5

Signal processing by biochemical reaction networks

ADAM P. ARKIN

5.1 Introduction

One cannot help but be impressed by the engineering, by evolution, of the cellular machinery. The cellular program that governs cell cycle and cell development does so robustly in the face of a fluctuating environment and energy sources. It integrates numerous signals, chemical and otherwise, each of which contains, perhaps, incomplete information of events that the cell must track in order to determine which biochemical subroutines to bring on- and off-line, or slow down and speed up. These signals, which are derived from internal processes, other cells and changes in the extracellular medium, arrive *asynchronously* and are *multi-valued*; that is, they are not merely 'on' or 'off' but also has a *memory* of signals that it has received in the past, and of its own particular history as written in the complement and concentrations of chemicals contained in the cell at any instant. These characteristics of robust, integrative, asynchronous, sequential and analog control are the hallmark of cellular control systems. Below, arguments will be made that there is also another characteristic of such control systems: there is often an *irreducible nondeterminism* in their function that, besides leading to differences in timing of cellular events across an otherwise genetically identical (isogenic) cell population, can also lead to profound differences in cell fate. The circuitry that implements these control systems is a network of interconnected chemical reactions. Included in these reactions are the genetic reactions involving: the gene expression reactions such as transcription initiation, transcript elongation, and translation; gene rearrangements such as DNA inversion reactions; and epigenetic control reactions such as DNA methylation. Enzymatic reactions, biosynthetic and mechanochemical interactions, and a host of other chemical reaction types also are central elements of this control system.

This chapter focuses on biochemical systems in which spatial concentration distributions and purely mechanical interactions are ignored or are not present. A full specification of cellular function is gained by (1) the determination of all the chemical parts making up the system, (2) the deduction of the mechanisms of interactions between these parts, and (3) the designation of the parameters necessary to describe the physics of all of these mechanisms. However, this specification, which theoretically could produce a computer simulation that exactly predicts the temporal behavior of all cellular constituents, represents a full understanding of the cellular system no more than a fully specified model of a Pentium chip gives us an understanding of the principles of its designs and function. In the case of the Pentium, understanding is best achieved by grouping individual transistors into logic gates, gates into devices such as counters, registers, and amplifiers, and then these devices into large devices such as arithmetic logic units, multiplexers/demultiplexers, clocks and bus controllers. The function of the chip can then be described by a relatively high-level programming language that makes clear the interactions between these composite devices and allows for a vastly simplified mathematical analysis. To achieve progress toward the description of cellular function on a similar basis is the ultimate motive for the work presented here.

5.1.1 Research goals

The work described herein represents efforts whose goal is the deduction of the engineering principles and logic of large biochemical reaction networks (BRNs). Specifically, the capacity of BRNs is explored (1) to sense and respond to multiple time-varying and conflicting signals (often chemical concentrations) in a robust and timely manner as well as (2) to execute internal developmental and behavioral programs. Two complementary types of analysis are presented. First, the utility of the circuit analogy for BRNs is examined and discussed. For this purpose, the device physics for a number of recurrent regulatory architectures is outlined to provide some background. In addition, the role of *thermal noise* in determining chemical reaction outcomes in cells is shown to be significant, especially for reactions involving genetic material. All these analyses assume that the individual components and their interactions have already been identified. However, this is often not the case. Thus, in addition, I briefly describe experimental methods for deducing BRN structures and assigning groups of chemicals into composite devices. The methods are designed to produce these deductions from measurements on the whole chemical reaction system rather than by breaking the system into small pieces.

Methods of network deduction and analysis are of special importance now that many genome projects are completing the inventory of all of the cellular proteins and genetic regulatory systems. If the full promise of these projects, i.e., to uncover the program of cellular life, is to be realized, it is necessary to compose these parts into functioning networks whose temporal behavior we may understand, whose properties we can control, and whose failures we can diagnose and ameliorate. Analytical tools such as the ones described herein, and in other contributions to this volume, lay the groundwork for this endeavor.

5.2. The circuit analogy and network analysis

In biology we are faced with often very complicated networks of interacting components. Ignoring atomic levels of detail, the lowest-level 'devices' in a BRN are often the individual chemicals and the set of reaction channels. Perhaps the best-characterized biochemical network of such devices is intermediary metabolism. Figure 1a is taken from Peter Karp's and Monica Riley's EcoCyc database of *Escherichia coli* (*E. coli*) metabolism, and is a representation of this network wherein every circle in the diagram represents one of the small organic molecules transformed in the course of metabolism, each black line represents a (possibly reversible) chemical reaction that converts one set of small organics into another (Karp et al., 1999). Each grey line indicates that the connected circles are the same species of molecule that appears in multiple pathways. The macromolecules and macromolecular complexes that catalyze these conversions are not shown nor are the regulatory interactions that allow combinatorial control of the rate of one reaction by a set of other chemicals in the network. Were these components to be included in the diagram it would resemble Figure 1b; the diagram would be black with interactions. This highlights the main difficulty in forming a qualitative understanding of how metabolism actually works: biochemical systems are highly nonlinear and interconnected and composed of large numbers of chemical components. It is natural, therefore, to look for other systems that have these properties and ask whether we can apply the tools developed for their analysis to biochemical systems. Because diagrams of BRNs bear some resemblance to *electronic* circuits, it is tempting to ask whether it is possible to map them onto analogous electrical or electronic circuits so that we may apply the well-developed methods of electrical circuit analysis, the theory of computation and Boolean algebra. The subsequent section discusses some current methodology for the analysis of BRNs, and similarities to and differences from electrical engineering analyses.

Mass Transfer in Metabolism
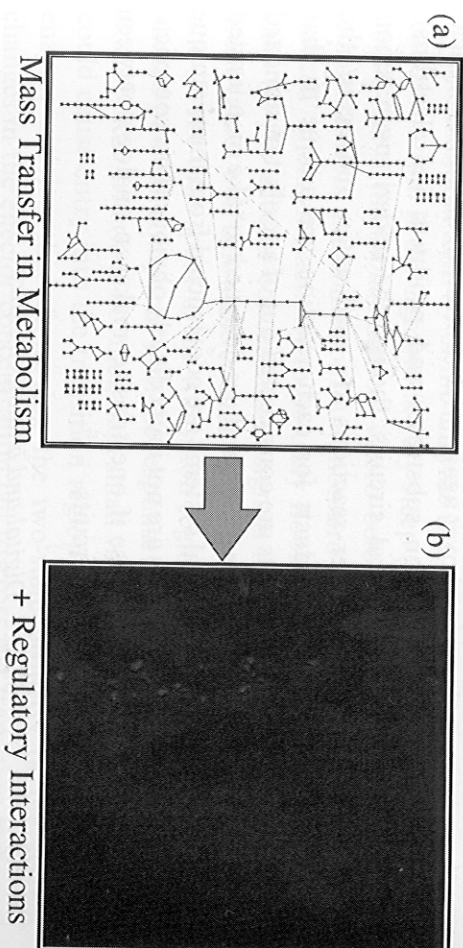
(a)

+ Regulatory Interactions

(b)

Figure 1. A biochemical reaction network. (a) The left panel shows all the mass transformation reactions of *E. coli* intermediary metabolism from EcoCyc (Karp et al., 1999). Details of the enzymatic reactions, the enzymes themselves and the regulatory interactions are not shown. If they were to be included, the vast number of interactions would make the diagram appear black as suggested in the right panel (b).

5.2.1 Comparisons of electrical and chemical networks

In this chapter, the words 'circuit' and 'network' are used somewhat interchangeably to mean a group of elements that have some property that is affected dynamically by interaction with other elements. In electrical circuits, these elements are, for example, resistors, capacitors, wires, and power sources. A signal in these networks is a voltage (or current) received at some node in the network, such as one lead of a resistor. Different signals are distinguished by the different points at which they impinge within the circuit and sometimes by their temporal pattern. In all cases, the currency of signals in electrical networks is carried by electrons. In chemical circuits, elements are, for example, enzymes, ions, reaction channels, and DNA. A signal in a chemical network is composed, most often, of the appearance of an amount of some chemical species at a point in the network such as at an allosteric regulatory site of an enzyme. The currency of signals in chemical networks is not uniform. It is the individual concentrations and chemical potentials of the particular species interacting at that point in the network. This lack of common currency for information transfer is one of the complications of BRNs: different chemical signals are often distinguished only by their interaction specificities with other members of the network.

This suggests another one of the significant complicating differences between electrical and chemical circuits. In electrical circuits, component types

(e.g., resistors) are used over and over again, with or without the same physical parameters, but sharing common substrates and products (i.e., electrons). Whereas, in (nonspatial) chemical circuits, though the underlying reaction motifs of first- and second-order reactions are used again and again, the parameters, substrates and products for each reaction are unique to that reaction. That is, electrical circuits are constructed out of a toolkit of standard parts whose physics are well understood and are designed to take on a limited set of values. In contrast, chemical systems are constructed from many unique pieces whose physical parameters are not immediately obvious. This complication in the analysis is even worse if one tries to make an analogy between chemical and digital electronic circuits.

5.2.1.1 Digital and analog circuitry

There are a number of instances when the behavior of a biochemical system is such that it suggests digital rather than analog circuitry. The distinguishing feature of a digital circuit is that the signals within a circuit take on discrete values rather than a continuum. Thus, the major benefit of a mapping of a chemical circuit to a digital one is the reduction in the number of states (state-space) of the constituent concentrations and activities that must be considered for analysis. The powerful machinery of Boolean algebra and digital circuit analysis can then be brought to bear on the problem. In addition, it is computationally more efficient to simulate Boolean networks than to simulate differential equations. At some level, all chemical signals might be considered digital, since their values are discrete; that is, their values are measured in numbers of molecules (per unit volume). However, the number of states available to a chemical signal in a kinetic network is usually far greater than two. We are usually concerned with large numbers of molecules so that individual reaction events cause changes in the number of molecules, which are very small, compared to the average, and thus can be approximated as infinitesimal or continuous changes of a concentration variable. (Important exceptions to this are discussed in Section 5.2.3.3). This is no different from disregarding the quantal nature of electrons flowing through an electrical circuit, because their numbers are so large in metal wires. Even neglecting the discrete nature of chemistry, the dynamics of a chemical system might be such that it is sometimes feasible to treat the system as a digital circuit.

