

Mathematical modeling as a tool for investigating cell cycle control networks

Jill C. Sible ^{*}, John J. Tyson

Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0406, USA

Accepted 11 July 2006

Abstract

Although not a traditional experimental “method,” mathematical modeling can provide a powerful approach for investigating complex cell signaling networks, such as those that regulate the eukaryotic cell division cycle. We describe here one modeling approach based on expressing the rates of biochemical reactions in terms of nonlinear ordinary differential equations. We discuss the steps and challenges in assigning numerical values to model parameters and the importance of experimental testing of a mathematical model. We illustrate this approach throughout with the simple and well-characterized example of mitotic cell cycles in frog egg extracts. To facilitate new modeling efforts, we describe several publicly available modeling environments, each with a collection of integrated programs for mathematical modeling. This review is intended to justify the place of mathematical modeling as a standard method for studying molecular regulatory networks and to guide the non-expert to initiate modeling projects in order to gain a systems-level perspective for complex control systems.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Cell cycle; Computational biology; Frog egg extracts; Mathematical modeling; Ordinary differential equations; Parameter estimation; Systems biology; *Xenopus laevis*

1. Introduction

Because a universal mechanism controlling DNA synthesis, mitosis, and cell division underlies the growth, development, and reproduction of all eukaryotes, an understanding of this molecular regulatory system is one of the most important goals of modern cell biology. As the complex network of cell cycle controls is uncovered, it becomes increasingly difficult to make reliable predictions about how modification of one component affects the system as a whole. However, such predictions are needed if we are to identify the host of mutations contributing to cancer or find within the molecular network novel targets for therapeutic intervention. Mathematical models provide powerful tools for managing the complexity of the cell cycle control system and of other signaling

networks. Models organize a large body of experimental data, describe the fundamental behaviors of the system as a whole, bridge gaps where experimental data are missing, and drive hypothesis-building for the next round of experimentation. The value of mathematical modeling in describing and predicting the behavior of complex systems has been well established in fields such as chemical engineering and meteorology, but its power has been underappreciated until recently in molecular cell biology.

Although mathematical models can be built to describe any signaling network, application of modeling tools to cell cycle regulation is particularly well suited and timely. First, the data in this field are vast, both providing a large body of information to build comprehensive models and creating the need for a tool to understand how these data fit together. Second, cell cycle signaling networks are modular, allowing models to be constructed in parts and then assembled and reassembled in various ways. Furthermore, many models of the network are comparable between different

^{*} Corresponding author. Fax: +1 540 231 9307.

E-mail address: siblej@vt.edu (J.C. Sible).

organisms (e.g., budding yeast and mammals) so that it is feasible to make relatively small changes to a model describing one particular system in order to apply it to another. Thus, each new model need not be constructed from scratch. Third, a reasonable amount of quantitative or semi-quantitative information can be extracted from the literature, facilitating the early phases of model building. By modifying established protocols (described in the accompanying articles), additional quantitative data can be generated to improve parameter estimation and experimental validation of models. Finally, despite the wealth of detailed information on cell cycle molecules and their specific interactions, we lack a systems-level perspective of this complex control network. Modeling can provide this perspective by helping to identify underlying regulatory principles. Where a specific experimental detail is missing, modeling can serve as bridge, enabling progress in building a systems-level view, and guiding the design and execution of future experiments.

Although the term “mathematical modeling” encompasses a wide range of computational approaches applicable to cell cycle studies, we focus this review on one branch of modeling: The construction of ordinary differential equations (ODEs) to describe protein interaction networks that regulate the cell cycle. This approach has a strong track record of yielding accurate, predictive and testable models [1–7], and in recent years, has been integrated well with experimental methodologies [8–12]. We will illustrate each step of the modeling process with an example: The core signaling network that controls entry into and exit from mitosis in frog (*Xenopus laevis*) egg extracts by regulating the activity of the mitotic cyclin-dependent kinase, Cdk1. We chose this example because the network is rela-

tively simple, many predictions of the model have been validated experimentally, and this module can be adapted to more complex cell cycles both in *Xenopus* and other eukaryotes.

