Modeling the Dynamic Behavior of Biochemical Regulatory Networks

John J. Tyson , Teeraphan Laomettachit , Pavel Kraikivski

PII: S0022-5193(18)30587-3
DOI: https://doi.org/10.1016/j.jtbi.2018.11.034
Reference: YJTBI 9736

To appear in: *Journal of Theoretical Biology*

Received date: 30 July 2018
Revised date: 12 November 2018
Accepted date: 27 November 2018

Please cite this article as: John J. Tyson , Teeraphan Laomettachit , Pavel Kraikivski , Modeling the Dynamic Behavior of Biochemical Regulatory Networks, *Journal of Theoretical Biology* (2018), doi: https://doi.org/10.1016/j.jtbi.2018.11.034

Highlights:

- Mathematical models are vital tools for understanding molecular regulatory networks
- Logical models give qualitative insights but lack quantitative details
- Piecewise-linear ODE models can be improved by smoothing the step function
- A ‘standard component’ approach is both flexible and accurate
- Stochastic differential equations can capture mRNA noise accurately and efficiently

Modeling the Dynamic Behavior of Biochemical Regulatory Networks

John J. Tyson[1,2], Teeraphan Laomettachit[3] & Pavel Kraikivski[1,2]

[1]Department of Biological Sciences, Virginia Tech, Blacksburg VA 24061, USA

[2]Division of Systems Biology, Academy of Integrated Science, Virginia Tech, Blacksburg VA 24061, USA

[3]Bioinformatics and Systems Biology Program, King Mongkut's University of Technology Thonburi, Bang Khun Thian, Bangkok 10150, Thailand

12 November 2018

Prepared for the Rene Thomas Memorial Issue of the Journal of Theoretical Biology

Abstract:

Strategies for modeling the complex dynamical behavior of gene/protein regulatory networks have evolved over the last 50 years as both the knowledge of these molecular control systems and the power of computing resources have increased. Here, we review a number of common modeling approaches, including Boolean (logical) models, systems of piecewise-linear or fully non-linear ordinary differential equations, and stochastic models (including hybrid deterministic/stochastic approaches). We discuss the pro's and con's of each approach, to help novice modelers choose a modeling strategy suitable to their problem, based on the type and bounty of available experimental information. We illustrate different modeling strategies in terms of some abstract network motifs, and in the specific context of cell cycle regulation.

Introduction

The physiological characteristics of living cells are governed by underlying molecular networks of bewildering complexity, from the well-known metabolic pathways described in biochemistry textbooks to the gene/protein regulatory networks available from WikiPathways (e.g., https://www.wikipathways.org/index.php/Pathway:WP414) and other web resources. These networks are, in essence, large systems of coupled biochemical reactions operating far from equilibrium, and, as such, they are dynamical systems, i.e., they describe how the concentrations of chemical species or the activities of macromolecules (genes and proteins) are changing in time and space within the cell.

There is a long history of modeling (bio)chemical reaction networks as dynamical systems of interacting components, dating back at least to Alfred Lotka at the onset of the 20th century (Lotka, 1920). The dynamic modeling approach, however, really took flight in the 1960s and '70s, as new experimental facts drew attention to the complex dynamical properties of biochemical networks within living cells. Of special importance in this context were the recognition of (1) genetic switching in bacterial operons (Jacob and Monod, 1961; Novick and Weiner, 1957), (2) sustained oscillations in yeast glycolysis (Chance et al., 1964; Hess et al., 1966) and in cyclic AMP signaling (Durston, 1974; Gerisch and Hess, 1974), and (3) temporal oscillations and spatial patterns in chemical reaction systems (Belousov, 1959; Winfree, 1972; Zhabotinsky, 1964). As the implications of these observations spread through the worldwide community of biochemists and physical chemists during an historical meeting in Prague in 1968 (Chance et al., 1973), three theoretical approaches were quickly adopted to the challenges posed by spatiotemporal organization in chemical reaction systems. One approach—pioneered by Motoyosi Sugita (1963), Stuart Kauffman (1969) and Rene Thomas (1973)—modeled genetic control systems as networks of Boolean switches. A second approach—pursued by Joseph Higgins (1967), Brian Goodwin (1965; 1966), J.S. Griffith (1968a; 1968b), Albert Goldbeter with Rene Lefever and Lee Segel (1972; 1977), John Tyson and Hans Othmer (1978) and others—modeled metabolic and signaling networks by systems of nonlinear ordinary differential equations. The third approach—expounded by the Brussels school of physical chemists, Ilya Prigogine and Paul Glansdorff (1971) and Gregoire Nicolis (1971; Prigogine and Nicolis, 1971)—used ideas from the field of non-equilibrium thermodynamics to describe the conditions under which chemical reaction systems can exhibit 'dissipative structures' (temporal oscillations, spatial patterns, etc.). This review article focuses on the first and second approaches; the third approach is reviewed by Nicolis & Nicolis in a contribution to this issue.

Dynamical processes in biochemical reactions are governed, in general, by systems of coupled 'rate' equations describing the turnover of chemical species in time and their transport in space. In this review, we will restrict our attention to spatially homogeneous systems, for which the governing rate equations might be written as

$$C_i(t + \Delta t) = C_i(t) + \Delta t * \sum_{j=1}^{M} \nu_{ij} R_j(C_1(t), \ldots, C_N(t)), \quad i = 1, \ldots, N \qquad (1)$$

where $C_i(t)$ is the concentration (or 'activity') at time t of the i-th chemical species in the reaction network, $R_j(\ldots)$ is the rate law of the j-th chemical reaction in the network (a function of the

activities of some subset of chemical species), and v_{ij} is the stoichiometric coefficient of the *i*-th chemical species in the *j*-th biochemical reaction ($v_{ij} < 0$ for reactants, $v_{ij} > 0$ for products, $v_{ij} = 0$ for enzymes and 'modulators' of a reaction, and $v_{ij} = 0$ for all other species that do not participate in the *j*-th reaction in any way).

There are many different ways to interpret Eq. (1). The time variable can be continuous ($0 \le t < \infty$) or discrete ($t = 0, 1, 2, 3, \ldots$); the concentration variables can be continuous ($0 \le C_i(t) \le C_{i,\max}$) or discrete (say, $C_i(t) = 0$ or 1); the rate laws can be continuous or discrete functions of their arguments; and the model itself may be deterministic or stochastic. **Table 1** catalogs some of these possibilities. In this review, we will discuss the properties—including the 'pros' and 'cons'—of modeling biochemical reaction networks by these different approaches. We use a few simple 'network motifs' and 'reaction mechanisms' to illustrate the general principles of each approach.

Network Motifs and Reaction Mechanisms

By a network motif, we have in mind the type of 'influence' diagram that one typically finds in an experimental publication (see **Fig. 1A**). In a network motif, the icons represent chemical species (genes, RNAs, proteins, metabolites), and the connectors represent the influences (barbed connector = activation, blunt connector = inhibition) of one species on another. These diagrams summarize the most basic type of experimental observations made by biochemists and molecular biologists: if one pushes up/down on species X and observes that species Y goes up/down, then one concludes that X has a positive influence ('activation') on Y, whereas if Y goes down/up, then X has a negative influence ('inhibition') on Y. A few simple motifs with two or three components, as illustrated in Fig. 1A, will provide examples for this review. For a more detailed review of network motifs in biochemical systems, see Tyson & Novak (2010), or the book-length treatment of the subject by Uri Alon (2007).

With additional experimental evidence, the positive and negative effects cataloged on influence diagrams can be assigned to specific types of biochemical reactions. For example, (1) a metabolite X may be an allosteric modifier (an activator or inhibitor) of an enzyme E that synthesizes Y from a precursor biochemical. (2) The expression of gene X may induce or repress the expression of gene Y. (3) The transcription of an mRNA molecule Y may be repressed by the expression of its anti-sense counterpart Y_{AS}. (4) A transcription factor X may activate or inhibit the expression of the gene encoding protein Y. (5) A protein kinase X may phosphorylate a protein Y and thereby change the activity of Y. (6) X may be a protease that degrades Y. Because of this variety of possible reaction mechanism for any sort of positive or negative 'influence', there are a variety of possible biochemical realizations of any particular network motif. For example, **Fig. 1B** shows four different ways to instantiate motif 1A(i) in biochemical reactions.

Boolean Network Models

One popular approach to model switching networks is to represent all gene (or protein) states by Boolean variables C_i. In this approach $C_i = 1$ if gene *i* is being expressed (or protein *i* is 'active'

or 'present') and 0 if gene *i* is silent (or protein *i* is 'inactive' or 'absent'). The update of gene (or protein) states happens at discrete time steps $t = \{0, 1, 2\ldots\}$ and depends on network interactions. For a gene (or protein) interacting with N other components, C_i is updated as

$$C_i(t+1) = B_i(C_1(t), C_2(t), \ldots, C_N(t)) \,, \tag{2}$$

where B_i is a Boolean function (i.e., B_i evaluates to 0 or 1) of its Boolean arguments, C_1, C_2, …, C_N. The final state of the regulatory network depends, of course, on the choices of the Boolean functions in Eq. (2).

In **Table 2**, column A, we represent the double-negative feedback motif (i) in Fig. 1A in terms of the Boolean functions $B_1 = \neg C_2$, and $B_2 = \neg C_1$ (where ¬ is logical negation). In this case, the states (1,0) and (0,1) are 'alternative' steady states, but the states (0,0) and (1,1) flip back-and-forth from one time step to the next. The back-and-forth solution is an artefact of the symmetry of the chosen Boolean functions and our choice to update both variables at the same time ('synchronous' updating). This symmetry is usually broken by assuming that only one variable may change in any time step ('asynchronous' updating). The variable that changes may be chosen at random (column B in Table 2) or by giving priority to updating one or the other variable (which is equivalent to choosing different Boolean functions, as in Columns C and D of Table 2). In each of these cases, the states (1,0) and (0,1) are alternative steady states, but the probability that the system ends up in one or the other steady state (row 3 of Table 2) depends on the choices we make in terms of updating the motif.

The Boolean functions, $B_i(C_1, C_2, \ldots, C_N)$ in Eq. (2), can be evaluated by 'truth' tables (https://en.wikipedia.org/wiki/Truth_table), but they are often evaluated numerically by using a Heaviside function,

$$B_i = \mathrm{Heav}\big(W_i(C_1, C_2, \ldots, C_N)\big) = \begin{cases} 0, \text{ if } W_i \le 0 \\ 1, \text{ if } W_i > 0 \end{cases} \tag{3A}$$

with a proper choice of coefficients in the function

$$W_i = \omega_{i0} + \textstyle\sum_{j=1}^{N} \omega_{ij} C_j \ (+ \text{ higher order terms, if necessary}). \tag{3B}$$

In this context, the coefficients in the function W_i have no particular significance or exceptionality; they are assigned values solely for the purpose of implementing a particular Boolean function. For example, to compute the Boolean functions $B_1 = \neg C_2$, and $B_2 = \neg C_1$ in column A of Table 2, we might choose

$$B_1 = \mathrm{Heav}(\omega_{10} - C_2),\ B_2 = \mathrm{Heav}(\omega_{20} - C_1), \tag{4A}$$

where ω_{10} and ω_{20} may be assigned any values in the interval (0,1). For case B in Table 2, we can write

$$B_1 = \mathrm{Heav}(W),\ B_2 = \mathrm{Heav}(-W);\ W = -r + C_1 - C_2 + pC_1C_2 + q(1-C_1)(1-C_2). \tag{4B}$$

In this case, p and q are fixed transition probabilities and r is a random number, drawn uniformly from [0,1) at each updating step. Often, in the absence of contravening evidence, people choose

the symmetric case $p = q = 0.5$. For cases C and D in Table 2, we can compute the Boolean functions as

$$B_1 = \text{Heav}(-\omega_{10} + \omega_{11}C_1 - C_2),\ B_2 = \text{Heav}(\omega_{20} + \omega_{22}C_2 - C_1),$$

with appropriate values for the coefficients; for example,

$\omega_{10} = 0.5,\ \omega_{11} = 2,\ \omega_{20} = 0.5,\ \omega_{22} = 0.1$, for Case C; (4C)

$\omega_{10} = 0.05,\ \omega_{11} = 0.25,\ \omega_{20} = 0.6,\ \omega_{22} = 0.5$, for Case D. (4D)

As we have said, the precise values of these coefficients have no significance, but the values we have chosen here will be convenient later in the review.

Boolean models have a number of appealing features. They are relatively easy to build and simulate, and their dynamic properties, though they may be unexpected, are usually easy to understand from an intuitive perspective. The structure of the model, i.e., the choice of Boolean functions, may often be derived directly from the topology of the underlying regulatory network, i.e., the constituent signaling motifs. And there are no kinetic parameters in the model that must be assigned numerical values based on experimental data. Boolean modeling is especially expedient when we have good reasons to draw the 'wiring diagram' (i.e., the motif structures) of a regulatory network but we lack the quantitative observations necessary to build a detailed biochemical representation of the network's topology (Assmann and Albert, 2009). In this sense, the 'parameter-free' nature of Boolean models is both their greatest strength and their greatest weakness. They can be excellent tools for developing an intuitive appreciation of the general characteristics of a regulatory-network hypothesis, but they are impotent to address detailed quantitative questions about mechanisms and cellular behaviors.

