Modeling Transcriptional Control in Gene Networks—Methods, Recent Results, and Future Directions

PAUL SMOLEN, DOUGLAS A. BAXTER AND JOHN H. BYRNE*
Department of Neurobiology and Anatomy, W.M. Keck Center for the
Neurobiology of Learning and Memory,
The University of Texas-Houston Medical School,
P.O. Box 20708,
Houston, TX 77225,
U.S.A.

E-mail: jbyrne@nba19.med.uth.tmc.edu

Mathematical models are useful for providing a framework for integrating data and gaining insights into the static and dynamic behavior of complex biological systems such as networks of interacting genes. We review the dynamic behaviors expected from model gene networks incorporating common biochemical motifs, and we compare current methods for modeling genetic networks. A common modeling technique, based on simply modeling genes as ON–OFF switches, is readily implemented and allows rapid numerical simulations. However, this method may predict dynamic solutions that do not correspond to those seen when systems are modeled with a more detailed method using ordinary differential equations. Until now, the majority of gene network modeling studies have focused on determining the types of dynamics that can be generated by common biochemical motifs such as feedback loops or protein oligomerization. For example, these elements can generate multiple stable states for gene product concentrations, state-dependent responses to stimuli, circadian rhythms and other oscillations, and optimal stimulus frequencies for maximal transcription. In the future, as new experimental techniques increase the ease of characterization of genetic networks, qualitative modeling will need to be supplanted by quantitative models for specific systems.

© 2000 Society for Mathematical Biology

1. INTRODUCTION

A variety of genome projects are rapidly leading to nearly complete characterizations of the genomes of diverse prokaryotes [e.g., *E. Coli*, Blattner *et al*. (1997); *M. tuberculosis*, Cole *et al*. (1998)] and eukaryotes [e.g., *C. elegans*, The *C. elegans* Sequencing Consortium (1998); and man, Collins *et al*. (1998)]. Moreover, temporal gene expression patterns, encompassing a large portion of the genome, are now being obtained for many cell types in response to specific stimuli, or during execution of developmental programs (Wen *et al*., 1998; Iyer *et al*., 1999). Such data can

---

*Author to whom correspondence should be addressed.

help in obtaining a detailed understanding of how groups of genes control cellular responses to environmental stimuli (Edwards, 1994), and execute stored programs governing such biological processes as development (Rossant and Hopkins, 1992) or the cell cycle (Okayama *et al.*, 1996). Understanding the genetic control of neural plasticity and learning (Byrne *et al.*, 1991; Abel *et al.*, 1998; Hevroni *et al.*, 1998) will also require integration of data obtained by a variety of biochemical, biophysical, and genetic techniques.

Achieving such understanding requires more than merely collecting large amounts of experimental data by gene expression assays. A framework for deriving and expressing the biochemical architecture of genetic systems, using experimental data, is required. The precision of mathematical language makes mathematical modeling a useful framework for conceptualizing and understanding complex biochemical systems. This review discusses and compares different types of mathematical models that can be used to analyse the dynamical behavior of genetic regulatory systems. It also illustrates some recent qualitative results, obtained by modeling, that characterize the types of behaviors genetic systems are likely to generate.

1.1. *The complexity of genetic regulatory systems.* Specific groups of genes may be activated by particular signals and, once activated, regulate a common process. The group members may regulate each other's transcription. Such groups may be termed genetic regulatory systems. Genetic regulatory systems are often activated by signal-transduction pathways in which stimuli (e.g., hormones or neurotransmitters) lead to second-messenger generation and to activation of transcription factors (TFs), often via phosphorylation. Activated TFs then bind to DNA sequences known as responsive elements and thereby regulate the transcription of specific nearby genes [for a review see Karin (1994)]. Responsive elements are also termed enhancers, if they activate transcription, or silencers, if they suppress transcription. The regulatory activity of TFs is modulated both by phosphorylation and intermolecular interactions. TFs often bind to DNA as homodimers, or as heterodimers of different TF family members. Responsive elements that bind TFs have been found in turn to affect the transcription of genes for TFs; such as Jun, Fos, and Ca^{2+}/cAMP-responsive element binding protein (CREB) (Meyer *et al.*, 1993; Karin, 1994; Sassone-Corsi, 1995). The potential complexity of genetic regulation has been increased by the finding that some TFs activate their own transcription. These include Jun (Angel *et al.*, 1988), myogenic factors of the helix–loop–helix family such as Myo-D or myogenin (Thayer *et al.*, 1989), and possibly CREB (Meyer *et al.*, 1993). Several homeotic genes are known to activate their own transcription (Serfling, 1989). Other TFs repress their own transcription, either directly as in the case of Fos (Sassone-Corsi *et al.*, 1988) or indirectly by activating the transcription of repressors. An example of such an inducible repressor is the inducible Ca^{2+}/cAMP-responsive early repressor (ICER) protein, whose transcription is increased upon binding of phosphorylated dimers of CREB to a nearby

Ca^{2+}/cAMP-responsive element (CRE) (Molina *et al.*, 1993). A negative-feedback interaction occurs because ICER suppresses its own transcription (Stehle *et al.*, 1993). This genetic regulatory system has been postulated to underlie oscillations of CREB mRNA levels (Walker *et al.*, 1995). Other eukaryotic TFs that function as transcriptional repressors include CREB2, a protein within the CREB family (Sassone-Corsi, 1995), and the Ca^{2+}—regulated DREAM protein, which binds to downstream regulatory element (DRE) silencers (Carrion *et al.*, 1999). Ubiquity of genetic autoregulation in even relatively simple organisms is suggested by an inventory of *E. Coli* σ^{70} promoter regulation that identifies 21 regulatory proteins that repress their own synthesis and four that activate their own synthesis (Collado-Vides *et al.*, 1991).

A conceptual problem arises of how to understand the operations and predict the functional properties of these complex systems. Their structure must be sufficiently elaborate to execute many stored 'programs' guiding development and differentiation, and to generate appropriate responses to diverse environmental stimuli. It has become increasingly evident that non-linear interactions, positive and negative feedback within signaling pathways, crosstalk between pathways, and time delays which may result from mRNA or protein transport, all need to be considered to understand the operation of genetic regulatory systems. Understanding the combined effects of these phenomena is often beyond the capacity of intuition. Mathematical modeling can provide the necessary conceptual framework to: (1) integrate these phenomena into a coherent picture of the operation of specific genetic regulatory systems, (2) identify design principles for the biochemical architecture of genetic systems, (3) understand the responses of both normal and mutant organisms to stimuli, or (4) verify the consistency and completeness of reaction sets hypothesized to describe specific systems. To illustrate these points further, recent qualitative modeling results will be reviewed below. These results illustrate the repertoire of dynamics to be expected from genetic regulatory systems embodying common biochemical motifs.

1.2. *Approximations and methods for modeling genetic systems.* As noted by Rosen (1968), two key approximations have historically been used to model genetic regulatory systems. These are: (1) control is exercised at the transcriptional level, and (2) the production of protein product is a continuous process, with the rate determined by the balance of gene activation vs repression. One point of this review will be to examine recent efforts, particularly in prokaryotic systems, to dispense with the second of these approximations by including the stochastic nature of production of individual protein molecules. This randomness is expected, qualitatively, to lead to considerable, irreducible variability in organism phenotypes. For at least one prokaryotic system, this variability has recently been modeled in detail, and the simulations are quantitatively consistent with experimental data (Arkin *et al.*, 1998). The first approximation, that control is exercised at the transcriptional level, has been bypassed by some modeling studies that consider translational con-

trol. Mechanisms considered include regulation of mRNA decay rates (Bazhan *et al*., 1995; Carrier and Keasling, 1997), regulation of proofreading accuracy, or competition between mRNAs for initiation of translation (Von Heijne *et al*., 1987). However, the number of studies focusing on translational control is much less than those concerned with transcriptional control. Also, there is virtually no overlap between these modeling efforts, in that there are few or no studies that model translational and transcriptional control in any specific genetic system. Translational control is of widespread importance and it is evident that more experimental and theoretical collaborations would be desirable to understand its mechanisms and consequences. In this review, however, we have limited our focus to transcriptional control. We also focus on control via binding of TFs to responsive elements, and do not treat other forms of transcriptional control such as attenuation, in which transcription termination is regulated by the amount of specific tRNAs (Koh *et al*., 1998).

Methods used to model transcriptional control within genetic regulatory systems include the 'Boolean' method that describes the states of genes simply as either ON or OFF, and the 'continuous' method that uses ordinary differential equations to describe the time courses of gene product concentrations. There is also a hybrid approach that blends elements of the Boolean and continuous approaches. The Boolean method is favored for ease of formulation and computation, whereas the continuous approach has the advantage of greater physical accuracy. In the following section, we discuss the advantages and disadvantages of these approaches to modeling genetic regulatory systems. We also summarize the extensions of the continuous approach necessary for modeling intracellular transport of the macromolecules involved in genetic regulation.

2. GENETIC SYSTEMS MODELED WITH BOOLEAN LOGIC, DIFFERENTIAL EQUATIONS, AND HYBRID APPROACHES

2.1. *The Boolean approach.* Genetic regulatory systems have been modeled as networks of Boolean logical elements. Assumptions of this approach are summarized in Somogyi and Sniegoski (1996). These are: (1) the state of each gene can be characterized as either ON (one) or OFF (zero), and (2) the regulatory control of gene expression can be approximated by Boolean logical rules. An example of such a rule is if genes A and B were ON at the preceding timestep and gene C was OFF, only then is gene D ON at the current timestep. Also, (3) all elements are most commonly assumed to update their ON and OFF states synchronously (i.e., at the beginning of each, relatively large timestep based on the state of the system during the previous timestep). Occasionally, asynchronous updating is considered instead (i.e., a defined sequence of updating the logical elements with the state of the system subsequent to updating each individual element used for updating the next element). By considering the mathematical properties of Boolean net-

works, Thomas and d'Ari (1990) and Thomas *et al*. (1995) have obtained significant insights concerning the expected dynamics of genetic systems. For example, these authors illustrated that negative feedback loops are, quite generally, important for maintaining homeostasis in levels of gene products, and that positive feedback loops are important for allowing multiple stable states of gene product levels (i.e., multiple network steady states). These conclusions are in agreement with earlier studies in which differential equations were used to model gene networks [e.g., Snoussi and Thomas (1993)] and therefore help build confidence that Boolean network representations can yield useful qualitative information about gene network dynamics. As a specific example of recent Boolean modeling, Boden (1997) has attempted to model comprehensively the early development of the *Drosophila* embryo. This model is only a first step, a suggestive conceptual picture (e.g., it can only perform one-dimensional simulations at this time), which needs to be supplemented by more data. The model can, however, begin to make predictions. For example, only a few changes—e.g., activation of a few genes considered redundant in *Drosophila* embryogenesis—were needed to qualitatively simulate embryogenesis of a short germ band in a different insect, such as a beetle or grasshopper. This similarity suggests that embryogenesis of these species may be quite similar, with differences amenable to experimental verification. Also, this model illustrates the way in which the Boolean method can allow for the conceptualization and simulation, in a reasonable amount of time on a personal computer, of the dynamics of many genes and interactions among them.

Methods for efficiently using experimental data to construct models of large genetic networks with the Boolean method have recently begun to be developed (Somogyi and Sniegoski, 1997; Wen *et al*., 1998). The essential experimental task is to determine regulatory relations from simultaneously measured expression time courses of sets of genes in a group of cells or a tissue sample. The time courses could be in response to an applied perturbation, such as a hormone, or time courses could be monitored during normal processes of development or differentiation. Use of distance matrices for pair-wise comparison of expression patterns allows a cluster analysis that groups together genes whose expression is similarly regulated (Eisen *et al*., 1998; Wen *et al*., 1998). Thus, possible regulatory interactions among genes can be identified based on the clustering together of expression profiles of similar shapes. A caveat, discussed further below, is that since these experimental techniques monitor populations of cells rather than single cells, their ability to delineate intracellular signaling pathways will be limited.

2.2. *The 'continuous' approach using differential equations.* The range of nonlinear behaviors exhibited by genetic regulatory systems can be more thoroughly understood with an approach in which ordinary differential equations are used to model reaction kinetics. This approach is computationally more intensive than the Boolean approach and it requires the assumption of specific kinetic schemes. However, it has two major advantages. Smaller timesteps and continuous rather than

logical variables allow, in principle, for a more accurate physical representation of the system. Also, a large body of theory and methodology, termed dynamical-systems theory, is available to characterize the dynamics produced by these models (Guckenheimer and Holmes, 1983; Wiggins, 1990). For example, by determining those conditions under which particular solutions cease to exist or come into existence, bifurcation analysis (Wiggins, 1990) can predict parameter values at which qualitative transitions in the behavior of a system should occur. For all but the simplest models, such analysis is done numerically by means of specialized software [e.g., Doedel (1981)].

Over the last few decades, considerable work has examined the properties of steady, periodic, and even chaotic solutions of sets of differential equations that use simple, generic functions to represent activation or repression of genes by effector molecules such as TFs. Early work was summarized and extended by Tyson and Othmer (1978). For example, these authors reviewed the 'operon' model of Jacob and Monod (1961), which remains a commonly used framework for the analysis of enzyme induction and repression. In this model, each structural gene that codes for an enzyme or protein is linked with an operator gene that regulates its transcription. Each operator gene is subject to repression by a repressor molecule whose binding to the operator is in turn modulated by an 'effector' molecule. Figure 1 presents an example of this architecture for the case of two structural genes. Here, a negative-feedback loop involves two enzymes, E1 and E2, which repress their own transcription in the presence of a metabolite M. M is converted by the enzymes to an effector molecule, *Eff*, which interacts with a regulatory protein, R. The interaction converts R to R', which represses the transcription of E1 and E2, closing the negative-feedback loop. If, instead, the protein R' activated transcription of E1 and E2, then a positive-feedback loop would exist, again dependent on the presence of the metabolite M. Jacob and Monod (1961) demonstrated that the 'operon' model helps to explain two examples of genetic regulation in *E. coli*. The first example is the repression of the enzyme tryptophan synthetase (and other enzymes in the metabolic pathway for tryptophan synthesis) in the presence of the metabolite tryptophan, and the second is the induction of the enzyme β-galactosidase in the presence of galactosides.