2. Approach

2.1. Construction of a wiring diagram

The first step in the modeling process is to organize the known interactions of the relevant molecules into a concept map or wiring diagram. A simple diagram representing the core machinery regulating mitotic transitions in *Xenopus* cell-free egg extracts is depicted in Fig. 1. The diagram depicts each biochemical entity with a separate icon. For example, the phosphorylated form of Wee1 is represented by a modified version of the icon for unphosphorylated Wee1. Solid arrows indicate chemical transitions between states, and dotted arrows represent a modulating signal on a biochemical reaction (often the catalytic influence of an enzyme). Because no universal conventions for building these diagrams exist, all symbols must be defined explicitly and used consistently. Hopefully, contributors to this field will adopt a common symbolism in the near future, facilitating the exchange of information between wiring diagrams and their affiliated mathematical models.

Despite the wealth of information regarding cell cycle signaling networks, gaps in knowledge may exist such that portions of the diagram cannot be wired with certainty. It is neither necessary nor advised to postpone the modeling process indefinitely until all molecular details of the wiring diagram are understood. Rather, the gaps in the wiring diagram can be filled with place-

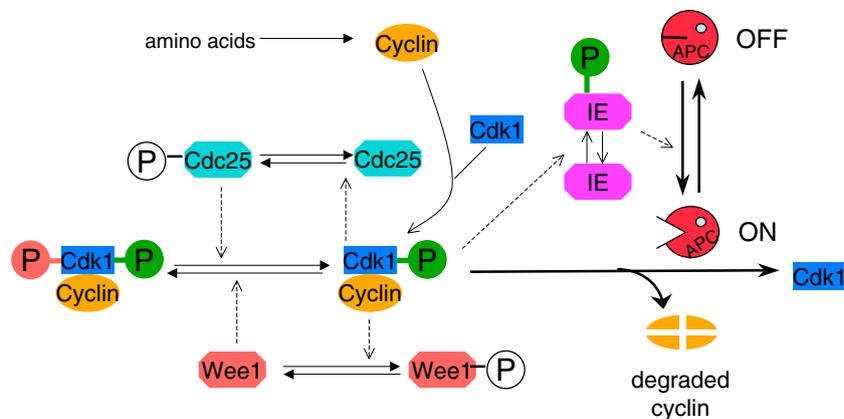


Fig. 1. Wiring diagram of the core mitotic cell cycle engine in *Xenopus* cell-free egg extracts. Central to the diagram is the active cyclin–Cdk1 complex (also referred to as MPF), with activating phosphorylation on Thr 161 indicated by the green ‘P’. Cyclin enters the system by de novo synthesis and then combines with Cdk1. The phosphorylation on Thr 161 is rapid and therefore not represented in the diagram or the mathematical equations. Active Cdk1 can be phosphorylated on Thr 14 and Tyr 15 to form inactive preMPF. The double arrows indicate the reversibility of this process. The inhibitory phosphates are represented by the red ‘P’; and the influences of the relevant kinase (Wee1) and phosphatase (Cdc25C) are indicated by the dotted arrows. MPF itself phosphorylates Cdc25 and Wee1, positively and negatively affecting their activities, respectively. These feedback loops are easily appreciated from the wiring diagram. MPF also participates in a negative feedback loop whereby it phosphorylates a component of the APC, which subsequently directs the polyubiquitination of cyclin, tagging it for degradation by the proteasome. Like cyclin synthesis, cyclin degradation is represented by a single unidirectional arrow because the process is irreversible. Although not indicated in the diagram, cyclin degradation by the APC applies equally to free cyclin monomers and to preMPF. IE indicates an intermediate component (see text).

holders. These placeholders can be generic ones for which parameters will be selected empirically to model the behavior of the system without representing the molecular mechanism in detail. This approach allows the modeling effort to continue in order to discover properties of the system as a whole as well as specific features of better-characterized portions of the diagram. Alternatively, a more specific placeholder can be created that makes some untested assumptions about the molecular mechanism. These assumptions in the wiring diagram will be tested in simulations of the model to determine whether they are consistent with known behaviors of the system. Eventually, the assumptions that work *in silico* must be tested experimentally when appropriate methodologies and reagents become available. Ideally, portions of the wiring diagram that are hypothetical should be clearly delineated from those that are grounded in experimentation.

Our example, a model of the cell cycle machinery regulating mitosis in frog egg extracts, demonstrates the usefulness of placeholders. When the wiring diagram depicted in Fig. 1 was first constructed [2], little information was known about the mechanisms regulating exit from mitosis. Experimental data suggested that MPF triggered the degradation of cyclin [13] and a previous mathematical model indicated that a time delay in this negative feedback loop could create sustained oscillations [3]. Based on this information, the wiring diagram was constructed with an intermediate component (called IE) that was phosphorylated by MPF, and then activated the APC, triggering the polyubiquitination and subsequent degradation of cyclin [2]. Several years later as the details of the ubiquitin–proteasome pathway governing cyclin degradation were discovered, IE was replaced by Fizzy/Cdc20 and components of the APC [14–16] but the fundamental features of the model remained the same.