Given a proper recognition of their strengths and weaknesses, Boolean models have been used extensively and profitably to investigate the properties of important biological regulatory networks. Some early examples of this approach are reviewed in the Proceedings of an EMBO course held in Brussels in 1977 (Thomas, 1979). More recent examples include: expression of the segment polarity genes during fruit fly embryogenesis (Albert and Othmer, 2003), T-cell differentiation (Mendoza and Xenarios, 2006; Naldi et al., 2010), cell cycle progression in budding yeast (Faure et al., 2008; Li et al., 2004) and fission yeast (Davidich and Bornholdt, 2008; Faure and Thieffry, 2009), and signaling networks in plant stomata (Sun et al., 2014).

Piecewise-Linear Models

After writing Eq. (2) as

$$C_i(t+1) - C_i(t) = B_i\big(C_1(t), C_2(t), \ldots, C_N(t)\big) - C_i(t), \quad (5)$$

we introduce a new time scale, $\hat{t} = \varepsilon t$, and continuous variables, $\hat{C}_i(\hat{t})$, with the property that $C_i(\hat{t}) = \text{Heav}(\hat{C}_i(\hat{t}) - \theta_i)$, where Heav($x$) is the Heaviside function defined in Eq. (3A). The parameters, θ_i, which lie in the interval (0,1) for all i, are 'thresholds' that convert the continuous variables, $\hat{C}_i(\hat{t})$, into discrete variables, $C_i(\hat{t})$, whenever a continuous variable crosses its specific

threshold; i.e., $C_i(\hat{t}) = 0$ whenever $0 \le \hat{C}_i(\hat{t}) \le \theta_i$, and $C_i(\hat{t}) = 1$ whenever $\theta_i < \hat{C}_i(\hat{t}) \le 1$. With these changes, Eq. (5) becomes

$$\hat{C}_i(\hat{t} + \varepsilon) - \hat{C}_i(\hat{t}) = \varepsilon\gamma_i\left[B_i\left(C_1(\hat{t}), C_2(\hat{t}), \ldots, C_N(\hat{t})\right) - \hat{C}_i(\hat{t})\right], \tag{6}$$

where the multiplicative factors $\varepsilon\gamma_i$ indicate that the continuous variables $\hat{C}_i(\hat{t})$ change very little over each small time step ε. After many small time steps, however, the continuous variable $\hat{C}_i(\hat{t})$ will approach either 0 or 1, if the Boolean function, $B_i(\ldots)$, doesn't change. Now we divide each side of Eq. (6) by ε and take the limit as $\varepsilon \to 0$. To make the resulting ordinary differential equation more readable, we swap hats on the C_i and t variables:

$$\begin{aligned} \frac{dC_i}{dt} &= \gamma_i\left[B_i(\hat{C}_1(t), \hat{C}_2(t), \ldots, \hat{C}_N(t)) - C_i(t)\right] \\ \hat{C}_i(t) &= \mathrm{Heav}(C_i(t) - \theta_i) = \begin{cases} 0, & \text{if } C_i(t) \le \theta_i \\ 1, & \text{if } C_i(t) > \theta_i \end{cases} \end{aligned} \tag{7}$$

The continuous variables $C_i(t)$ take values between 0 and 1, representing the activities of species i=1,…,N. Each discrete variable $\hat{C}_i(t)$ adopts the value 0 or 1 at any time t, and flips between 0 and 1 only after many small steps in the slow time variable t, when its continuous counterpart $C_i(t)$ is driven by the underlying dynamics to cross the threshold θ_i. The rate constant γ_i governs how fast the variable $C_i(t)$ approaches its steady state value $B_i(\ldots) = 0$ or 1, depending on the current values of all the discrete variables $\hat{C}_i(t)$.

The system of ODEs (7) is the piecewise linear version of a Boolean model, first introduced by (Glass and Kauffman, 1973) and studied in a series of pioneering papers (Glass, 1975; Glass and Pasternak, 1978). The elegant results that can be derived from this approach are expertly reviewed in the paper by Glass & Edwards in this issue. The nice features of PWL-ODEs have been used to advantage by many authors in addition to Leon Glass, for example, Henry McKean's PWL 'caricatures' of Nagumo's equation (in the theory of nerve impulse propagation) (McKean, 1970) and in Uri Alon's magisterial review of molecular regulatory networks in systems biology (Alon, 2007).

In the PWL-ODE (7), the threshold surfaces, $C_i(t) = \theta_i$, divide state space into 2^N orthants, and within each orthant we are solving a set of linear ODEs, where some of the variables are tending toward 0 and some toward 1. The system evolves along these trajectories until it hits a boundary of the orthant, where it abruptly changes direction (see **Figs. 2A,B** for examples). The flow of trajectories in state space bears a dual relation to the state-transitions observed in the discrete Boolean version of the model (compare Fig. 2A to Table 2, column D). In the particular case of Fig. 2A, if we were to sample initial conditions uniformly distributed over the unit square [0,1]×[0,1], then the probability of ending in the state (1,0) would be 0.276 and the probability of ending in the state (0,1) would be 0.724. These probabilities are comparable to an asynchronous Boolean model (Table 2, column B) with p+q = 0.1.

For the special case when all γ_i's are identical, the trajectories are straight lines (as in Fig. 2A), and full solutions of the PWL-ODEs can be evaluated analytically by solving a set of linear algebraic equations that piece together these linear trajectories across the discontinuities on the orthant boundaries. If the γ_i's are different, then the trajectories are curved (as in **Fig. 2B**), and the piecing-together of solutions involves solving sets of transcendental algebraic equations at the orthant boundaries.

An alternative way to write a PWL-ODE model is to replace the Boolean function of Boolean variables, $B_i(\hat{C}_1, \hat{C}_2, \ldots, \hat{C}_N(t))$ in Eq. (7), by a Heaviside function of a linear function of the continuous variables:

$$\frac{dC_i}{dt} = \gamma_i \left[\text{Heav}\left(W_i(C_1(t), C_2(t), \ldots, C_N(t))\right) - C_i(t)\right] \qquad (8)$$

where W_i is given by Eq. (3). In this case, there is no longer a constant threshold, θ_i for each variable, where $\hat{C}_i(t)$ flips between 0 and 1, but rather a 'null' hyperplane, $W_i(C_1,C_2,\ldots,C_N) = 0$ for each variable, where dC_i/dt changes sign between − (C_i decreasing toward 0) and + (C_i increasing toward 1). For example, in Fig. 2B we show the phase plane portrait for the 'Case C' realization of Motif 1A(i) that we introduced in Table 2, column C, using Eq. (8) to implement the PWL model. Notice that the lines along which $W_i(C_1,C_2) = 0$ play the role of 'nullclines' in the PWL-ODEs. In this case, if we were to take a sample of initial conditions uniformly distributed over the unit square [0,1]×[0,1], then the probability of ending in the state (1,0) would be 0.51 and the probability of ending in the state (0,1) would be 0.49.

Nonlinear ODE Models

Traditional ODE Models. The traditional approach to ODE modeling of molecular interaction networks is to use the basic principles of biochemical reaction kinetics to write a set of nonlinear ODEs for each time-dependent variable in the reaction network. On the right-hand-side of each ODE is a sum of terms representing the rates of accumulation and the rates of removal of each substance; something like this:

$$\frac{dC_1}{dt} = \nu_{11}k_1f_1(C_1,\ldots) + \nu_{12}k_2f_2(C_1,\ldots) + \cdots - \nu_{13}k_3f_3(C_1,\ldots) - \nu_{14}k_4f_4(C_1,\ldots) - \cdots \qquad (9)$$

where ν_{1j} is the stoichiometric coefficient of species C_1 in the j-th reaction, k_j is the rate constant of the j-th reaction, and $f_j(C_1,\ldots)$ is the rate law for the j-th reaction (e.g., mass-action rate law, Michaelis-Menten rate law, Hill function, etc.). In the case of Eq. (9) we are assuming that C_1 is a product of reactions 1,2,… (i.e., in the reaction network diagram there is a reaction arrow that points toward C_1) and a reactant of reactions 3,4,… (an arrow that points away from C_1 in the reaction network diagram). From modest beginnings in the 1960's, there are now thousands of such models in the molecular-cell biology literature. The field has been thoroughly reviewed in research monographs by Segel (1980), Murray (1989), Goldbeter (1996), Keener and Sneyd (2009), and in textbooks by Edelstein-Keshet (1988), Klipp (2009), Ingalls (2013), DiStefano (2014), and Voit (2018).

The major goal of these models has been to understand how the coupling of basic biochemical interactions can yield the complex types of 'information processing' exhibited by living cells. Examples include (1) 'perfect adaptation' of signaling networks (Barkai and Leibler, 1997; Bray et al., 1993; Levchenko and Iglesias, 2002; Segel et al., 1986; Tang and Othmer, 1994), (2) 'alternative fates' of differentiating cells (Busse et al., 2010; Hong et al., 2015b; Tian et al., 2013; Wang et al., 2014), (3) 'oscillatory' hormone secretion (Goldbeter et al., 2000; Khadra and Li, 2006; Martiel and Goldbeter, 1987; Pratap et al., 2017; Vidal and Clement, 2010), and (4) 'autonomous oscillations' in circadian rhythms (Forger and Peskin, 2003; Goldbeter, 1995; Kim and Forger, 2012; Leloup and Goldbeter, 2003; Leloup and Goldbeter, 1998; Relogio et al., 2011; Tyson et al., 1999). From decades of modeling with traditional biochemical rate laws, some basic principles of complex signal-processing by molecular regulatory networks have become clear (Novak and Tyson, 2008). First of all, the topology of the reaction network is crucial; for example, incoherent feedforward loops for perfect adaptation, positive feedback (mutual activation or mutual inhibition) for multiple stable steady states, negative feedback loops for oscillations. Secondly, the molecular interactions in these networks must be 'sufficiently' nonlinear. For example, a sufficiently steep sigmoidal function for gene expression may be described by a Hill function with a large exponent (say, $n \geq 4$). Substrate processing described by Michaelis-Menten rate laws exhibits 'ultrasensitivity' when the Michaelis constants are small with respect to the substrate concentration (the so-called Goldbeter-Koshland function) (Goldbeter and Koshland Jr., 1981). Protein activation can exhibit steep sigmoidal responses when the protein is phosphorylated on multiple amino acid sites (Kapuy et al., 2009; Salazar and Hofer, 2007). Third, the rate constants and binding constants governing the rates of these constituent reactions must be set within distinct limits to yield the desired outcomes.

There is essentially no limitation to the breadth and molecular detail that can be represented by models of this type, when the reaction network, kinetic rate laws, and rate constant values are known in sufficient detail. These models can be extremely powerful and informative. Nevertheless, they can be difficult to build, to parametrize and to comprehend. Some simpler approaches that strike a balance between the intuitive appeal—but limited applicability—of Boolean-type models and the limitless potential—but demanding constraints—of comprehensive biochemical models may be warranted.

Soft-Heaviside ODE Models. PWL-ODE models occupy this middle ground, but the discontinuities inherent in the piecewise-linear approach preclude the application of many analytical methods from the theory of (continuous) nonlinear ODEs. To take advantage of the powerful tools of nonlinear dynamical systems, we need to smooth out the discontinuities of PWL-ODE models; for example, by replacing the Heaviside function, Heav(W_i) in Eq. (8), by a smooth sigmoidal function, $H(W_i)$,

$$\frac{dC_i}{dt} = \gamma_i \left[H\left(W_i\left(C_1(t), C_2(t), \ldots, C_N(t)\right)\right) - C_i(t) \right] \tag{10}$$

where W_i is given by Eq. (3B) (usually linear terms suffice). In developmental biology, this 'connectionist' approach was pioneered by Mjolness et al. (1991). A convenient 'smooth sigmoidal function' to use in this approach is the 'soft Heaviside' function

$$H(\sigma_i W_i) = \frac{1}{1+\exp(-\sigma_i W_i)}\,. \qquad (11)$$

The redundant parameter σ_i is introduced to modulate the 'steepness' of the soft Heaviside function, independently of the parameters defining W_i.

An advantage of the soft-Heaviside function is that it can be parametrized to closely resemble the nonlinearity of a Hill function or zero-order ultrasensitivity or multisite phosphorylation. Hence, a soft-Heaviside model can be expected to give results comparable to traditional modeling of standard biochemical reaction mechanisms. In situations where the origin of nonlinearity in a mechanism is not obvious, then it is as reasonable to use a soft-Heaviside function as a Hill function or a Goldbeter-Koshland function. Indeed, it may be desirable to use a generic sigmoidal function, like Eq. (11), rather than a specific function like Hill or Goldbeter-Koshland. Notice that $H(\sigma_i W_i)$ approaches the discontinuous Heaviside function in Eq. (8) in the limit as $\sigma_i \to \infty$.

