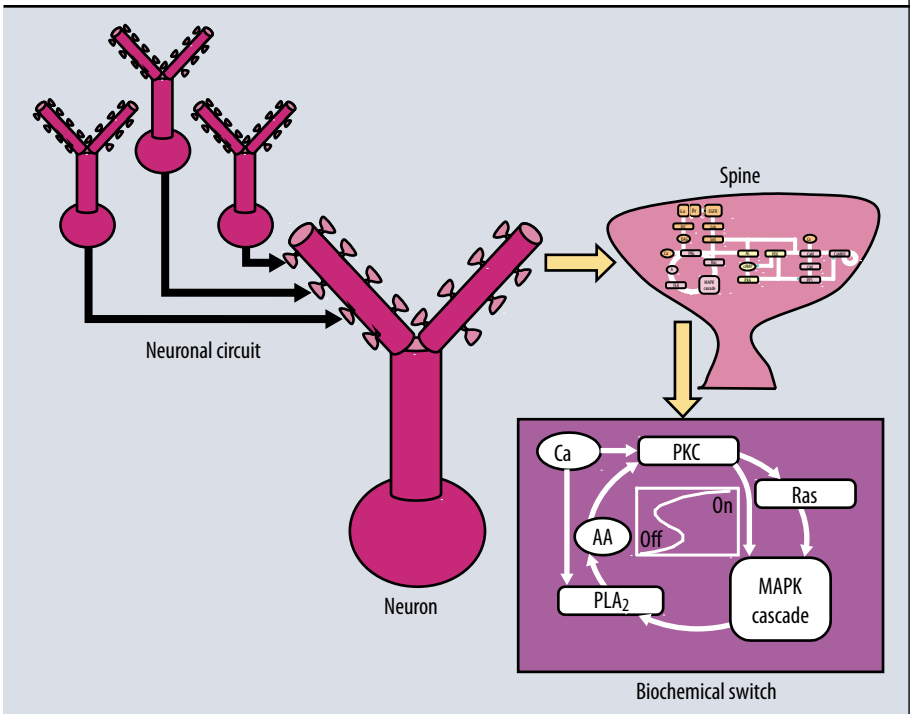


Figure 2. The role of chemical switches in memory. Neuronal circuits involve interconnections between neurons, and memories can be stored in the "weights" of these connections. Connections terminate on synapses, which are often located on 0.5-micron diameter mushroom-shaped structures called synaptic spines. Each spine houses a complex chemical circuit, which includes different kinds of putative bistable biochemical switches. Specific patterns of synaptic input can change the state of these switches, leading to a cascade of chemical changes that alter the synaptic weight.

Just as switches arise in networks with positive feedback loops, oscillations arise in biochemical reaction networks with negative feedback and some sort of memory.[6] Negative feedback is actually a good design for homeostasis—buffering steady-state levels against changes in input signals or network constants. However, if the delay in the negative feedback loop is too long or the loop has some memory due to an embedded positive feedback loop, the negative feedback loop might break out into unexpected oscillations. If homeostasis is the goal, natural selection will weed out the pathological oscillatory states. But, just as likely, evolution could select for oscillations in situations where temporal periodicity is beneficial—in predicting the time of sunrise or sunset, for example.

A variety of mechanisms can achieve negative feedback in protein regulatory networks. As Figure 3 shows, a protein may repress the transcription of its own gene. Alternatively, protein X might activate protein Y, which promotes the degradation of protein X, as is the case for the tumor-supressor protein p53 (X) and its regulator Mdm2 (Y), and for mitosis promoting factor (MPF = X) and its regulator Cdc20 (Y). "Delay" can be provided by a chain of intermediate states, such as p53 in nucleus → MDM2 mRNA in nucleus → MDM2 mRNA in cytoplasm → Mdm2 protein in cytoplasm → Mdm2 in nucleus → p53 degradation in nucleus. A positive feedback loop such as MPF → Cdc25 → MPF can provide "memory," operating in conjunction with a negative feedback loop (MPF → Cdc20 ⊣ MPF). The positive feedback loop creates a bistable switch that the negative feedback loop flips ON and OFF periodically.

Pattern selectivity

There is more richness to protein computation than toggle switches and oscillators. Pattern selectivity refers to the ability of cellular responses to be tuned to specific stimulus patterns. Embryogenesis, the process by which an embryo develops and specializes into a complex organism with diverse body parts, relies heavily on pattern selectivity. Similarly, in neuronal circuits, the primary input a synapse receives is a pattern of electrical impulses from an upstream neuron, and it must be able to decode this input pattern.