Cooperative enzyme activity The typical example often used to justify Boolean approximations in biochemistry is the cooperative enzyme. For simplicity, consider the cooperativity to be in the action of an allosteric effector on the maximal activity of the enzyme. In this case, for various models of cooperativ-

ity, the maximal activity of the enzyme is a sigmoidal function of the effector concentration. If the sigmoid is very steep (high cooperativity), then it looks much like a threshold function – zero below a critical value of the effector concentration, and some constant, nonzero value above it. However, for this 'switch' in enzyme activity to be truly Boolean (digital) two criteria must be met: (1) the effector concentration signals must cause the effector to transit from values well below the critical value to values well above it (or vice versa), and (2) the time spent in the transition region of the sigmoidal activity curve should be small compared with the response time of the receptive system. If the first of these criteria is not met, then small variations in effector concentration could cause entry into the transition region resulting in large fluctuations in enzyme activity, thus destroying the two-state property of the system. If again the two-state behavior is ill defined. If the second criterion is unfulfilled, then the integrated activity of the enzyme during the transition time might become significant to the receptive process. This means that these intermediate activity values cannot be ignored and so this 'switch' becomes multi-valued at best. (Sometimes this criterion is violated in very fast electronic circuits where in certain components are fast enough to 'see' the transition time of a transistor. Digital designs that fail to take this into account will fail.)

Multiple signal encoding Finally, another complication of applying digital signal analysis to chemical networks of even 'switch-like' reaction mechanisms is the following: unlike most electronic digital systems, for each signal-receptive mechanism, the value of an 'On'-signal is, in general, different from other systems that also might receive that same signal. For example, one kind of enzyme might become active above a calcium ion concentration, $[Ca^{2+}]$, of 10 μM, and another one only above 100 μM. Functionally, then, there are at least three significant values for the $[Ca^{2+}]$ signal: below 10 μM, between 10 and 100 μM, and above 100 μM. The dynamics of the $[Ca^{2+}]$ signal might be such that it is driven very rapidly from 0 to greater than 100 μM; in which case the early sensitivity of the first enzyme is only functionally important if the difference in activation time between the two enzymes is significant. In chemical systems, where there are many different chemical signals whose value ranges and time scales are all unique, discerning the 'logic' of the network is even more difficult and dependent on the exact parameters of the system.

5.2.1.2 Synchronous and asynchronous design

A final point of contrast between engineering principles of electrical or, more specifically, electronic circuits and biochemical ones is the use of synchronous

designs. The majority of sequential (as opposed to combinational) digital designs rely on a *system clock* for synchronization of processes. In digital design, clock synchronization is considered desirable for two reasons. First, susceptibility to noise is improved, since the transient dynamics of component devices that occur before the edge of a synchronizing clock pulse do not affect circuit function. Second, the different delays of various signals through a circuit to their respective outputs can be ignored, since the time between clock-pulses in a synchronous circuit is generally set to be longer than the longest delay; thus outputs are not read until all signals have reached their destination.

With very few exceptions, biochemical circuits are unclocked, that is, *asynchronous*. Even when there seems to be a central oscillator, such as that underlying the timing of the cell cycle, microscopic examination of the process reveals a large variability in timing of the oscillations. Progressive dephasing arises both due to noise in the underlying processes (see Section 5.2.3.3) and the fact that the cell cycle oscillator has check points so that the cycle does not proceed until all necessary subprocesses have completed their work. Since cellularly uncontrolled variables such as externally available nutrients control how fast certain of these processes can be executed, the cell cycle is designed to be tolerant to these metabolically induced large changes in timing. Interestingly, electronic asynchronous design (traditionally used for interface circuits) has become increasingly popular as circuit size and complexity has increased. The reasons cited by electrical engineers for asynchronous circuit design are precisely the reasons a biological circuit would be expected to be asynchronous. Five such reasons are stated by Myers (1995):

Average case performance: The clock period for synchronous systems must be set long enough so that the circuit can accommodate the slowest operation possible even though the average delay of an operation is often much shorter. Asynchronous circuit designs allow the speed of the circuit to change dynamically. The speed of the circuit is, therefore, governed by average case delay.

Adaptivity to processing and environmental conditions: Since variables such as temperature change with the environment, circuit up-time and processing rate, and circuit component speeds can be greatly affected by such changes, synchronous designs must be simulated under a wide range of conditions and the clocking set so that the circuit functions under the widest range of variation. Asynchronous designs, in contrast, are adaptive and speed up and slow down as necessary.

Component modularity: In asynchronous systems, components (functional subcircuits) can be interfaced without the difficulties of synchronizing clocks necessary in synchronous systems. Also, when a new version of a component with different timing is developed, the old component can often be replaced without requiring any

other changes in the rest of the system. In other words, the system is robust to (some) changes in its component circuitry.

Elimination of clock-skew: In large digital circuits, the time it takes a clock pulse to reach different parts of the chip can be different, leading to loss of synchronization. To minimize this skew in arrival times, a great deal of extra circuitry must be designed in. (Nearly a third of the silicon area is required for clock distribution in a DEC Alpha microprocessor.) Asynchronous circuits are tolerant to signal timing differences among components.

Lower system power requirements: Since they do not require all the extra clock circuitry, asynchronous circuits reduce synchronization power. They can also be easily adjusted to make use of dynamic power supplies.

The advantages of asynchrony have to do, then, with robustness to changes in the circuit environment and in the dynamical state of its various components and efficiency both in speed and energy. The noise-filtering behavior of synchronous design is an advantage only because clocks make rejection of noise and transients relatively easy to design. Asynchronous circuits can be designed to be as stable to spurious signals. It is likely that biochemical circuits have evolved for this robustness, efficiency and adaptability to environmental changes. It is not surprising, therefore, that most biochemical circuits are found to be asynchronous. However, even wholly digital, asynchronous circuits are notoriously difficult to analyze. Thus one can expect similar difficulties for analog biochemical circuits. On the other hand, study of biological circuits may provide unthought of stable electronic asynchronous circuit designs and any analytical tools developed for the biological circuits may have application to the electronic ones and vice versa.

5.2.2 Device function and state

Metabolic charts like the one in Figure 1a are daunting in their complexity, but perhaps no more daunting in topological complexity than the schematic for a modern computer chip. The difference between these two interaction maps is that the device physics for every element on the chip schematic are fairly well characterized. The behavior of the circuit is fully specified by these physics and the functioning and reliability of the chip can be probed efficiently using simulation tools such as the *SPICE* software package (Tuinenga, 1995). Even better, because of the precisely designed physical characteristics of these elements, their function may be partially abstracted into a higher-level language: *digital Boolean logic*. Thus, most analyses do not need to include the detailed differential equations that most completely describe transistor function. Instead, the device details can be abstracted to a higher level, i.e. to perform as

logic gates. Circuits composed of such gates can be grouped together to form higher-level devices whose input/output behavior can be derived and used without reference to the exact mechanism from which this behavior is derived. This type of grouping of subnetworks into functional components greatly facilitates the analysis of the larger circuit. One challenge, then, for the analysis of biochemical and genetic networks is to dissect complex networks into individually analyzable devices that can be hooked back together to predict the total system behavior.

What, then, constitutes a device? In electrical circuits, elementary devices are objects such as wires, resistors, capacitors and inductors. In digital circuits, elementary devices are parts such as transistors and gates. Perhaps, the defining property of a device is not that it may be separated physically from its network, but rather that the physics of the device may be derived for a general case without reference to the precise dynamics of the rest of the network in which it might be embedded. Resistors, for example, must always obey Ohm's law and Kirchoff's laws, no matter the circuit in which they are used (within broad limits). These laws, along with perhaps some equations for effects of dissipation on the resistivity of the material, fully specify the device function. Practically, this results in the need for only a single parameter to characterize a resistor, the *resistance*. This value is the same no matter how the circuit elements up- and downstream are functioning.

A single chemical reaction step is an elementary device in a chemical network in much the same way as a resistor is an elementary device in an electrical circuit. A single number may characterize the behavior of the reaction: the *rate constant*. Though it is possible to describe the reaction event in much more detail, via quantum mechanics, collision theory, etc., it is generally not necessary. Just as with the resistor, the rate constant for a given elementary reaction does not depend, to first approximation, on the other reactions going on around it. Vast networks of chemical reactions such as in metabolism or during signal transduction, then, are circuits of these elementary devices in which each device accepts chemical concentrations as inputs, and outputs *chemical fluxes*. This is conceptually different from digital devices that accept voltage and output voltage. Electronic devices, on the one hand, simply feed their output voltage to the 'voltage receptor' on the downstream device. Chemical reactions, on the other hand, convert concentration to flux; the output must be reconverted to concentration for input to the next device. In the elementary case, it is both the upstream and downstream elements together that dynamically determine this conversion of the upstream flux to instantaneous chemical concentration and, thence, the downstream flux.

5.2.2.1 Elementary electronic and chemical devices

The main advantage of a device description is that composite devices may be constructed out of a 'basis' set of elementary devices. A simple, but informative, example from electronics is the *voltage divider*, the most primitive of power supplies (Figure 2a). The output of the voltage divider is a voltage and current, and the voltage is given by $V_{in} R_2/(R_1 + R_2)$. Next to the divider is represented a *chemical analogy* (Figure 2b) that, here, let us call the 'A' buffer. If we assume that the steady-state concentration value of reaction species A, [A], is the output to this device, then its value is given by $k_1[B]/k_2$ (this circuit can both divide and amplify the signal [B]).

The voltage divider can be considered a device only if the circuits driven from its output do not affect its function. This is only the case if the downstream devices, the load, have very high impedances compared to R_2. Figure 2d shows these devices as a single load with resistance, R_L, connected from the output of the divider to ground. This arrangement puts R_L in parallel with R_2; thus the two resistances can be combined into one (as guaranteed by Thevinin's theorem) that has an effective resistance of $R_2 R_L/(R_2 + R_L)$. Thus, when $R_L \gg R_2$ the effective resistance is equal to R_2. In this case, the voltage divider remains an intact device. However, as the load resistance decreases, the value at the output of the device becomes more and more dependent on the properties of the devices to which it is attached. Similarly, in the 'A' buffer (Figure 2b), if A is consumed by a third reaction then this reaction rate must be very small compared to k_2 in order for the chemical device to remain intact. If many other reactions consume A, then the sum of their rate constants must be much less than k_2. On the other hand, consider the case when A is an allosteric effector of a set of enzymes downstream. Assuming binding to the enzyme is a reversible step, the steady-state [A] is unaffected by the interaction with the downstream enzymes. Seemingly, then, this chemical device remains intact when connected to the rest of the network in this way. However, the time it takes to achieve the steady-state value of [A] after 'turning on' the device (by, for example, adding a catalyst required for B-to-A conversion) increases as the concentrations and binding constants of the downstream enzymes increase.
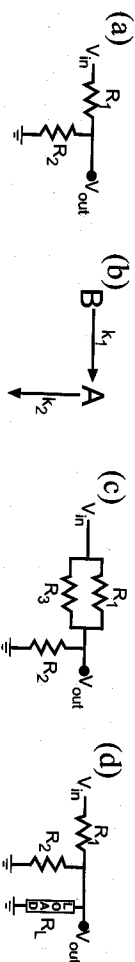
(a) V_{in} — R_1 — V_{out}, R_2 (b) $B \xrightarrow{k_1} A$, k_2 (c) V_{in} — R_1 — V_{out}, R_3, R_2 (d) V_{in} — R_1 — V_{out}, R_2, R_L

Figure 2. Examples of electrical and chemical 'voltage dividers'. For explanations see text in Section 5.2.1.