In this formalism, the coefficients in the function W_i take on a new meaning. They are no longer somewhat arbitrary numbers introduced simply to generate a Boolean function of the C_i's when W_i is run through the Heaviside filter. Now they can be interpreted as 'model parameters' whose values characterize the topology of the regulatory motif and the strengths of the interactions. Thus, $\omega_{ij} > 0$ if species j activates species i, $\omega_{ij} < 0$ if species j inhibits species i, $\omega_{ij} = 0$ if species j has no effect on species i. The magnitudes of the ω_{ij}'s indicate the relative strengths of the interactions. The value of ω_{i0} determines whether species i is active or inactive when it is receiving no inputs from any species in the network.

The PWL models of Motif 1A(i) in **Figs. 2A, B** are compared in **Figs. 2C, D** with the soft-Heaviside ('soft-H' hereafter) versions of the same models. The soft-H models smooth out the 'kinks' in the PWL models. In **Fig. 3A, B** we present soft-H models of Motifs 1A(ii) and 1A(iii).

Motif 1A(ii) is often used to model cell-differentiation processes e.g., Hong et al. (2011; 2012), where the alternative stable steady states ($C_1 \approx 1$, $C_2 \approx 0$ and $C_1 \approx 0$, $C_2 \approx 1$) represent two alternative differentiated states (e.g., epithelial vs mesenchymal cells, or granulocytes vs monocytes). Typically, this motif has two other (potentially stable) steady states: a naïve, undifferentiated state ($C_1 \approx 0$, $C_2 \approx 0$) and a hybrid, dual-expressing state ($C_1 \approx 1$, $C_2 \approx 1$). When a naïve, undifferentiated stem cell is exposed to a certain cocktail of signaling molecules, it can be induced to differentiate into one or another of the three terminally differentiated states. In the example in **Fig. 3A**, a population of naïve cells (in the lower left corner) is exposed to a strong differentiation signal that supports three stably differentiated states. The fate of any particular naïve cell depends (in the case of this deterministic model) on its precise initial conditions. Under other assumptions, the fate may depend on small differences in parameter values in

individual cells, or on stochastic fluctuations in molecular abundances during the differentiation process.

By modeling molecular regulatory networks with continuous, nonlinear ODEs, such as a soft-H model, we gain access to all the powerful analytical tools of dynamical systems theory (Edelstein-Keshet, 1988; Kuznetsov, 1995; Strogatz, 1994), such as phase-plane analysis (as in Figs. 3A,B) and bifurcation diagrams. These tools reveal how the dynamical properties of a model depend on the values of numerical parameters in the model. One of the most powerful of these tools is the one-parameter bifurcation diagram, which displays the dependence of the asymptotic states of a dynamical variable (both steady-state and periodic solutions) on the value of a 'bifurcation parameter' in the model. Typically, the bifurcation parameter is chosen to be a quantity that is under experimental control (such as the concentration of an 'inducer' in the bathing medium of the cells, or the level of expression of a gene with an inducible or repressible promoter), and the dynamical variable is chosen to be a protein with a major controlling influence on cell behavior (such as a 'master' transcription factor or a 'checkpoint' protein). The one-parameter bifurcation diagram shows how the output—the 'response'—of the control system (the activity of major controlling protein) depends on the input—the 'signal'—to the control system. In this interpretation, a one-parameter bifurcation diagram is the precise mathematical representation of a typical 'signal-response' curve measured experimentally.

To illustrate the utility of one-parameter bifurcation diagrams we use them to characterize the behaviors of Motifs 1A(ii) and 1A(iii) in Figs. 3C, D. In **Fig. 3C** we show how the steady-state activity of the response protein, C_1, depends on the strength of the differentiation signal, S. At S=0, the only stable state is the naïve state (both C_1 and C_2 small). As S increases, two alternatively differentiated states (C_1 large, C_2 small; and C_1 small, C_2 large) are created at saddle-node bifurcations (at S=0.04), but initially the naïve state is very stable to perturbations. (For a brief description of some common types of bifurcation points, see Table 4.) Only when the naïve state disappears, at a pitchfork bifurcation at S=0.28, will the stem cells be forced to differentiate to one or the other of the alternatively differentiated states. (In this simulation, because of the symmetry of the parameter values we have chosen, the naïve cells have a 50:50 chance of going in one direction or the other.) For larger values of signal strength (S > 0.43), a stable dual-expressing state (both C_1 and C_2 large) appears, and the phase-plane portrait of the system has the appearance of panel A, until S exceeds 0.67, after which the only stable steady state is the dual-expressing state. The soft-H modeling approach has the advantage of being able to trace out, using one-parameter bifurcation diagrams, all the subtle relations among differentiated states and signal strengths, and to explore the effects of broken symmetries on the behavior of the regulatory network (Hong et al., 2011; 2012). Two-parameter bifurcation diagrams can be used to illustrate the response of a control system to the simultaneous manipulation of two signals, for example, a primary differentiation signal, S_1, and a secondary 'polarizing' signal, S_2, as in Hong et al. (2015a).

Motif 1A(iii) is often used to model spontaneous oscillations (e.g., Fig. 3B) in systems with positive and negative feedback loops (Novak and Tyson, 2008; Tsai et al., 2008). In **Fig. 3D** we use a one-parameter bifurcation diagram to show a typical scenario for the 'onset' and

'demise' of oscillations in motif 1A(iii). As signal strength, *S*, increases, stable limit cycle oscillations arise at a Hopf bifurcation at *S*=0.825. The oscillations quickly grow to large amplitude and persist until *S*=1.3, where the limit cycle is extinguished at a saddle-node homoclinic bifurcation, where the period of the oscillation goes to infinity. Motif 1A(iii) is especially prone to exhibit bistability as well as oscillations, and to a host of complex bifurcation scenarios generated by potential interactions between bistability and oscillations. For example, the complex 'tableau' of phase-plane portraits presented by Guckenheimer (1986) in his figures 7 and 8 are readily recreated by this soft-H model.

Soft-H models have many advantages over Boolean and PWL-ODE models. They are readily formulated, because the structure of the ODEs is derived directly from the topology of the motif through the W_i functions. The ω_{ij} parameters now have a biophysical meaning in terms of the relative strengths of the interactions between species *i* and *j*. These parameter values can now be adjusted to fit the output of the model to quantitative and qualitative experimental data (e.g., measured time courses of certain species $C_i(t)$, or phenotypes of genetically mutated strains). Once the parameter values have been estimated from experimental data, the model can be used to make quantitative predictions about the outcome of novel experiments.

Soft-H models of this sort have been used by many investigators in the past. In an early, innovative approach, Mendoza and Xenarios (2006) combined discrete dynamical modeling (a Boolean network) with continuous ODEs (with soft-H nonlinearities) to study a complex model of T-cell differentiation. Nelander et al. (2008) and Molinelli et al. (2013) used soft-H models in their ambitious study of signaling networks in cancer cell lines. Fu et al. (2012) used soft-H ODEs to dissect the topologies of molecular networks that underlie priming and tolerance of cells of the innate immune system.

Soft-H models need not be limited to the form of Eq. (10), with a single soft-Heaviside function $H(\sigma_i W_i)$. It is certainly within the 'spirit' of soft-H modeling to employ multiple soft-Heaviside functions as sums and/or products to represent more complex network topologies. Nonetheless, there are certain limitations to this formalism that are addressed in the next section.

'Standard-Component' Models. A disadvantage of Soft-H models is that each component of the model is described by an ODE of the form Eq. (10), which is an unnecessarily restrictive constraint. There is no reason why we cannot mix soft-H ODEs with more traditional biochemical rate equations, such as ODEs for protein synthesis and degradation ($\rightarrow X_j \rightarrow$):

$$\frac{dX_j}{dt} = k_{\text{syn}} F_j(C_i, X_k) - k_{\text{deg}} G_j(C_i, X_k) X_j \qquad (12)$$

or ODEs for bimolecular complex formation and dissociation ($X_j + X_k \leftrightarrow Z_{jk}$):

$$\frac{dZ_{jk}}{dt} = k_{\text{assoc}} X_j X_k - k_{\text{dissoc}} Z_{jk} \qquad (13)$$

We call this sort of hybrid model, mixing ODEs of the form of Eqs. (10), (12) and (13), a 'standard-component' model (SCM) (Laomettachit et al., 2016). (Note: it is not the modeling approach itself that is 'standard' but the components of the model that are 'standardized'.) Underlying the idea of an SCM is the notion of a natural separation of time scales: that protein

synthesis and degradation, Eq. (12), occurs on a longer time scale (10 min, say) than post-translational protein modifications (1 min or less), as described by Eq. (10), which in turn are slower than the formation of protein complexes, Eq. (13), which associate and dissociate in seconds.

In the SCM paradigm, we often treat complex formation by a pre-equilibrium assumption:

$$k_{\text{assoc}} X_j X_k = k_{\text{dissoc}} Z_{jk}, \tag{14}$$

which we can write as

$$\left(X_{j\text{T}} - Z_{jk}\right)\left(X_{k\text{T}} - Z_{jk}\right) = K_{\text{d}} Z_{jk}, \tag{15}$$

where $X_{j\text{T}} = X_j + Z_{jk}$, $X_{k\text{T}} = X_k + Z_{jk}$, and $K_{\text{d}} = k_{\text{dissoc}}/k_{\text{assoc}}$ is the 'dissociation constant' for the complex Z_{jk}. Equation (13) is a quadratic equation in Z_{jk}, whose solution is

$$Z_{jk} = \frac{1}{2}\left(X_{j\text{T}} + X_{k\text{T}} + K_{\text{d}} - \sqrt{\left(X_{j\text{T}} + X_{k\text{T}} + K_{\text{d}}\right)^2 - 4X_{j\text{T}}X_{k\text{T}}}\right) . \tag{16}$$

In the limit of very strong binding, $K_{\text{d}} << \min(X_{j\text{T}}, X_{k\text{T}})$, we can approximate Eq. (16) by

$$Z_{jk} = \min(X_{j\text{T}}, X_{k\text{T}}) . \tag{17}$$

To illustrate the SCM approach, we study a simple motif (**Fig. 4A**) for regulation of entry into and exit from mitosis in eukaryotic cells. The model has four components: 'CycB' is the regulatory subunit of cyclin B-dependent protein kinase, the principal driver of mitosis; 'Cdh1' is a regulatory subunit of the Cyclosome, which degrades CycB as cells exit mitosis; Cdh1:Cyclosome remains active throughout G1 phase of the cell cycle; 'CAP' is a generic 'counter-acting phosphatase' that must be activated to turn on Cdh1; and 'INH' is a generic stoichiometric 'inhibitor' of CAP. Notice that CycB and Cdh1 are locked in a mutually antagonistic 'embrace' (motif 1A(i)), which creates a bistable switch (Cdh1 ON and CycB OFF in G1 phase, CycB ON and Cdh1 OFF in S-G2-M phase of the cell cycle). Furthermore, CycB ---| INH ---| CAP → Cdh1 ---| CycB is a 4-component negative feedback loop, rather like motif 1A(iv). Based on general properties of these four components in the eukaryotic cell-cycle control network (Barik et al., 2010), we use Eq. (12) to describe the synthesis and degradation of CycB, Eq. (10) to describe the phosphorylation and dephosphorylation of Cdh1 and INH, and Eq. (17) to describe the formation of stoichiometric complex between INH and CAP (**Fig. 4B**). Simulations of this SCM are illustrated in **Fig. 4C**, and more details are given in the legend. The bifurcation diagram in **Fig. 4D** shows how oscillations and bistability in this network depend on cell size, $M(t)$. A small, newborn cell ($M_{\text{birth}} \approx 1$) is captured by the G1 steady state (Cdh1 ON and CycB OFF) and persists therein until it grows to a mass of 1.34, where the G1 steady state is lost at a saddle-node bifurcation. Thereafter, the cell skips onto a large amplitude limit cycle oscillation in CycB(t), which drives the cell through S and G2 phases into M phase. As the cell comes out of mitosis, CycB(t) drops below a threshold value, CycB(t) = 0.3, and the cell divides, ($M_{\text{division}} \approx 2$) → ($M_{\text{birth}} \approx 1$), and the newborn cell resets to G1 phase.

More elaborate versions of the SCM approach have been used to model, in great detail, the molecular mechanism regulating cell cycle progression in budding yeast (Kraikivski et al., 2015; Laomettachit et al., 2016). It is instructive to compare the SCM approach to this problem to other deterministic approaches: (1) Boolean models in Li et al. (2004), Davidich and Bornholdt (2008), and Faure et al. (2009); (2) traditional, nonlinear ODE models in Novak and Tyson (1993), Qu et al. (2003), Chen et al. (2004), Csikasz-Nagy et al. (2006), Gerard and Goldbeter (2009); and (3) a hybrid Boolean-ODE approach by Singhania et al. (2011).