Decoding happens in at least three ways. Some patterns simply pass through the synapse to the nerve cell body, where they are combined with other synaptic inputs to determine whether the nerve will remain at rest or

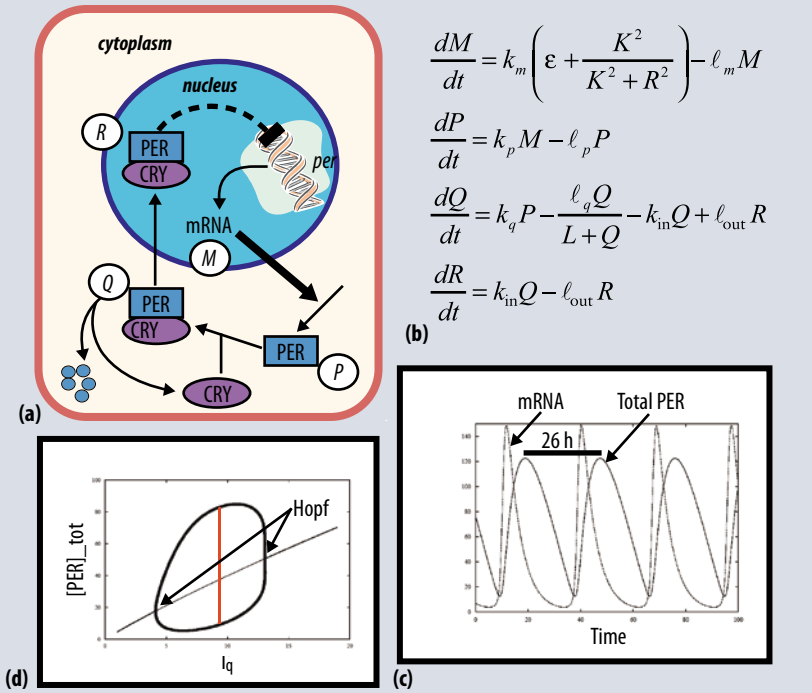


Figure 3. The *per* oscillator in a fruit fly. (a) Expression of the *per* gene drives production of the PER protein in the cytoplasm, where it binds to CRY. The dimer moves back into the nucleus, where it represses transcription of the *per* gene. (b) This simple negative-feedback loop can be modeled by four ODEs, for M = [mRNA], P = [PER], Q = $[PER:CRY]_{cytoplasm}$, and R = $[PER:CRY]_{nucleus}$. (c) The k's and l's are rate constants, and K and L are dissociation constants. For an appropriate choice of the rate constants and dissociation constants, these ODEs support sustained oscillations with a period close to 24 hours. (d) The steady state level of $[PER]_{tot}$ = P + Q + R as a function of the rate constant for degradation of CRY.

generate an output action potential. Other patterns, in addition to passing along to the nerve cell body, also cause sustained increases in synaptic weights, possibly by turning on the synaptic plasticity switches. A third category of patterns does the converse, weakening synapses, possibly by toggling the switch off.

These options are only a small subset of what a synapse can do. It also can handle short-term changes that play a role in working memory, and it can take into account the activities of other synapses and of the cell as a whole. As with synaptic switches, the emerging picture shows many different players in the molecular network that processes patterns of synaptic inputs to give rise to different forms of synaptic plasticity. Each may handle some part of the tuning to specific patterns and some subset of output pathways. Some networks, including the ubiquitous MAP kinase pathway, have been shown to exhibit temporal tuning to bursts of activity that repeat on a five-minute timescale.[7] Others, including calcium-responsive molecules upstream of CaMKII, may be tuned to strong but brief inputs on the timescale of seconds.