Tyson and Othmer (1978) discussed how, for the case of one structural gene, the 'operon' model can usefully be expressed as the following set of differential equations

$$\frac{dx_1}{dt} = f(x_n) - k_1 x_1, \tag{1}$$

$$\frac{dx_j}{dt} = x_{j-1} - k_j x_j, \qquad j = 2, \ldots, n.$$

Here, if the structural gene product is an enzyme that produces the effector molecule directly, there are three x_i (i.e., mRNA, enzyme, and effector molecule). The function $f(x_n)$ embodies either activation or repression by increasing x_n, and is often
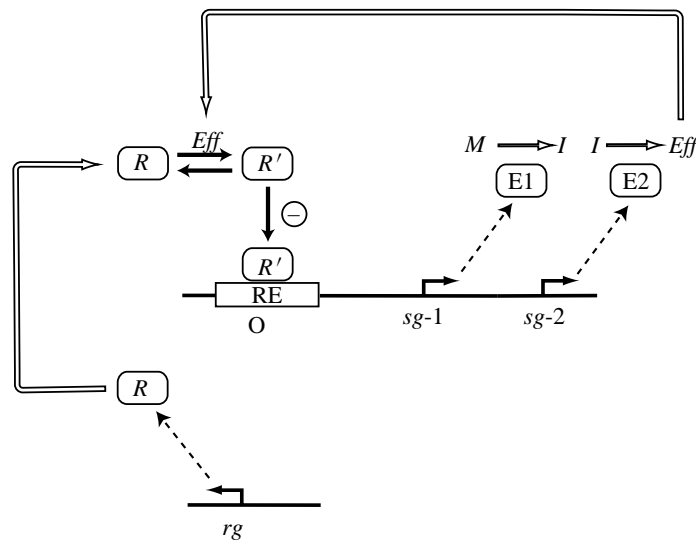
Figure 1. Diagram of the 'operon' model of Jacob and Monod (1961) for the case of two 'structural' genes, denoted *sg-1* and *sg-2*. Enzymes E1 and E2 are transcribed from these genes. E1 converts metabolite *M* to an intermediate compound *I*, and enzyme E2 converts *I* to an effector compound, *Eff*. A regulatory gene *rg* produces a protein *R*, which can reversibly convert to the form R' in the presence of *Eff*. Only R' has a regulatory effect. It represses the transcription of both *sg-1* and *sg-2* by binding to a responsive element, RE, in the region of 'operon' DNA denoted O. Overall, this scheme embodies a negative feedback of E1 and E2 on their own transcription, with the negative feedback only operating in the presence of the metabolite *M*.

taken to be a sigmoid. These equations are often termed the 'Goodwin' model (Goodwin, 1965). The properties of equations (1) have been extensively studied, e.g., by Griffith (1968a,b) and Tyson and Othmer (1978). The multiplicity and stability of steady states have been analysed, and conditions for the existence of periodic solutions have been determined. For the case $n = 3$ with $f(x_n)$ repressive, Bliss *et al.* (1982) carried out a detailed analysis. These authors allowed time delays between the production of x_3 (effector molecule) and its effect on the production of x_1 (mRNA), and also between the production of x_1 and the subsequent production of x_2 (gene product). A non-linear degradation of x_3 (effector molecule) was also included. Bliss *et al.* (1982) determined a condition on parameters that ensures the stability of the unique steady state of this model. These authors then chose parameters that allowed the model to describe the tryptophan operon of *E. coli*. With these parameters, a stable steady state was obtained, but with parameters corresponding to a mutant with reduced repression, the stability condition was violated and oscillations occurred. These simulations were in qualitative agreement with experiment for normal and mutant *E. coli* strains. We note that the analysis of Bliss *et al.* (1982), and the condition on parameters required to ensure stability of the steady state, also applies for the case $n = 2$. This case could describe direct repression of a gene by its own product.

Subsequently, Smith (1987a) generalized equations (1) to describe a network of N genes with repressive interactions between genes. In this network, the product of each gene is the first in a linear series of biochemical intermediates. The last intermediate in each series suppresses the transcription of the next gene in the network, except for the last intermediate in the Nth series, which suppresses the transcription of the first gene. Time delays are included to allow for the time required for transcription, translation, and transport. For strong repression (Hill coefficient >1), Smith (1987a) demonstrated a qualitative difference in the dynamics for even vs odd N in this generalized system. For even N, the negative-feedback interactions tend to 'cancel out', and the system can actually behave like a positive-feedback system, with multiple stable steady states coexisting. In contrast, traditional negative-feedback dynamics (i.e., periodic solutions) tend to exist for odd N, because the negative-feedback interactions cannot 'cancel out'. For a Hill coefficient of 1, Banks and Mahaffy (1978) had earlier found a single, globally asymptotically stable steady state of this system. In another study, Smith (1987b) considered the case $N = 1, n > 1$, and $f(x_n)$ cooperative (activating). Time delays were included between the production of each x_j and its effect on the production of x_{j+1}. For this case, Smith (1987b) found that no stable periodic solutions existed, and he determined the stability of the steady states.

Another frequently used type of model (Snoussi and Thomas, 1993; Mestl *et al.*, 1995a, 1996; Plahte *et al.*, 1998) that describes gene networks is:

$$\frac{dx_i}{dt} = a_i(\mathbf{Z}) - b_i(\lambda)x_i, \qquad i = 1, \ldots, n. \tag{2}$$

Here, x_i represents the concentration of the product of the ith gene. $\mathbf{Z}$ is a vector of stepped or steeply sloped threshold functions, thus the a_i often exhibit a series of steps at different hypersurfaces within x_i-space. λ is often a constant vector, although it can express hyperbolic or sigmoidal kinetics of gene product degradation. Vector components are Z_i(or occasionally λ_i) $= S_i(x_{1,\ldots,n}, \eta_i)$. S_i denotes a threshold function with threshold η_i. If, in equations (2), the b_i are constant and the a_i are taken to be linear functions of x_{i-1} except for the case $i = 1$, then one obtains a special case of equations (1).

In recent work that focuses on analytic determination of steady states and periodic orbits in equations (2), the generic threshold functions S_i have been taken as: (1) step functions, generating piecewise-linear (PL) differential equations (Snoussi and Thomas, 1993; Mestl *et al.*, 1996), (2) logoid functions, which rise linearly from 0 to 1 over a defined range of the effector (Mestl *et al.*, 1995a), or (3) steep sigmoidal functions (Plahte *et al.*, 1998). In most of this work each S_i is a function of only one gene product concentration. For PL equations, the steady states can be determined analytically. The existence and stability of periodic orbits can be partially analysed by examining Poincaré maps, although some numerical simulations may still be required to assess stability (Mestl *et al.*, 1996). Methods to classify chaotic dynamics and suggest when they could be expected have also been developed for

PL equations (Mestl *et al.*, 1996). However, the practical utility of these methods may be limited, insofar as it has not been demonstrated that chaotic dynamics operate, or would be advantageous, in any specific genetic system. Indeed, such dynamics should probably be avoided if consistency of response to external or internal stimuli is desired. Differential equations using logoids have been suggested to be superior to PL equations because they can incorporate non-additivity of the effects of multiple TFs in a way that appears more biochemically reasonable (Mestl *et al.*, 1995a). For these equations, the steady states can also be determined analytically (Mestl *et al.*, 1995a) and progress has been made in locating periodic solutions (Mestl *et al.*, 1995b). For differential equations using steep sigmoids, an analytic method for determining the steady states has recently been developed (Plahte *et al.*, 1994, 1998). However, few results have been obtained for determining periodic solutions of these equations, or for characterizing more complex dynamics. Recently, the methods of Plahte *et al.* (1994, 1998) have been applied to analyse the existence and location of equilibria in a model of cellular iron homeostasis (Omholt *et al.*, 1998). This model is based on translational rather than transcriptional control, but is still cast as a set of differential equations using steep sigmoids.

In the future, computerized versions of the methods of Plahte *et al.* (1998), Mestl *et al.* (1996), Mestl *et al.* (1995a), or Snoussi and Thomas (1993) could prove a valuable adjunct to numerical simulations. These methods could make tentative identifications, subject to experimental confirmation, of steady and periodic dynamic solutions for gene product concentrations in specific genetic regulatory systems. However, there is a caveat which greatly limits the utility of these analytic methods in predicting dynamics on the timescale of a few hours or less. Compartmentalization and transport of macromolecules is not generally included within these simplified model frameworks. The dynamics of model genetic systems are strongly dependent on the mechanism and rate of intracellular transport of mRNA and protein (Mahaffy and Pao, 1984; Mahaffy *et al.*, 1992; Smolen *et al.*, 1999a). Transport, however, cannot be included in models without eliminating the applicability of the above analytic methods for finding steady and periodic solutions.

Transport can be modeled as diffusive or active in nature. If active (energy-requiring) it can be modeled with a time delay. The time delay can be discrete, which assumes each macromolecule takes the same length of time to translocate from its place of synthesis to the location where it exerts an effect. If this simplification is too drastic, another approach is to assume a distributed delay. Here, the derivative of a variable, which can be the concentration of a macromolecule, depends on an integral of a function of one or more variables over a specified range of previous time. For example, a general distributed delay for one variable takes the form

$$\frac{dX(t)}{dt} = F(X_{\text{del}}), \text{ with}$$

$$X_{\text{del}} \equiv \int_{-\infty}^{0} X(t - \tau)G(X(t - \tau))d\tau, \text{ and with } \int_{-\infty}^{0} G(X(t - \tau))d\tau = 1.$$

The last equation expresses a normalization condition imposed for biological realism (MacDonald, 1989). In principle, this approach is general enough to model any mechanism of macromolecular transport. For example, if movement of mRNA from a transcription site to translation sites is an active process with a significant range of transport times for individual molecules, a distributed delay would be the proper modeling framework. The simplest biochemically reasonable distributed delay would be an average of $X(t)$ over an interval of width δ, centered at a delay τ. For movement of mRNA to be translated, τ might be on the order of 1–2 h, with δ taken as a percentage of τ. δ can be increased from zero to determine the value for which significant changes in simulated dynamics occur.

There is considerable evidence for active transport of specific mRNAs and proteins in eukaryotic cytoplasm. Examples include microtubule-dependent movement of $Vg1$ mRNA in *Xenopus* oocytes (Yisraeli *et al.*, 1990), microtubule-dependent localization of several maternal mRNAs for proteins that direct embryonic development (such as *bicoid, bicuadal-D*, and *oskar*) in *Drosophila* oocytes (King, 1996), movement of tubulin monomers from the cell body to the ends of growing axons, in a variety of vertebrate neuronal types, at rates well in excess of those that could be explained by diffusion (Sabry *et al.*, 1995); and, more generally, axonal and dendritic transport of various types. The above studies have not, however, provided a sense of the extent to which a 'typical' eukaryotic mRNA is actively transported. Active macromolecular transport in prokaryotic cells can also be important. For example, it has been observed that specific intracellular localization of proteins determine the relative fates of daughter cells after cell division (Shapiro and Losick, 1997; Jacobs and Shapiro, 1998). Either these proteins, or other macromolecules that bind them, must be actively transported in order to generate their intracellular concentration gradients.

2.3. *Comparison of the Boolean and continuous approaches.* Although the continuous approach is more computationally intensive, it is often preferred because it is more physically accurate. The limitations of the Boolean approach have been known for some time. Although a steady state of a Boolean representation of a genetic system will correspond, qualitatively, to an analogous steady state in the 'continuous' representation with ordinary differential equations, the converse is not necessarily true. Not all steady states of the continuous representation are required to appear as steady solutions in the Boolean representation (Glass and Kauffman, 1973). Also, periodic solutions in the Boolean representation may not correspond to periodic solutions in the continuous system. Several examples are given in Glass and Kauffman (1973) and Bagley and Glass (1996). For example, Glass and Kauffman (1973) considered the simple feedback loop in which X_1 induces the production of X_2 and conversely X_2 represses the production of X_1. Using 1 to refer to a gene being ON and 0 to refer to a gene being OFF, the Boolean representation of this system has four possible states: (1,0), (0,1), (1,1) and (0,0). Assuming synchronous updating of the states of X_1 and X_2, the Boolean

representation has a periodic solution through the four states in succession (i.e., if the system starts in (1,1), then repression will make the next state (0,1), then lack of activation will make the next state (0,0), then lack of repression allows (1,0), then induction gives (1,1) again, and so forth). However, the continuous representation of this system, using steep sigmoidal functions for induction and repression, does not have stable oscillations. Rather, there is a globally stable steady state. Only if an additional complication is added (i.e., if X_1 and X_2 are assumed to be synthesized in separate compartments) is a stable limit cycle found for some parameter values. Glass and Kauffman (1973) gave a procedure for determining whether limit cycles in low-dimensional (2–4 variable) Boolean systems are 'spurious' in the sense that they would not be seen in the continuous representation. However, no such procedure appears to have been developed for higher-dimensional systems.

Recent work indicates that these considerations continue to limit the utility of Boolean modeling of gene networks, and that this problem cannot be fully resolved by techniques such as comparing results obtained by assuming asynchronous with synchronous updating of gene states in the Boolean representation (Mestl *et al*., 1996). For example, an accurate Boolean model should ensure that the order in which gene product concentrations cross 'threshold' levels for exerting biological effects is the same as the order determined with a more biochemically accurate model based on differential equations. However, neither synchronous nor asynchronous updating rules have been found that can guarantee this outcome.

Also, a considerable amount of work [e.g., Mahaffy and Pao (1984), Mahaffy *et al*. (1992), Smolen *et al*. (1999a)] has made it evident that the dynamics of model genetic systems are strongly dependent on whether, and how, intracellular transport of mRNA and protein is modeled. In prokaryotic cells, it may sometimes be less necessary to model transport in detail. Translation of mRNA often begins immediately upon its synthesis, even while transcription is still ongoing (Koh *et al*., 1998). Also, if transport is diffusive, the small dimensions of many prokaryotic cells and the lack of internal compartments might imply that, on the timescale of minutes, concentrations of many mRNAs and proteins can be regarded as homogenous. However, recent studies have demonstrated that the homogeneity assumption does not always hold. For example, in several cases it is now established that localization of proteins within specific regions in a parent bacterial cell determines the relative fates of daughter cells after division (Shapiro and Losick, 1997; Jacobs and Shapiro, 1998). Thus, homogeneity of macromolecules in prokaryotic cells cannot be casually assumed.

Macromolecular transport often cannot be incorporated into a Boolean model. If transport of macromolecules is described by a discrete time delay, this could be incorporated by considering the delay to be part of the relatively large timestep used in simulations or by propagating its effect across several timesteps. However, this incorporation does not appear possible if transport is described by diffusion or by a distributed delay. In contrast, models based on differential equations that incorporate transport (either via distributed delays or via diffusion) have been analysed

extensively. For example, early work summarized in Tyson and Othmer (1978) established conditions for a diffusive, spatially non-uniform system containing a single negative-feedback loop to have stable steady states. Distributed delays, which could represent macromolecular transport or slow biochemical reactions not modeled explicitly, are included in the classes of gene network models analysed by Smith (1987a,b), which generalize equations (1). His results concerning the existence and stability of multiple steady states, and the existence and stability of periodic solutions, hold with these distributed delays.

The disadvantage of models based on ordinary differential equations is that they typically take much more computer time to integrate over simulated times of perhaps hours or days than Boolean models do. Thus, for modeling large genetic regulatory systems, or lengthy processes such as development, Boolean networks might represent the only practical alternative.

2.4. *A hybrid Boolean—continuous approach.* A hybrid modeling approach has been put forth in which biochemical processes characterized by sharp thresholds, such as the activation of genes over a narrow concentration range of an effector molecule, are represented as Boolean ON–OFF switches. However, activation functions not characterized by sharp thresholds are modeled as continuous input–output relations. Either continuous time courses, or switch-like functions, can be assumed for hormonal or other inputs. In this approach, macromolecular concentrations are numerically integrated as ordinary differential equations. Time delays can also be incorporated.

McAdams and Shapiro (1995) used the hybrid approach to model the *E. Coli–λ* phage lysis–lysogeny genetic switch. Figure 2 illustrates their hybrid model. Here, G1–8 denote distinct computational elements, such as AND or NOT gates, or the key bistable two-promotor (P_R and P_{RM}) switch which integrates the output of the remainder of the circuit to make the lysis vs lysogeny decision. The symbol 'kinetic model required' refers to activation or switching functions not characterized by sharp thresholds and therefore modeled as continuous input–output relations. Model simulations qualitatively matched experimental data for the order of promotor activation following phage infection, and also demonstrated that the time delays associated with transcription and translation of key proteins were essential for correct sequencing of circuit functions and for circuit decision-making.