Stochastic Models

Whenever we model molecular interaction networks within cells, we must eventually come to terms with the fact that living cells are very small, with limited numbers of macromolecules: one or two genes per cell, 5-50 mRNA molecules per gene per cell, and 500-5000 or more specific types of protein molecules per cell. Furthermore, living cells are spatially non-uniform environments, with distinct compartments (nucleus, organelles, cortex, membranes, cytoskeleton, etc.). Consequently, ODEs are not a realistic mathematical framework, ultimately, for modeling network dynamics and cell physiology. Spatial non-uniformities must be modeled with distinct compartments and intracellular transport processes (diffusion, advection, motor proteins, nuclear and cytoplasmic localization signals, etc.), and thermal fluctuations among small numbers of molecules must be modeled by stochastic equations. Spatial non-uniformities are beyond the scope of this review, but we will make a few observations about stochastic modeling.

The most accurate way to handle stochastic fluctuations in chemical reaction systems is by Gillespie's stochastic simulation algorithm (SSA) (Gillespie, 1976; Gillespie, 1977). In this formalism, mass-action kinetic terms are no longer treated as 'deterministic rates of reaction' but rather as 'stochastic propensities for a reaction to occur over a small interval of time'. Given these propensities for an arbitrarily complex reaction network and a particular starting point in terms of the number of molecules of each species $C_i(t)$ in the network at time t, Gillespie's SSA computes the time interval Δt until the next reaction occurs, and the identity of the next reaction to occur (i.e., its index j). With this information, the network state is updated according to $C_i(t + \Delta t) = C_i(t) + \nu_{ij}$, and the procedure is repeated.

Strictly speaking, Gillespie's SSA is valid only for elementary reaction mechanisms expressed in terms of mass-action rate laws. Nonetheless, some authors (ourselves included) have applied SSA to non-elementary reaction steps because phenomenological rate laws (like Michaelis-Menten and/or Hill functions) are commonly used in deterministic models and it is tempting to turn a deterministic simulation into a stochastic realization simply by reinterpreting the kinetic rate law as a propensity of reaction. Rao and Arkin (2003) have carefully studied the conditions under which this 'reinterpretation' gives sufficiently accurate results. They found that the Michaelis-Menten rate law can be reinterpreted in this way under conditions under which it is a valid approximation in the deterministic setting, i.e., when total enzyme concentration is much less than total substrate concentration, $[E]_0 << [S]_0$. But the approximation is very misleading when $[E]_0 \geq [S]_0$. Similarly, in a gene-expression model (autoactivation of the Cro transcription factor) with a Hill-like function for promoter occupancy, Rao and Arkin found that the statistics

of gene expression were well approximated by the reinterpretation method when dimerization of Cro was rapid (~seconds) compared to protein synthesis (~min) but not if the dimerization rates were comparable to the protein synthesis rate. These observations are especially relevant to gene-protein interaction networks, where 'enzymes' and 'substrates' are often proteins of comparable concentrations in a cell, and the rates of transcription factor interactions with each other and with binding sequences on DNA are often unknown and unlikely to exhibit the requisite separation of time scales. Indeed, in our studies of gene-protein regulatory networks (Barik et al., 2008; Ciliberto et al., 2007; Kapuy et al., 2009; Kar et al., 2009; Sabouri-Ghomi et al., 2008), we have run into these problems time and time again, and we have come to the conclusion that it is not prudent to use these phenomenological modeling-and-simulation approximations at all, because the risk of computing unreliable results far outweighs the modest gains in computational efficiency. We recommend 'unpacking' phenomenological rate laws into elementary reaction steps before applying Gillespie's SSA to compute realizations of a stochastic chemical reaction network.

The drawbacks of Gillespie's method are manifold. Most seriously, one must have a detailed biochemical reaction mechanism in terms of elementary steps, one must know the rate constant characterizing each step, and one must know the numbers of molecules of every species in the reaction mechanism. This sort of detailed knowledge is rarely available. Furthermore, Gillespie's method is computationally intensive because it spends most of its time computing fluctuations of the fastest reactions in the network, which are often the least informative (e.g., rapid association and dissociation of protein complexes to establish a state of pseudo-equilibrium). Nonetheless, we have been able to carry out 'brute force' Gillespie simulations of a detailed model of the molecular mechanism governing progression through the budding yeast cell cycle, with very satisfying results (Barik et al., 2016).

Seeking a reasonable way to speed up SSA calculations without sacrificing accuracy, we make note of the fact that stochastic fluctuations are most likely to affect the numbers of mRNA molecules in a cell, because they are present at levels (N = 5-50 molecules per gene per cell) that are most susceptible to thermal fluctuations (standard deviation $\approx N^{1/2} \approx$ 2-7). The number of genes per cell, though small, is not subject to stochastic fluctuations, and the numbers of protein molecules per cell may be large enough to buffer thermal noise, although we must consider that protein abundances will magnify fluctuations in mRNA numbers. To see how this comes about, consider the effects of thermal noise on mRNA turnover in a living cell. If mRNA synthesis and degradation, $\rightarrow$ mRNA $\rightarrow$, is governed by a simple birth-death process (k_s = rate of synthesis, molecule second^{-1}; k_d = rate of degradation, second^{-1}), then the distribution of mRNA molecules, N_M, in a cell follows a Poisson distribution (Thattai, 2016),

$$P(N_M = k) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{18}$$

where $\lambda = k_s/k_d$. For a Poisson distribution, mean = $\langle N_M \rangle = \lambda$ and var = $\langle N_M^2 \rangle - \langle N_M \rangle^2 = \lambda$; hence, the coefficient of variation of N_M = var^{½}/mean = mean$^{-½}$. Accurate counts of mRNA molecules in budding yeast cells give mean ≈ 5-10 molecules per cell (Ball et al., 2013; Zenklusen et al., 2008); hence, the CV for mRNA expression in yeast cells is expected to be 32-

45%. The referenced experimental papers show that, in budding yeast cells, for genes that are constitutively expressed, mRNA counts indeed obey a Poisson distribution.

The corresponding case for protein molecules per cell (mean = 500-5000) would give a CV for protein expression of 1.4-4.5%. This expectation, however, is not valid, because protein turnover is not governed by a simple birth-death process. Rather, the rate of synthesis of protein molecules is dependent on the abundance of mRNA molecules in a cell, and mRNA numbers are themselves fluctuating, as described in the previous paragraph. As a result of transcription-translation coupling (Pedraza and Paulsson, 2008; Swain et al., 2002), the expected CV of protein numbers in a cell is

$$CV_{\mathrm{P}} \approx \sqrt{\frac{1}{\langle N_{\mathrm{P}} \rangle} + \frac{\tau_{\mathrm{M}}}{\tau_{\mathrm{M}}+\tau_{\mathrm{P}}} \frac{1}{\langle N_{\mathrm{M}} \rangle}} \tag{19}$$

where τ_{M} and τ_{P} are the half-lives of these particular mRNA and protein molecules in the cell. Taking representative numbers for a budding yeast cell ($<N_{\mathrm{P}}> = 2000$, $<N_{\mathrm{M}}> = 5\text{-}10$), we find a tolerable value for protein-number fluctuations, $CV_{\mathrm{P}} \approx 20\%$, if $0.25 < \tau_{\mathrm{M}}/\tau_{\mathrm{P}} < 0.67$, i.e., if mRNA molecules are significantly less stable than protein molecules in a yeast cell (as we might expect).

Equation (19) suggests that we must include mRNA dynamics in our molecular mechanism of a regulatory network, in order to get an accurate estimate of protein fluctuations in a cell and, consequently, to get a realistic appraisal of the effects of thermal noise on the behavior of the regulatory network. Models that are expressed solely in terms of protein interactions (folding gene expression terms and mRNA levels into the rate of synthesis of the encoded protein) will underestimate the magnitude of protein-number fluctuations and will give completely unreliable predictions about the effects of noise on the physiological properties of single cells. It is not difficult, of course, to add mRNA variables to a protein interaction network, but doing so exacerbates the computational demands of Gillespie's SSA.

We have proposed two ways to circumvent this problem. The first is a 'hybrid' computational approach, following the lead of Hazeltine & Rawlings (2002). Our basic idea (Liu et al., 2012) is to use Gillespie's SSA to simulate stochastic fluctuations in species that are in low abundance and participate in 'slow' (i.e., infrequent) reactions, and to approximate the dynamics of 'fast', high-abundance species by nonlinear ODEs. For gene-mRNA-protein regulatory networks, an effective 'partitioning' strategy is to simulate mRNA synthesis and degradation by SSA and all subsequent protein-interaction processes by nonlinear ODEs. We have shown (Ahmadian et al., 2017; Liu et al., 2012; Wang et al., 2016) that this approach is 50-100 times faster than 'brute force' SSA, without sacrificing accuracy of the stochastic results.

The second approach adopts the strategy of a chemical Langevin equation (Gillespie, 2000), with the additive 'white noise' terms designed to account for transcription-translation coupling. Within the context of an SCM, we write (Laomettachit et al., 2016)

$$\frac{\Delta X_j}{\Delta t} = F_j - G_j X_j + \sqrt{F_j + G_j X_j} \cdot \frac{\xi_1}{\sqrt{\Delta t}} + X_j \sqrt{\frac{2G_j}{\langle m_j \rangle} \cdot \frac{G_j}{G_j + k_{\mathrm{dm}j}}} \cdot \frac{\xi_2}{\sqrt{\Delta t}} \tag{20}$$

$$\frac{\Delta C_i}{\Delta t} = \gamma_i[H(\sigma_i W_i) - C_i(t)] \quad (21)$$

$$Z_{jk} = \min(X_{j\mathrm{T}}, X_{k\mathrm{T}}) \; . \quad (22)$$

In Eq. (20), which is derived in Suppl. Text S2 of (Laomettachit et al., 2016), ξ_1 and ξ_2 are independent random variables, each chosen from a normal distribution with mean = 0 and variance = 1. The two 'noise' terms correspond to random fluctuations at the protein level (ξ_1) and at the mRNA level (ξ_2). In the 'mRNA-inherited' noise term, $k_{\mathrm{dm}j}$ is the rate constant for degradation of the mRNA encoding protein X_j, and $\langle m_j \rangle = k_{\mathrm{sm}j}/k_{\mathrm{dm}j}$ = mean number of these mRNA molecules (which may be time-dependent, if the expression of the gene encoding X_j is regulated by the reaction mechanism). We do not add 'noise' terms to Eqs. (21) and (22) because we assume that these reactions are fast enough to average over any molecular fluctuations.

We illustrate the second approach with a stochastic simulation of the 'cell cycle control system' in Fig. 4A. The `xpp.ode` file for this stochastic model is given in the Appendix, and a representative simulation is provided in **Fig. 4E**. In this model, the total abundances of Cdh1 and INH are fluctuating around their steady state levels of 1000 and 3000 molecules, respectively (due to transcription-translation coupling). From Eq. (19) we calculate that CV_{Cdh1T} = 22% and CV_{INHT} = 13%, which values agree quite well with the magnitudes of the fluctuations observed in the turquoise curves in Fig. 4E.

Conclusions

In this review article we have described a number of different strategies for building mathematical models of the temporal dynamics of molecular regulatory networks in living cells. This is a problem of fundamental importance in cell biology, because the physiology of a living cell is determined by its underlying molecular control systems, and these control systems are staggeringly complex and defy understanding by informal, intuitive reasoning. Accurate mathematical models are necessary to deduce the consequences of our hypotheses about the molecular components and interactions of these control networks. By comparing quantitative and qualitative consequences of the mathematical models to the observed behavior of real cells under a variety of conditions and perturbations, we can assess the reliability of our network hypotheses, identify weaknesses in our understanding, improve on our models, and eventually come to an understanding of molecular control processes that is satisfying in its scope, accuracy and predictive power. Accurate mathematical modeling is the tool that makes these deductions possible.

The challenge to mathematical modeling is that different aspects of cell physiology are characterized, at an experimental level, in widely divergent details. In some cases (e.g., the cell division cycle in budding yeast), the molecular control system is known in great detail and there is a wealth of qualitative and quantitative observations about cell cycle progression in wild-type and mutant yeast cells, in single cells as well as cell populations, under a variety of unperturbed and perturbed conditions. In this happy circumstance, we can build detailed, accurate, reliable mathematical models of the cell cycle control system. In other cases (e.g., the growth, proliferation and spread of cancer cells in mammals), the control systems are much more

complex and the experimental details much less comprehensive and reliable. Under these circumstances, it is difficult to build trustworthy mathematical models, but modeling can still be a useful tool in testing and refining our admittedly incomplete hypotheses about the underlying molecular control systems.