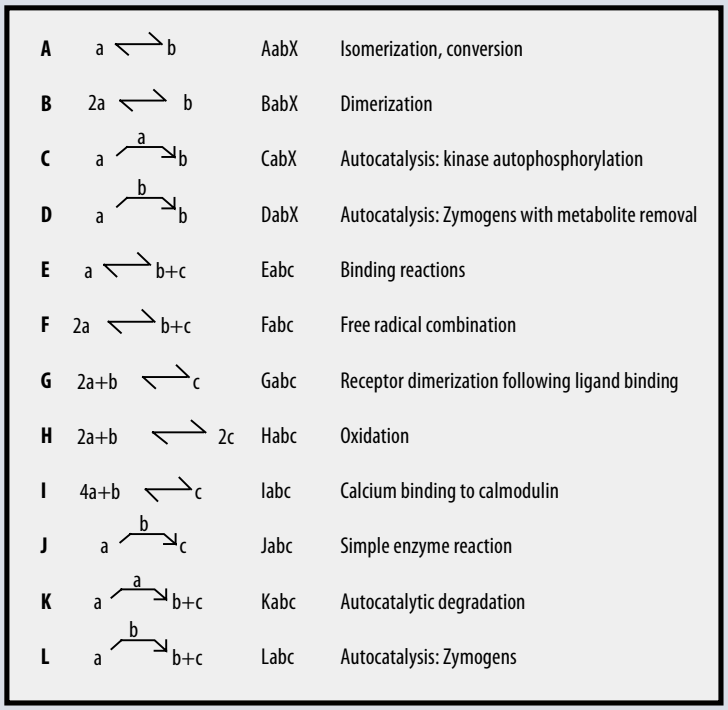
In describing some simple examples of computations performed by protein regulatory circuits, as shown in Figures 1 through 3, we have gotten by with simple analysis and simulation tools. But as the circuits get more complicated, traditional tools become cumbersome and they are plagued by human error. Fortunately, researchers are currently developing a variety of new algorithms to facilitate the construction, simulation, and analysis of molecular regulatory networks.

One class of tools focuses on a specific mathematical formalism (such as ODEs) and automates everything from model building to simulation, parameter estimation, and comparison to experimental datasets. Examples include Copasi (www.copasi.org), Cell Designer (www.celldesigner.org), Virtual Cell (www.nrcam.uchc.edu), SBW (http://sbw.sourceforge.net), and JigCell (http://jigcell.biol.vt.edu). The development and adoption of the Systems Biology Markup Language (SBML) (www.sbml.org) standard for representing biochemical networks have greatly aided interoperability between software components. Another class of tools emphasizes "network theories" that, similar to electrical circuit theory, analyze the dynamic behavior emerging from a network of reactions among proteins.[8] The development of biological circuit module catalogs[9,10] provides yet another trend, thereby supplying a parts manual for understanding biological function, much as electrical engineers hook together simple electrical circuits to achieve complex functions. With the advent of massively parallel clusters, the parts catalog approach has become particularly attractive.

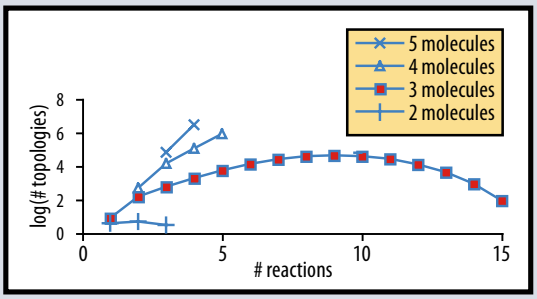
ELEMENTS OF BIOCHEMICAL COMPUTATION

How many different ways can a group of (bio)chemical reactions come together to realize a particular functional element, such as a switch or an oscillator? The research community knows that positive feedback loops are necessary but insufficient for realizing bistability, but how exactly are these loops implemented in real biochemical circuits? A recent study [10] sheds some light on how chemical reactions might join to form a functional switch.

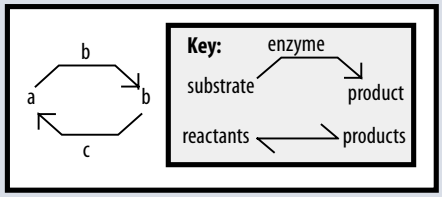
As Figure 4 shows, we adopt a building-block approach to exploring the mechanisms of biochemical switches by using 12 basic biochemical reactions widely known to



(a)



(b)



(c)

Figure 4. Identifying switches by searching through a space of chemical reactions. (a) The 12 possible chemical reactions considered in the study. (b) For a given number of molecules, the number of legal topologies first increases with the number of reactions, then declines due to symmetries and incompatibilities. (c) The smallest switch possible from the reactions in (a) involves three protein molecules—a, b, and c—and two reactions, types D and J.

occur in living cells to exhaustively generate all unique, chemically consistent combinations of reactions. We then test each resulting circuit for switch-like behavior. This approach gives us an unbiased view of how simple chemical reactions might possibly group together to realize emergent bistable behavior.