Thus, some aspects of the device function are changed by connection to the rest of the network. In summary, any analysis of devices dissected from the rest of a large network, therefore, relies on recognizing when these changes are significant and when they are not.

Composite devices like the voltage divider (or the 'A' buffer) serve not only as recognizable units of function but also as a means for *simplifying* circuit analyses. For many analyses of circuits containing the voltage divider, the two voltages (V_{in} and V_{out}), the current (I), and the two resistor parameters (R_1 and R_2) can be replaced by the single parameter, V_{out}. There are no approximations in this simplification, the single parameter is derived directly from the device physics of the underlying components. If we see two resistors and a power supply in the same configuration as shown in Figure 2a, and we can see that the downstream impedances are high, then we know that we need only measure the output voltage in order to determine the central functionality of the circuit. That is, we do not need to determine the particular resistances R_1 and R_2, or the properties of the power supply to obtain the circuit function. This ability to reduce the number of physical measurements that must be performed on the system is of special importance in the biochemical case. Here, it is often at great expense in time and resources that a particular variable can be determined quantitatively and *in vivo*.

Often, reductions in number of parameters or in the dimensionality of a dynamical system are fundamental steps in analyzing the overall function of the circuit. If these reductions are derived directly from the device physics as above, then much of the circuit behavior is retained in the new simplified circuitry. But sometimes, especially in biology, such reductions remove important experimental features of the system. Again, an analogy from electronics provides the simplest explanation. Consider the voltage divider in Figure 2c: here, R_1 has been placed in parallel with R_3. Application of the parallel resistor rule allows us to replace this circuit with one identical in structure with that shown in Figure 2a in which the top resistance is $R_1R_3/(R_1 + R_3)$. However, the *reliability* of the reduced circuit is much different from that of the full circuit. Failure of the top resistor is catastrophic for circuit function in the reduced circuit, whereas failure of both R_1 and R_3 is necessary to completely destroy the function of the circuit in Figure 2c. The chemical version, in which the parallel resistor paths are two different reaction channels that convert B to A, shows the same sort of sensitivity. This has important implications: for example, biologists know that debilitating mutations in a protein, assumed here to control one of the reaction channels, may not be lethal to the function of the whole network.

5.2.2.2 Definition of state in electronic and chemical networks

Finally, an important concept for circuit analysis of chemical networks is the definition of *state*. We distinguish between the state of a particular input or output, the *local* state, and the state of the *system*, the *global* state. In digital electronics, local states can take on only two values, 0 and 1. The state of the system is a vector of the local states for each distinguishable input and output. In chemical systems, nominally the local state is often the value of a particular concentration that may take on any positive-indefinite number. The global state is the vector of concentrations of all chemically distinguishable species in the system. Each global state is also associated with properties such as dynamical stability and type (stable node, stable focus, limit cycle, etc.) This theoretically infinite state-space of a chemical system makes its analysis extremely difficult compared with digital systems. However, the dynamical and stoichiometric structure of the system may strongly restrict the range of concentrations that can be reached by any particular chemical species.

Switching in an enzymatic futile cycle As an example consider the circuit shown in Figure 3a. This is a standard futile cycle in which a protein, here labeled A, is phosphorylated by another protein, called a kinase (B), and then subsequently dephosphorylated by a phosphatase or by hydrolysis. It is called a futile cycle because it takes energy (usually in the form of adenosine triphosphate, ATP) to achieve the unidirectional phosphorylation step only to have it seemingly wasted when the protein spontaneously dephosphorylates. The total amount of A and A-p (A_{tot}) remains constant. This is the first restriction

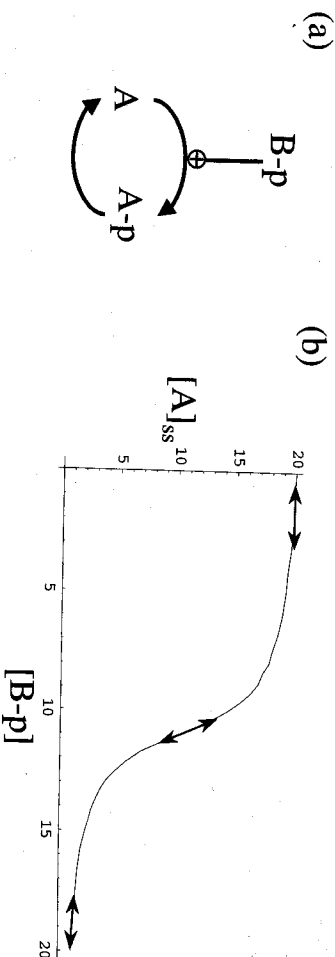(a)

B-p

A — A-p

(b)

[A]$_{ss}$ vs [B-p]

Figure 3. (a) A simple futile cycle in an enzyme-based reaction system. This configuration of reactions (i.e., a regulatory architecture or motif) is a ubiquitous control structure found in many prokaryotic and eukaryotic signal transduction circuits. (b) The stationary-state concentration of A, [A]$_{ss}$, as a function of B-p concentration, [B-p]. For details see text in Section 5.2.2.2.

on the state of the system: concentrations of A, [A], and A-p, [A-p], are restricted between 0 and A_{tot}. As shown in Figure 3b, the kinetics of the system represented in Figure 3a are such that there is a sigmoidal transition from high [A] to low [A] as a function of the concentration of B-p, [B-p]. The steepness of this sigmoid is largely dependent on the fraction of the [B-p] range in which the kinase and phosphatase reactions are both saturated (thus, causing the system to enter a state called 'zero-order ultrasensitivity'; ZOU). Thus the smaller the dissociation constants of A and A-p from their respective enzymes, the more the curve in Figure 3b resembles a Boolean step function. In the case of high ZOU, then, it may be reasonable to say that, in the steady state, the variable A takes on two states, low and high, whose physical values are roughly 0 and A_{tot}, respectively. However, the applicability of this simplification depends ultimately on the dynamics of B-p. Even if the changes in [B-p] were slow enough, compared with the dynamics of the futile cycle, such that the cycle was always near the steady state, [A] only *functionally* has two states if two further conditions are met. First, the controlling physiological changes in [B-p] must cross the threshold region of the [A] steady-state curve completely, and, second, some downstream targets of A activity respond differentially to the high and low states of [A] (or [A-p]).

Bistability and hysteresis The chemical switch represented by Figure 3a is a 'soft-switch'. That is, it is not a true *bistable* state. Rather, [B-p] is a control parameter that smoothly transforms the single steady-state solution of the kinetic equations from a high to a low value. However, one change in the circuit topology (the addition of another reaction) and small quantitative changes in the circuit parameters convert this soft-switch to a 'hard-switch', a truly bistable system. The reaction in Figure 4a is identical with that of Figure 3a, with the exception of a positive feedback that allows A-p to catalyze the phosphorylation of A. When the strength of this positive feedback is low, i.e., when the maximal rate is a small fraction of the maximal rate of B-p-catalyzed phosphorylation, the switch behaves nearly identically to the 'soft-switch' (compare Figures 3b and 4b). However, relatively small changes in the strength of the feedback cause a strong qualitative change in the behavior of the switch.

Figure 4c shows the case where the feedback strength has been doubled. The switch now exhibits *hysteresis*: the [B-p] at which A switches from high to low is different from the [B-p] at which A switches from low to high. In Figure 4d the feedback strength has been doubled yet again and now the switch is *irreversible*. Once a switch changes from high to low, switching back from low to high is now physically impossible with B-p as the sole control parameter. These qualitative changes in behavior can have profound effects on the func-
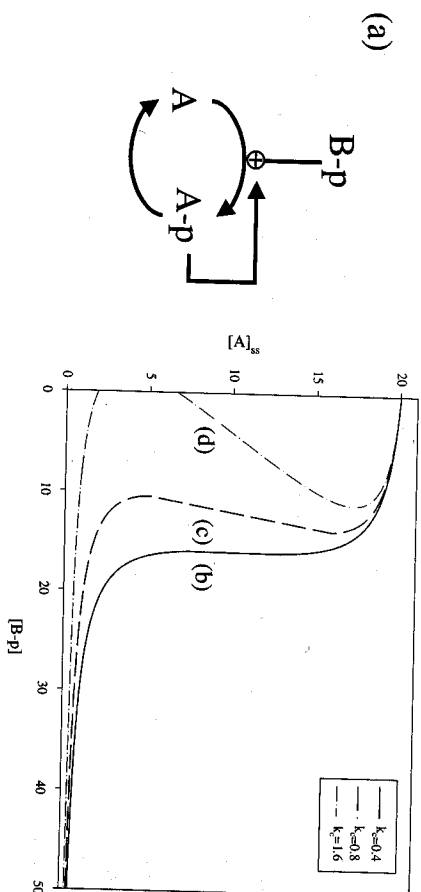
(a)

Figure 4. A biochemical 'switch'. (a) The same futile cycle as in Figure 3 is shown; however, here the phosphorylated from of A catalyzes its own production. Depending on the exact value of the feedback strength, this system can behave either exactly like the futile cycle in Figure 3, or generate true bistability with its attendant hysteresis and 'memory' at higher feedback strengths. (b–c) A family of stationary-state concentration curves for A as a function of B-p concentration. For details see text in Section 5.2.2.

tion of the rest of the network in which this 'hard-switch' is embedded. These qualitative changes may result from changes in kinetic parameters of well under an order of magnitude. The addition of the positive feedback in Figure 4a may seem to be a large perturbation to the system represented in Figure 3a, but such topological changes in a network structure can be found in control processes that occur in 'real' biological systems. For example, the pp125 focal adhesion kinase (FAK), a cytoplasmic tyrosine kinase-transducing signal initiated by integrin engagement and G-protein-coupled receptors, is alternatively spliced (and more highly expressed) in brain tissue. Some alternative splices that are preferentially expressed in brain tissue have an increased autophosphorylation activity, suggesting that FAK may have properties that are specific to neurons. It has been suggested that these isoforms of FAK may play an increased role in turn over of point contacts in motile or invasive cells.