The 'art' of modeling is to choose the right approach to the biological problem under consideration. Poorly characterized systems should be modeled by methods that can be implemented quickly, with minimal requirements for mechanistic details, and that provide general, qualitative insights that can be useful in guiding our intuition about the control system and designing the next set of experiments to be pursued. Well characterized systems should be modeled by accurate methods that provide reliable, quantitative behaviors of the hypothetical control systems; properties that can be compared in detail to observed cellular behaviors and that provide rigorous, testable predictions for future experiments.

To provide a context in which the next generation of molecular cell biologists can practice this art of mathematical modeling, we have described a selection of mathematical approaches for modeling the temporal dynamics of molecular regulatory systems. We have selected approaches that have proved, over the past 50 years, to be useful in a variety of circumstances, and we have described how these approaches are related to each other in a sequence of increasing complexity (more molecular details, more requirements for quantitative experimental data, and more potential for accurate qualitative and quantitative predictions). The methods we have discussed, along with their major advantages and disadvantages, are summarized in **Table 5**. We sincerely hope that this overview will be useful to newcomers to the field of molecular systems/computational biology, and we encourage you to use these tools wisely and effectively in your future studies.

Acknowledgements

The ideas and methods described in this review have been developed and perfected over the past 50 years by (an initially small) cadre of brilliant scientists, many of whom have passed away in recent years. This paper is dedicated to their memory; most especially to Rene Thomas, a computational cell biologist who pioneered the 'logical modeling' approach to molecular systems biology. In addition to being a scientist of great talent and creativity, Rene was widely respected for his humanity and gentlemanly demeanor. JJT would also like to recognize other deceased friends and colleagues who made fundamental contributions to theoretical biology: Joel Keizer, Arthur Winfree, Lee Segel, George Oster and Garry Odell. Preparation of this review article was supported in part by an NIH grant (5R01GM078989-12) to JJT.

References

Ahmadian, M., Wang, S., Tyson, J., Cao, Y., 2017. Hybrid ODE/SSA Model of the Budding Yeast Cell Cycle Control Mechanism with Mutant Case Study. Acm-Bcb' 2017: Proceedings of the 8th Acm International Conference on Bioinformatics, Computational Biology,and Health Informatics, 464-473, doi:10.1145/3107411.3107437.

Albert, R., Othmer, H. G., 2003. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. J Theor Biol 223, 1-18.

Alon, U., 2007. An introduction to systems biology : design principles of biological circuits. Chapman & Hall/CRC, Boca Raton, FL.

Assmann, S. M., Albert, R., 2009. Discrete dynamic modeling with asynchronous update, or how to model complex systems in the absence of quantitative information. Methods Mol Biol 553, 207-25, doi:10.1007/978-1-60327-563-7_10.

Ball, D. A., Adames, N. R., Reischmann, N., Barik, D., Franck, C. T., Tyson, J. J., Peccoud, J., 2013. Measurement and modeling of transcriptional noise in the cell cycle regulatory network. Cell Cycle 12, 3203-3218, doi:10.4161/cc.26257.

Barik, D., Ball, D. A., Peccoud, J., Tyson, J. J., 2016. A Stochastic Model of the Yeast Cell Cycle Reveals Roles for Feedback Regulation in Limiting Cellular Variability. Plos Computational Biology 12, e1005230, doi: 10.1371/journal.pcbi.1005230.

Barik, D., Paul, M. R., Baumann, W. T., Cao, Y., Tyson, J. J., 2008. Stochastic simulation of enzyme-catalyzed reactions with disparate time scales. Biophys. J. 95, 3563-3574.

Barik, D., Baumann, W. T., Paul, M. R., Novak, B., Tyson, J. J., 2010. A model of yeast cell cycle regulation based on multisite phosphorylation. Mol. Syst. Biol. 6, 405.

Barkai, N., Leibler, S., 1997. Robustness in simple biochemical networks. Nature 387, 913-917.

Belousov, B. P., 1959. A periodic chemical reaction and its mechanism. Sb. Ref. Radiats. Med. (Collections of Abstracts on Radiation Medicine), Medgiz, Moscow., 145-147.

Bray, D., Bourret, R. B., Simon, M. I., 1993. Computer simulation of the phosphorylation cascade controlling bacterial chemotaxis. Mol. Biol. Cell 4, 469-482.

Busse, D., de la Rosa, M., Hobiger, K., Thurley, K., Flossdorf, M., Scheffold, A., Hofer, T., 2010. Competing feedback loops shape IL-2 signaling between helper and regulatory T lymphocytes in cellular microenvironments. Proc Natl Acad Sci U S A 107, 3058-63, doi:10.1073/pnas.0812851107.

Chance, B., Hess, B., Betz, A., 1964. DPNH oscillations in a cell-free extract of S. carlsbergensis. Biochem Biophys Res Commun 16, 182-7.

Chance, B., Ghosh, A. K., Pye, E. K., Hess, B., 1973. Biological and Chemical Oscillators Proceedings of a Conference in Prague, 1968), Academic Press, New York, 534.

Chen, K. C., Calzone, L., Csikasz-Nagy, A., Cross, F. R., Novak, B., Tyson, J. J., 2004. Integrative analysis of cell cycle control in budding yeast. Mol. Biol. Cell 15, 3841-3862, doi:10.1091/mbc.E03-11-0794.

Ciliberto, A., Capuani, F., Tyson, J. J., 2007. Modeling netowrks of coupled enzymatic reactions using the total quasi-steady state approximation. PLoS Comput. Biol. 3, e45.

Csikasz-Nagy, A., Battogtokh, D., Chen, K. C., Novak, B., Tyson, J. J., 2006. Analysis of a generic model of eukaryotic cell cycle regulation. Biophys. J. 90, 4361-4379.

Davidich, M. I., Bornholdt, S., 2008. Boolean network model predicts cell cycle sequence of fission yeast. PLoS One 3, e1672, doi:10.1371/journal.pone.0001672.

DiStefano, J. I., 2014. Dynamic Systems Biology Modeling and Simulation, Academic Press.

Durston, A. J., 1974. Pacemaker activity during aggregation in Dictyostelium discoideum. Dev Biol 37, 225-35.

Edelstein-Keshet, L., 1988. Mathematical Models in Biology. Random House, New York.

Faure, A., Thieffry, D., 2009. Logical modelling of cell cycle control in eukaryotes: a comparative study. Molecular Biosystems 5, 1569-1581, doi:10.1039/b907562n.

Faure, A., Naldi, A., Lopez, F., Chaouiya, C., Ciliberto, A., Thieffry, D., 2008. Modular logical modeling of the budding yeast cell cycle. Mol. Biosyst. 5, 1787-1796.

Forger, D., Peskin, C. S., 2003. A detailed predictive model of the mammalian circadian clock. Proc. Natl. Acad. Sci. USA, in press.

Fu, Y., Glaros, T., Zhu, M., Wang, P., Wu, Z., Tyson, J. J., Li, L., Xing, J., 2012. Network topologiesw and dynamics leading to endotoxin tolerance and priming in innate immune cells. PLoS Comput. Biol. 8, e1002526.

Gerard, C., Goldbeter, A., 2009. Temporal self-organization of the cyclin/Cdk network driving the mammalian cell cycle. Proc Natl Acad Sci U S A 106, 21643-8, doi:10.1073/pnas.0903827106.

Gerisch, G., Hess, B., 1974. Cyclic-AMP-controlled oscillations in suspended Dictyostelium cells: their relation to morphogenetic cell interactions. Proc Natl Acad Sci U S A 71, 2118-22.

Gillespie, D. T., 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comput. Phys. 22, 403-434.

Gillespie, D. T., 1977. Exact Stochastic Simulation of Coupled Chemical-Reactions. Abstracts of Papers of the American Chemical Society 173, 128-128.

Gillespie, D. T., 2000. The chemical Langevin equation. J. Chem. Phys. 113, 297-306.

Glansdorff, P., Prigogine, I., 1971. Thermodynamic theory of structure, stability and fluctuations. Wiley-Interscience, London, New York,.

Glass, L., 1975. Classification of biological networks by their qualitative dynamics. J Theor Biol 54, 85-107.

Glass, L., Kauffman, S., 1973. The logical analysis of continuous non-linear biochemical control networks. J. Theor. Biol. 39, 103–129.

Glass, L., Pasternak, J. S., 1978. Prediction of limit cycles in mathematical models of biological oscillations. Bull. Math. Biol. 40, 27-44.

Goldbeter, A., 1995. A model for circadian oscillations in Drosophila period protein (PER). Proc. Roy. Soc. Lond. B. Biol. Sci. 261, 319-324.

Goldbeter, A., 1996. Biochemical Oscillations and Cellular Rhythms. Cambridge University Press, Cambridge.

Goldbeter, A., Lefever, R., 1972. Dissipative structures for an allosteric model: Application to glycolytic oscillations. Biophys. J. 12, 1302-1315.

Goldbeter, A., Segel, L. A., 1977. Unified mechanism for relay and oscillation of cyclic AMP in Dictyostelium discoideum. Proc. Natl. Acad. Sci. USA 74, 1543-1547.

Goldbeter, A., Koshland Jr., D. E., 1981. An amplified sensitivity arising from covalent modification in biological systems. Proc. Natl. Acad. Sci. USA 78, 6840-6844.

Goldbeter, A., Dupont, G., Halloy, J., 2000. The frequency encoding of pulsatility. Novartis Found Symp 227, 19-36; discussion 36-45.

Goodwin, B. C., 1965. Oscillatory behavior in enzymatic control processes. Adv Enzyme Regul 3, 425-38.

Goodwin, B. C., 1966. An entrainment model for timed enzyme synthesis in bacteria. Nature 209, 479-481.

Griffith, J. S., 1968a. Mathematics of cellular control processes: II. Positive feedback to one gene. J. Theor. Biol. 20, 209-216.

Griffith, J. S., 1968b. Mathematics of cellular control processes. I. Negative feedback to one gene. J. Theor. Biol. 20, 202-208.

Guckenheimer, J., 1986. Multiple bifurcation problems for chemical reactors. Physica 20D, 1-20.

Hazeltine, E. L., Rawlings, J. B., 2002. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. J. Chem. Phys. 117, 6959-6969.

Hess, B., Brand, K., Pye, K., 1966. Continuous oscillations in a cell-free extract of S. carlsbergensis. Biochem Biophys Res Commun 23, 102-8.

Higgins, J., 1967. The theory of oscillating chemical reactions. Ind. Eng. Chem. 59, 18-62.

Hong, T., Oguz, C., Tyson, J. J., 2015a. A Mathematical Framework for Understanding Four-Dimensional Heterogeneous Differentiation of T Cells. Bulletin of Mathematical Biology 77, 1046-1064, doi:10.1007/s11538-015-0076-6.

Hong, T., Xing, J., Li, L., Tyson, J. J., 2011. A mathematical model for the reciprocal differentiation of T helper 17 cells and induced regulatory T cells. PLoS Comput. Biol. 7, e1002122.

Hong, T., Xing, J., Li, L., Tyson, J. J., 2012. A simple theoretical framework for understanding heterogeneous differentiation of CD4+ T cells. BMC Syst. Biol. 6, 66.

Hong, T., Watanabe, K., Ta, C. H., Villarreal-Ponce, A., Nie, Q., Dai, X., 2015b. An Ovol2-Zeb1 Mutual Inhibitory Circuit Governs Bidirectional and Multi-step Transition between Epithelial and Mesenchymal States. PLoS Comput Biol 11, e1004569, doi:10.1371/journal.pcbi.1004569.

Ingalls, B. P., 2013. Mathematical modeling in systems biology an introduction. MIT Press,, Cambridge, Massachusetts, pp. 1 online resource (xiv, 408 pages).

Jacob, F., Monod, J., 1961. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 3, 318-56.

Kapuy, O., Barik, D., Domingo Sananes, M. R., Tyson, J. J., Novak, B., 2009. Bistability by multiple phosphorylation of regulatory proteins. Prog. Biophys. Mol. Biol. 100, 47-56.

Kar, S., Baumann, W. T., Paul, M. R., Tyson, J. J., 2009. Exploring the roles of noise in the eukaryotic cell cycle. Proc. Natl. Acad. Sci. USA 106, 6471-6476.

Kauffman, S. A., 1969. Metabolic stability and epigenesis in randomly constructed genetic nets. J Theor Biol 22, 437-67.

Keener, J. P., Sneyd, J., 2009. Mathematical physiology I: Cellular Physiology. Springer.

Khadra, A., Li, Y. X., 2006. A model for the pulsatile secretion of gonadotropin-releasing hormone from synchronized hypothalamic neurons. Biophys J 91, 74-83, doi:10.1529/biophysj.105.080630.

Kim, J. K., Forger, D. B., 2012. A mechanism for robust circadian timekeeping via stoichiometric balance. Mol Syst Biol 8, 630, doi:10.1038/msb.2012.62.

Klipp, E., 2009. Systems biology : a textbook. Wiley-VCH, Weinheim.