Figure 4a depicts the 12 basic reactions used in our study. The molecule names—such as a, b, and so on—are simply placeholders for actual protein names. Thus, reaction type C means that a reactant (here a) serves to catalyze its own conversion to a product (here b). The a's and b's in this reaction can be uniformly interchanged and we would continue to have an instance of reaction type C, except now the reactant is b and the product is a. When we put these reactions together, as Figure 4b demonstrates, the number of possible, unique, topologies first rises and then sharply declines due to symmetries and constraints on interactions between reactions.

We then converted each unique configuration to an ODE model and tested for the possibility of two or more steady states, using Monte Carlo sampling over a wide range of possible parameters. In addition to brute-force simulation of the ODEs from a variety of initial conditions, we also employed a numerical homotopy-continuation method that solves the underlying system of nonlinear algebraic equations.[10]

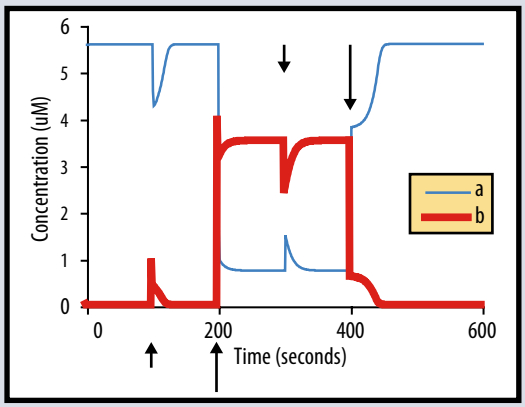


Figure 5. Exploring the switch-like behavior of the circuit in Figure 4c. The two states of the switch are modeled through the concentrations of the molecules a and b. The switch begins in an initial state, is switched to a new state at the 200-second mark, and then switched back at 400 seconds.