Limiting assumptions and caveats The analysis of the futile-cycle switches above assumes that they may be treated as *self-contained* subcircuits whose dynamics may be analyzed without reference to the rest of the network in which they are embedded. The first caveat to this assumption arises from the ambiguity of the *functional* definition of the state of A and A-p discussed above. For example, simply because a bifurcation analysis predicts that a system is bistable does not mean that both states are used by the biological system. The

second caveat comes from the concept of chemical device impedance discussed above. Connection of this device to downstream targets (by reaction with A or A-p) should not disrupt the function of the device. However, for this device it is not so easy. Consumption of A or A-p, or rapid equilibrium binding to downstream targets, can destroy or greatly alter the bistable behavior of this circuit. This particular type of switch suffers more from the fan-out problem than even the voltage divider in the circuitry discussed above (Figure 2b). Because the total amount of A and A-p is conserved, a connection with enough downstream targets can cause the partition of A_{tot} into the target-bound forms, thus driving the circuit out of the bistable region, and resulting in the destruction of the switch-like properties.

However, once one is convinced that it is reasonable to analyze a chemical subunit as a self-contained device, there are many available methods for predicting the classes of possible circuit behaviors. For example, full solution of the differential equations, bifurcation analysis, and stoichiometric network analysis, all provide means for predicting the range of qualitatively different states that the circuit dynamics may achieve. The sets of kinetic parameters required to switch between each of these states, and control the exact position within, may sometimes be derived as well.

One of the criticisms often leveled at quantitative analysis of biochemical and genetic networks is that one needs measurements of all the mechanisms and of the possibly hundreds of kinetics parameters for those mechanisms, and that obtaining these data is nearly impossible *in vitro*, let alone *in vivo*. The response to this is three-fold: (1) using the analyses just mentioned, it is possible to derive limited classes of behaviors that even relatively roughly measured networks may express; (2) these same analyses can sometimes yield sets of parameter estimates each of which specifies a range of parameters necessary to achieve each qualitatively different behavior; and (3) since most of the BRNs that govern cellular function must be robust to often large fluctuations in the environment and to molecular noise in their own apparatus, chemical circuit behavior should not be overly sensitive to the exact values of each of the kinetics parameters; otherwise there would be a high rate of cell failure. The detailed study of these circuit motifs yields a better qualitative understanding of how a biological pathway is controlled and, as shown below, may point to biologically important physical phenomena that have not yet been fully considered by bench biologists. Further, experimental observation of a particular dynamic behavior may suggest that one or few types of regulatory motifs are responsible for the behavior. Knowledge of how the different motifs achieve a particular behavior can then lead to targeted experiments to differentiate among them.

5.2.3 *Regulatory architecture, motifs, and circuit elements*

In this section, I outline some of the work we have done on identifying common control architectures and elements in biochemical systems. It is an underlying assumption that these elements have evolved to perform one or more specific functions that are useful to an organism. One indicator that these elements are 'functions' is that their architectures recur across organisms and across pathways within a single organism. Not only does identification of the elements simplify the analysis of larger BRNs but they also provide a basis set from which researchers might possibly construct custom networks that perform *novel* functions. The following is far from a complete set of such devices. They are chosen simply to illustrate what sorts of network function can be realized by biochemical systems.

5.2.3.1 *Single enzymes and enzyme networks*

Enzymes and other proteins are examples of how a single molecule or a small molecular complex may be a fairly complicated chemical device. The presence of an enzyme that catalyzes a simple Michaelis–Menten-type reaction is described by an already-composite chemical device composed of three elementary reaction steps. If an enzyme is considered a device in itself, then its inputs are (at least) its substrates and effectors, and its outputs are the rates of product generation. These output rates are usually, but not always, monotonically increasing or decreasing, saturating functions of the various substrates, inhibitors and activators. That is to say, the outputs are often sigmoidal functions of the inputs. In the extreme case, sigmoid functions look like step functions and, therefore, it is tempting to use a Boolean truth-table to describe their function rather than the full enzymological description (Arkin and Ross, 1994). Even if the sigmoid functions are not so steep, they resemble various models of computational 'neurons'. Thus, networks of such enzymes resemble these formal neuronal networks (Bray, 1990). In fact, Hjelmfelt and Ross have demonstrated the equivalence of a particular parameterization of the example given in Figure 3a to a McCulloch–Pitts artificial neuron, and they showed how to make various computational circuits out of interconnected networks of these elements (Hjelmfelt *et al.*, 1991, 1992; Hjelmfelt and Ross, 1992). On the basis of these results, in collaboration with F. Schneider, they experimentally implemented a chemical neuronal network made out of bistable chemical reactions with a dynamical function similar to the futile cycle, as discussed in Section 5.2.2 (Hjelmfelt *et al.*, 1993). Furthermore, Bray (1990) has suggested that the neuronal network-like properties of chemical parallel-distributed processes may help to explain, in part, the reliability and evolutionary adaptability of these networks.

Finally, it is worth noting that some protein-based devices have dynamics fundamentally different from the standard enzymological mechanisms. Molecular machines such as polymerases, ribosomes, and kinesin can be very intricate molecular devices. New techniques are allowing the quantitative measurement of the microscopic and mesoscopic dynamics of motion of these molecular complexes on their macromolecular substrates (Guthold et al., 1994; Yin et al., 1995; Bustamante and Rivetti, 1996; Wang et al., 1997). However, consistent models of the in vivo dynamics of these machines are still in their infancy. The initial attempts to treat mathematically the dynamics of various of these molecular motors and their input/output behavior have laid a good foundation against which to test future measurements and models (Peccoud and Ycart, 1995; Peskin and Oster, 1995; Astumian and Bier, 1996; Duke and Leibler, 1996; McAdams and Arkin, 1997; Arkin et al., 1998; Goss and Peccoud, 1998). One of the interesting common dynamical phenomena found in these machines is that their operation is fundamentally stochastic. A similar observation is made when examining the function of, for example, ion channels (Collins et al., 1995). Noise in the operation of these devices, as is discussed further below, necessitates a consideration of robustness and reliability in the design of cellular signal-processing networks.

5.2.3.2 Biochemical oscillators

Biochemical oscillators are found to play a number of roles in the control of cellular and organismal behavior (Berridge and Rapp, 1979; Rapp, 1979). The most pervasive form of oscillator in biology is the cell cycle oscillations that underlie repeated patterns of cell growth and division (Borisuk and Tyson, 1998; Novak et al., 1998). Though the chemical network that drives a particular cell cycle is usually not a 'free running' oscillator in that it is regulated by checkpoints that can stop, slow or even redirect the cycle (e.g., in order to wait for unsynchronized processes to catch up, to deal with damages and stresses in the cell or to change the chemical pathway responsible for the cell cycle in different cell types), at root is a chemical system capable of repeatedly leaving and very nearly restoring an initial condition. Biochemical oscillations are also found in mitochondrial volume, in yeast glycolytic flux (Jonnalagadda et al., 1982), in GTP/G-protein activity, in cytoplasmic calcium concentrations, in neuronal signaling, in circadian rhythms (Goto et al., 1985; Ouyang et al., 1998) and in certain reconstituted enzyme systems such as horseradish peroxidase (Stemwedel et al., 1994; Hung and Ross, 1995). These oscillations have many different functional roles in the cells in which they are found. Timing and synchronization are the most obvious ones. However, there is some evidence that oscillatory dynamics can reject noise while propagating signals,

and that the frequency and amplitude of an oscillation can carry information that can be decoded by chemical frequency filters like the ones discussed further below. Thus, these oscillators are signal generators whose output can be modulated in amplitude, frequency and phase by chemical, thermal and/or light inputs.

Ross and co-workers have attempted to classify chemical and biochemical oscillators into a finite set of classes distinguished by their network topology and their responses to various experimental perturbations (Eiswirth et al., 1991a,b). Chemical species within such oscillator devices are classified as essential or nonessential depending on whether or not quenching of their oscillatory behavior destroys the overall ability of the network to support oscillation.

These classification methods demonstrate a number of the advantages of a device analysis. They provide a theoretical framework for understanding the different ways in which chemical systems can provide oscillatory signals. They also provide an ordered set of diagnostic experiments by which a novel oscillatory chemical species may be classified in a small number of experimental steps. This classification, then, severely restricts the underlying mechanisms and their parameterizations that give rise to the experimental observations.

5.2.3.3 Genetic regulatory circuits

In 1961, at the Cold Spring Harbor Symposium in Quantitative Biology, Jacob and Monod first outlined a circuit theory of genetic control in prokaryotes (Monod and Jacob, 1961). The basic theory describing combinational control of transcription initiation, expression of polycistronic operons and feedback control as a basis for control of metabolism, growth and development remains largely unchanged today. Most of the basic mechanisms proposed are used in prokaryotes and eukaryotes alike, although eukaryotic gene control has a few more levels of complexity to it. The central process is the transcription of DNA to RNA via the multiprotein complex RNA polymerase (RNAP), and then the translation of RNA to protein via transfer RNA and ribosomes. Transcription can be broken up into at least two processes: transcription initiation and transcript elongation. Translation can be broken into three processes: translation initiation, protein elongation and transcript degradation. Each of these processes may, in turn, be regulated by cellular signals.

Transcription initiation The best characterized of these controls is the regulation of transcription initiation. Initiation begins from a region of DNA called the promoter, upstream (at the 5' end of DNA) from the genes of interest. In prokaryotes, this is most often accomplished by the binding of proteins,

transcription factors, to sites on the DNA called *operator* sites. The pattern of transcription factors bound to sites can modulate both the strength with which RNAP binds to the promoter and the rate at which it begins transcription. The number of patterns (states) of the promoter and the rate at which it begins transcription. The number of patterns (states) of the promoter can be quite large. For example, the λ phage P_R/P_{RM} control region is composed of two promoters and three operator sites (shown schematically in Figure 5). The operator sites can bind homodimers of two proteins, Cro and CI with different affinities. The region can have 40 different configurations of RNAP, Cro, and CI with different affinities. The region can have 40 different configurations of RNAP, Cro_2, and CI_2 bound, each of which is characterized by its stability (free energy) and its transcriptional activity (Ackers et al., 1982; Shea and Ackers, 1985).