A second recent example of the hybrid approach is the model of Yuh *et al.* (1998) for regulation of the developmental gene *endo16* in sea urchin embryo. Upstream from the basal *endo16* promoter, seven distinct modules contain a variety of sites for binding of TFs that regulate *endo16* transcription. The occupancies of some of these modules (or, alternatively, their presence or absence in mutated constructs) were modeled by Boolean switching functions whose sums or products determine the final transcriptional activity. The occupancies of three sites for TF binding were modeled in more detail by continuous functions of time. One of these sites integrates the regulatory effects of all the others to determine the 'output'—*endo16*
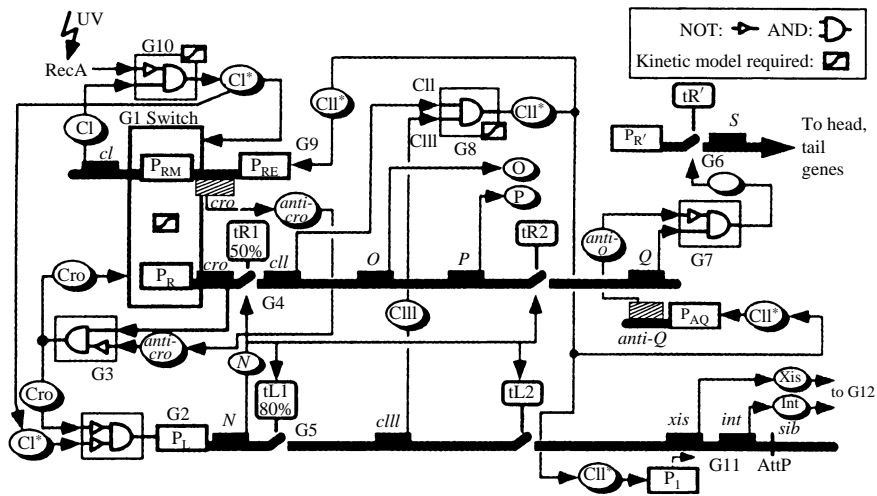
Figure 2. A model of the genetic circuit governing the phage λ lysis–lysogeny decision in *E. Coli*. Early λ genes are shown configured in operons relative to controlling promoters. Arrows connect genes with the sites of action of their protein products. Rectangles enclosing Boolean logic symbols identify control logic amenable to approximation as Boolean logic. The sigmoid curve-in-rectangle symbol identifies control logic requiring continuous kinetic modeling (as opposed to Boolean threshold functions). Reproduced with permission from McAdams and Shapiro (1995). ➡️, DNA; [P_L], Promoter; ▬, Gene; ▨, Antisense gene; CI, Gene product; CI*, Postreaction gene product; ◣▪, Terminator; ➝, Signal path.

transcription—which is therefore also a continuous function of time. The Boolean switching functions sometimes act to scale the output without changing its shape, and other times to set the output to zero. This model emphasizes the analog computational nature of the *endo16* regulatory scheme. It integrates complex environmental information via occupancies of TF binding sites to produce, via Boolean switches dependent on these occupancies, a well-defined output profile of gene expression.

A particular advantage of the hybrid approach is, therefore, that a clear conceptual distinction is drawn between elements of the system which can, because of sharp thresholds of activation, be modeled as Boolean logical variables, and those which cannot. Also, the computational time required for simulations using this approach can be significantly less than with the continuous approach.

In the following section, we have chosen to review some representative recent characterizations of dynamics expected from genetic regulatory systems. These results are generally obtained using the continuous approach, with small sets of ordinary differential equations that embody typical schemes of positive and negative feedback by transcriptional activators and repressors. Such models are valuable for qualitatively illustrating the types of dynamics larger genetic systems that incorporate similar feedback loops can be expected to exhibit.

3. SOME RECENT INSIGHTS CONCERNING THE DYNAMIC PROPERTIES OF GENETIC REGULATORY SYSTEMS

3.1. *Time delays and multistability lead to history-dependent responses to perturbations.* As discussed above, the concept that positive feedback could give rise to multistability (i.e., multiple stable steady states of concentrations of gene products and other system variables) has been appreciated since early models of genetic regulation (Griffith, 1968b; Tyson and Othmer, 1978). Recent modeling studies have continued to delineate the biochemical architectures that could give rise to multistability. For example, Keller, using biochemically realistic differential equations, determined the number and stability of steady states given typical schemes of activation or repression by multiple or oligomeric TFs (Keller, 1994, 1995). Keller (1994) applied these results by suggesting that the choice among multiple steady states of gene product concentrations could constitute a form of epigenetic information, mitotically inheritable, and possibly sufficient to distinguish different cell types within a multicellular organism.

Wolf and Eeckman (1998) have pointed out that the simplest genetic systems capable of bistability and perturbation-induced state transitions are: (1) a single gene with two binding sites for a monomeric TF, or (2) a gene with a single binding site for a dimeric TF. These authors also use the term 'soft switch' to denote the configuration of two genes, both regulated by a single binding site which can accept either monomeric gene product, with each gene product repressing the other gene. This system has only a single steady state. The location of the steady state can be quite sensitive to parameters, leading the authors to suggest a switch-like behavior. However, such a 'soft' switch does not actually seem to constitute a true switch, insofar as the system always returns to the original steady state after the cessation of the imposed perturbation.

From a more theoretical perspective, it has been demonstrated that positive-feedback loops are necessary for multistability, in both the Boolean and continuous approaches (Snoussi and Thomas, 1993; Thomas, 1994; Thomas *et al.*, 1995). Such a positive-feedback loop can be direct, as when a TF activates its own transcription. It can also be indirect, and somewhat counter-intuitive, as when two TFs repress each other's transcription. In that case, an increase in the first TF, by repressing the second, indirectly favors a further increase in the first (Keller, 1995).

It is evident that the response of a multistable model gene network to perturbations will depend on the state of the network when the perturbation is applied. However, in a model gene network with delay differential equations, the response to a given perturbation can be dramatically different depending not only on the levels of gene products present immediately prior to the perturbation, but also on the history of gene product levels over the range of previous time encompassed by the delays. We have recently investigated generic models that illustrate this concept. The simplest model (Smolen *et al.*, 1998, 1999a) incorporates a single transcription factor, TF-A, that activates its own transcription [Fig. 3(a)]. TF-A

forms a homodimer that activates transcription by binding to enhancers (TF-REs). A rapid equilibrium is assumed between monomeric and dimeric TF-A. The *tf-a* gene incorporates a TF-RE. The transcription rate saturates with TF-A dimer concentration to a maximal rate k_f, which is proportional to TF-A phosphorylation. Responses to stimuli are modeled by varying the degree of TF-A phosphorylation. A basal synthesis rate R_{bas} at negligible [TF-A] is present, as well as a first-order process for degradation (k_d). Transport of macromolecules is modeled simply, by incorporating a discrete time delay $\tau(\sim 1-2$ h) between any change in the level of nuclear TF-A and the appearance in the nucleus of further TF-A synthesized in response to the nuclear TF-A. τ lumps together the two delays τ_1 and τ_2 indicated in Fig. 3(a). The model is described by a single delay differential equation

$$\frac{d[\text{TF-A}]}{dt} = \left\{ \frac{k_f[\text{TF-A}]^2}{[\text{TF-A}]^2 + K_d} \right\}(t - \tau) - k_d[\text{TF-A}] + R_{bas}. \tag{3}$$

Equation (3) exhibits bistability in [TF-A] [Fig. 3(b)]. For a range of parameters, there is one solution with [TF-A] low and its synthesis rate close to R_{bas}, and another with [TF-A] high and its synthesis rate close to k_f. Moreover, perturbations can switch the model between these states. Perturbations, conceived as signal-dependent activation of protein kinases or phosphatases, were modeled as transient changes in the maximal transcription rate k_f. Induced state transitions could correspond physiologically to brief stimuli, such as exposure to a hormone, leading to long-lasting changes in the levels of particular proteins. Following an increase in k_f (or R_{bas}), Fig. 4(a) illustrates that the model would undergo a 'staircase' transition between steady states of [TF-A]. There is a series of steps in the transcription rate of *tf-a* mRNA, and therefore in [TF-A]. Each step is separated by approximately the delay τ. The steps are caused by successive increases in positive feedback due to previous steps in [TF-A].

Figure 4(b) illustrates two ways in which the transient responses of this system are history-dependent. First, a brief increase in the maximal transcription rate k_f, applied at arrow *a* when the system is in the upper steady state, gives a subsequent large excursion of [TF-A] above the upper steady state following the delay τ, at $t = 240$ min. Then, starting at $t = 340$ min (bar), the baseline value of k_f is decreased so that the system falls to a lower steady state. Afterwards, the same brief increase in k_f, applied at arrow *c* when the system is in the lower steady state, gives only a negligible subsequent excursion of [TF-A], at $t = 870$ min (within the dashed circle, excursion is too small to see). This difference in response of [TF-A] is a model for 'priming' of a system, by a longer lasting stimulus sufficient to cause a state transition, to respond more or less vigorously to subsequent stimuli. Second, Fig. 4(b) illustrates that the delay τ gives the system a type of 'memory'. Following the beginning of the sustained decrease in k_f (bar), the excursion in [TF-A] due to a subsequent brief change in k_f depends strongly on whether the interval between these two changes is less than or greater than τ. A brief increase in k_f applied at arrow *b*, shortly after the sustained decrease in k_f, gives a subsequent large excursion
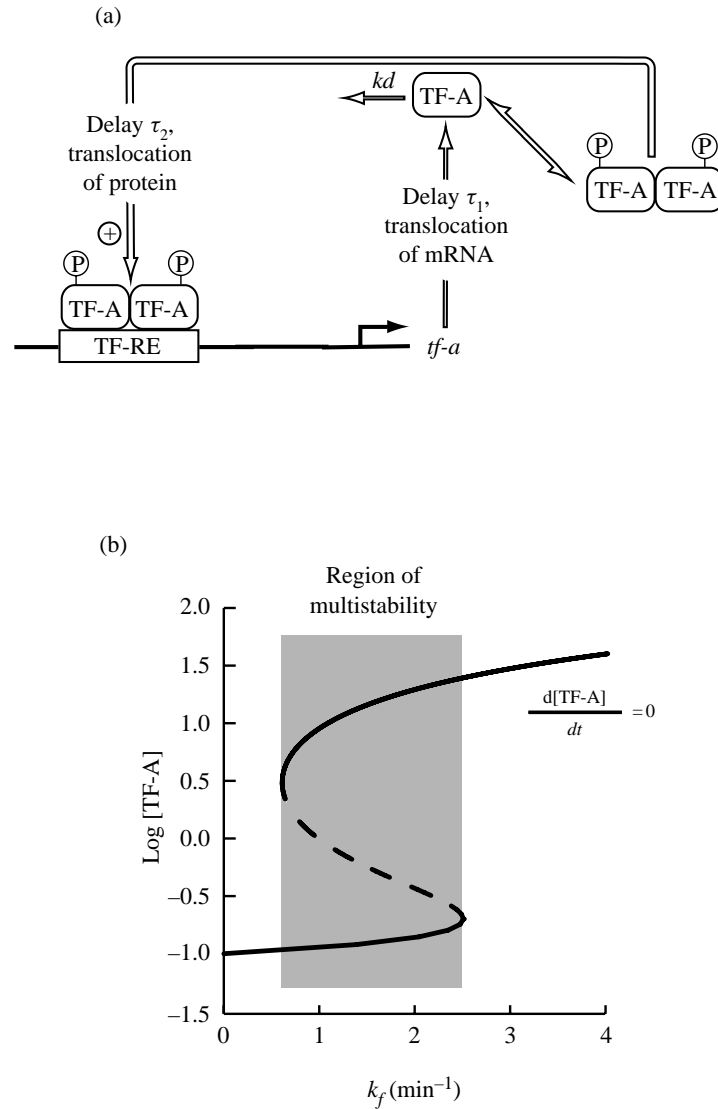
(a)

(b)

Figure 3. (a) Phosphorylated dimers of TF-A activate *tf-a* transcription when bound to specific DNA sequences (TF-REs). Degradation (k_d) is also indicated. (b) Bistability in the model of (a). For 0.6 min^{-1} < k_f < 2.5 min^{-1}, two stable steady-state solutions of [TF-A] exist (lower and upper portions of $\frac{d[\text{TF-A}]}{dt} = 0$ curve) with an unstable solution between (middle, dashed portion). Outside this region there is a single steady-state solution. Modified from Smolen *et al.* (1998).
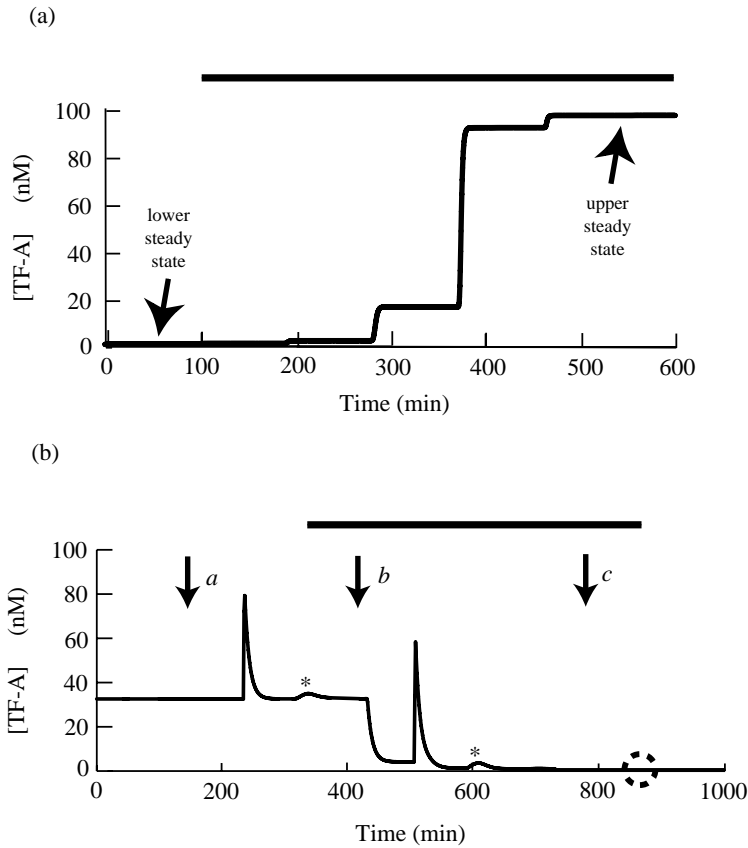
(a)

(b)

Figure 4. Responses to stimuli in the model of Fig. 3(a). (a) A transition between a lower steady state and an upper steady state is elicited by a long-lasting perturbation (bar). The fraction of TF-A considered to be phosphorylated, which is proportional to the maximal transcription rate k_f, is increased at $t = 100$ min. After a length of time equal to the delay τ(90 min) a small step in [TF-A] occurs. This is followed by successive, larger steps to a new steady state. (b) State-dependent responses to perturbations. The model is initially in a steady state of high TF-A. As in part (a), $\tau = 90$ min. At $t = 150$ min (arrow labeled a) k_f is increased for 2 min, generating an increase in [TF-A] at $t = 240$ min. At $t = 340$ min a transition to the lower steady state was elicited by decreasing k_f to a new baseline (bar). After the delay, at $t = 430$ min, [TF-A] decreases to the low steady state. At $t = 410$ min (arrow labeled b) k_f is increased for 2 min. A large excursion in [TF-A] results at $t = 500$ min. At $t = 780$ min (arrow labeled c) k_f is again increased for 2 min. Only a negligible excursion in [TF-A] occurs at $t = 870$ min (circle). The small increases in [TF-A] that are labeled by a * are 'echoes' of the larger excursions (see text). Perturbations $a–c$ are all identical. Modified from Smolen *et al.* (1999a).

of [TF-A] at $t = 500$ min, whereas the same brief increase applied later at arrow c gives a negligible excursion of [TF-A] at $t = 870$ min. This 'memory' is due to the delay required for a change in transcription rate to cause a change in nuclear protein level. The perturbation at arrow b occurred when [TF-A] was still high. The high [TF-A] combined with the increased k_f to strongly activate *tf-a* transcription.

Also, following a single brief increase in transcription rate, the model of Fig. 3(a) often exhibits one or more brief subsequent increases ['echoes' in the legend of Fig. 4(b)] in [TF-A] and in transcription rate (Smolen *et al.*, 1999a). Finally, similar results were observed when distributed delays replaced discrete delays (Smolen *et al.*, 1999a).