Kraikivski, P., Chen, K. C., Laomettachit, T., Murali, T. M., Tyson, J. J., 2015. From START to FINISH: computational analysis of cell cycle control in budding yeast. npj Systems Biology and Applications 1, 15016, doi:10.1038/npjsba.2015.16.

Kuznetsov, Y. A., 1995. Elements of applied bifurcation theory. Springer Verlag, New York.

Laomettachit, T., Chen, K. C., Baumann, W. T., Tyson, J. J., 2016. A Model of Yeast Cell-Cycle Regulation Based on a Standard Component Modeling Strategy for Protein Regulatory Networks. PLoS One 11, e0153738, doi:10.1371/journal.pone.0153738.

Leloup, J., Goldbeter, A., 2003. Toward a detailed computational model for the mammalian circadian clock. Proc. Natl. Acad. Sci. USA 100, 7051-7056.

Leloup, J. C., Goldbeter, A., 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.

Levchenko, A., Iglesias, P. A., 2002. Models of eukaryotic gradient sensing: application to chemotaxis of amoebae and neutrophils. Biophys J 82, 50-63, doi:10.1016/S0006-3495(02)75373-3.

Li, F., Long, T., Lu, Y., Ouyang, Q., Tang, C., 2004. The yeast cell-cycle network is robustly designed. Proc Natl Acad Sci U S A 101, 4781-6, doi:10.1073/pnas.0305937101.

Liu, Z., Pu, Y., Li, F., Shaffer, C. A., Hoops, S., Tyson, J. J., Cao, Y., 2012. Hybrid modeling and simulation of stochastic effects on progression through the eukaryotic cell cycle. Journal of Chemical Physics 136, 034105, doi:10.1063/1.3677190.

Lotka, A. J., 1920. Undamped oscillations derived from the law of mass action. J. Am. Chem. Soc. 42, 1595.

Martiel, J. L., Goldbeter, A., 1987. A model based on receptor desensitiztion for cyclic-AMP signaling in Dictyostelium cells. Biophys. J. 52, 808-828.

McKean, H. P. J., 1970. Nagumo's Equation. Adv. Math. 4, 209-223.

Mendoza, L., Xenarios, I., 2006. A method for the generation of standardized qualitative dynamical systems of regulatory networks. Theor Biol Med Model 3, 13, doi:10.1186/1742-4682-3-13.

Molinelli, E. J., Korkut, A., Wang, W., Miller, M. L., Gauthier, N. P., Jing, X., 2013. Perturbation Biology: Inferring Signaling Networks in Cellular Systems. PLoS Comput Biol 9, e1003290.

Murray, J. D., 1989. Mathematical Biology. Springer-Verlag, Heidelberg.

Naldi, A., Carneiro, J., Chaouiya, C., Thieffry, D., 2010. Diversity and plasticity of Th cell types predicted from regulatory network modelling. PLoS Comput Biol 6, e1000912, doi:10.1371/journal.pcbi.1000912.

Nelander, S., Wang, W., Nilsson, B., She, Q. B., Pratilas, C., Rosen, N., Gennemark, P., Sander, C., 2008. Models from experiments: combinatorial drug perturbations of cancer cells. Mol Syst Biol 4, 216, doi:10.1038/msb.2008.53.

Nicolis, G., 1971. Dissipative structures in open systems far from equilibrium. Adv. Chem. Phys. 19, 209-324.

Novak, B., Tyson, J. J., 1993. Numerical analysis of a comprehensive model of M-phase control in *Xenopus* oocyte extracts and intact embryos. J. Cell Sci. 106, 1153-1168.

Novak, B., Tyson, J. J., 2008. Design principles of biochemical oscillators. Nature Rev. Mol. Cell Biol. 9, 981-991.

Novick, A., Weiner, M., 1957. Enzyme Induction as an All-or-None Phenomenon. Proc Natl Acad Sci U S A 43, 553-66.

Pedraza, J. M., Paulsson, J., 2008. Effects of molecular memory and bursting on fluctuations in gene expression. Science 319, 339-43, doi:10.1126/science.1144331.

Pratap, A., Garner, K. L., Voliotis, M., Tsaneva-Atanasova, K., McArdle, C. A., 2017. Mathematical modeling of gonadotropin-releasing hormone signaling. Mol Cell Endocrinol 449, 42-55, doi:10.1016/j.mce.2016.08.022.

Prigogine, I., Nicolis, G., 1971. Biological order, structure and instabilities. Q Rev Biophys 4, 107-48.

Qu, Z., MacLellan, W. R., Weiss, J. N., 2003. Dynamics of the cell cycle: checkpoints, sizers and timers. Biophys. J. 85, 3600-3611.

Rao, C. V., Arkin, A. P., 2003. Stochastic chemical kinetics and the quasi-steady-state assumption: application to the Gillespie algorithm. J. Chem. Phys. 118, 4999-5010.

Relogio, A., Westermark, P. O., Wallach, T., Schellenberg, K., Kramer, A., Herzel, H., 2011. Tuning the mammalian circadian clock: robust synergy of two loops. PLoS Comput Biol 7, e1002309, doi:10.1371/journal.pcbi.1002309.

Sabouri-Ghomi, M., Ciliberto, A., Kar, S., Novak, B., Tyson, J. J., 2008. Antagonism and bistability in protein interaction networks. J. Theor. Biol. 250, 209-218.

Salazar, C., Hofer, T., 2007. Versatile regulation of multisite protein phosphorylation by the order of phosphate processing and protein-protein interactions. FEBS J 274, 1046-61, doi:10.1111/j.1742-4658.2007.05653.x.

Segel, L. A., 1980. Mathematical Models in Molecular and Cellular Biology. Cambridge University Press, Cambridge, UK.

Segel, L. A., Goldbeter, A., Devreotes, P. N., Knox, B. E., 1986. A mechanism for exact sensory adaptation based on receptor modification. J. Theor. Biol. 120, 151-179.

Singhania, R., Sramkoski, R. M., Jacobberger, J. W., Tyson, J. J., 2011. A hybrid model of mammalian cell cycle regulation. PLoS Comput. Biol. 7, e1001077.

Strogatz, S. H., 1994. Nonlinear Dynamics and Chaos. Addison-Wesley Co., Reading, MA.
Sugita, M., 1963. Functional analysis of chemical systems in vivo using a logical circuit equivalent. II. The idea of a molecular automation. J Theor Biol 4, 179-92.
Sun, Z., Jin, X., Albert, R., Assmann, S. M., 2014. Multi-level modeling of light-induced stomatal opening offers new insights into its regulation by drought. PLoS Comput Biol 10, e1003930, doi:10.1371/journal.pcbi.1003930.
Swain, P. S., Elowitz, M. B., Siggia, E. D., 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc. Natl. Acad. Sci. USA 99, 12795-12800.
Tang, Y., Othmer, H. G., 1994. A G protein-based model of adaptation in Dictyostelium discoideum. Math Biosci 120, 25-76.
Thattai, M., 2016. Universal Poisson Statistics of mRNAs with Complex Decay Pathways. Biophysical Journal 110, 301-305, doi:10.1016/j.bpj.2015.12.001.
Thomas, R., 1973. Boolean formalization of genetic control circuits. J Theor Biol 42, 563-85.
Thomas, R., 1979. Kinetic logic : a Boolean approach to the analysis of complex regulatory systems : proceedings of the EMBO course "Formal analysis of genetic regulation," held in Brussels, September 6-16, 1977. Springer-Verlag, Berlin ; New York.
Tian, X. J., Zhang, H., Xing, J., 2013. Coupled reversible and irreversible bistable switches underlying TGFbeta-induced epithelial to mesenchymal transition. Biophys J 105, 1079-89, doi:10.1016/j.bpj.2013.07.011.
Tsai, T. Y. C., Choi, Y. S., Ma, W. Z., Pomerening, J. R., Tang, C., Ferrell, J. E., 2008. Robust, tunable biological oscillations from interlinked positive and negative feedback loops. Science 321, 126-129, doi:10.1126/science.1156951.
Tyson, J. J., Othmer, H. G., 1978. The dynamics of feedback control circuits in biochemical pathways. In Progress in Theoretical Biology (Academic Press, New York) 5, 2-62.
Tyson, J. J., Novak, B., 2010. Functional motifs in biochemical reaction networks. Annu. Rev. Phys. Chem. 61, 219-240.
Tyson, J. J., Hong, C. I., Thron, C. D., Novak, B., 1999. A simple model of circadian rhythms based on dimerization and proteolysis of PER and TIM. Biophys. J. 77, 2411-2417.
Vidal, A., Clement, F., 2010. A dynamical model for the control of the gonadotrophin-releasing hormone neurosecretory system. J Neuroendocrinol 22, 1251-66, doi:10.1111/j.1365-2826.2010.02055.x.
Voit, E. O., 2018. A first course in systems biology. Garland Science.
Wang, P., Song, C., Zhang, H., Wu, Z., Tian, X. J., Xing, J., 2014. Epigenetic state network approach for describing cell phenotypic transitions. Interface Focus 4, 20130068, doi:10.1098/rsfs.2013.0068.
Wang, S., Ahmadian, M., Chen, M. H., Tyson, J., Cao, Y., 2016. A Hybrid Stochastic Model of the Budding Yeast Cell Cycle Control Mechanism. Proceedings of the 7th Acm International Conference on Bioinformatics, Computational Biology, and Health Informatics, 261-270, doi:10.1145/2975167.2975194.
Winfree, A. T., 1972. Spiral waves of chemical activity. Science 175, 634-6, doi:10.1126/science.175.4022.634.
Zenklusen, D., Larson, D. R., Singer, R. H., 2008. Single-RNA counting reveals alternative modes of gene expression in yeast. Nat. Struct. Mol. Biol. 15, 1263-1271.
Zhabotinsky, A. M., 1964. Periodic process in the oxidation of malonic acid in solution: study of the kinetics of Belousov's reaction. Biofizika 9, 1306.

Table 1. Variant models of the dynamic properties of biochemical reaction networks*.

Deterministic Models

Time	Concen/Activ	Rate Laws	System of Rate Equations
Discrete	Discrete (0,1)	Discrete	Boolean: $C_i(t+1) = B_i(C_1(t), \ldots, C_N(t))$
Continuous	Continuous	Discrete	Piecewise-linear ODEs: $\frac{dC_i}{dt} = \gamma_i\left[B_i\left(\hat{C}_1(t), \ldots, \hat{C}_N(t)\right) - C_i(t)\right]$
Continuous	Continuous	Continuous	Nonlinear ODEs: $\frac{dC_i}{dt} = \sum_{j=1}^{M} \nu_{ij} R_j(C_1, \ldots, C_N)$

Stochastic Models

Time	Concen/Activ	Rate Laws	System of Rate Equations
Continuous	Continuous	Discrete	Piecewise-linear SDE: $\frac{dC_i}{dt} = \gamma_i\left[B_i\left(\hat{C}_1(t), \ldots, \hat{C}_N(t)\right) - C_i(t)\right] + \sigma_i \frac{\xi_i}{\sqrt{\Delta t}}$
Continuous	Continuous	Continuous	Nonlinear SDEs (Langevin formalism): $\frac{dC_i}{dt} = \sum_{j=1}^{M} \nu_{ij} R_j(C_1, \ldots, C_N) + \left[\sum_{j=1}^{M} \lvert\nu_{ij}\rvert R_j\right]^{1/2} \frac{\xi_i}{\sqrt{\Delta t}}$
Continuous	Discrete	Continuous	Gillespie SSA: $C_i(t + \Delta t) = C_i(t) + \nu_{ij}$ where Δt is the time interval until the next reaction and j is the index of the next reaction

*Explanatory footnotes to Table 1.

Abbreviations: ODE, ordinary differential equation; SDE, stochastic differential equation; SSA, stochastic simulation algorithm; PWL, piecewise linear.

$B_i(C_1, C_2, \ldots, C_N)$ is a Boolean function (i.e., 0 or 1) of Boolean variables.

$\hat{C}_i(t) = 0$, if $0 \leq C_i(t) \leq \theta_i$; $=1$, if $\theta_i < C_i(t) \leq 1$.

$R_j(C_1, C_2, \ldots, C_N)$ is a continuous function (rate law) of continuous variables.

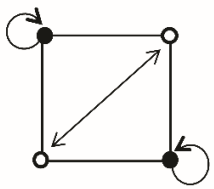
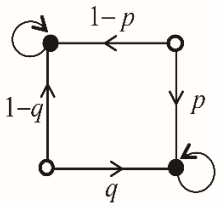
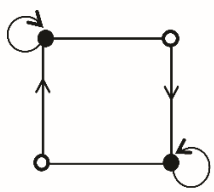
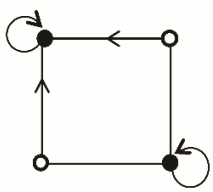
ν_{ij} is the stoichiometric coefficient of species i in reaction j.