It is tempting to think of these states as 40 different logical states of a complex Boolean switch that transmits a set of RNA signals when particular sets of transcription factors are present or absent. In some cases, this may indeed be a good approximation, but a number of issues need to be addressed before such an abstraction is made. Most important is probably the time that a given configuration of transcription factors and RNAP at a given promoter persists. In many cases, in prokaryotes, these proteins are assumed to be in rapid equilibrium with their respective binding sites. The binding dynamics of RNAP to its promoter, especially, is likely more complicated than this; however, empirically this does not seem to be a bad approximation in most cases. In this approximation, the 40 molecular configurations of P_R/P_{RM} are sampled many times in between transcription initiation events. The total time spent in any configuration is related to its stability. Thus an average transcriptional activity may be calculated for any instantaneous concentration of proteins. The thermodynamic and kinetic parameters for the P_R/P_{RM} promoters have been determined (Ackers et al., 1982; Shea and Ackers, 1985). The graph of P_{RM} activity as a function of CI_2 and Cro_2 concentrations, $[CI_2]$ and $[Cro_2]$,
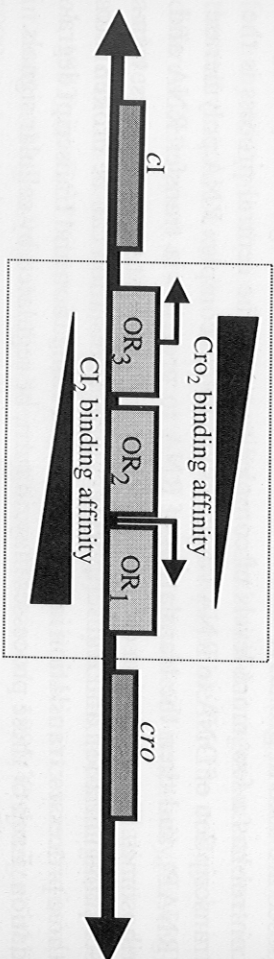

Figure 5. The organization of the P_R and P_{RM} divergent promoters from bacteriophage λ. The gene products of *cI* and *cro* dimerize, then bind to operator sites (OR₁, OR₂, and OR₃) in the promoter region with differential affinities. The pattern of CI_2, Cro_2, and RNA polymerase binding to the operator and promoter sites determines the frequency of transcription initiation from P_{RM} and P_R.
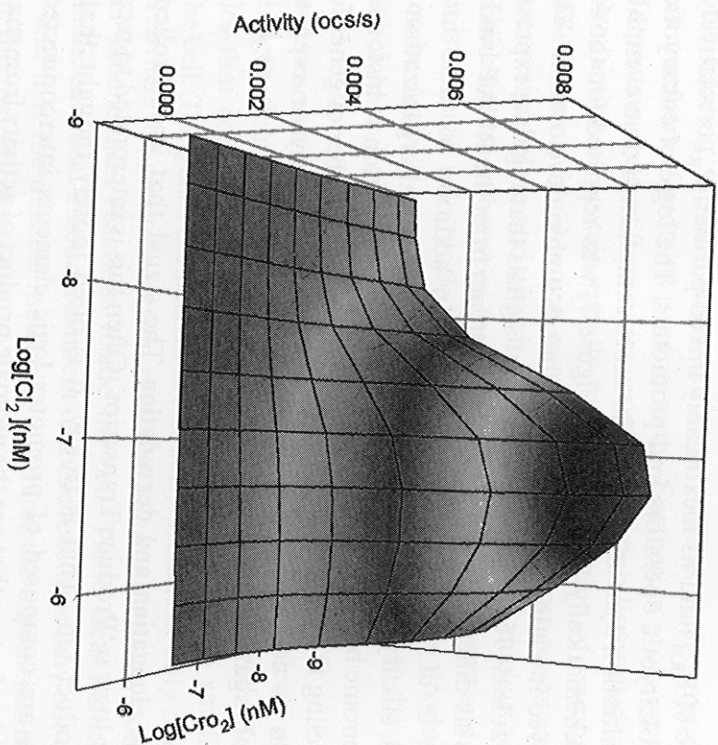
Figure 6. A plot of the activity of the P_{RM} promoter as function of the concentration of its two transcription factors, CI_2 and Cro_2. Activity is expressed in units per time of open-complex formation (ocs/s). For details see Section 5.2.3.3.

respectively, is shown in Figure 6 for a constant (available) RNAP concentration of 30 nM. This plot summarizes the 'control logic' for transcription initiation at P_{RM}. The 40 'logical' states of the promoter region are not directly visible in this activity curve. The smoothness of this curve arises because of the rapid equilibrium assumption. Were the transcription factors' 'sticky' (i.e., were their off-rates from DNA comparable with or slower than the rate of transcription initiation), then the timing (order) of CI and Cro binding would become important and a time-independent control surface could not be plotted. Further, the activity of P_{RM} as a function of CI_2 is not a monotonically saturating function. Instead, as $[CI_2]$ increases, the activity of P_{RM} at first increases, then decreases. There are, then, at least three functionally different regions of $[CI_2]$ and a Boolean abstraction of this 'switch' could not necessarily map $[CI_2]$ into a single binary variable. In addition, as discussed above, the appropriateness of a Boolean approximation to this curve is dependent on the time it takes the effector molecules to traverse from their initial to final values and vice versa.

The logic of P_{RM} (and its inextricably linked partner, P_R) is relatively complex for prokaryotic operators and promoters. The logic of eukaryotic transcription initiation dynamics can be far more complicated than even this. One example, schematically illustrated in Figure 7, was reported for the *endo-16* gene involved in endodermal formation in sea urchin development. The promoter has at least 15 different protein input signals that regulate expression by binding to six different binding regions upstream from the RNAP binding site (Davidson et al., 1998; Yuh et al., 1998). In addition to these modes of regulation, eukaryotes also can regulate the more global organization of their nuclear genome by, for example, controlling the acetylation of histones, thereby remodeling the chromatin structure. The larger number of genes found in eukaryotes as compared to prokaryotes is not the only or even the best measure of organismal complexity.

Transcript elongation and degradation The signal that is controlled by the promoter logic is 'Produce Transcript'. Often this is interpreted to mean 'Turn Gene Product On'. That is to say, it is often loosely thought that genetic networks are composed of promoter logic elements interconnected by the transcription factors that are the ultimate product of activity from many of the constituent promoters. But there are other factors that must be considered before such an abstraction can be made. After transcription initiation, there are numerous mechanisms of elongation control, including terminators and antiterminators (regions of DNA at which a transcribing RNAP can fall off the template, or at which RNAP can be modified to be resistant to such termination, respectively), downstream binding sites for proteins that block a processive RNAP, and RNAP pause sites. In polycistronic operons, mostly found in prokaryotes, these can lead to strong polarity effects in which there is higher expression of transcripts proximal to the promoter than for those that are distal. Further, each gene in the transcript may or may not have its own
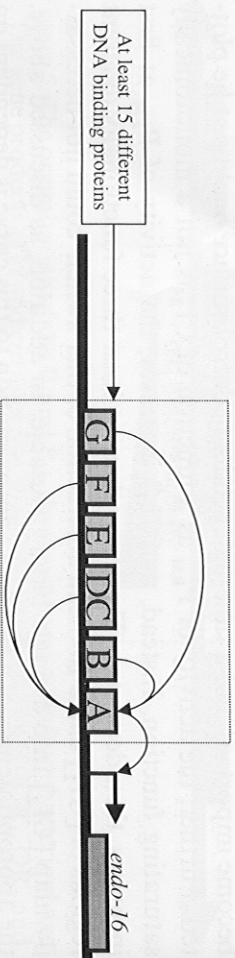
At least 15 different DNA binding proteins

G F E DC B A

endo-16

Figure 7. Diagram of the promoter control of the sea urchin *endo-16* gene involved in early embryogenesis. Derived from Yuh et al. (1998).

ribosome binding site and degradation rate. Thus each gene can express different numbers of proteins per transcript. Further, in prokaryotes, translation of a protein product from a transcript is often tightly coupled to transcription. Production of protein rapidly follows production of transcript. In eukaryotes, this coupling is much weaker, since many processes can act on a given transcript before and after it is exported from the nucleus to be translated. The correlation between the concentration of transcript and its protein product is lower in eukaryotic systems. In all cases, protein products may or may not be actively degraded. All these processes have to be taken into account before realistic models can be constructed.

Stochastic processes in gene regulation There is one further level of complexity to the genetic machinery that must be carefully considered. Many of the molecules that control gene expression are often present in small numbers inside the cell. The genes themselves are usually present in only one or a few copies. Further, the genetic reaction rates are often rather slow compared to the other biochemical reactions inside the cell. These facts indicate that a deterministic chemical kinetic treatment of these reactions may not always be possible and that the discrete molecular nature of the expression machinery and their thermal fluctuation must be taken into account as well (Kampen, 1981; Ko, 1991, 1992; Peccoud and Ycart, 1995; McAdams and Arkin, 1997; Arkin et al., 1998; Goss and Peccoud, 1998). Consider, for example, the rate of transcription initiation from P_{RM}. The maximum activity of P_{RM} shown in Figure 6 is about 0.007 open-complexes/s; that is, one transcript initiation about every 2.5 min on average. This occurs at $[CI_2]$ of the order of 200 nM, and at $[Cro_2]$ of zero. However, during early λ phage development, $[Cro_2]$ and $[CI_2]$ are generally less than 100 nM. In *E. coli*, which has a cell volume of approximately one femtoliter, 1 nM corresponds to about one molecule only. Given that a well-fed *E. coli* cell has a cell cycle time of about 20 min, there are, on average, fewer than ten transcription initiations from a fully activated P_{RM} per cell division. However, the actual number of transcription initiations from P_{RM} is most likely a stochastic process. Further, the number of proteins produced per transcript is also probably a random process. The 'back-of-the-envelope' explanation for this is as follows. We can roughly divide the gene expression process into four stages: (1) RNAP binding to the promoter, (2) transcription initiation, (3) RNAP arrival at the end of a gene, and (4) competitive binding of ribosomes and RNA degrading enzymes to the RNA transcript. The probability of RNAP being bound at its promoter is determined by the partition function of operator/promoter states that enters into the calculation of the curve in Figure 6. The probability of transcription

initiation is very roughly a first-order rate process whose rate constant can be read off the figure. The time until transcription initiation, after RNAP has bound to its promoter in a given state, is distributed approximately exponentially (McAdams and Arkin, 1997). Elongation then proceeds by a series of independent steps each exponentially distributed in time. The arrival time at the end of a gene is, therefore, the sum of a set of independent exponential distributions (one for each nucleotide in the sequence) which has the form of a Γ-distribution.