3.2. ***Conditions permitting stable oscillations in models with feedback loops are continuing to be delineated.*** Models of genetic regulation have been developed to explain a variety of periodic biological phenomena. These include periodic hormonal secretion (Liu *et al.*, 1997), and the circadian rhythm, which appears in many organisms to be based on a negative-feedback loop of one or two core genes repressing their own expression (Reppert, 1998). It is of interest to delineate the structure of the simplest types of models that could exhibit periodic solutions stable to small perturbations. Models such as that of Fig. 3(a) have only one type of feedback—either positive or negative. In the absence of a time delay, such a simple model cannot support stable oscillations. More generally, oscillations are not possible with only positive feedback, whereas with a single negative-feedback loop, oscillations are only possible if the loop contains at least three variables and if the feedback term contains a high power of the concentration of the repressing species (Thomas *et al.*, 1995). However, if a delay is present, due for example to macromolecular transport, stable oscillations can be obtained even with a single negative-feedback loop containing one element [e.g., a single transcription factor that directly inhibits its own transcription (Thomas *et al.*, 1995)]. As noted above, Smith (1987a) gave conditions for periodic solutions to exist that apply to a class of gene network models incorporating a single negative-feedback loop, with or without delays. What if only positive feedback is present with a delay—can there be stable oscillations? As discussed above, Smith (1987b) ruled stable oscillations out for the system of equations (1) with a cooperative feedback function $f(x_n)$ and with time delays for transcription and translation included. Smolen *et al.* (1999a) used a somewhat simpler analysis to rule out stable periodic solutions for models similar to Fig. 3(a), with a single TF and a single positive-feedback loop. It appears unlikely that a reasonable model for a genetic network, incorporating only positive feedback, could yield periodic behavior.

As a simple example of the effect of both negative and positive feedback within the same genetic system, an additional gene *tf-r* can be introduced into the system of Fig. 3(a) (Smolen *et al.*, 1999a). The rate of *tf-r* transcription is increased by binding the TF-A dimer to a TF-RE. TF-R monomer represses transcription of the genes *tf-a* and *tf-r* by competitively inhibiting binding of TF-A dimers to TF-REs

[Fig. 5(a)]. Delays are included between changes in the transcription rate of either gene and changes in the concentration of the corresponding protein, giving

$$\frac{d[\text{TF-A}]}{dt} = \left\{ \frac{k_{1,f}[\text{TF-A}]^2}{[\text{TF-A}]^2 + K_{1,d}(1 + [\text{TF-R}]/K_{R,d})} \right\}(t - \tau)$$
$$- k_{1,d}\text{TF-A} + R_{\text{bas}} \tag{4}$$

$$\frac{d[\text{TF-R}]}{dt} = \left\{ \frac{k_{2,f}\text{TF-A}^2}{\text{TF-A}^2 + k_{2,d}(1 + [\text{TF-R}]/K_{R,d})} \right\}(t - \tau)$$
$$- k_{2,d}[\text{TF-R}]. \tag{5}$$

Parameters in equations (4) and (5) are analogous to those in equation (3). $K_{R,d}$ is the dissociation constant of TF-R monomers from TF-REs. Robust oscillations in the rates of both *tf-a* and *tf-r* transcription are readily generated by this model when $\tau = 0$ [Fig. 5(b)]. Discrete delays of an order reasonable for macromolecular transport ($\tau \approx 2$ h) eliminate this periodic solution. However, a new limit cycle, robust to perturbations, and with a period on the order of the delay, is created.

An important example of oscillatory transcription is the circadian rhythm found in most organisms adapted to day–night cycles. Oscillators that generate circadian rhythms are emerging as a key group of genetic systems where a core negative-feedback loop underlies periodicity. The basic 'clock' appears to involve recurrent repression by one or a few core gene products of their own expression, with the core genes in turn regulating the transcription of output genes more directly tied to behavior. Goldbeter (1995) and Leloup and Goldbeter (1998) have used models based upon single negative-feedback loops to simulate circadian rhythms in *Drosophila*. In *Drosophila*, and apparently also in mammals, accumulation of the PER protein represses transcription of its own core gene, *per*. The TIM protein dimerizes with PER and, in *Drosophila* but not mammals, exhibits a circadian rhythm in phase with PER. Progressive, slow phosphorylation of PER and TIM is followed by rapid degradation and a surge of renewed *per* expression. In *Neurospora* yeast, the FRQ gene product exhibits a qualitatively similar cycle (Merrow *et al.*, 1997; Luo *et al.*, 1998). Even in cyanobacteria, genes that constitute an analogous negative-feedback loop have been found (Ishiura *et al.*, 1998).

Some modeling of circadian rhythms, including recent efforts [e.g., Scheper *et al.* (1999)], is essentially phenomenological in that little biochemical justification is given for the equations used. However, biochemically realistic models of circadian rhythmicity in *Drosophila*, which should apply without great modification to mammals, have been developed recently by Goldbeter and coworkers. These models (Goldbeter, 1995; Leloup and Goldbeter, 1998) have a single negative-feedback loop and generate robust oscillations by effectively including a biochemical time delay—slow, obligatory phosphorylation of PER protein prior to its degradation. These models were fitted to some experimental data. The observed lag between the concentration peaks of *per* mRNA and PER-TIM heterodimer was fitted, and
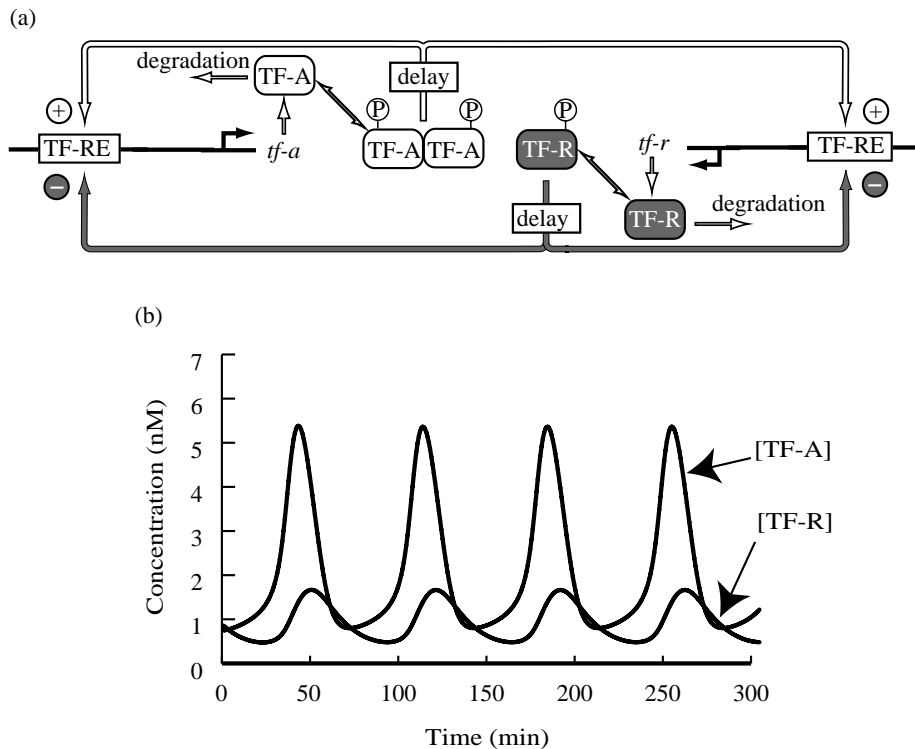
(a)

(b)

Figure 5. (a) Schematic resulting from the addition of a second transcription factor, TF-R, to the model of Fig. 3(a). TF-R represses transcription by competing with the TF-A dimer for binding to TF-REs. Delays between transcription and appearance of functional protein are indicated for TF-A and TF-R. (b) Sustained oscillations of [TF-A] and [TF-R] produced by the model of (a). Modified from Smolen *et al.* (1998).

an experimental phase-response curve for the *Drosophila* rhythm shift in response to brief light pulses applied at various times throughout the day was qualitatively reproduced (Leloup and Goldbeter, 1998). Recent data should allow more quantitative and comprehensive models for *Drosophila* or mammalian circadian rhythms. For example, it now seems that the PER-TIM heterodimer represses *per* transcription indirectly, by binding with an activating TF termed CLOCK. Measurements of the time courses of PER, TIM, and CLOCK protein and mRNA levels have recently become available (Lee *et al.*, 1998). Also, it seems timely to develop a more quantitative model (Smolen *et al.*, 1999b) based on experimental data concerning the *Neurospora* oscillator. This oscillator may depend only on one core oscillating gene, the *frq* gene, although its autorepression may also be indirect. Recent experiments have quantified the time course of repression by the *frq* gene product following its own induction, and the time required for degradation of the *frq* gene product (Merrow *et al.*, 1997). This data should greatly help in constructing a quantitative model.

Finally, we briefly consider modeling of cell cycle oscillators. Models are beginning to incorporate complex biochemical features such as controls that determine when mitosis initiates (Novak *et al.*, 1998a) and DNA synthesis (Novak *et al.*, 1998b), but explicit processes of genetic activation or repression have not yet been included. An integrated model of prokaryotic cell cycle regulation, including genetic regulation, may be feasible in the near future, because the relevant genetic systems controlling chromosome replication, cell division, and cell growth are becoming relatively well characterized [e.g., Donachie (1993); Roberts *et al.* (1996)].

3.3. ***Dynamics can be very different when macromolecular transport is modeled with a time delay vs diffusion.*** In studies using time delays, the most common objective is to determine, for given kinetic parameters, values of the delay at which steady states lose stability and periodic solutions are concurrently formed. Several studies [e.g., Mahaffy and Pao (1984); Busenberg and Mahaffy (1985)] have extended this approach by analysing model gene networks that contain a negative-feedback loop and that include both diffusive macromolecular transport and active macromolecular transport modeled as a delay. The study by Busenberg and Mahaffy (1985) is of particular interest in that it demonstrates how to represent diffusion as a distributed delay. In these models, if transport was slowed by increasing the delays, steady states were often found to lose stability to periodic solutions. However in contrast, if transport was slowed by reducing diffusion coefficients, oscillations were damped. The parameter regime in which both diffusion and delays contributed significantly to total transport time could only be analysed by numerical simulations. These simulations showed that for some parameter values, modest reductions in diffusion coefficients could also be destabilizing and create oscillations (Busenberg and Mahaffy, 1985). Nevertheless, a major conclusion of these studies is that slowing transport by decreasing diffusion coefficients is much less likely to destabilize steady states than is slowing transport by increasing time delays.

When might intracellular macromolecular transport be appropriately described by a discrete, as opposed to a broadly distributed, time delay? If transport of macromolecules is primarily active (i.e., driven by motor proteins such as kinesin or dynein) we have argued (Smolen *et al.*, 1999a) that the variance in the time taken for individual molecules to move a given distance could be quite small compared with the mean—much smaller than if transport was diffusive. Then, a discrete time delay could be an appropriate description of transport from, for example, a transcription site to a particular cytoplasmic locale. Briefly, our argument considered recent quantitative data describing transport of latex beads by kinesin (Schnitzer and Block, 1997). These data allowed an estimate that, for typical intracellular distances of transport, the variance in time taken could be several orders of magnitude less for kinesin-driven transport than for diffusion. Even considering caveats (e.g., translocation of mRNA from its place of transcription to that of translation is not likely to be along a single continuous cytoskeletal element), this estimate suggests

it might be plausible to model active transport by a discrete or narrowly distributed time delay. However, it is not plausible to model diffusion in this manner.

We have examined whether the dynamic phenomena exhibited by the model of Fig. 3(a) with discrete or narrowly distributed delay—multiple stable states and repeated perturbations of [TF-A] following a single stimulus—are preserved in the case of diffusive transport. To address this issue, a model cell divided into spherical shells was considered, with the nucleus a small sphere at the center (Smolen *et al.*, 1999a). A total cell radius of 10 μm was assumed for most simulations, divided into 10 shells of 1 μm thickness. In each shell, the model of Fig. 3(a) was extended by assuming separate differential equations for the rates of change of mRNA and protein concentrations. mRNA and protein were assumed to diffuse between the shells. Equations from Blumenfeld *et al.* (1992) were used to model radial diffusion.

Bistability was still present for this extended model. That is, for physiologically reasonable diffusion coefficients, there were two stable intracellular profiles of [TF-A], each stable to small perturbations, with one characterized by a much higher average [TF-A] than the other. However, in response to a large perturbation that sufficed to cause a state transition, the time course of protein and mRNA concentrations never exhibited a 'staircase' pattern of steps of [TF-A] or of *tf-a* transcription rate. Nor were there ever repeated transients of [TF-A] in response to a single perturbation [i.e., 'echoes' as in the legend of Fig. 4(b)]. Diffusive transport, as opposed to transport characterized by a time delay, acts to 'spread out' any excess of mRNA or protein in one region, which eliminates the distinct spatial peaks necessary for steps of [TF-A], or for 'echo' perturbations in [TF-A].

Diffusive transport of mRNA and protein was similarly incorporated into the model of Fig. 5(a), with both transcriptional activation and repression (Smolen *et al.*, 1999a). With this model, if diffusion coefficients were assumed not to depend on shell number, then oscillations similar to those of Fig. 5(b) were readily obtained for physiologically reasonable values of diffusion coefficients. Very small diffusion coefficients were required to suppress oscillations. However, simulations were also performed that crudely simulated the existence of a nuclear membrane by reducing the diffusion coefficient between the second and third spherical shell. This inhomogeneity strongly suppressed oscillations. For example, diffusion coefficients for protein and mRNA were reduced to 0.03 μm^2 s^{-1}, which corresponds to assuming that a molecule moves one shell width, crossing the membrane, in a time on the order of 3 min. This reduction sufficed to abolish oscillations when other parameter values were similar to those used in the simulation of Fig. 5(b). Together, the above simulations demonstrate that with physiologically reasonable diffusion coefficients and a crude representation of a nuclear membrane, many dynamic phenomena that were exhibited by the models of Figs 3(a) and 5(a) when transport was neglected are not now preserved. Oscillations and repeated perturbations of [TF-A] following a single stimulus are abolished, although multistability is still present.

3.4. *Intermolecular interactions contribute non-linearities important for a variety of complex dynamic behaviors.* Early modelers of negative and positive feedback realized that oligomerization might contribute to increasing the Hill coefficient of feedback, thus favoring the occurrence of multistability in the case of positive feedback or periodicity in the case of negative feedback [e.g., Griffith (1968a,b)]. However, recent models have provided additional examples of dynamic complexity for which oligomerization of transcription factors may be essential, such as the creation of moving 'stripes' of transcription essential for embryonic development (Kerszberg and Changeux, 1994; Kerszberg, 1996). Many, if not most, eukaryotic transcription factors bind to their target sequences as homo- or heterodimers, or in larger oligomers. For example, CREB and related TFs function as oligomers (Sassone-Corsi, 1995), as do members of the Fos–Jun family (Kouzarides and Ziff, 1988). Keller (1995) developed a simple method for deriving the differential equations corresponding to a given regulatory scheme. He analysed multistability in four typical genetic regulatory schemes in which homo- or heterodimerization of TFs is essential. However, oligomerization is not essential for multistability in all genetic regulatory schemes. Schemes with more than one binding site for monomeric TFs can also exhibit multistability. Keller (1995) analysed multistability in two such schemes. In the first, a monomeric transcription factor binds to two distinct sites in its own promoter, with both sites activating transcription. The resulting equation for the rate of change of TF concentration, denoted by X, was derived as

$$\frac{dX}{dt} = \frac{\lambda(aX + bX^2)}{1 + cX + bX^2} + \varepsilon - \eta X.$$

The parameter b is of interest in that it can be either positive or negative, it denotes the strength and sign of synergism between TF bound to site 1 and site 2. In this model, two stable steady states exist for a range of parameter values, with an unstable steady state in between. The second model consists of two monomeric TFs, each repressing the transcription of the other at a single site in its promoter. The equations are more complex, but the result is the same. For a range of parameters there are two stable steady states separated by an unstable steady state.