$\gamma_i > 0$ are rate constants that govern the rate at which species i approaches its steady-state value, $C_i = B_i(\ldots)$.

ξ_i is a random number chosen from a Gaussian distribution with mean = 0 and variance = 1.

$\sigma_i > 0$ are amplitudes for the white-noise terms in the PWL-SDEs.

Table 2. Boolean Realizations of Motif 1A(i)

		A	B	C	D
	C_1 C_2	B_1 B_2	$_1$ B_2	B_1 B_2	B_1 B_2
Row 1	1 1 1 0 0 1 0 0	0 0 1 0 0 1 1 1	1 0 with prob p 0 1 with prob $1-p$ 1 0 0 1 1 0 with prob q 0 1 with prob $1-q$	1 0 1 0 0 1 0 1	0 1 1 0 0 1 0 1
Row 2	State Space				
	Asymptotic state	Probability of asymptotic states*			
Row 3	Steady state (1,0)	0.25	0.25+0.25(p+q)	0.5	0.25
	Steady state (0,1)	0.25	0.75–0.25(p+q)	0.5	0.75
	Oscillation (00 ↔11)	0.5	0	0	0

* The four different versions of motif 1A(i) make different predictions about the probability that the system will end up in a particular asymptotic state given that the system is started from an ensemble of initial conditions uniformly distributed over the four possible states of the system {(0,0), (1,0), (0,1), (1,1)}.

Table 3. Parameter values for ODE models of motifs in Fig. 1A.

Fig. # Motif #	γ_1 γ_2	σ_1 σ_2	ω_{10} ω_{20}	ω_{11} ω_{21}	ω_{12} ω_{22}
2A 1A(i):A	1 1	- -	0.4 0.6	0 −1	−1 0
2B 1A(i):C	1 2	- -	−0.5 0.5	2 −1	−1 0.1
2C 1A(i):A	1 1	15 15	0.4 0.6	0 −1	−1 0
2D 1A(i):C	1 2	8 8	−0.5 0.5	2 −1	−1 0.1
3A, C 1A(ii)	1 1	5 5	$-0.9+1.4S$ $-0.9+1.4S$	1.4 −0.6	−0.6 1.4
3B, D 1A(iii)	1 0.2	10 20	$0.4+0.1S$ 0.3	1.1 1	−0.5 0

Table 4. Common types of bifurcation points

Name	**Defining characteristics**
Saddle-node (SN)	The coalescence of a stable node and a saddle point, leading to the annihilation of both steady states, forcing the system to evolve to some other attractor, e.g., a different stable node. Examples: Fig. 3C at $S\approx0.04$ and $S\approx0.67$; Fig. 3D at $S\approx1.30$ and $S\approx3.64$ (where the saddle point coalesces with an unstable node). Close to an SN point, the speed at which the dynamical system departs from the vicinity of the recently annihilated steady states is very slow; an experimentally noticeable trait called 'critical slowing down'.
Pitchfork bifurcation (PF)	The point where a steady-state solution exchanges stability with a pair of alternative steady-state solutions. Close to PF point the loci of the three steady states has the appearance of a 'pitchfork'. Examples: Fig. 3C at $S\approx0.28$ and $S\approx0.43$. PF bifurcations are indicative of 'symmetries' in the dynamical system's governing equations. If the symmetry is broken, the pitchfork splits into a continuous branch of steady states and a SN bifurcation connecting the two other 'tines' of the pitchfork.
Hopf bifurcation (HB)	The point where a steady state changes from slowly damped oscillations to oscillations of increasing amplitude that are 'captured' by a periodic limit cycle oscillation. Close to an HB point, the amplitude of the periodic solutions is small, and the frequency of the oscillations is nearly constant. Example: Fig. 3D at $S\approx0.825$ (in this case the amplitude does indeed start off at zero but rapidly increases as S increases above 0.825).
Saddle-node on an invariant circle (SNIC)	The coalescence of a stable node and a saddle point that are connected by a trajectory (an 'invariant circle') that proceeds out of the saddle point, makes a long loop through the state space and comes back to the saddle-node along one of its attracting directions. Example: Fig. 3D at $S\approx1.30$ (where the large amplitude limit cycle oscillations end precisely at the SN bifurcation point). Close to a SNIC bifurcation, the amplitude of the limit cycle oscillations is large and nearly constant, and the frequency of the oscillations is small (i.e., the period approaches infinity). Hence, the 'signatures' of limit cycle oscillations arising from HB's and SNIC's are precisely opposite to each other.

Table 5. Summary of modeling strategies: advantages/disadvantages

Method	Variables	Time	Advantages	Disadvantages
Boolean networks	0 or 1	$t = 0,1,2,\ldots$	Easy to implement. Easy to understand.	No quantitative details.
ODEs	Real positive numbers	t = real number	Quantitative details. Predictive potential. Powerful tools for simulation & analysis.	Rate constants must be estimated from quantitative experimental data.
Stochastic	Positive integers	t = real number	Effects of molecular noise in single cells. SSA is an accurate, reliable simulation algorithm.	Requires even more quantitative experimental data. Computationally intensive.
Hybrid Det/Stoch	Real positive numbers	t = real number	Computationally efficient and sufficiently accurate.	Not very 'intuitive'. Partitioning must be done correctly.

Figure Legends

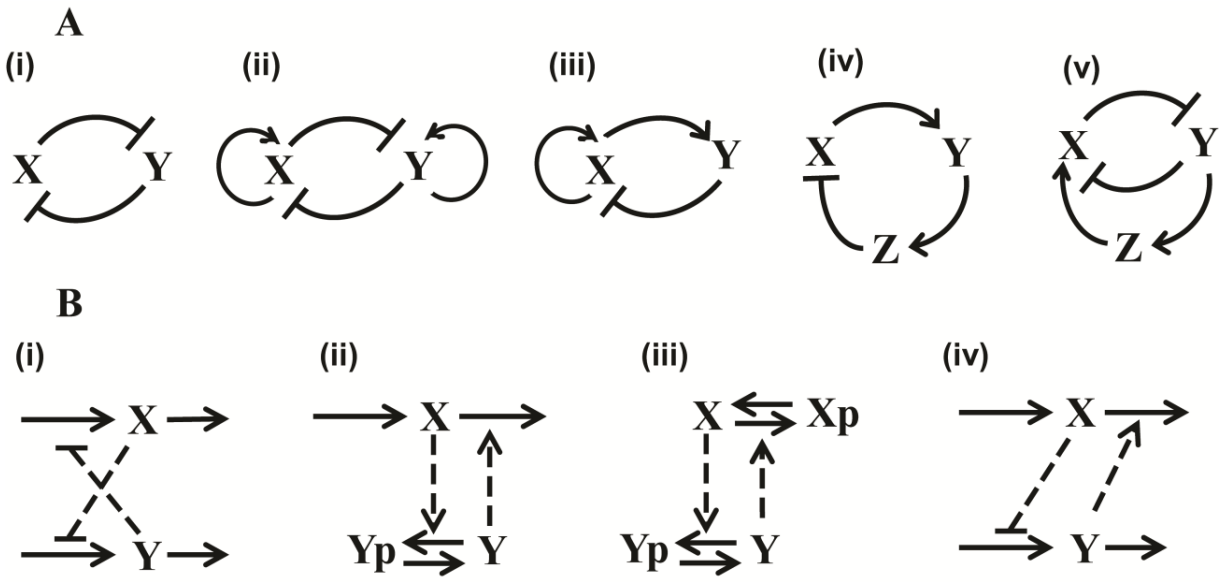


Figure 1. Regulatory network diagrams. (**A**) Five representative topological motifs. X, Y and Z are proteins. A barbed connector represents the 'activation' of one protein by another (perhaps by itself). A blunt connector represents 'inhibition'. (**B**) Four possible biochemical reaction mechanisms that implement motif (i) in panel A. In the first case, X and Y inhibit each other's production. In the second case, Y is a protease that degrades X, whereas X is a kinase that phosphorylates and inactivates Y. In the third case, two kinases phosphorylate and inactive each other. In the fourth case, X inhibits the production of Y and Y activates the degradation of X.

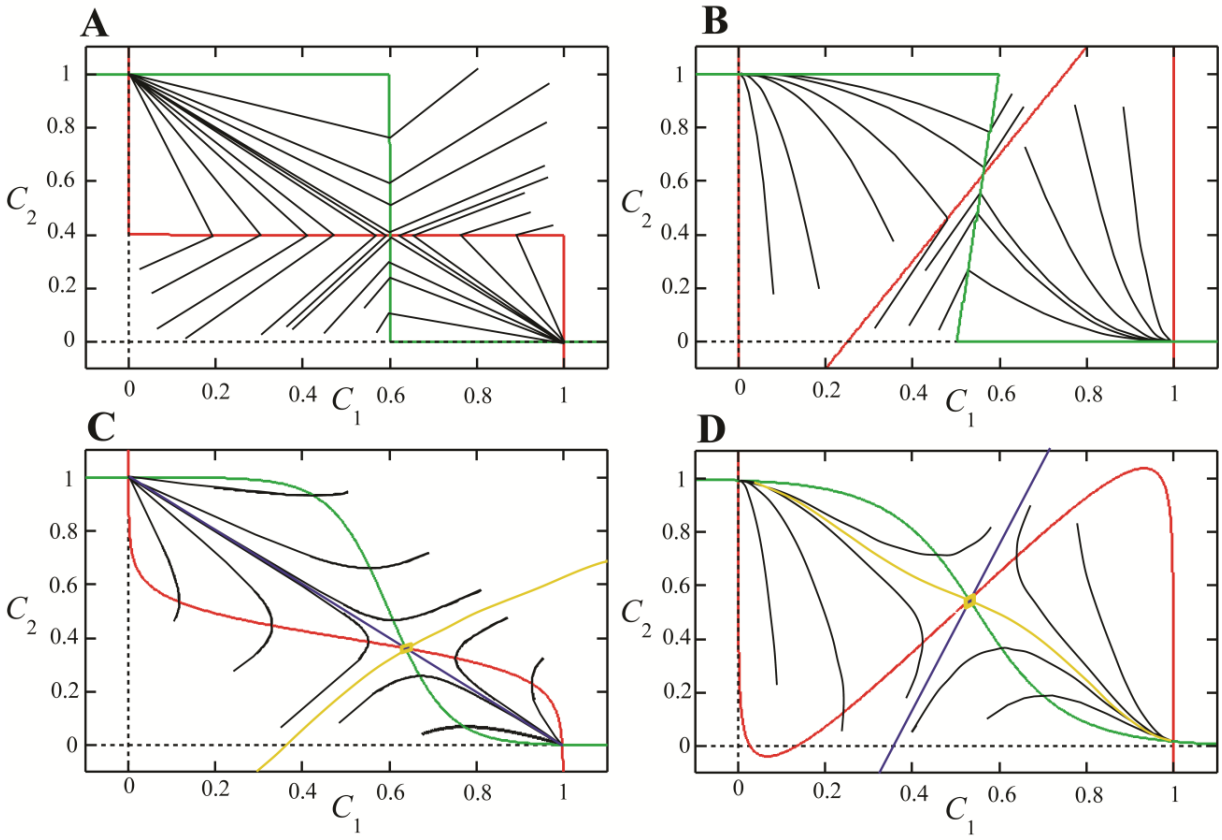


Figure 2. Phase-plane portraits of piece-wise linear (PWL) and smooth, nonlinear (soft-H) ODE models of motifs discussed in the text. (**A**) PWL model of Motif 1A(i)-Case A, with $\gamma_1 = \gamma_2$. (**B**) PWL model of Motif 1A(i)-Case C, with $\gamma_2 = 2\gamma_1$. (**C**) Soft-H model of Motif 1A(i)-Case A, with $\gamma_1 = \gamma_2$. (**D**) Soft-H model of Motif 1A(i)-Case C, with $\gamma_2 = 2\gamma_1$. Parameter values are given in Table 3. In all four panels, the black lines are 'trajectories' starting from a variety of initial locations in state space and proceeding to one or the other of two stable steady states, at C_1=1, C_2=0 or at C_1=0, C_2=1. In cases C and D, the red curve is the C_1-nullcline (where $dC_1/dt = 0$) and the green curve is the C_2-nullcline (where $dC_2/dt = 0$). The red and green curves intersect at the third steady state, an unstable saddle point, in the interior of the unit square. The curve passing through the saddle point (yellow in panel C) separates the 'domains of attraction' of the two stable steady states.