Finally, the number of proteins per transcript is determined by how many ribosomes can bind to the ribosome binding site on the transcript before the transcript is degraded by an RNase protein. These two processes are often competitive, so ribosome binding temporarily protects the transcript from degradation. Thus, the question arises of how many ribosomes can bind before degradation by RNase occurs. This is analogous to asking how many heads does one get before one gets a tail when flipping a biased coin. Such processes are described by a geometric distribution. Each of these distributions can be rather broad and skewed. Consequently, the pattern of protein production from a single promoter can be expected to be burst-like and erratic.

When all of the above arguments are put forth in a chemically more rigorous fashion, the dynamics of gene expression may be described by a chemical master equation (McAdams and Arkin, 1997). Figure 8 shows the pattern of protein production from a model of the λ phage P_R promoter (McAdams and Arkin, 1997). Each curve is one realization of the stochastic gene expression process started from the *same* initial conditions in *identical* cells. Our theoretical model indicates that individual cells can have quite different expression patterns. There is ample experimental evidence that this is indeed the case in cell populations (Novick and Weiner, 1957; Ko, 1992; Ross et al., 1994; Siegele and Hu, 1997). The implications of this noise for the control of cellular behavior and development, and for the engineering of reliable genetic circuitry, has been discussed in detail (McAdams and Arkin, 1997, 1998, 1999; Arkin et al., 1998).

5.2.3.4 Electrical and chemical frequency filters

Since biological signals can be *periodic*, as described in Section 5.2.3.2, and *noisy*, as described in Section 5.2.3.3, a consideration of the frequency-dependent responses of chemical reactions to time-varying chemical signals is in order. A frequency dependence can be considered to be a type of *filtering*. In an electrical context, frequency filters are devices that accept a time-dependent input and differentially pass on some frequencies in the signal, while suppressing others to different degrees.
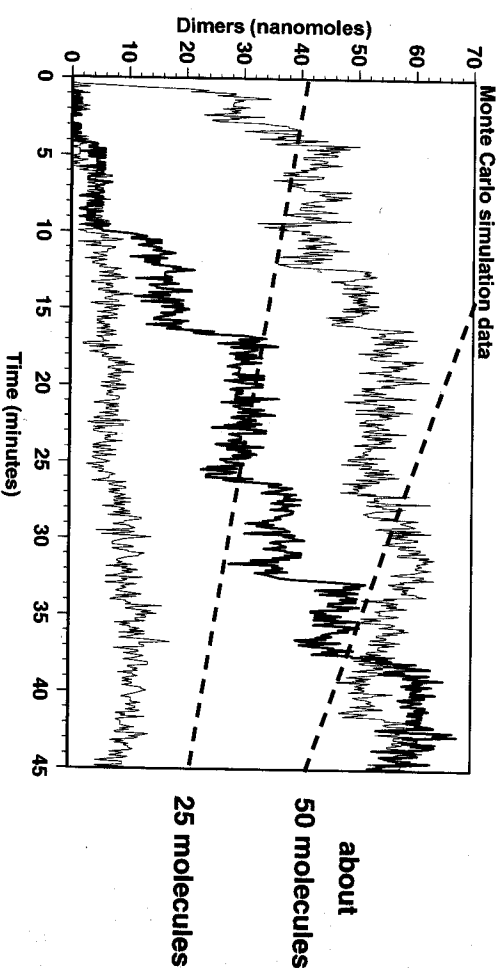
Figure 8. Master equation simulation of stochastic gene expression from a P_R-like promoter in three initially identical cells. For details see Section 5.2.3. (From McAdams and Arkin, 1997.)

The two basic, passive electrical filters, the *low-pass* filter and the *high-*pass filter, are represented in Figure 9. The amplitude of each frequency component of the output signal is always less than, or equal to, the corresponding amplitude in the input signal. These small circuits are composed of 'linear' elements (resistors and capacitors) and thus are noise filters as well as frequency filters. The spectrum of the output signal is the superposition of the filtered amplitudes of each frequency component. The filter causes no interference among the different components of the input signal to arise in the output signal.

Chemical low-pass filter A chemical version of the low-pass filter is shown in Figure 10 (top). The input is the amplitude of time-varying (positive) input of a chemical species; the output is the amplitude of the concentration of species A. The frequency response function has the same fall-off as that of the low-pass filter. However, this 'filter' can (at low frequencies) *amplify* a signal as well, since it has a factor of 'κ' in the numerator. In fact, any network of such linear chemical reactions (with one input) is a low-pass filter. The additional reactions between the input and output change the phase retardation of the signal as well as the exact shape of the monotonically decreasing filtering profile.

Chemical band-pass filter As soon as a *nonlinear* chemical reaction is considered, there exists the possibility of *band*-pass filtering (all chemical systems, like all electronic systems, are low-pass at high enough frequencies). As an
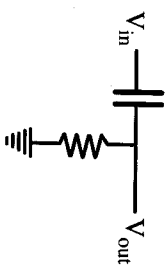
Low-Pass Filter
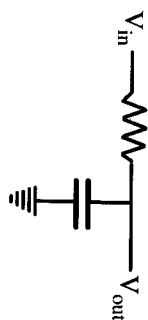
$$\frac{V_{in}}{V_{out}} = \frac{1}{(1+\omega^2 R^2 C^2)^{1/2}}$$
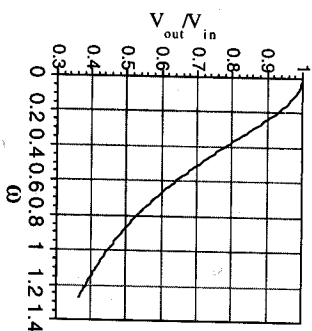
High-Pass Filter

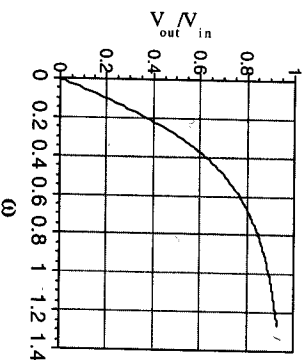$$\frac{V_{in}}{V_{out}} = \frac{R}{(R^2 + 1/(\omega^2 C^2))^{1/2}}$$

Figure 9. Electrical frequency filters. (Top) A filter that passes only low frequencies (a low-pass filter). (Bottom) A filter that passes only high frequencies (a high-pass filter). For details see text in Section 5.2.3.4.
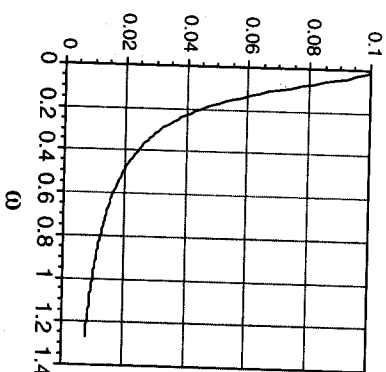
example, consider the bimolecular reaction shown in Figure 10 (bottom). In this reaction scheme, γ must be less than or equal to P, which, in turn, must be equal to RB. Solving the Riccati equation that describes this system analytically is at best difficult, but very good asymptotic solutions may be found. The filtering profile shown in Figure 10 (bottom) shows a pronounced band-pass region between the critical frequency, ω_c and k_2. This particular chemical circuit is admittedly artificial, since the constraint that 'P match RB' is not likely to be met in biological systems. The example demonstrates, however, that chemical systems can be both band-pass and low-pass filters. Note, however, that this does not describe a linear filter, and if the driving signal has many frequency components, then there are in fact (very small) interference bands that appear at frequencies not contained in the original signal.

More complex chemical filters The above types of mechanism can be easily linked in serial or parallel fashion to form very complicated filtering profiles

First-Order Reaction

$$\gamma(1+\sin(\omega t)) \longrightarrow A \xrightarrow{k}$$

$$\frac{|A|}{|\gamma \sin(\omega t)|} = \frac{k}{(1+\omega^2/k^2)^{1/2}}$$

Second-Order Reaction

$$P+\gamma\sin(\omega t) \longrightarrow A$$
$$RB \longrightarrow B$$

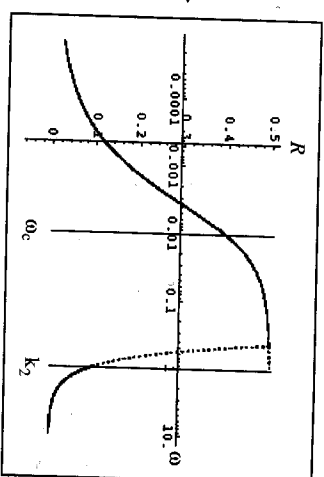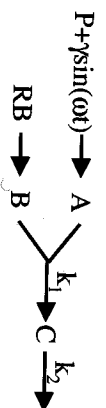$$A, B \xrightarrow{k_1} C \xrightarrow{k_2}$$

Figure 10. The frequency filtering properties of two different chemical mechanisms. *Top*: A reaction pathway with linear (first-order) kinetics acts as a low-pass filter. Here, the concentration of the species A is driven by a sinusoidal influx of material (amplitude = γ). The amplitude of the oscillation in A is plotted as a function of the frequency of the influx species. The analytical equation describing this curve is shown below the reaction mechanism. *Bottom*: As soon as a simple nonlinearity is found in the reaction (second-order kinetics) then the system can behave as a band-pass filter. Here, there is influx of material from independent sources into species A and B. Species A is driven by a sinusoidal signal that has an amplitude equal to γ around an average influx value of P (the pedestal). The amplitude of the oscillation in C is plotted as a function of the driving frequency of the A influx. The equation for this curve is too complicated to show (see Samoilov, 1997).

indeed, such as *notch* filters that do not pass some intermediate band of frequencies. The ability to easily construct complex filters out of relatively simple chemical reactions suggests that such filtering could be used by biological systems to respond differentially to the types of oscillatory input discussed above. This is particularly interesting when considering, for example, that in T lymphocytes some expression of some transcription factors are stimulated by much lower frequencies of cytosolic Ca^{2+} oscillations than others (Dolmetsch et al., 1997, 1998). Also, although this has not yet been directly observed, chemical frequencies can allow the demultiplexing of multiple signals carried

in the frequency spectrum of a signal chemical concentration. For example, perhaps one hormone could induce one frequency of Ca^{2+} spiking and another could superimpose another. Thus the presence or absence of two external signal species could be carried throughout the cell using only one species; downstream chemical filters could then decompose the signal at the respective sites of action.