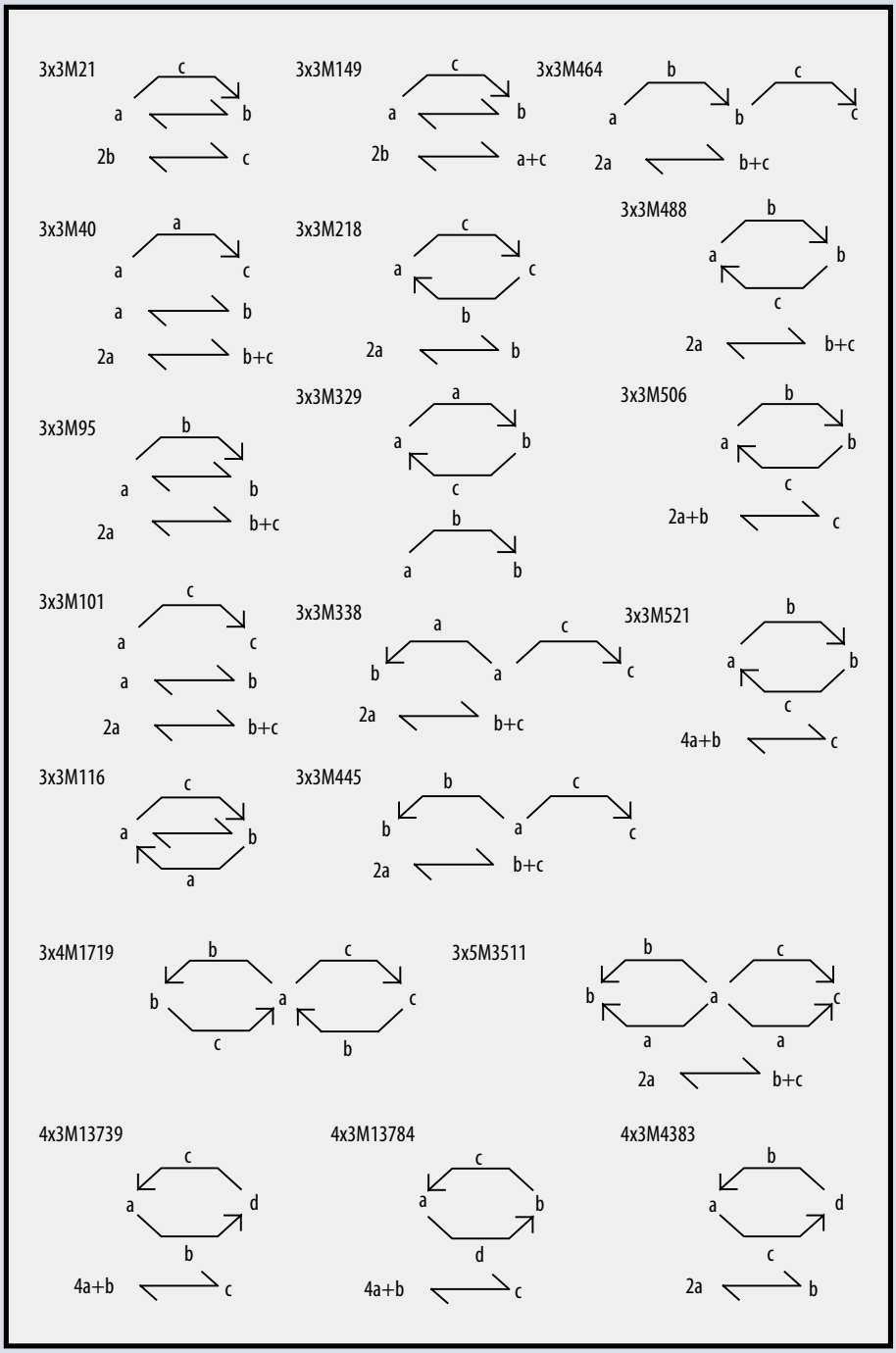


Figure 6. A few of the thousands of switches obtained through computational exploration. Each switch is composed of the basic reactions from Figure 4a and receives a unique identifier. For example, the identifier 3x5M3511 refers to system 3511 for the model class of networks involving three molecules and five reactions.

We evaluated more than 40,000 configurations in this manner, a massive computation that took more than 100 CPU years on Virginia Tech's System X supercomputer. Figure 4c depicts the simplest switch discovered through this study, one that involves three molecules and just two reactions. The concentrations of molecules a and b encode the switch's state—thus (a HIGH, b LOW) encodes a 1 and (a LOW, b HIGH) encodes a 0.

Small perturbations, shown as small arrows in Figure 5, return to the originating stable state. Large perturbations, shown as large arrows in Figure 5, cause state transitions.

Our entire catalog of switches revealed nearly 4,500 topologies, some of which we show in Figure 6. The diversity of switches found in this study reveals that while positive feedback is a good metaphor for bistability, the actual ways in which biochemical circuits realize bistability can be quite intricate.

One striking observation from our study is that the switches we discovered, although different, relate to one another in unexpected ways. We visualized the relationships among our switches by drawing an edge from one to another if one mechanism could be considered an extension of another by the addition of one or more reactions. These relationships constitute a directed acyclic graph (DAG)—that is, a multiply-rooted banyan-tree-like structure such as the one shown in Figure 7. This demonstrates that many complex switches actually derive from simpler ones, an observation that suggests bistable systems involving small numbers of molecules can form the architectural core of more complex bistable reaction networks. Further, the DAG suggests that the incremental addition and deletion of reactions could be the mechanism whereby

evolution might have stumbled upon the complex bistable switches observed in present-day organisms.

RESEARCH AGENDA

Protein-based computers are ripe for spectacular scientific advances, because we know a great deal about these circuits' molecular components. But we are only beginning to understand how they process information and make decisions. Protein regulatory networks have novel information-processing capabilities that might inspire new control paradigms in traditional engineering contexts such as biomimetics. Moreover, they are foundational to many physiological processes of fundamental importance in human health, biotechnology, and defense, such as trauma, wound healing, immune responses, and circadian rhythms.