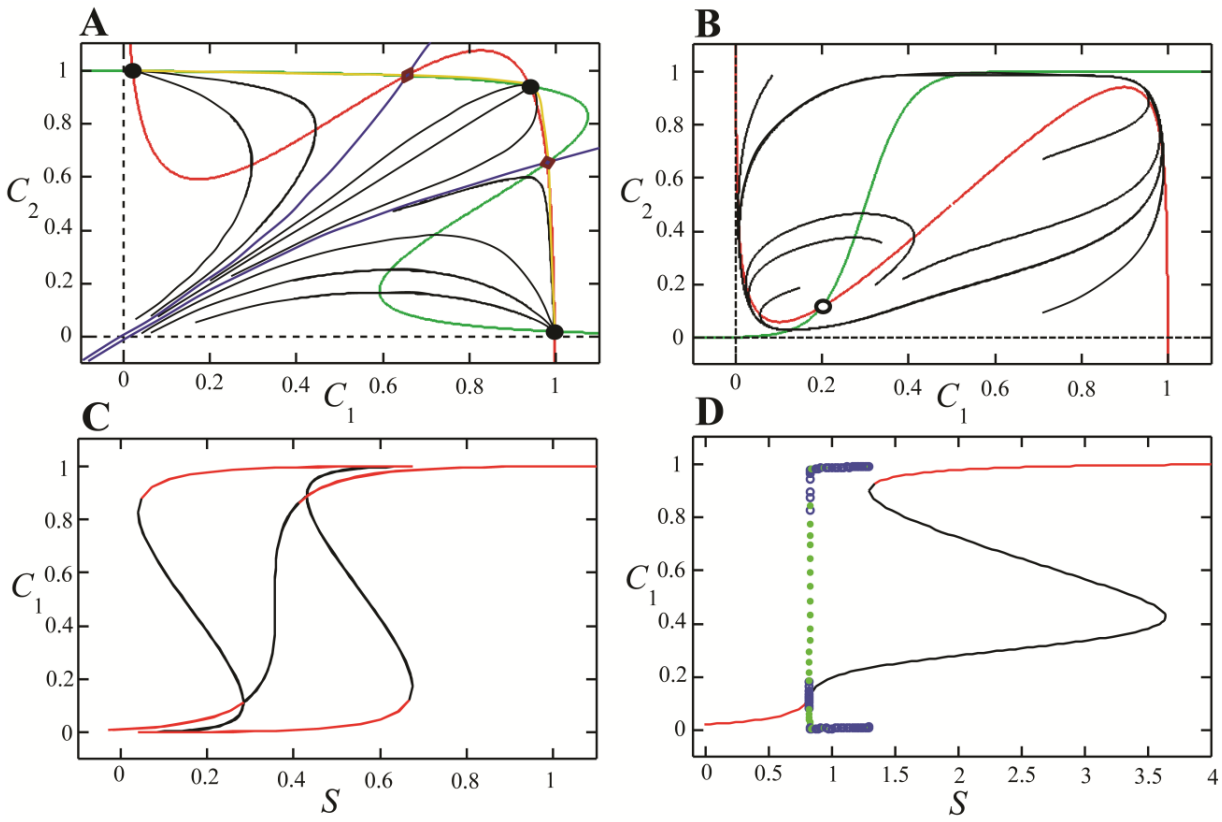


Figure 3. Phase-plane portraits (**A, B**) and one-parameter bifurcation diagrams (**C, D**) of soft-H ODE models of Motifs 1A(ii) (panels A and C) and 1A(iii) (panels B and D). Parameter values are given in Table 3; (A) $S = 0.5$, (B) $S = 1$, (C and D) S is the bifurcation parameter. In panel C we plot the steady state value of C_1 as a function of S. At $S = 0$, there is a single steady state (the 'naïve' state) at $C_1 = C_2 \approx 0$. At $S = 0.04$, the system undergoes a pair of saddle-node bifurcations that create two new, 'differentiated' states (one at $C_1 \approx 0.82$, $C_2 \approx 0$ and the other at $C_1 \approx 0$, $C_2 \approx 0.82$) in addition to the naïve state. At $S = 0.28$, the naïve state is lost at a pitchfork bifurcation. The two differentiated states persist for $0.04 < S < 0.67$; however, at $S = 0.43$, a second pitchfork bifurcation creates a new, stable steady state (called a 'dual expressing' state) for which $C_1 = C_2 > 0.88$. In panel D the system executes stable limit cycle oscillations for signal strengths between the Hopf bifurcation at $S = 0.825$ and a saddle-node-on-an-invariant-circle (SNIC) bifurcation at $S = 1.30$. There is a second saddle-node bifurcation at $S = 3.64$, but the node is unstable so the bifurcation is unobservable in experiments. (For a brief explanation of the terminology of bifurcation points, see Table 4.)

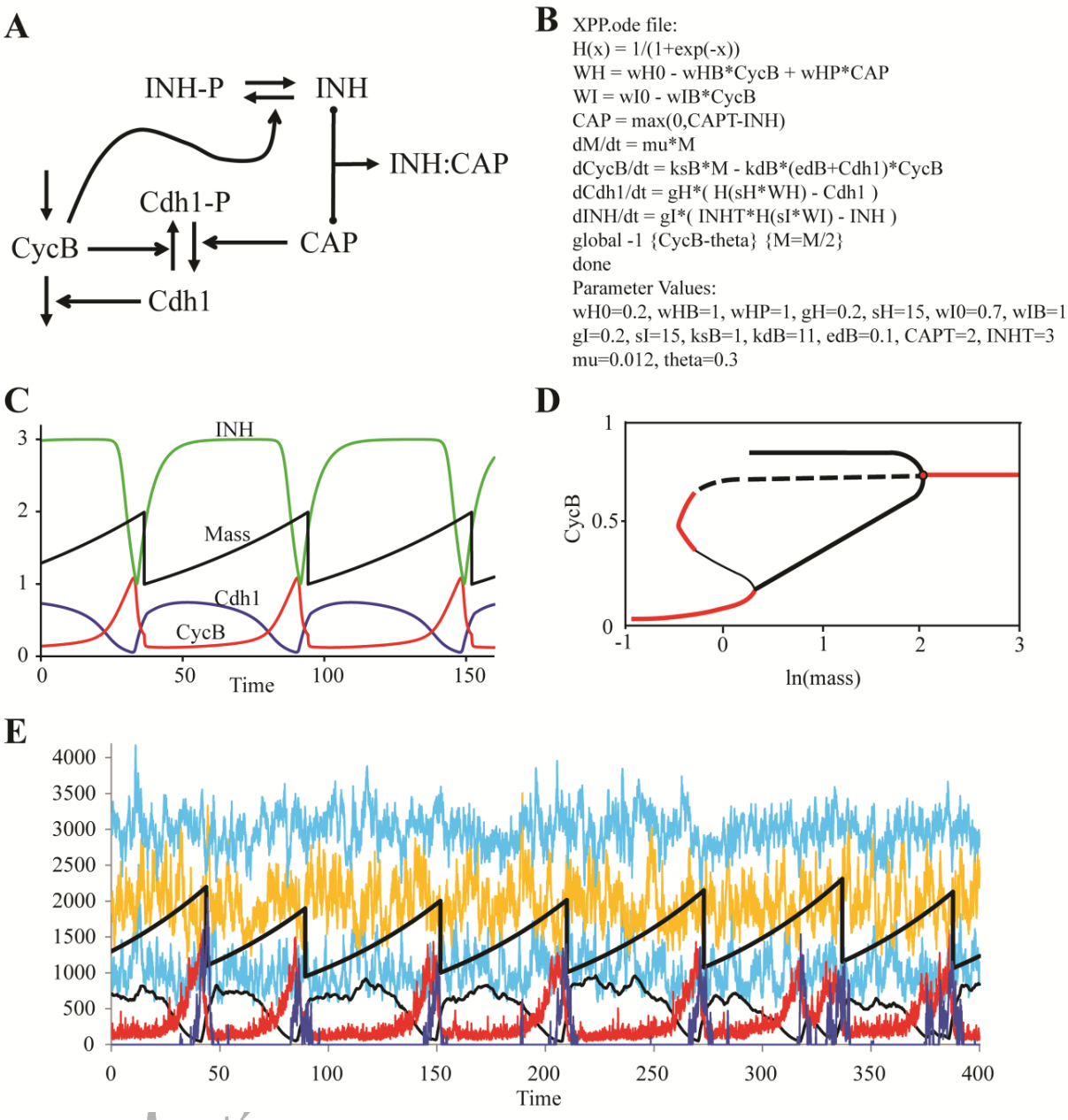


Figure 4. Standard component model of mitotic control in a 'generic' eukaryotic cell. (**A**) Generic model. CycB = cyclin B-dependent kinase, which drives cell into mitosis; Cdh1 = active form of the cyclosome, which degrades cyclin B and drives cell out of mitosis; CAP = counter-acting phosphatase, which activates Cdh1 during exit from mitosis and entry into G1 phase; INH = a stoichiometric inhibitor of CAP. (**B**) XPP-ode file for an SCM implementation of the generic model in panel A. CycB is governed by an ODE like Eq. (12); Cdh1 and INH are governed by ODEs like Eq. (10); and CAP is given by a tight-binding approximation like Eq. (17). Progression through the cell cycle (G1-S-G2-M) is driven by growth. The variable M(t) is cell

mass, which increases exponentially at a specific growth rate 'mu'. The mass-doubling time for a cell is 0.693/mu. The cell divides (M → M/2) when CycB level drops below a threshold, theta. (**C**) Simulation of the SCM for the parameter values given in panel B. The simulated cell undergoes periodic divisions every 58 time units, which is exactly the mass-doubling time. The simulated cell persists in G1 phase for ~38 time units, and then proceeds through S-G2-M in ~20 time units. For a mammalian cell, we might take 1 time unit ≈ 20 min; for a yeast cell, we might take 1 time unit ≈ 2 min. (**D**) One-parameter bifurcation diagram (schematic) for the SCM, treating 'mass' as a parameter instead of a variable. Saddle-node bifurcations at $M = 0.6055$ and 1.3355 bound a region of bistability. Limit cycle oscillations are evident between a saddle-loop bifurcation at $M \approx 1.3$ and a Hopf bifurcation at $M \approx 7.4$. (**E**) Stochastic simulation of the model studied by a deterministic SCM in panel C. The turquoise curves show stochastic fluctuations in $Cdh1_{total}$ and INH_{total}; the orange curve, CAPT. The red curve is CycB abundance. The black curve shows the fluctuating activity of Cdh1 (high in G1 phase, low in S-G2-M), and the blue curve shows how CAP is released from binding to INH as the cell exits mitosis. The saw-tooth black curve is 1000×mass. Notice that cell size at division and cell cycle time are stochastic properties of cell cycle progression in this model.

Appendix: XPP.ode file for stochastic simulation of SCM for CycB-Cdh1-INH-CAP model.

```
# SCM example of stochastic simulation of cell cycle progression.

# First, compute fluctuating levels of total Cdh, INH, and CAP.
wiener wienH
wiener wienI
wiener wienP
sigmaH = CdhT*sqrt((2*kdH/mRNAH)*(kdH/(kdH+kdmH)))
sigmaI = INHT*sqrt((2*kdI/mRNAI)*(kdI/(kdI+kdmI)))
sigmaP = CAPT*sqrt((2*kdP/mRNAP)*(kdP/(kdP+kdmP)))
dCdhT/dt = ksH*mRNAH - kdh*CdhT + sigmaH*wienH
dINHT/dt = ksI*mRNAI - kdI*INHT + sigmaH*wienI
dCAPT/dt = ksP*mRNAP - kdP*CAPT + SigmaP*wienP

# Then solve the SCM of CycB, Cdh, INH and CAP.
wiener wienB
FB = ksb*mRNAB*M
GB = kdB*(edB+Cdh/1000)
sigmaB = CycB*sqrt((2*GB/mRNAB)*(GB/(GB+kdmB)))
dCycB/dt = FB - GB*CycB + sigmaB*wienB
softH(x) = 1/(1+exp(-x))
WH = wH0 - wHB*CycB/1000 + wHP*CAP/1000
WI = wI0 - wIB*CycB/1000
dCdh/dt = gH*( CdhT*softH(sH*WH) - Cdh )
dINH/dt = gI*( INHT*softH(sI*WI) - INH )
CAP = max(0,CAPT-INH)

# Cell grows exponentially and divides when Cdh activity rises above a
# threshold as cell exits mitosis.
dM/dt = mu*M
global +1 {Cdh-theta} {M=M/2}

# Define auxiliary variables for plotting.
aux CAP=CAP
aux mass=1000*M

# Set parameter values.
par ksH=250, kdH=1, ksI=250, kdI=1, ksP=250, kdP=1
par mRNAH=4, kdmH=4, mRNAI=12, kdmI=4, mRNAP=8, kdmP=4
par gH=.2, sH=15, gI=.2, sI=15
par wH0=.2, wHB=1, wHP=1
par wI0=.7, wIB=1
par ksb=167, kdb=11, edB=.1, mRNAB=6, kdmB=10
par theta=400, mu=0.012

# Set initial conditions.
init CdhT=1000, INHT=3000, CAPT=2000, CDH=700, INH=3000
init CycB=100, M=1.3

# Euler's method is used to solve the Langevin equations.
@ meth=euler, dt=.01
done
```