Kerszberg and colleagues (Kerszberg and Changeux, 1994; Kerszberg, 1996) developed simple, elegant models in which dimers of TFs are used to interpret morphogenetic gradients in terms of embryonic gene transcription patterns. The basic model (Kerszberg and Changeux, 1994) uses a gradient of one transcription factor, M, which is itself a morphogen or responsive to a morphogen. M forms heterodimers with a second transcription factor, V. Homodimers of M and V also can be formed. Depending upon which of these three oligomeric species activates transcription of V, the model can generate a gradient of V, a sharp boundary for V, a stationary stripe of V, or a moving stripe of V. It was suggested that the *Drosophila* bicoid and hunchback proteins might function analogously to M and

V, respectively. Subsequently, Kerszberg (1996) have elaborated this model to describe more specifically transduction by retinoic acid receptors of a gradient of retinoic acid into patterns of embryonic gene expression. Here, the RXR and RAR receptors play roles similar to *M* and *V*, respectively. The receptors were assumed to regulate an output gene, which could encode, for example, a protein necessary for cell differentiation. The output gene was assumed to be activated by the RXR homodimer. A moving stripe of transcription of the output gene was seen for a range of parameter values, and this phenomenon was robust to changes of about 20% in any kinetic parameter.

Finally, the result of the following section—a possible maximum in transcription at a particular stimulus frequency—also depends on TF-A acting as a dimer (Smolen *et al.*, 1998). All of these results help to emphasize that protein oligomerization can provide an important mechanism for generating complex dynamics of transcription.

3.5. *Optimal stimulus frequencies for transcription could be generated by competition between transcriptional activators and repressors.*

Some recent experiments have demonstrated optimal stimulus frequencies for activation, or repression, of transcription. Transcription of the cell adhesion molecule L1 in cultured neurons is strongly repressed by imposed continual 0.1 Hz electrical stimulation, but not by 0.3 Hz stimulation (Itoh *et al.*, 1995). Also, *c-fos* transcription in cultured neurons is enhanced almost 200% by bursts of six 10 Hz electrical stimuli with an interburst interval of 1 min, but not by bursts of 12 stimuli with an interburst interval of 2 min (Sheng *et al.*, 1993). Continuous 0.1 Hz stimuli gave a 70% enhancement.

These results might be explained by a model in which an intermediate intensity or frequency of stimulation phosphorylated and activated one TF, which activated the transcription of a target gene, whereas a higher frequency of stimulation activated a second TF that repressed transcription of the target gene. Moreover, a similar model might help explain an apparent optimum stimulus frequency for the formation of long-term memory (LTM). The relationship between stimulus frequency and the amount of LTM formation, and by inference amount of transcription, can show an optimum in *Drosophila* (Tully *et al.*, 1994). Insofar as protein synthesis is essential for the formation of LTM, this result could reflect an optimal stimulus frequency for transcription of an essential gene. Optimal stimulus frequencies appear also to exist for some task learning by humans (Kientzle, 1946). Based on the result of Tully *et al.* (1994), Yin *et al.* (1995) hypothesized that the optimal stimulus frequency might predominantly phosphorylate a TF that activates transcription of a gene essential for LTM formation, with a countervailing transcriptional repressor being phosphorylated at stimulus frequencies above optimal.

To illustrate this phenomenon, we extended the model of Fig. 5(a) to incorporate dynamically varying modification of transcriptional activator and repressor efficacy by stimulus-dependent phosphorylation (Smolen *et al.*, 1998). Different kinetic pa-
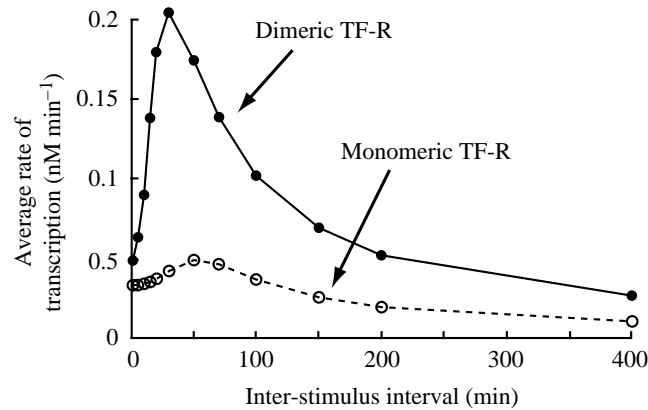
Figure 6. Non-monotonic dependence of transcription rate upon stimulus frequency with kinetic parameters consistent with the hypothesis of Yin *et al*. (1995). Top curve, a maximum in transcription rate occurs when dimeric TF-R is assumed to repress transcription. Bottom curve, if only monomeric TF-R represses, the maximum is almost abolished. Modified from Smolen *et al*. (1998).

rameters of phosphorylation and dephosphorylation for the two TFs were assumed to allow for different sensitivities to stimuli, such that stimuli frequent enough to saturate the phosphorylation of the activator, TF-A, may not be frequent enough to saturate the repressor, TF-R. An 'output' gene, whose transcription was regulated by both TFs, was added to the model. Transcription of this gene was assumed to be activated by binding of dimeric TF-A to a responsive element, and binding of TF-A was competitively inhibited by dimeric TF-R. As Fig. 6 demonstrates, this model predicts an optimal stimulus frequency for maximal transcription of the 'output' gene. The average rate of transcription during stimuli of a given frequency increases with the frequency, through a maximum, and then falls sharply as TF-R becomes phosphorylated by very frequent stimuli. This peak in transcription rate is almost abolished if monomeric, as opposed to dimeric, TF-R is assumed to be responsible for repression of the 'output' gene.

Similar dynamics were obtained with a model without TF-R, using phosphorylated TF-A to activate the 'output' gene, and including competing kinase and phosphatase activities with the kinase activated by lower stimulus frequencies than the phosphatase. Consideration of these results allows the inference that the existence of two competing processes, such as activator and repressor phosphorylation, that have different sensitivity to stimuli, could provide a quite general mechanism for tuning the response of a genetic system to an optimum stimulus frequency.

3.6. *Stochastic fluctuations in the numbers of macromolecules can introduce significant variability into the dynamics of genetic regulation.* Stochastic fluctuations in the numbers of macromolecules due to the random timing of individual synthesis, degradation, and transport events are expected to be important when, as

is often the case in genetic systems, small (tens to hundreds) average copy numbers of important molecules are present (Keizer, 1987). Stochasticity in chemical reactions, including macromolecular synthesis and degradation, can be simulated exactly using the Monte Carlo algorithm of Gillespie (1977). Denote the average time interval between occurrences of a particular biochemical reaction by T_{avg} (T_{avg} can be determined from deterministic kinetic analysis, neglecting stochasticity). If a particular biochemical reaction occurs at $t = 0$, the probability P(t) that the *next* reaction of that type will occur within a specific short time interval Δt centered at a later time t is (Gillespie, 1977)

$$P(t) = \frac{\Delta t}{T_{avg}} \exp\left(\frac{-t}{T_{avg}}\right). \tag{6}$$

Recently, McAdams and Arkin (1997) applied this algorithm in simulations that suggest different phenotypes of prokaryotic systems could be selected by stochastic switching between alternative dynamic states. Figure 7(a) reproduces the kinetic scheme used for the majority of their simulations. Here, protein P_1 dimerizes and the dimer represses its own transcription. This is the only feedback loop in the circuit, although the P_1 homodimer is also assumed to respectively activate and repress the transcription of proteins P_2 and P_3. With physiologically plausible parameters characteristic of a prokaryotic genetic system, Fig. 7(b) demonstrates that the amount of time taken for the P_1 gene product to accumulate to a 'threshold' level, here chosen arbitrarily as either 25 or 50 molecules of homodimer, is subject to great stochastic variability. For example, assuming a gene dosage of 1 and a threshold level of 25 homodimers, 100 simulations were performed that were identical except for different initializations of the random number generator controlling stochasticity. From Fig. 7(b), the time required for these simulations to reach threshold varied from less than 10 min to more than 45 min. Thus, if P_1 also regulated other genes and if its accumulation was triggered by an environmental stimulus of a given duration, only in some cells might the stimulus suffice to raise the P_1 concentration above a level necessary to activate transcription of a downstream gene and to yield an observable phenotypic effect.

Essentially the same point was made earlier by Ko (1991) who used a much simpler model to argue that stochastic effects yield an irreducible variability in phenotype among cells. More specifically, a given environmental stimulus can yield a transient response, such as a growth spurt, which is quite variable because of fluctuations at the genetic level. To support this contention, Ko (1991) simulated the variability of the amount of transcription of an inducible gene over a fixed time interval. Several series of simulations examined the variability as a function of the ratio of dissociation and association rate constants for binding of an obligatory transcriptional activator to the promoter of the target gene. This ratio was taken as being equivalent to the ratio of the probabilities of dissociation and association per simulation timestep. In this simple model, these probabilities were the only parameters. Great variability in the amount of transcription was seen if association and
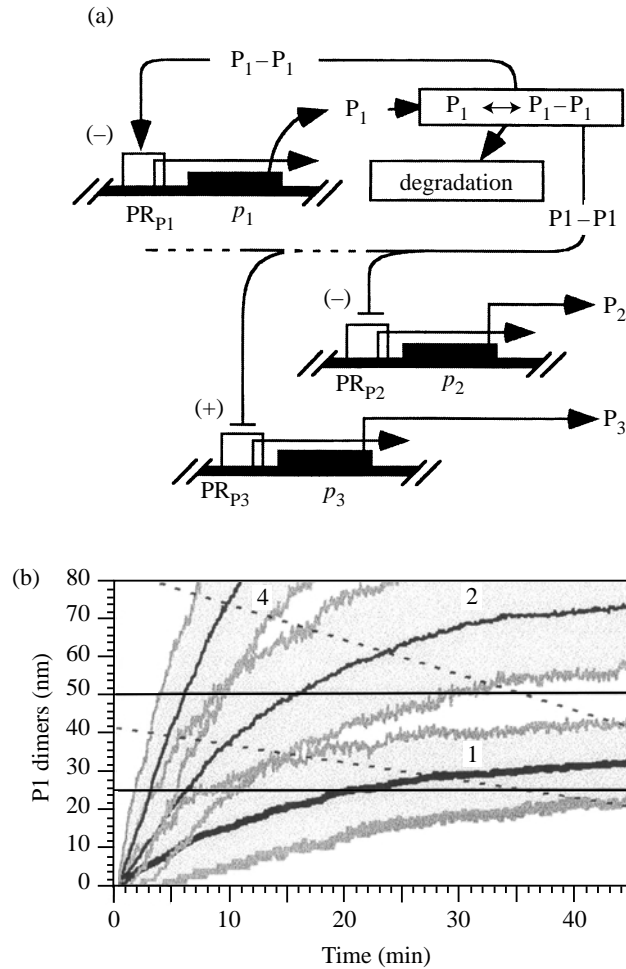
Figure 7. (a) The kinetic scheme used for the majority of simulations presented in McAdams and Arkin (1997). (b) Mean and $\pm 1\sigma$ (± 1 standard deviation, gray regions around black mean time courses) results for 100 stochastic simulations of the onset of P_1 dimer production assuming the kinetic scheme of (a), at P_1 gene dosages of 1, 2, and 4. P_1 concentration has units of nM. The 'σ' values plotted are the 16th and 84th concentration percentiles at each time point. Great variability in the amount of time required for individual simulations to reach 'threshold' protein concentrations of 25 or 50 molecules (horizontal lines) is evident. For parameter values see McAdams and Arkin (1997). Reproduced with permission from McAdams and Arkin (1997).

dissociation probabilities were such that the half-life of a bound factor-promoter complex was considered to be relatively long (i.e., if the half-life was comparable to the simulation time, which can be thought of as the time for which a stimulus is applied). On the other hand, little variability was seen if the half-life is relatively short so that many individual binding events occurred during the simulated time period, with their effects averaging out.

Such variability might be expected to have observable consequences in multi-cellular organisms, due to stochastic fluctuations in gene expression at the earliest stage of development when the organism consists of only one or a few cells. As a possible empirical example, systematic efforts to reduce variability in inbred mice have uncovered an apparently irreducible minimum variability, not attributable to environmental difference, which has been suggested to be due to stochasticity in the timing of expression events occurring at or before fertilization (Gartner, 1990).

Arkin *et al*. (1998) have recently included stochasticity of macromolecular synthesis and degradation in a detailed model for the *E. coli*–λ phage genetic switch. A schematic diagram of this switch was presented in Fig. 2. Two regulatory proteins (Cro and Cl in Fig. 2) acting at low concentrations competitively control the switch point (G1 in Fig. 2) that determines whether an infected *E. coli* cell follows a lyso-genic vs a lytic path. Fluctuations in the numbers of protein molecules randomize the choice of path for individual cells. A comprehensive model was constructed, using a large amount of extant experimental data to estimate all parameter values. Simulations of the lysogenic vs lytic decision-making process were compared with experimental data for the fraction of cells choosing the lytic path as a function of a parameter which could also be adjusted in the model—the ratio of the phage vs cell concentrations. Good agreement between simulations and experiment was found.

Previously, genetic switches have been described that are based on competition between transcription factors for binding to specific promoters [e.g., Hicks and Grossman (1996)] or on random inversion of DNA segments [e.g., Robert-son (1992); Dorman (1995)]. These are all bistable switches that can commit the cell to one of two quite different phenotypes. Although detailed models of these switches have not yet been developed, these types of switches are also expected to exhibit stochasticity in their outcomes, and to contribute to random variability in phenotypes. Ko (1991) found that variability is maximized if the individual stochastic biochemical events (binding and unbinding of TF molecules) contribut-ing to the observable outcome are few in number. Therefore, a conjecture can be made. Those genetic switches whose contribution to phenotype depends on only one or a few elementary random events (such as inversion of a DNA segment) should contribute more to variability than do switches whose contribution depends on many individual random events whose effects would tend to average (such as rapid bindings and unbindings of transcription factors). Arkin *et al*. (1998) dis-cussed possible mechanisms to ensure that, in the face of stochastic variation in macromolecular concentrations and in the timing of genetic events, cells still man-age to achieve the regulatory determinism necessary for reliably carrying out key

functions. These authors suggest two possibilities. First, the biochemical architecture of many pathways might be such that only the kinetics of one, or a few, rate-limiting steps determine the overall flux, in which case much stochastic variability, not directly associated with those steps, might be rendered irrelevant. Second, in many cases, only the proper sequencing of cellular events, and not their duration (within broad limits), may be important.

3.7. *Stochastic fluctuations also have the potential to destabilize and thereby mask some steady states of genetic regulatory systems.* We recently investigated, with a modified version of the model of Fig. 3(a) (Smolen *et al.*, 1999a), whether, with physiologically reasonable parameter values, stochastic fluctuations in molecule numbers might be expected to induce transitions between stable states of TF-A concentration. We also investigated whether fluctuations would tend to 'mask' particular steady states by making it very unlikely the system would be found at or near these states.

To construct a stochastic model, we considered explicitly the dynamics of both *tf-a* mRNA and TF-A protein. By rescaling parameters, the variables [*tf-a* mRNA] and [TF-A] were reinterpreted as molecule numbers rather than as concentrations. Deterministic rates for synthesis and degradation processes are given by

$$R_{\text{transcription}} = \frac{k_{1,f}(\text{TF-A}_{\text{dimer}})}{(\text{TF-A}_{\text{dimer}}) + K_d} + R_{\text{bas}} \tag{7}$$

$$R_{\text{degRNA}} = k_{1,d}[tf - a \text{ mRNA}] \tag{8}$$

$$R_{\text{translation}} = k_{2,f}\langle[tf - a \text{ mRNA}]\rangle \tag{9}$$

$$R_{\text{degP}} = k_{2,d}(\text{TF-A}_{\text{monomer}}). \tag{10}$$

The expression for $R_{\text{translation}}$ incorporates a distributed delay for macromolecular transport. An average of [*tf-a* mRNA] over an interval of a width of 20–30 min, centered at a delay on the order of 2 h, was used. In order to include fluctuations due to association and dissociation of TF-A monomers, we dropped the assumption that TF-A dimer concentration is proportional to [TF-A]^2, and deterministic rates for association and dissociation were assigned, giving equations for dimer formation and dissociation rates,

$$R_{\text{monomer–dimer}} = k_f(\text{TF-A}_{\text{monomer}})^2 \tag{11}$$

$$R_{\text{dimer–monomer}} = k_b(\text{TF-A}_{\text{dimer}}). \tag{12}$$

The reciprocals of equations (7)–(12) are the average time intervals between individual biochemical reactions of specific types [e.g., between individual dimer dissociations in the case of equation (12)]. At each timestep Δt of the simulation, a separate random number was chosen for each elementary process of synthesis

or degradation of mRNA or protein, and TF-A association and dissociation. Each random number was drawn from a uniform distribution on {0,1}. For any elementary process, if the random number was less than the product of Δt and the average rate [equations (7)–(12)], we assumed the process occurred and altered the appropriate molecular population by 1 (except that association and dissociation events change the TF-A monomer population by 2). If the products of Δt and the average rates are kept small, then this simple algorithm constitutes an approximation of the Gillespie algorithm, equation (6).