Nonlinear chemical filters do not behave, as was mentioned above, like linear filters. For example, interference effects can introduce new frequency components into the output spectrum. Thus these macroscopic kinetic circuits are not really noise filters, but they may perform more general transformations. Also, these treatments of these circuits assume that the fluctuations in chemical reactions are negligible. How a given chemical kinetic mechanism filters its own internal noise is still an open question. However, it is certain that filter mechanisms such as molecular dimerization, and other forms of cooperativity, filter molecular noise to some extent and can therefore lend increased reliability to the genetic circuits described above.

More complex chemistry can have ever more exotic and interesting behaviors. These include analytical delays and strong band-pass effects, quasiperiodicity and chaos. When the input signal is noisy, some complex chemical networks can exhibit a phenomenon called *stochastic resonance* in which the noise signal improves the detection of very small periodic signals in the system over some region of power in the noise spectrum (Collins et al., 1995, 1996; Braun et al., 1997; Astumian and Moss, 1998; Jung et al., 1998). These examples only underscore the point that biological systems are non-stationary dynamical systems whose signal-processing machinery may be far more sophisticated than is generally understood.

5.3 Comments on the parameterization of models, nonlinear systems and cellular reliability

One of the concerns often voiced about the business of biological systems modeling concerns the fact that most mathematical descriptions of biological processes contain a good many parameters most of which cannot easily be experimentally measured. Further, these descriptions are most often sets of coupled nonlinear differential equations that may, in general, show an extreme sensitivity to parameters (and initial conditions). Since it is currently difficult, at best, to measure all of the kinetic parameters for even isolated network components, such as enzymes, and since such measurements of *in vitro* kinetics are not guaranteed to be the same as those that are obtained *in vivo*, it might seem a hopeless task to construct confirmatory and predictive models of

complex BRNs. This fear is valid; however, there are a number of phenomenological observations and mathematical facts that argue that the situation is not so bleak as might be initially thought.

The first trivial observation is that all that is nonlinear is not necessarily sensitive to parameter changes. Extreme sensitivity, such as that found in deterministically chaotic systems, at least so far, has proven to be a relatively rare phenomenon in biological and biochemical systems. Though chaos has certainly been detected at the tissue level such as in heart and brain dynamics, and on the chemical level in reconstituted and forced peroxidase–oxidase enzyme systems, it is certainly not the rule in even very complex, nonlinear biochemical systems. The reason for this must reside in the engineering specifications for good biological function. Most cells operate in fluctuating environments (wherein variables such as temperature, pressure, volume and ionic strength can change unpredictably) and must both detect and use chemical components that are at very low concentrations and whose kinetics, therefore, is likely to exhibit large fluctuations in reaction rates. In addition, there is a finite chance that a given component of a regulatory network may fail, due either to this noise in its components or to more extreme processes such as mutations. In order for a cell to survive under such conditions, the function of its regulatory networks cannot be so sensitive to their parameters (which are sensitive to these fluctuations) and must be reliable in the face of individual component failure and mutation. In order to achieve this robustness, cells use *functional redundancy* and *feedback stabilization* among other design strategies to obtain reliable operation.

Barkai and Liebler have suggested that this very insensitivity to parameters might be one criterion for judging whether or not a particular biochemical model is reasonable (Barkai and Liebler, 1997). As an example, they have analyzed various models of exact adaptation in bacterial chemotaxis. This is a phenomenon in which the ratio of clockwise to counter-clockwise rotation of the flagellum initially decreases upon cellular exposure to a step of chemoattractant, but then returns exactly to the initial basal value, even under continued (constant) exposure to the attractant. Since this behavior is judged to be important to the cell's fitness for survival, Barkai and Liebler (1997) have argued that the regulatory network that controls this behavior should be *insensitive* to changes in its parameters. They propose a schematic model of adaptation that maintains 'exactness' despite order-of-magnitude changes in one or more parameters, whereas some previous models of adaptation seemed to require a highly tuned parameter set. However, although the model by Barkai and Liebler exhibits high reliability in exact adaptation, the time to recovery is less robust. Experimental measurements on the dispersion of times

to exact recovery in different mutants of chemotactic *E. coli* should further constrain the class of models that can explain the chemotactic behavior (Alon et al., 1999). An important side-note is that the argument that cells should not be too sensitive to small perturbations assumes that the perturbations are natural (i.e. commonly occurring ones). Specific toxins and pharmaceuticals, hard radiation, and other such 'artificial' perturbations are rarely encountered during the normal course of a particular organism's evolution. Its BRNs may therefore be sensitive to very low exposures to these types of attacks.

Biochemical models, therefore, are often subject to a number of restrictive global functional constraints such as robustness and, in some cases, homeostasis as well as an often large list of experimental data that greatly restricts the class of models that can explain and predict organismal behavior. If the basic stoichiometric network of reactions is known, this provides a further restriction on the class of behaviors and the feasible sets of parameters that can reproduce experimental observations (Clarke, 1981).

5.4 Summary and outlook

The challenges of understanding how these incredibly complicated biological systems function to the point where we can predict their behavior, control them and rationally design modifications into them are clear. The chemical and physical systems that underlie their function operate in regimes of which we do not yet have a full theoretical facility. They operate *asynchronously*, *asymmetrically* and *nonlinearly* in fluctuating environments with less than fully reliable parts. In addition, the systems are rather large and highly interconnected networks that operate over a large range of time and space scales. Practically, it is not feasible to derive the equations for each microscopic event that occurs within and among cells. Some higher levels of abstraction will be necessary to make useful and rapid analyses. The work presented here has considered a bottom-up approach that starts with the detailed kinetics of networks of chemical reactions and attempts to derive when and where such networks may be dissected into self-contained 'devices'. It was a hypothesis of this chapter, and one that is confirmed in part by the literature, that these devices are *regulatory motifs* that recur within different pathways of the same organism and across organisms. The motifs may be realized using different (or related) biochemical species, but their functions may be the same. The level of abstraction from the detailed molecular kinetics of each device will be different, and it is a central challenge to develop methods for making models that can combine such heterogeneous submodels in a physically consistent way. Also, in analogy to finding the basis set of protein folds in order to understand the

principles of protein folding, it would be advantageous to identify a basis set of such devices, and their restricted class of functions, from which many BRNs may be constructed.

Meeting this challenge is especially crucial in light of the accelerated onslaught of essentially raw data that has fallen out of new high-throughput biological measurement devices and their resultant 'projects'. Genome projects provide partial parts lists for the cellular machinery. Information technologies are providing large numbers of hypotheses for predicting protein activity, structure/function predictions and even network hypotheses. Gene chip technology and two-dimensional protein gel/mass spectrometry methods are beginning to provide quantitative measurements of the condition- and time-dependent variations in concentrations of mRNA-transcript proteins. Advanced microscopy and other cell measurement devices are beginning to create large databases of spatial information, cell motion and cellular interaction data that can be related to changes in ion concentrations and gene expression. In addition to these relatively new stores of data, there are all the data generated from the standard biochemical and genetic research communities as well as massive amounts of clinical and medical diagnostic data. It is one central challenge to deduce from these data the responsible regulatory networks. Such reverse engineering methods are in their very early days (Arkin and Ross, 1995; Arkin et al., 1997; Liang et al., 1998; Thieffry and Thomas, 1998).

Theoretical and computational tools developed to dissect and analyze complex biological systems are essentially tools to make more rigorous the process of hypothesis formation that every biologist must conduct before and after performing such experiments. These tools provide a central structure for organizing the data generated by the above techniques. They help to yield new insights and new biological principles, some of which are discussed above. Several of these insights have profound implications for biological processes such as development, facultative infection and other diseases. Finally, these tools are beginning to aid in the design of novel functions into cells. Since the engineering principles by which such circuitry is constructed in cells comprise a super-set of that used in electrical engineering, it is, in turn, possible that we will learn more about how to design asynchronous, robust electronic circuitry as well.

References

Ackers, G. K., Johnson, A. D. and Shea, M. A. (1982) Quantitative model for gene regulation by phage-1 repressor. *Proc. Natl. Acad. Sci. USA* **79**: 1129–1133.

Alon, U., Surette, M. G., Barkai, N. and Leibler, S. (1999) Robustness in bacterial chemotaxis. *Nature* **397**: 168–171.

Arkin, A. P. and Ross, J. (1994) Computational functions in biochemical reaction networks. *Biophys. J.* **67**: 560–578.

Arkin, A. P. and Ross, J. (1995) Statistical construction of chemical mechanisms from measured time series. *J. Phys. Chem.* **99**: 970–979.

Arkin, A. P., Ross, J. and McAdams, H. H. (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage 1-infected *E. coli* cells. *Genetics* **149**: 1633–1648.

Arkin, A. P., Shen, P.-D. and Ross, J. (1997) A test case of correlation metric construction of a reaction pathway from measurements. *Science* **277**: 1275–1279.

Astumian, R. D. and Bier, M. (1996) Mechanochemical coupling of the motion of molecular motors to ATP hydrolysis. *Biophys. J.* **70**: 637–653.

Astumian, R. D. and Moss, F. (1998) Overview: the constructive role of noise in fluctuation driven transport and stochastic resonance. *Chaos* **8**: 533–538.

Barkai, N. and Leibler, S. (1997) Robustness in simple biochemical networks. *Nature* **387**: 913–917.

Berridge, M. J. and Rapp, P. E. (1979) A comparative survey of the function, mechanism and control of cellular oscillators. *J. Exp. Biol.* **81**: 217–279.

Borisuk, M. T. and Tyson, J. J. (1998) Bifurcation analysis of a model of mitotic control in frog eggs. *J. Theor. Biol.* **195**: 69–85.