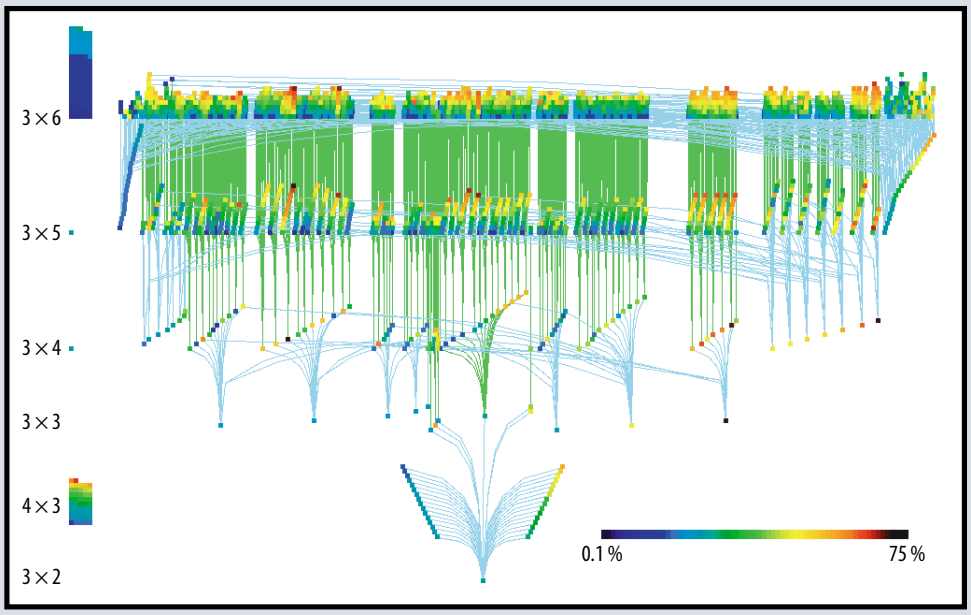


Figure 7. Relationships between switches can be represented as a directed acyclic graph (DAG). Each square in this figure denotes a biochemical switch. Switches are connected if one can be viewed as an extension of the other through the addition of one or more reactions. Node colors refer to the configuration's propensity to function as a switch, based on its parameters. Observe how the DAG is multiply rooted but most of the larger configurations derive from smaller ones.

It may soon be possible to reengineer biochemical circuits to our own specifications. The emerging subdiscipline of synthetic biology[11] promotes such a biology-as-engineering approach. Using the methods of genetic engineering in bacteria, pioneering scientists have created an artificial genetic toggle switch[12] and an artificial negative-feedback oscillator.[13] This technology, now quite well advanced in bacteria, is rapidly developing in yeast and higher organisms. These early successes of synthetic biology provide prime examples of the utility of mathematical models for gene-protein regulatory networks, as surveyed here. We propose the following key research tasks for the next decade.

Improve experimental methods

First, we need improved experimental methods to poke into a cell and ascertain its molecular state so that we can obtain a more complete picture of cellular processes' molecular dynamics. Quantitative microscopy of fluorescently labeled proteins is beginning to generate exactly this sort of information. From this data, theoreticians will be able to create more realistic mathematical models of the underlying circuitry and close the gap between models and reality. System identification techniques from control theory will be valuable here in systematically probing networks to yield a critical understanding of their functionalities.

Design better computational tools and algorithms

Second, there is an acute need for better computational tools and algorithms for biochemical network construction, simulation, analysis, comprehension, and comparison to experiment. As a community of systems biologists, we need to build upon our initial successes to design better theories and tools for comprehending massively complex circuits. We must continually raise the bar of complexity for which we can understand the information processing machinery of the living cell. A combination of data mining, numerical simulation, and modeling capabilities will be necessary to provide a comprehensive set of tools for this purpose.

Understand the nature of biochemical computation

Finally, unlike computations in well-engineered circuits, there is sloppiness, cross-talk, and noise in the computations that cells carry out. A key challenge will be to understand how a cell survives and carries out robust decision-making in the face of such distractions. New modeling tools that acknowledge and even exploit the uncertain, stochastic, and imprecise nature of biochemical computation will see increasing demand.

We hope to have conveyed our excitement for studying the information-processing capacities of biochemical circuits and our desire to see computer scientists give serious attention to their unique capabilities as computational elements. We have detailed only a few basic elements of biochemical circuitry here, but they serve as useful metaphors for comprehending complex circuits, such as the eukaryotic cell cycle.[14]

Advancing the agenda proposed here will require overcoming some cultural impediments. Although computational molecular biology—bioinformatics—has been a highly successful collaboration of molecular biologists and computer scientists, the field of computational cell biology—mathematical models of the molecular regulatory systems underlying cell physiology—remains a contentious area of experimentalists, theoreticians, and computer scientists' divergent opinions. Heated debates flare about the merits of high-throughput data collection versus focused experimentation, about top-down versus bottom-up modeling, and about ODE models versus discrete models such as Boolean networks or Petri nets. Some influential cell biologists even consider the whole modeling enterprise a colossal waste of time and money.