Simultaneously, the analogous, deterministic model that neglected stochastic fluctuations was constructed. This was done by using equations (7)–(12) to formulate ordinary differential equations for the rates of change of [*tf-a* mRNA], [TF-A]$_{monomer}$, and [TF-A]$_{dimer}$. The same values of rate and dissociation constants were used in both the deterministic and stochastic models. Simulations with the deterministic model were compared with simulations with the stochastic model that included fluctuations.

The deterministic model was found to exhibit bistability for a significant range of parameter values. One stable solution has [*tf-a* mRNA] low and its synthesis rate close to the minimum synthesis rate R_{bas} [equation (3)], and the other has [*tf-a* mRNA] high and its synthesis rate close to the fully activated rate, $k_{1,f}$. Figure 8(a) illustrates an example in which the deterministic model is stable in the lower steady state until a temporary increase in R_{bas} switches it to a new stable state. In the corresponding simulation, the stochastic variant of the model was initialized near the lower steady state. Within ~ 40 h of simulated time, without any alteration in parameter values, random fluctuations accumulated and enabled a transition to the upper state [Fig. 8(b)]. This behavior occurred with physiologically reasonable parameter values, similar to those used by McAdams and Arkin (1997). For example, in the simulation of Fig. 8(b), a maximal mean transcription rate $k_{1,f}$ of 3.8 mRNA molecules min^{-1} was assumed, which is reasonable for a strongly activated promoter. A reverse transition, from the upper to the lower state, was not seen even after 400 hr of simulated time. Similar results were obtained with five different initializations of the random number generator used for calculating fluctuations. Thus, the lower stable state can be 'masked' by fluctuations, in that the system would not ordinarily be observed near it. To understand why the lower stable state is preferentially masked, one can consider that, since the lower stable state is characterized by a smaller number of macromolecules, the ratio of the fluctuation amplitude to the average molecule number is expected to be greater in the lower state (Keizer, 1987). Therefore, the lower state is more readily destabilized by fluctuations than is the upper stable state.

The results from the above studies suggest that models of gene networks should consider stochastic fluctuations to the extent practicable. Predictions of the existence or stability of steady states and periodic solutions that are made by considering purely deterministic models may not be relevant to systems in which some steady states, and presumably other types of dynamic solutions, could be destabi-
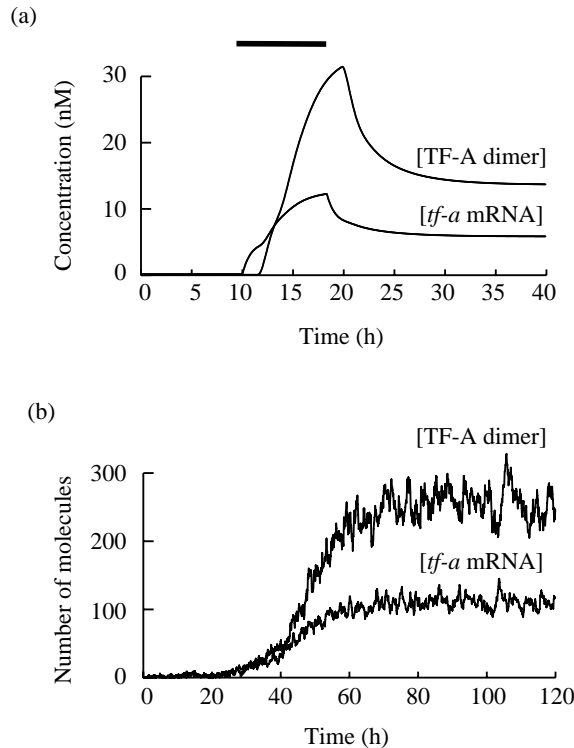
(a)

(b)

Figure 8. Stochastic fluctuations in molecule numbers can destabilize steady states of genetic regulatory systems. (a) Without fluctuations, the model of equations (7)–(12) is bistable. Initial levels of RNA and protein are low (<1 nM) and steady. At $t = 10$ h (arrow) the basal (unstimulated) rate of transcription is increased for 500 min. This causes a transition to a stable state of high RNA and protein levels. (b) The initial state of Fig. 8(a) is spontaneously destabilized when stochastic fluctuations are incorporated. Parameters are otherwise as in Fig. 8(a) except for a scaling factor to convert concentrations to molecule numbers. Modified from Smolen *et al.* (1999a).

lized and masked by fluctuations. Since most modeling of gene networks has been deterministic, the necessity for considering fluctuations will necessitate a paradigm shift of much investigation. However, deterministic modeling with ordinary differential equations or Boolean networks will remain important because of insufficient data to construct a stochastic model in most cases and because of limits on computing power. In particular, the temporal resolution of models constructed with the Boolean network approach is generally insufficient to allow inclusion of fluctuations. At least, however, if the data used to construct a Boolean model consists of large and reproducible responses to stimuli, then one can reasonably hope that, even with the neglect of fluctuations, the Boolean model might still reliably predict responses to novel stimuli of similar strength to the stimuli used in constructing the model. Similar considerations lead one to conclude that deterministic models based on ordinary differential equations will also have considerable predictive power.

We now return to a discussion of the point that detailed modeling of specific genetic systems is still at an early stage. This is due primarily to a lack of data. The necessary experimental techniques to simultaneously monitor the expression time courses of many genes in response to imposed perturbations have only recently been developed. However, a few systems, mostly prokaryotic, have been sufficiently well characterized to allow detailed modeling. Some of these are discussed in the following section.

4. EXPERIMENTAL DATA TO SUPPORT DETAILED MODELING OF SPECIFIC GENETIC SYSTEMS IS OFTEN LACKING

For eukaryotic regulation in particular, the biochemical data necessary for the determination of parameter values, and often even for determining the proper forms of equations, is usually not present. For prokaryotic regulatory systems, a significant number of specific models exist, because the greater simplicity and lack of differentiation of prokaryotic cells makes it somewhat easier to delineate macromolecular interactions and rates of processes. For example, the lytic vs lysogenic decision circuit for λ phage integrated in the *E. coli* chromosome, which is the most thoroughly studied genetic switch (Meyers and Friedland, 1984; Ptashne, 1992), has been modeled in detail as discussed above (Arkin *et al.*, 1998; McAdams and Arkin, 1998). Experimentally based estimates for all model parameter values have been developed. Thus, for this system, modeling can be said to have moved beyond 'elucidation of qualitative principles'. Other modeled prokaryotic or viral genetic regulatory systems include *lac* operon regulation (Wong *et al.*, 1997), the *E. coli* phosphate starvation response (Vandien and Keasling, 1998), regulation of HIV-1 gene expression (Hammond, 1993), and bacterial chemotaxis (Hauri and Ross, 1995; Spiro *et al.*, 1997). The above examples were all fitted to considerable sets of experimental data, thus, modeling of these systems has also progressed to a point where the models can provide somewhat detailed and quantitative descriptions.

One eukaryotic example is, perhaps, the recent model of the regulation by TFs of a sea urchin gene important for development, *endo16* (Yuh *et al.*, 1998). This model elucidates part of the complex regulation of the gene, but remains incomplete in ways that illustrate typical limitations of experimental knowledge. There are seven distinct regulatory modules, termed A–G, in the DNA sequence upstream from the gene and its basal promoter, and at least 15 TFs regulate transcription through these modules. Module A integrates the effects of all the other modules. Experimentally, Yuh *et al.* (1998) synthesized constructs in which reporter genes were attached to regulatory regions containing one or more of these modules. The constructs were then introduced into sea urchin eggs and expression of the reporter gene monitored during execution of the egg's normal developmental program. Comparison of the time courses of expression of different constructs allowed the assignment of roles to the modules (e.g., module B can 'take over'

module A so that module A acts only to amplify the time course of transcription created by module B alone). However, many aspects of *endo16* regulation remain to be experimentally characterized. For example, the roles of the majority of the TFs that affect *endo16* transcription appear not to have been established yet. Also, dose-response curves do not exist for the effects of TFs. The only data that exist are binary in nature (i.e., the data consist of expression time courses for reporter constructs in which a given module is present or absent). The binary nature of the data means that the model cannot predict the effects of stimuli that, for example, might simultaneously activate transcription factors specific for competing modules such as module F (which represses transcription) and module B (which enhances transcription). Finally, the data characterizing activation of transcription by modules D–F was obtained during conditions that may not reflect the *in vivo* situation. Nevertheless, despite the incompleteness of the current understanding of *endo16* regulation, it is evident that if this degree of regulatory complexity is common for single eukaryotic genes, gathering data sufficient to characterize the regulation of groups of genes and constructing models that provide a reasonable representation of this regulation will be a difficult challenge.

When characterizing a genetic regulatory system, whether prokaryotic or eukaryotic, some types of biochemical parameters important for model development are often more straightforward to estimate than others are. For example, binding coefficients between macromolecules can often be measured *in vitro* by a variety of methods [e.g., changes in fluorescence of labeled oligonucleotides upon binding of a transcription factor (Richards *et al.*, 1996)], although corrections may then have to be estimated for macromolecular crowding *in vivo* because such crowding tends in general to strongly promote macromolecular association and oligomerization (Zimmerman and Minton, 1993). More difficult to obtain are rate constants appropriate for *in vivo* reactions. For example, experimenters will often be satisfied to know that a given transcription factor is phosphorylated to an appreciable extent subsequent to an imposed stimulus. It is more laborious, and less common, to obtain sufficient data to construct a moderately accurate time course for phosphorylation and subsequent dephosphorylation. Estimation of rate constants, however, requires such data. Time courses of total protein amount, or of total mRNA amount, are yet more difficult to determine. As a result, there are, for example, few instances in which the fraction of transcription factor phosphorylated by a stimulus of given strength is determined.

Also important is experimental evidence for the rate and mechanism of intracellular macromolecular transport. Such data is not often available. One of the few studies that begins to quantify the relative contributions of active vs diffusive transport is that of Femino *et al.* (1998) who visualized the formation and movement of single β-actin mRNA transcripts in cultured fibroblasts by fluorescent *in situ* hybridization. Here, transcripts often left the transcription site along specific tracks. This finding suggests an active transport mechanism in which motor proteins direct mRNAs along cytoskeletal elements. However in contrast, in at least half of the

cases, the transcripts appeared to simply diffuse away from their site of formation, suggesting that a combination of passive and active transport was used.

In the final section, we point out issues which, in the near term, could serve as foci for further investigation, by cooperation of simulation and experiment.

5.   SPECIFIC ISSUES WHERE FURTHER INVESTIGATION IS NEEDED

Modeling of stochasticity and of alternative biochemical architectures is far from complete. One issue of interest is whether stochastic fluctuations in molecule numbers tend to be significantly smaller, or to have a different frequency spectrum, when diffusive transport dominates over active transport or *vice versa*. It is likely that if stochastic simulations are to predict the degrees of variability in the behavior of genetic systems, stochasticity in transport will have to be included. A second issue is that the dynamic possibilities associated with many biochemical architectures observed in genetic regulatory systems have not yet been assessed by modeling. Examples of such architectures are convergence of different signaling pathways through distinct transcription factors onto a particular gene (Howard and Maurer, 1995) heterodimerization of transcription factors in two separate pathways with a common third transcription factor (Hunter *et al*., 1996), and conditional regulation by a single TF [e.g., the TF YY1 mediating both enhancement of basal transcription and suppression of cAMP-induced transcription of the dopamine β-hydroxylase gene (Seo *et al*., 1996)]. Such architectures may be expected to provide alternative mechanisms for generating dynamic phenomena such as multistability and oscillations.

Aspects of genetic organization and transcriptional regulation that are unique to eukaryotes have not been discussed at length in this review, because few modeling studies have focused on them. In eukaryotes, TFs must gain access to DNA in chromatin. DNA must dissociate, to some extent, from nucleosomes to allow TFs and RNA polymerase to function. One recent study (Polach and Widom, 1996) models cooperative binding of TFs to nucleosomal target sites. The cooperativity is not due to protein–protein interactions, but rather, the first TF to bind partially dissociates DNA from the nucleosome, making it easier for subsequent TFs to bind. The authors also considered experiments (Adams and Workman, 1995) in which nucleosomes were constructed containing binding sites for unrelated TFs and cooperativity was observed in the binding of TF pairs. The model succeeded in quantitatively reproducing the experimental data. Polach and Widom (1996) pointed out that their model suggests a new mechanism of transcriptional regulation. The binding of a particular TF to DNA could be varied by changing the concentration or activity of another TF that binds to a site within the same nucleosome. Transcriptional regulation could also involve modification of the binding of DNA to nucleosomes by acetylation of histones. Regions of chromatin can alter

their packing structure to allow easier access of proteins to DNA. To develop models of specific eukaryotic genetic regulatory systems, we expect that these features will often need to be considered.

Not all regulated genes can be expected to exhibit the behaviors illustrated by the models discussed in this review. However, the diversity of transcription factors and their interactions suggests behaviors such as multistability and periodicity will be identified. Possible physiological consequences of multistability in genetic regulatory systems have been suggested. Keller (1994) suggested that the choice among multiple steady could be sufficient to distinguish different cell types within a multicellular organism. MacLeod (1996) has recently proposed that epigenetic, heritable changes in gene expression following exposure to chemicals might play a role in carcinogenesis. Such changes would correspond dynamically to switching of genetic regulatory systems from one steady state to another by brief stimuli. Thomas *et al*. (1995) noted that cell differentiation is essentially epigenitic in nature (i.e., differentiation may be the biological modality of multistability). As discussed above, a positive-feedback loop is necessary for multistability. Therefore, Thomas *et al*. (1995) suggested that the study of genes which exert a positive control (direct or indirect) on their own expression could help identify key genes involved in differentiation.

5.1. *Modeling and experimentation are needed to understand a key genetic regulatory system important for learning and memory.* A specific genetic network, based on CREB and related transcription factors, has been repeatedly implicated in the formation of LTM (Tully *et al*., 1994; Dash and Moore, 1996; Yin and Tully, 1996; Guzowski and McGaugh, 1997; Lamprecht *et al*., 1997; Bartsch *et al*., 1998; Glazewski *et al*., 1999). However, no attempt has yet been made to integrate experimental data concerning the kinetics of activation of these transcription factors into a model. CREB can bind to Ca^{2+}/cAMP-responsive elements (CREs) to induce the transcription of genes crucial for neuronal plasticity (Bartsch *et al*., 1995; O'Leary *et al*., 1995; Yin and Tully, 1996; Femino *et al*., 1998). Proteins related to CREB, such as CREB2, are transcriptional repressors that bind to the same CRE sequence as CREB (Sassone-Corsi, 1995). They are generally phosphorylated by the same signals that phosphorylate CREB. The functional relevance of repressor phosphorylation has not yet been established. However, it has recently become apparent that essential elements of the model of Fig. 5(a)—positive and negative feedback via competition by a transcriptional activator and a repressor for determining the expression of target genes—are present. Positive feedback exists via binding CREB to CREs affecting its own transcription. Negative feedback exists in the form of a repressor protein, inducible Ca^{2+}/cAMP-responsive early repressor (ICER), whose transcription is induced by CREB. These feedback loops could create multistability or periodic behavior. There is some empirical indication for complex dynamics mediated by this architecture. Oscillations in CREB mRNA have been reported in mammalian secretory cells and the above feedback loops

have been proposed as essential components of the oscillatory mechanism (Walker *et al.*, 1995).