Braun, H. A., Schafer, K., Voigt, K., Peters, R., Bretschneider, F., Xing, P., Wilkens, L. and Moss, F. (1997) Low-dimensional dynamics in sensory biology. I. Thermally sensitive electroreceptors of the catfish. *J. Comput. Neurosci.* **4**: 335–347.

Bray, D. (1990) Intracellular signalling as a parallel distributed process. *J. Theor. Biol.* **143**: 215–231.

Bustamante, C. and Rivetti, C. (1996) Visualizing protein–nucleic acid interactions on a large scale with the scanning force microscope. *Annu. Rev. Biophys. Biomol. Struct.* **25**: 395–429.

Clarke, B. L. (1981) Complete set of steady states for the general stoichiometric dynamical system. *J. Chem. Phys.* **75**: 4970–4979.

Collins, J. J., Chow, C. C. and Imhoff, T. T. (1995) Stochastic resonance without tuning. *Nature* **376**: 236–238.

Collins, J. J., Imhoff, T. T. and Grigg, P. (1996) Noise-enhanced information transmission in rat SA1 cutaneous mechanoreceptors via aperiodic stochastic resonance. *J. Neurophysiol.* **76**: 642–645.

Davidson, E. H., Cameron, R. A. and Ransick, A. (1998) Specification of cell fate in the sea urchin embryo: summary and some proposed mechanisms. *Development* **125**: 3269–3290.

Dolmetsch, R. E., Lewis, R. S., Goodnow, C. C. and Healy, J. I. (1997) Differential activation of transcription factors induced by Ca^{2+} response amplitude and duration. *Nature* **386**: 855–858.

Dolmetsch, R. E., Xu, K. and Lewis, R. S. (1998) Calcium oscillations increase the efficiency and specificity of gene expression. *Nature* **392**: 933–936.

Duke, T. and Leibler, S. (1996) Motor protein mechanics: a stochastic model with minimal mechanochemical coupling. *Biophys. J.* **71**: 1235–1247.

Eiswirth, M., Freund, A. and Ross, J. (1991a) Mechanistic classification of chemical oscillators and the role of species. *Adv. Chem. Phys.* **LXXX**: 127–199.

Eiswirth, M., Freund, A. and Ross, J. (1991b) Operational procedure toward the classification of chemical oscillators. *J. Phys. Chem.* **95**: 1294–1299.

Goss, P. J. and Peccoud, J. (1998) Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. *Proc. Natl. Acad. Sci. USA* **95**: 6750–6755.

Goto, K., Laval-Martin, D. L. and Edmunds, Jr, L. N. (1985) Biochemical modeling of an autonomously oscillatory circadian clock in *Euglena. Science* **228**: 1284–1288.

Guthold, M., Bezanilla, M., Erie, D. A., Jenkins, B., Hansma, H. G. and Bustamante, C. (1994) Following the assembly of RNA polymerase–DNA complexes in aqueous solutions with the scanning force microscope. *Proc. Natl. Acad. Sci. USA* **91**: 12927–12931.

Hjelmfelt, A. and Ross, J. (1992) Chemical implementation and thermodynamics of collective neural networks. *Proc. Natl. Acad. Sci. USA* **89**: 388–391.

Hjelmfelt, A., Schneider, F. W. and Ross, J. (1993) Pattern recognition in coupled chemical kinetic systems. *Science* **260**: 335–337.

Hjelmfelt, A., Weinberger, E. D. and Ross, J. (1991) Chemical implementation of neural networks and Turing machines. *Proc. Natl. Acad. Sci. USA* **88**: 10983–10987.

Hjelmfelt, A., Weinberger, E. D. and Ross, J. (1992) Chemical implementation of finite state machines. *Proc. Natl. Acad. Sci. USA* **89**: 383–387.

Hung, Y.-F. and Ross, J. (1995) New experimental methods towards the deduction of the mechanism of the oscillatory peroxidase–oxidase reaction. *J. Phys. Chem.* **99**: 1974–1979.

Jonnalagadda, S. B., Becker, J. U., Sel'kov, E. E. and Betz, A. (1982) Flux regulation in glycogen-induced oscillatory glycolysis in cell-free extracts of *Saccharomyces carlsbergensis. Biosystems* **15**: 49–58.

Jung, P., Cornell-Bell, A., Moss, F., Kadar, S., Wang, J. and Showalter, K. (1998) Noise sustained waves in subexcitable media: from chemical waves to brain waves. *Chaos* **8**: 567–575.

Kampen, N. G. (1981) *Stochastic Processes in Physics and Chemistry.* New York: North-Holland.

Karp, P. D., Riley, M., Paley, S. M., Pellegrini-Toole, A. and Krummenacker, M. (1999) Eco Cyc: encyclopedia of *Escherichia coli* genes and metabolism. *Nucl. Acids Res.* **27**: 55–58.

Ko, M. S. H. (1991) A stochastic model for gene induction. *J. Theor. Biol.* **153**: 181–194.

Ko, M. S. H. (1992) Induction mechanism of a single gene molecule: stochastic or deterministic? *BioEssays* **14**: 341–346.

Liang, S., Fuhrman, S. and Somogyi, R. (1998) Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pac. Symp. Biocomput.* **4**: 18–29.

McAdams, H. H. and Arkin, A. P. (1997) Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA* **94**: 814–819.

McAdams, H. H. and Arkin, A. P. (1998) Simulation of prokaryotic genetic networks. *Annu. Rev. Biophys. Biomol. Struct.* **27**: 199–224.

McAdams, H. H. and Arkin, A. P. (1999) Genetic regulation at the nanomolar scale: it's a noisy business! *Trends Genet.* **15**(2): 65–69.

Monod, J. and Jacob, F. (1961) General conclusions: teleonomic mechanisms in cellular metabolism, growth and differentiation. *Cellular Regulatory Mechanisms*, vol. XXVI, pp. 389–401. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Myers, C. J. (1995) Computer-aided synthesis and verification of gate-level timed circuits. Ph.D. thesis, Stanford University.

Novak, B., Csikasz-Nagy, A., Gyorffy, B., Chen, K. and Tyson, J. J. (1998) Mathematical model of the fission yeast cell cycle with checkpoint controls at the G1/S, G2/M and metaphase/anaphase transitions. *Biophys. Chem.* **72**: 185–200.

Novick, A. and Weiner, M. (1957) Mixed population response to induction of *lac* promoter. *Proc. Natl. Acad. Sci. USA* **43**: 553–566.

Ouyang, Y., Andersson, C. R., Kondo, T., Golden, S. S. and Johnson, C. H. (1998) Resonating circadian clocks enhance fitness in cyanobacteria. *Proc. Natl. Acad. Sci. USA* **95**: 8660–8664.

Peccoud, J. and Ycart, B. (1995) Markovian modelling of gene product synthesis. *Theoret. Population Biol.* **48**: 222–234.

Peskin, C. S. and Oster, G. F. (1995) Force production by depolymerizing microtubules: load–velocity curves and run–pause statistics. *Biophys. J.* **69**: 2268–2276.

Rapp, P. E. (1979) An atlas of cellular oscillators. *J. Exp. Biol.* **81**: 281–306.

Ross, I. L., Browne, C. M. and Hume, D. A. (1994) Transcription of individual genes in eukaryotic cells occurs randomly and infrequently. *Immunol. Cell Biol.* **72**: 177–185.

Samoilov, M. S. (1997) Reconstruction and functional analysis of general chemical reactions and reaction networks. Ph.D. thesis, Stanford University.

Shea, M. A. and Ackers, G. K. (1985) The O_R control system of bacteriophage lambda: a physical-chemical model for gene regulation. *J. Mol. Biol.* **181**: 211–230.

Siegele, D. A. and Hu, J. C. (1997) Gene expression from plasmids containing the *araBAD* promoter at subsaturating inducer concentrations represents mixed populations. *Proc. Natl. Acad. Sci. USA* **94**: 8168–8172.

Stemwedel, J. D., Schreiber, I. and Ross, J. (1994) Formulation of oscillatory reaction mechanisms by deduction from experiments. *Adv. Chem. Phys.* **LXXXIX**: 327–387.

Thieffry, D. and Thomas, R. (1998) Qualitative analysis of gene networks. *Pac. Symp. Biocomput.* **4**: 77–88.

Tuinenga, P. W. (1995) *Spice: A Guide to Circuit Simulation and Analysis Using Pspice*. Upper Saddle River, NJ: Prentice Hall.

Wang, M. D., Yin, H., Landick, R., Gelles, J. and Block, S. M. (1997) Stretching DNA with optical tweezers. *Biophys. J.* **72**: 1335–1346.

Yin, H., Wang, M. D., Svoboda, K., Landick, R., Block, S. M. and Gelles, J. (1995) Transcription against an applied force. *Science* **270**: 1653–1657.

Yuh, C. H., Bolouri, H. and Davidson, E. H. (1998) Genomic *cis*-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science* **279**: 1896–1902.

Part II

Nonlinear sensitivity of biological systems to electromagnetic stimuli

Electromagnetic stimuli represent a special class of external perturbations that are discussed in almost all of the remaining book chapters. Part II therefore provides important information in regard to the biophysical foundations of interactions between biological processes and electric or magnetic fields. In addition, experimental examples are described that demonstrate the nonlinear sensitivity to electromagnetic stimuli of enzymes, single cells and tissues. The principles of *electric field* interactions and the functional role of bioelectric fields are reviewed in Chapter 6 by Paul Gailey. He discusses the remarkable electrosensitivity of selected biological systems and how oscillating electric fields may be detected and amplified by biological structures. The chapter concludes with the description of a model based on the concept of long-range coherence, which may explain how relatively weak electric fields may effectively interact with excitable cellular assemblies in the presence of noise. An electric field-sensitive cellular oscillator in cells of the immune system is the subject of Chapter 7 by Howard Petty. He discusses experiments that have led to the discovery of coherent metabolic oscillations in human neutrophils, and describes the response of these cells to time-varying chemical and electric fields. His work shows the critical importance of the phase relationship between internal cellular oscillations and the externally applied field oscillations in the induction of cellular responses. Direct interactions between *magnetic field* stimuli and biological activity is the main theme of Chapter 8 by Jan Walleczek and Clemens Eichwald. Their chapter presents a brief history of research in this area and then describes work showing that enzyme activity, including oscillatory enzyme dynamics, may serve as an effective magnetic field coupling target. Further, they present results from nonlinear modeling studies that propose mechanisms by which biological processes may become sensitive to the frequency of oscillating magnetic field perturbations. Finally, Chapter 9, contributed by Stefan Engström and collaborators,