We propose that the more heated these debates become, the less light they shed on the subject. The field of quantitative cell biology is expansive and vibrant enough to accommodate a range of different approaches. The future is promising but the path forward remains unclear. A variety of experimental and modeling methods will likely be needed to light the way. We have only begun to understand the intricacies of protein-based computing machines. C

References

1. J.R. Whitlock et al., "Learning Induces Long-Term Potentiation in the Hippocampus," *Science*, Aug 2006, pp. 1093-1097.
2. D. Bray, "Protein Molecules as Computational Elements in Living Cells," *Nature*, vol. 376, July 1995, pp. 307-312.
3. P. Miller et al., "The Stability of a Stochastic CaMKII Switch: Dependence on the Number of Enzyme Molecules and Protein Turnover," *PLoS Biology*, vol. 3, no. 4, Mar. 2005, e107, 13 pp.
4. U.S. Bhalla, P.T. Ram, and R. Iyengar, "MAP Kinase Phosphatase as a Locus of Flexibility in a Mitogen-Activated Protein Kinase Signaling Network," *Science*, Aug. 2002, pp. 1018-1023.
5. K. Tanaka and G.J. Augustine, "A Positive Feedback Signal Transduction Loop Determines Timing of Cerebellar Long-Term Depression," *Neuron*, Aug. 2008, pp. 608-620.
6. B. Novak and J.J. Tyson, "Design Principles of Biochemical Oscillators," *Nature Reviews Molecular Cell Biology*, vol. 9, Dec. 2008, pp. 981-991.
7. S.M. Ajay and U.S. Bhalla, "A Role for ERKII in Synaptic Pattern Selectivity on the Time-Scale of Minutes," *European J. Neuroscience*, Nov. 2004, pp. 2671-2680.
8. G. Craciun, Y. Tang, and M. Feinberg, "Understanding Bistability in Complex Enzyme-Driven Reaction Networks," *PNAS*, June 2006, pp. 8697-8702.
9. J.J. Tyson, K.C. Chen, and B. Novak, "Sniffers, Buzzers, Toggles and Blinkers: Dynamics of Regulatory and Signaling Pathways in the Cell," *Current Opinion in Cell Biology*, vol. 15, no. 2, 2003, pp. 221-231.
10. N. Ramakrishnan and U.S. Bhalla, "Memory Switches in Chemical Reaction Space," *PLoS Computational Biology*, vol. 4, no. 7, July 2008, e1000122, 9 pages.
11. E. Andrianantoandro et al., "Synthetic Biology: New Engineering Rules for an Emerging Discipline," *Molecular Systems Biology*, vol. 2, no. 2006.0028, May 2006, 14 pp.
12. T.S. Gardner, C.R. Cantor, and J.J. Collins, "Construction of a Genetic Toggle Switch in Escherichia Coli," *Nature*, vol. 403, 2000, pp. 339-342.
13. M.B. Elowitz and S. Leibler, "A Synthetic Oscillatory Network of Transcriptional Regulators," *Nature*, vol. 403, 2000, pp. 335-338.
14. J.J. Tyson, A. Csikasz-Nagy, and B. Novak, "The Dynamics of Cell Cycle Regulation," *BioEssays*, vol. 24, no. 12, 2002, pp. 1095-1109.

Naren Ramakrishnan *is a professor and the associate head for graduate studies in the Department of Computer Science at Virginia Tech. His research interests include computational science, especially computational biology, mining scientific data, and information personalization. Ramakrishnan received a PhD in computer sciences from Purdue University. He is a member of the IEEE Computer Society, the ACM, and the AAAI, and serves on* Computer's *editorial board. Contact him at naren@cs.vt.edu.*

Upinder S. Bhalla *is an associate professor at the National Centre for Biological Sciences, India. His research interests are in systems biology and experimental and computational neuroscience. Bhalla received a PhD in biology from the California Institute of Technology. He is a member of the Society for Neuroscience and the International Society for Systems Biology and a Fellow of the Indian Academy of Sciences. Contact him at bhalla@ncbs.res.in.*

John J. Tyson *is a university distinguished professor in the Department of Biological Sciences at Virginia Tech. His research interests are network dynamics and cell physiology, especially regulation of the cell division cycle. Tyson received a PhD in chemical physics from the University of Chicago. He is a past president of the Society for Mathematical Biology and a former co-chief editor of the* Journal of Theoretical Biology. *Contact him at tyson@vt.edu.*