The hypothesis of Yin *et al.* (1995) for generation of maximal transcription at a particular stimulus frequency could be realized in genetic networks regulated by the CREB family of proteins if a repressor, such as CREB2, was phosphorylated at a somewhat higher frequency than an activator, such as CREB. However, recent data, obtained in *Aplysia*, actually suggests that phosphorylation of the repressor CREB2 *reduces* its repressing activity (Martin *et al.*, 1997). An optimal stimulus frequency for transcription in such a gene network might still, however, be explainable in terms of two competing processes that have different sensitivity to stimuli and opposing effects on transcription. For example, CREB has both activating and inhibiting phosphorylation sites (Ser 133 and Ser 142, respectively), and if the inhibiting site only became significantly phosphorylated by very frequent stimuli, then an optimal stimulus frequency could result.

5.2. ***A variety of methods can be applied to examine, by experimentation in cooperation with modeling, whether the biochemical architecture and parameters of particular genetic systems*** **in vivo** ***permit specific dynamic behaviors.*** These methods include obtaining the time course of transcription of transfected reporter gene constructs, which can yield estimates of important kinetic parameters governing transcription onset and termination and mRNA degradation (Castano *et al.*, 1996). Obtaining simultaneous expression time courses for up to several thousand genes subsequent to a stimulus, or during organism development, is feasible with the use of 'DNA chips' [i.e., high-density arrays of oligonucleotides which each hybridize to a specific mRNA sequence (Winzeler *et al.*, 1998)]. Systems of cluster analysis are being developed that can be applied to expression data obtained with DNA chips (Eisen *et al.*, 1998). These systems use standard statistical methods to classify genes according to similarity in expression profiles, a necessary aid to identifying genes that function together. Also, cluster analysis identifies novel genes with functions similar to known genes. Iyer *et al.* (1999) used a DNA chip to determine expression time courses for 8600 human genes following the application of fetal bovine serum to cultured fibroblasts. Cluster analysis divided these genes into groups. The first, rapidly induced group contained many known or suspected transcription factors. Other TFs had their levels repressed. Throughout the response, a recurring theme was coordinated regulation of groups of genes whose products act to modulate a common process, such as synthesis of a macromolecule. Over 200 previously unknown genes were discovered as a result of serum altering their baseline expression. As discussed earlier, methods for efficiently using such large sets of expression time courses to construct models of genetic networks with the Boolean method have recently begun to be developed (Somogyi and Sniegoski, 1997; Wen *et al.*, 1998). Very recently, principal component analysis of gene expression time series, which were obtained from DNA chip hybridization, has also been used to identify patterns of expression alteration be-

tween tamoxifen-sensitive and tamoxifen-resistant human breast cancer cells, and between estrogen-stimulated and unstimulated cells (Hilsenbeck *et al.*, 1999).

Another technique that yields similar data is PCR amplification and quantitation of specific mRNAs from tissue samples (Wen *et al.*, 1998). From a series of time points for specific mRNA amounts, one can back-calculate to estimate the time courses of expression of the corresponding genes in the tissue. By PCR amplification of tissue samples followed by statistical cluster analysis of the expression profiles, Wen *et al.* (1998) detected distinct waves of synchronized expression during development of cervical spinal cord tissue. These authors were able to classify a large number of genes into distinct clusters of highly correlated expression. Such experimental techniques would also be well suited for detecting long-lasting state transitions in response to brief perturbations. Such state transitions would suggest underlying multistability, as was discussed earlier for the simple model of Fig. 3(a). Furthermore, such state transitions would allow a prediction that a positive-feedback loop was present in the genetic network being studied, since as discussed earlier, positive feedback is a general prerequisite for multistability (Thomas *et al.*, 1995).

A caveat that limits the amount of information given by application of either DNA chips or PCR amplification is that since these techniques monitor populations of cells rather than single cells, their ability to delineate intracellular signaling pathways will be limited. Details of single-cell dynamics may be obscured because observed responses are averaged over many cells. If most or all cells are of the same type, then this lack of resolution can be partially overcome if all cells are exposed simultaneously to an applied perturbation for which, at least qualitatively, the response should be similar throughout the population. If the mRNA is unavoidably a mixture from cells of different types (e.g., neurons and glia in a tissue sample) this further complicates the analysis. However, time courses of 'marker' genes specific to one cell type can be associated, via cluster analysis, to other time courses of similar shape. If the expression time courses are obtained during a process such as development or differentiation, these associated genes are likely to share common regulatory inputs with the 'marker' genes and to be expressed in the same cell type. The results of Wen *et al.* (1998) illustrate that much useful information can be obtained by considering associations of marker and other genes. The authors identified a high degree of order among the expression of 112 selected genes. Five distinct waves of expression were seen during development. Wave 1 contained indicators for dividing neuroglial progenitor cells, wave 2 was indicative of neurogenesis as evidenced by the appearance of specific neuronal markers, and wave 3 was composed almost exclusively of neurotransmitter signaling genes and neuronal markers. In several cases, previously unsuspected and tight associations were found between the expression of genes with known function and genes with unknown function. Such association suggests common regulatory inputs to the genes, and may suggest functional relationships. Other qualitative observations included pronounced asymmetries of the structure revealed by principal compo-

nent analysis of the gene expression time series. Such asymmetry emphasizes that gene expression was overall a strongly constrained and ordered process. Wen *et al.* (1998) compared the PCR methodology to that based on DNA. They concede that more tissue genes could be assayed at once by DNA chip technology. However, the authors contend that with the use of robotics and capillary electrophoresis PCR techniques could be scaled up to cover as many genes as DNA chips. However, for most investigators, this would prove to be a tedious undertaking. It would prove more feasible to obtain DNA chips from vendors such as Affymetrix or Incyte (Gerhold *et al.*, 1999). There is a trade-off, in that DNA chip technology may be easier to implement, but may not yield as much dynamic range as PCR techniques. The latter can provide a log-linear assay over as much as 7 orders of magnitude of amount of starting mRNA (Wen *et al.*, 1998).

Given a stable baseline for the expression levels within a specific gene network in cultured cells, a mathematical approach discussed by Arkin *et al.* (1997) might be useful to infer the regulatory inter-connectivity of the system. To apply the method, one would begin by computing a correlation matrix between a series of expression time courses subsequent to distinct perturbations, such as brief applications of different hormones. The method was implemented by Arkin *et al.* (1997) to infer the connectivity of a well-studied biochemical network with regulatory interactions—the glycolytic pathway. The inferred connectivity was the same as that known to exist, although to deduce this connectivity some heuristic biochemical reasoning had to be used in combination with the mathematical analysis. The authors mention the possible applicability to genetic networks. However, their biochemical implementation was within a continuously stirred *in vitro* reaction vessel. It remains to be seen whether inputs and outputs for a genetic regulatory system operating in the context of intact cells (for which there is generally no non-cellular analog) could be sufficiently well defined for their method to be useful. Also, the method requires a baseline, or unperturbed state, that does not itself depend on time. It is not clear whether the baseline of a genetic system operating in intact cells would be sufficiently stable for the long time (order of days) that would be required for imposition of multiple perturbations onto a genetic system and relaxation of the system to baseline after each perturbation. Some, perhaps a large percentage, of perturbations might result in permanent changes in gene expression patterns.

Fluorescent *in situ* hybridization of mRNA (Femino *et al.*, 1998) could profitably be applied to visualize the mRNAs of possibly autoregulatory transcription factors, such as Jun or Fos. The transcription rate could be quantitated, and after a sustained change in a stimulus such as the level of a hormone, any distinct 'steps' in transcription rate during the approach to a new steady state could be visualized. Also, the technique could reveal any repeated perturbations in transcription rate in response to a single brief stimulus. If mRNA was actively transported to its site of translation, and if newly synthesized transcription factor was actively transported to the vicinity of its own gene, simulations discussed above (Smolen *et al.*, 1999a) suggest that these phenomena might be observed. Such observations would con-

stitute significant evidence that active transport was important in the system under study.

As experiments such as those discussed above become less labor-intensive with advances in technology, it may be anticipated that construction of quantitative models of gene networks important for controlling key biological processes such as differentiation, memory, healing of injuries, or responses to drugs, will occur more routinely. Mechanistic hypotheses specific to particular genetic systems will be readily formulated and tested within the context of such models. This methodology, when applied in conjunction with experiments, promises to help predict and analyse the responses of tissues and organisms to physiological stimuli as well as to biologically active environmental contaminants or novel pharmaceutical agents.

REFERENCES

Abel, T., K. C. Martin, D. Bartsch and E. R. Kandel (1998). Memory suppressor genes: inhibitory constraints on the storage of long-term memory. *Science* **279**, 338–341.

Adams, C. C. and J. L. Workman (1995). Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Mol. Cell. Biol.* **15**, 1405–1421.

Angel, P., K. Hattori, T. Smeal and M. Karin (1988). The *jun* proto-oncogene is positively autoregulated by its product, Jun–AP-1. *Cell* **55**, 875–885.

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

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

Bagley, R. J. and L. Glass (1996). Counting and classifying attractors in high dimensional dynamical systems. *J. Theor. Biol.* **183**, 269–284.

Banks, H. and J. M. Mahaffy (1978). Stability of cyclic gene models for systems involving expression. *J. Theor. Biol.* **74**, 323–334.

Bartsch, D., A. Casadio, K. A. Karl, P. Serodio and E. R. Kandel (1998). CREB1 encodes a nuclear activator, a repressor, and a cytoplasmic modulator that form a regulatory unit critical for long-term facilitation. *Cell* **95**, 211–223.

Bartsch, D., M. Ghirardi, P. Skehel, K. A. Karl, S. Herder, M. Chen, C. Bailey and E. R. Kandel (1995). *Aplysia* CREB2 represses long-term facilitation: relief of repression converts a transient facilitation into long-term functional and structural change. *Cell* **83**, 979–992.

Bazhan, S. I., V. A. Likhosvai and O. E. Belova (1995). Theoretical analysis of the regulation of interferon expression during priming and blocking. *J. Theor. Biol.* **175**, 149–160.

Blattner, F. R. *et al*. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1462.

Bliss, R. D., P R. Painter and A. G. Marr (1982). Role of feedback inhibition in stabilizing the classical operon. *J. Theor. Biol.* **97**, 177–93.

Blumenfeld, H., L. Zablow and B. Sabatini (1992). Evaluation of cellular mechanisms for modulation of Ca^{2+} transients using a mathematical model of fura-2 Ca^{2+} imaging in *Aplysia* sensory neurons. *Biophys. J.* **63**, 1146–1164.

Boden, J. (1997). Programming the *Drosophila* embryo. *J. Theor. Biol.* **188**, 391–445.

Busenberg, S. and J. M. Mahaffy (1985). Interaction of spatial diffusion and delays in models of genetic control by repression. *J. Math. Biol.* **22**, 313–333.

Byrne, J. H. *et al*. (1991). Neural and molecular bases of nonassociative and associative learning in *Aplysia*. *Ann. New York Acad. Sci.* **627**, 124–149.

The *C. elegans* Sequencing Consortium (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018.

Carrier, T. A. and J. D. Keasling (1997). Mechanistic modeling of prokaryotic mRNA decay. *J. Theor. Biol.* **189**, 195–209.

Carrion, A. M., W. A. Link, F. Ledo, B. Mellstrom and J. R. Naranjo (1999). DREAM is a Ca^{2+}-regulated transcriptional repressor. *Nature* **398**, 80–84.

Castano, J., R. Kineman and L. S. Frawley (1996). Dynamic monitoring and quantification of gene expression in single, living cells: a molecular basis for secretory cell heterogeneity. *Mol. Endocrinol.* **10**, 599–605.

Cole, S. T. *et al*. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544.

Collado-Vides, J., B. Magasanik and J. D. Gralla (1991). Control site location and transcriptional regulation in *Escherichia coli*. *Microbiol. Rev.* **55**, 371–394.

Collins, F. S., A. Patrinos, E. Jordan, A. Chakravati, R. Gesteland and L. Walters (1998). New goals for the U.S. human genome project: 1998–2003. *Science* **282**, 682–689.

Dash, P. K. and A. N. Moore (1996). Characterization and phosphorylation of CREB-like proteins in *Aplysia* central nervous system. *Brain Res. Mol. Brain Res.* **39**, 43–51.

Doedel, E. (1981). AUTO: a program for the automatic bifurcation analysis of autonomous systems. *Congr. Num.* **30**, 265–284.

Donachie, W. D. (1993). The cell cycle of *Escherichia coli*. *Ann. Rev. Microbiol.* **47**, 199–230.

Dorman, C. J. (1995). DNA topology and the global control of bacterial gene expression: implications for the regulation of virulence gene expression. *Microbiology (Reading)* **141**, 1271–1280.

Edwards, D. R. (1994). Cell signaling and the control of gene transcription. *Trends Pharmacol. Sci.* **15**, 239–244.

Eisen, M. B., P. T. Spellman, P. O. Brown and D. Botstein (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868.

Femino, A. M., F. S. Fay, K. Fogarty and R. H. Singer (1998). Visualization of single RNA transcripts *in situ*. *Science* **280**, 585–590.

Gartner, K. (1990). A third component causing variability beside environment and geno-type. A reason for the limited success of a 30 year long effort to standardize laboratory animals? *Lab. Anim.* **24**, 71–77.

Gerhold, D., T. Rushmore and C. T. Caskey (1999). DNA chips: promising toys have become powerful tools. *Trends Biochem. Sci.* **24**, 168–173.

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **61**, 2340–2361.

Glass, L. and S. A. Kauffman (1973). The logical analysis of continuous, non-linear bio-chemical control networks. *J. Theor. Biol.* **39**, 103–129.

Glazewski, S., A. L. Barth, H. Wallace, M. McKenna, A. Silva and K. Fox (1999). Impaired experience-dependent plasticity in barrel cortex of mice lacking the α and δ isoforms of CREB. *Cereb. Cortex* **9**, 249–256.

Goldbeter, A. (1995). A model for circadian oscillations in the *Drosophila period* (PER) protein. *Proc. R. Soc. Lond. Ser. B* **261**, 319–324.

Goodwin, B. C. (1965). Oscillatory behavior of enzymatic control processes. *Adv. Enzyme Reg.* **3**, 425–439.

Griffith, J. S. (1968a). Mathematics of cellular control processes. I. Negative feedback to one gene. *J. Theor. Biol.* **20**, 202–208.

Griffith, J. S. (1968b). Mathematics of cellular control processes. II. Positive feedback to one gene. *J. Theor. Biol.* **20**, 209–216.

Guckenheimer, J. and P. Holmes (1983). *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, New York: Springer-Verlag.

Guzowski, J. and J. L. McGaugh (1997). Antisense oligodeoxynucleotide-mediated dis-ruption of hippocampal cAMP response element binding protein levels impairs consol-idation of memory for water maze training. *Proc. Natl. Acad. Sci. USA* **94**, 2693–2698.

Hammond, B. J. (1993). Quantitative study of the control of HIV-1 gene expression. *J. Theor. Biol.* **163**, 199–221.

Hauri, D. C. and J. Ross (1995). A model of excitation and adaptation in bacterial chemo-taxis. *Biophys. J.* **68**, 708–722.

Hevroni, D. *et al.* (1998). Hippocampal plasticity involves extensive gene induction and multiple cellular mechanisms. *J. Mol. Neurosci.* **10**, 75–98.

Hicks, K. A. and A. D. Grossman (1996). Altering the level and regulation of the major sigma subunit of RNA polymerase affects gene expression and development in *Bacillus subtilus. Mol. Microbiol.* **20**, 201–212.

Hilsenbeck, S. G., W. E. Friedrichs, R. Schiff, P. O'Connell, R. K. Hansen, C. K. Osborne and S. A. Fuqua (1999). Statistical analysis of array expression data as applied to the problem of tamoxifen resistance. *J. Natl. Cancer Inst.* **91**, 453–459.

Howard, P. and R. Maurer (1995). A composite Ets/Pit-1 binding site in the prolactin gene can mediate transcriptional responses to multiple signal transduction pathways. *J. Biol. Chem.* **270**, 20930–20936.

Hunter, J., A. Kassam, C. Winrow, R. Rachubinski and J. Capone (1996). Crosstalk be-tween the thyroid hormone and peroxisome proliferator-activated receptors in regulating peroxisome proliferator-responsive genes. *J. Mol. Cell. Endocr.* **116**, 213–221.

Ishiura, M., S. Kutsuna, S. Aoki, H. Iwasaki, C. Andersson, A. Tanabe, S. Golden, C. Johnson and T. Kondo (1998). Expression of a gene cluster *kaiABC* as a circadian feedback process in cyanobacteria. *Science* **281**, 1519–1523.

Itoh, K., B. Stevens, M. Schachner and R. D. Fields (1995). Regulated expression of the neural cell adhesion molecule L1 by specific patterns of neural impulses. *Science* **270**, 1369–1372.

Iyer, V. *et al*. (1999). The transcriptional program in the response of human fibroblasts to serum. *Science* **283**, 83–87.

Jacobs, C. and L. Shapiro (1998). Microbial asymmetric cell division: localization of cell fate determinants. *Curr. Opin. Gen. Dev.* **8**, 386–391.

Jacob, F. and J. Monod (1961). On the regulation of gene activity. *Cold Spring Harb. Symp. Quant. Biol.* **26**, 193–211, 389–401.

Karin, M. (1994). Signal transduction from the cell surface to the nucleus through the phosphorylation of transcription factors. *Curr. Opin. Cell Biol.* **6**, 415–424.

Keizer, J. (1987). *Statistical Thermodynamics of Equilibrium Processes*, New York: Springer-Verlag.

Keller, A. (1994). Specifying epigenetic states with autoregulatory transcription factors. *J. Theor. Biol.* **170**, 175–181.

Keller, A. (1995). Model genetic circuits encoding autoregulatory transcription factors. *J. Theor. Biol.* **172**, 169–185.

Kerszberg, M. (1996). Accurate reading of morphogen concentrations by nuclear receptors: a formal model of complex transduction pathways. *J. Theor. Biol.* **183**, 95–104.

Kerszberg, M. and J. Changeux (1994). A model for reading morphogenetic gradients: autocatalysis and competition at the gene level. *Proc. Natl. Acad. Sci. USA* **91**, 5823–5827.

Kientzle, M. J. (1946). Properties of learning curves under varied distributions of practice. *J. Exp. Psychol.* **36**, 187–211.

King, M. L. (1996). Molecular basis for cytoplasmic localization. *Dev. Genet.* **19**, 183–189.

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

Koh, B. T., R. B. Tan and M. G. Yap (1998). Genetically structured mathematical modeling of *trp* attenuator mechanism. *Biotech. Bioeng.* **58**, 502–509.

Kouzarides, T. and E. Ziff (1988). The role of the leucine zipper in the fos–jun interaction. *Nature* **336**, 646–651.

Lamprecht, R., S. Hazvi and Y. Dudai (1997). cAMP response element-binding protein in the amygdala is required for long- but not short-term conditioned taste aversion memory. *J. Neurosci.* **17**, 8443–8450.

Lee, C., K. Bae and I. Edery (1998). The *Drosophila* CLOCK protein undergoes daily rhythms in abundance, phosphorylation, and interactions with the PER-TIM complex. *Neuron* **21**, 857–867.

Leloup, J. C. and A. Goldbeter (1998). A model for circadian rhythms in *Drosophila* incorporating the formation of a complex between the PER and TIM proteins. *J. Biol. Rhythms* **13**, 70–87.

Liu, B. Z., J. H. Peng, Y. C. Sun and Y. W. Liu (1997). A comprehensive dynamical model of pulsatile secretion of the hypothalamo-pituitary-gonadal axis in man. *Comput. Biol. Med.* **27**, 507–513.

Luo, C., J. Loros and J. C. Dunlap (1998). Nuclear localization is required for function of the essential clock protein FRQ. *EMBO J.* **17**, 1228–1235.

MacDonald, N. (1989). *Biological Delay Systems: Linear Stability Theory*, Cambridge: Cambridge University Press.

MacLeod, M. (1996). A possible role in chemical carcinogenesis for epigenetic, heritable changes in gene expression. *Mol. Carcinog.* **15**, 241–250.

Mahaffy, J. M. (1984). Cellular control models with linked positive and negative feedback and delays. I: The models. *J. Theor. Biol.* **106**, 89–102.

Mahaffy, J. M., D. A. Jorgensen and R. L. van der Heyden (1992). Oscillations in a model of repression with external control. *J. Math. Biol.* **30**, 669–691.

Mahaffy, J. M. and C. V. Pao (1984). Models of genetic control by repression with time delays and spatial effects. *J. Math. Biol.* **20**, 39–57.

Martin, K., D. Michael, J. Rose, M. Barad, A. Casadio, H. Zhu and E. R. Kandel (1997). MAP kinase translocates into the nucleus of the presynaptic cell and is required for long-term facilitation in *Aplysia. Neuron* **18**, 899–912.

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

McAdams, H. and A. Arkin (1998). Simulation of prokaryotic genetic circuits. *Ann. Rev. Biophys. Biomed. Struct.* **27**, 199–224.

McAdams, H. and L. Shapiro (1995). Circuit simulation of genetic networks. *Science* **269**, 650–656.

Merrow, M. W., N. Garceau and J. C. Dunlap (1997). Dissection of a circadian oscillation into discrete domains. *Proc. Natl. Acad. Sci. USA* **94**, 3877–3882.

Mestl, T., C. Lemay and L. Glass (1996). Chaos in high-dimensional neural and gene networks. *Physica D* **98**, 33–52.

Mestl, T., E. Plahte and S. W. Omholt (1995a). A mathematical framework for describing and analyzing gene regulatory networks. *J. Theor. Biol.* **176**, 291–300.

Mestl, T., E. Plahte and S. W. Omholt (1995b). Periodic solutions in systems of piecewise linear differential equations. *Dyn. Stab. Systems* **10**, 179–183.

Meyer, T., G. Waeber, J. Lin, W. Beckmann and J. Habener (1993). The promoter of the gene encoding cAMP response element binding protein contains cAMP response elements: evidence for positive autoregulation of gene transcription. *Endocrinology* **132**, 770–780.

Meyers, S. and P. Friedland (1984). Knowledge-based simulation of genetic regulation in bacteriophage λ. *Nucleic Acids Res.* **12**, 1–9.

Molina, C., N. Foulkes, E. Lalli and P. Sassone-Corsi (1993). Inducibility and negative autoregulation of CREM: an alternative promoter directs the expression of ICER, and early response repressor. *Cell* **75**, 875–886.

Novak, B., A. Csikasz-Nagy, B. Gyorffy, K. Chen and J. J. Tyson (1998a). 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.

Novak, B., A. Csikasz-Nagy, B. Gyorffy, K. Chen and J. J. Tyson (1998b). Model scenarios for evolution of the eukaryotic cell cycle. *Philos. Trans. R. Soc. Lond. Ser. B* **353**, 2063–2076.

Okayama, H., A. Nagata, S. Jinno and H. Murakami (1996). Cell cycle control in fission yeast and mammals: identification of new regulatory mechanisms. *Adv. Cancer Res.* **69**, 17–62.

O'Leary, F., J. Byrne and L. Cleary (1995). Long-term structural remodeling in *Aplysia* sensory neurons requires *de novo* protein synthesis during a critical time period. *J. Neurosci.* **15**, 3519–3525.

Omholt, S. W., X. Kefang, O. Andersen and E. Plahte (1998). Description and analysis of switchlike regulatory networks exemplified by a model of cellular iron homeostasis. *J. Theor. Biol.* **195**, 339–350.

Plahte, E., T. Mestl and S. W. Omholt (1994). Global analysis of steady points for systems of differential equations with sigmoid interactions. *Dyn. Stab. Syst.* **9**, 275–291.

Plahte, E., T. Mestl and S. W. Omholt (1998). A methodological basis for description and analysis of systems with complex switch-like interactions. *J. Math. Biol.* **36**, 321–348.

Polach, K. J. and J. Widom (1996). A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. *J. Mol. Biol.* **258**, 800–812.

Ptashne, M. (1992). *A Genetic Switch: Phage λ and Higher Organisms*, Cambridge: Cell Press/Blackwell.

Reppert, S. M. (1998). A clockwork explosion! *Neuron* **21**, 1–4.

Richards, J., H. Bachinger, R. Goodman and R. Brennan (1996). Analysis of the structural properties of cAMP-responsive element binding protein (CREB) and phosphorylated CREB. *J. Biol. Chem.* **271**, 13716–13723.

Roberts, R. C., C. D. Mohr and L. Shapiro (1996). Developmental programs in bacteria. *Curr. Top. Dev. Biol.* **34**, 207–257.

Robertson, B. D. (1992). Genetic variation in pathogenic bacteria. *Trends Genet.* **8**, 422–427.

Rosen, R. (1968). Recent developments in the theory of control and regulation of cellular processes, in *International Review of Cytology*, G. H. Bourne (Ed.), New York: Academic Press.

Rossant, J. and N. Hopkins (1992). Of fin and fur: mutational analysis of vertebrate embryonic development. *Genes Dev.* **6**, 1–13.

Sabry, J., T. O'Connor and M. W. Kirschner (1995). Axonal transport of tubulin in Ti1 Pioneer neurons *in situ*. *Neuron* **14**, 1247–1256.

Sassone-Corsi, P. (1995). Transcription factors responsive to cAMP. *Ann. Rev. Cell Dev. Biol.* **11**, 355–377.

Sassone-Corsi, P., J. C. Sisson and I. M. Verma (1988). Transcriptional autoregulation of the proto-oncogene *fos*. *Nature* **334**, 314–319.

Scheper, T., D. Klinkenberg, C. Pennartz and J. van Pelt (1999). A mathematical model for the intracellular circadian rhythm generator. *J. Neurosci.* **19**, 40–47.

Schnitzer, M. J. and S. M. Block (1997). Kinesin hydrolyses one ATP per 8-nm step. *Nature* **388**, 386–389.

Seo, H., C. Yang, H. Kim and K. Kim (1996). Multiple protein factors interact with the *cis*-regulatory elements of the proximal promoter in a cell-specific manner and regulate transcription of the dopamine β-hydroxylase gene. *J. Neurosci.* **16**, 4102–4112.

Serfling, E. (1989). Autoregulation—a common property of eukaryotic transcription factors? *Trends Gen.* **5**, 131–133.

Shapiro, L. and R. Losick (1997). Protein localization and cell fate in bacteria. *Science* **276**, 712–718.

Sheng, H., R. D. Fields and P. Nelson (1993). Specific regulation of immediate-early genes by patterned neuronal activity. *J. Neurosci. Res.* **35**, 459–467.

Smith, H. (1987a). Oscillations and multiple steady states in a cyclic gene model with repression. *J. Math. Biol.* **25**, 169–190.

Smith, H. (1987b). Monotone semiflows generated by functional differential equations. *J. Diff. Eqs.* **66**, 420–442.

Smolen, P., D. A. Baxter and J. H. Byrne (1998). Frequency selectivity, multistability, and oscillations emerge from models of genetic regulatory systems. *Am. J. Physiol.* **274**, C531–C542.

Smolen, P., D. A. Baxter and J. H. Byrne (1999a). Effects of macromolecular transport and stochastic fluctuations on the dynamics of genetic regulatory systems. *Am. J. Physiol.,* **277**, C777–C790.

Smolen, P., D. A. Baxter and J. H. Byrne (1999b). Modeling clarifies the role of delays and feedback in circadian oscillators. *Soc. Neurosci. Abstr.,* **25**, 867.

Snoussi, E. H. and R. Thomas (1993). Logical identification of all steady states: The concept of feedback loop characteristic states. *Bull. Math. Biol.* **55**, 973–991.

Somogyi, R. and C. Sniegoski (1996). Modeling the complexity of genetic networks: understanding multigenic and pleiotropic regulation. *Complexity* **1**, 45–63.

Somogyi, R. and C. Sniegoski (1997). The gene expression matrix: towards the extraction of genetic network architectures, *Proceedings of the Second World Congress of Nonlinear Analysis*, Elsevier Science.

Spiro, P. A., J. Parkinson and H. Othmer (1997). A model of excitation and inhibition in bacterial chemotaxis. *Proc. Natl. Acad. Sci. USA* **94**, 7263–7268.

Stehle, J. H., N. Foulkes, C. Molina, V. Simonneaux, P. Pevet and P. Sassone-Corsi (1993). Adrenergic signals direct rhythmic expression of transcriptional repressor CREM in the pineal gland. *Nature* **365**, 314–320.

Thayer, M. J., S. J. Tapscott, R. L. Davis, W. E. Wright, A. B. Lassar and H. Weintraub (1989). Positive autoregulation of the myogenic determination gene MyoD1. *Cell* **58**, 241–248.

Thomas, R. (1994). The role of feedback circuits: positive feedback circuits are a necessary condition for positive real eigenvalues of the Jacobian matrix. *Ber. Bunsenges. Phys. Chem.* **98**, 1148–1151.

Thomas, R. and R. d'Ari (1990). *Biological Feedback*, Boca Raton, FL: CRC Press.

Thomas, R., D. Thieffry and M. Kauffman (1995). Dynamical behaviour of biological regulatory networks—I. Biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bull. Math. Biol.* **57**, 247–276.

Tomita, M. *et al.* (1999). E-CELL: a software environment for whole-cell simulation. *Bioinformatics* **15**, 72–84.

Tully, T., T. Preat, S. Boynton and M. Del Vecchio (1994). Genetic dissection of consolidated memory in *Drosophila melanogaster*. *Cell* **79**, 35–47.

Tyson, J. and H. G. Othmer (1978). The dynamics of feedback control circuits in biochemical pathways. *Prog. Theor. Biol.* **5**, 2–62.

Vandien, S. J. and J. D. Keasling (1998). A dynamic model of the *E. coli* phosphate-starvation response. *J. Theor. Biol.* **190**, 37–49.

Von Heijne, G., C. Blomberg and H. Liljenstrom (1987). Theoretical modeling of protein synthesis. *J. Theor. Biol.* **125**, 1–4.

Walker, W., L. Fucci and J. Habener (1995). Expression of the gene encoding transcription factor CREB: regulation by follicle-stimulating hormone-induced cAMP signaling in primary rat sertoli cells. *Endocrinology* **136**, 3534–3545.

Wen, X. L., S. Fuhrman, G. Michaels, D. Carr, S. Smith, J. Barker and R. Somogyi (1998). Large-scale temporal gene expression mapping of central nervous system development. *Proc. Natl. Acad. Sci. USA* **95**, 334–339.

Wiggins, S. (1990). *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, Heidelberg: Springer-Verlag.

Winzeler, E. A. *et al.* (1998). Direct allelic variation scanning of the entire yeast genome. *Science* **281**, 1194–1197.

Wolf, D. M. and F. H. Eeckman (1998). On the relationship between genomic regulatory element organization and gene regulatory dynamics. *J. Theor. Biol.* **195**, 167–186.

Wong, P., S. Gladney and J. D. Keasling (1997). A mathematical model of the *lac* operon-inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. *Biotech. Prog.* **13**, 132–143.

Yagil, G. and E. Yagil (1971). On the relation between effector concentration and the rate of induced enzyme synthesis. *Biophys. J.* **11**, 11–27.

Yin, J., M. Del Vecchio, H. Zhou and T. Tully (1995). CREB as a memory modulator: induced expression of a dCREB2 activator isoform enhances long-term memory in *Drosophila*. *Cell* **81**, 107–115.

Yin, J. and T. Tully (1996). CREB and the formation of long-term memory. *Curr. Opin. Neurobiol.* **6**, 264–267.

Yisraeli, J. K., S. Sokol and D. A. Melton (1990). A two-step model for the localization of maternal mRNA in *Xenopus* oocytes: involvement of microtubules and microfilaments in the translocation and anchoring of *V g*1 mRNA. *Development* **108**, 289–298.

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

Zimmerman, S. and A. P. Minton (1993). Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Ann. Rev. Biophys. Biomol. Struct.* **22**, 27–65.