2.2. Writing the ordinary differential equations

Once the model is represented diagrammatically, the biochemical relationships need to be converted to the language of mathematics in the form of nonlinear ODEs. A rate equation is written for each biochemical entity whose concentration changes over time. The right side of the equation includes a positive term for every arrow that directs toward that entity and a negative term for every arrow that directs away from the entity. The kinetics selected should be appropriate for the type of biochemical transition represented. In general, mass-action kinetics are appropriate for most biochemical reactions except when $[\text{substrate}] \gg [\text{enzyme}]$, in which case Michaelis–Menten kinetics should be applied.

In our example of the core mitotic machinery, the rate equation for the concentration of cyclin monomers is based on mass-action kinetics:

$$\frac{d}{dt}[\text{Cyclin}] = k_1 - k_2[\text{Cyclin}] - k_3[\text{Cyclin}][\text{Cdk}]$$

where k_1 is the rate constant for cyclin synthesis, k_2 is a function that describes the activity of the APC in promoting cyclin degradation, and k_3 is the rate constant for association of cyclin monomers and Cdk monomers to form MPF dimers. These dimers are phosphorylated on threonine 161 to make active MPF. Because this reaction is fast and essentially irreversible, we are justified in neglecting the T161-unphosphorylated forms of MPF.

The rate equation for [MPF] is likewise based on mass-action kinetics:

$$\begin{aligned} \frac{d}{dt}[\text{MPF}] &= k_3[\text{Cyclin}][\text{Cdk}] - k_2[\text{MPF}] \\ &\quad - k_{\text{wee}}[\text{MPF}] + k_{25}[\text{preMPF}] \end{aligned}$$

A similar equation pertains to [preMPF]. [Cdk], the concentration of Cdk monomers, is governed not by an ODE but rather by a conservation condition:

$$[\text{Total Cdk}] = [\text{Cdk}] + [\text{MPF}] + [\text{preMPF}] = \text{constant.}$$

We assume that [Total Cdk] = 100 nM in a typical frog egg extract.

The full set of differential equations for the model of M-phase control is shown in Fig. 2. This early version of the model uses Michaelis–Menten kinetics to describe the rate changes of the regulatory enzymes, Cdc25 and Wee1 [2]. Because of the positive feedback between Cdc25 and MPF and between Wee1 and MPF, the assumption that $[\text{enzyme}] \ll [\text{substrate}]$ may be invalid. In more recent iterations of the model, these equations have been rewritten with mass-action kinetics, and parameter sets that reproduce the fundamental behaviors of the system have been identified (Dravid and Tyson, unpublished data).

2.3. Parameterizing the model

The next step in the modeling process is the most challenging. A preliminary set of parameters (rate constants for each reaction) must be selected. The challenges arise because for many models, few if any of these parameters have been measured directly. Therefore, the initial set of parameters must be selected based on semi-quantitative data, indirect measurements and typically, a fair amount of guesswork. In an iterative process, each set of parameters must be “run” through the model (described below) and modified to bring the output of the model into better and better agreement with observed behaviors of the system.

Even though few direct measurements of rate constants or concentrations of cell-cycle regulatory proteins have been published, valuable information for parameter estimation can be mined from the literature, and a thorough effort to do so is recommended before one undertakes the task of measuring parameters *de novo*. A reasonable body of experimental literature has informed the model depicted in Figs. 1 and 2, although the initial parameter set was almost purely guesswork, based on rough qualitative features of mitotic control in frog eggs and extracts [2].

1. $\frac{d}{dt} [\text{Cyclin}] = k_1 - k_2[\text{Cyclin}] - k_3 [\text{Cyclin}] [\text{Cdk}]$
2. $\frac{d}{dt} [\text{MPF}] = k_3 [\text{Cyclin}] [\text{Cdk}] - k_2 [\text{MPF}] - k_{\text{wee}}[\text{MPF}] + k_{25} [\text{preMPF}]$
3. $\frac{d}{dt} [\text{preMPF}] = -k_2 [\text{preMPF}] + k_{\text{wee}}[\text{MPF}] - k_{25} [\text{preMPF}]$
4. $\frac{d}{dt} [\text{Cdc25P}] = \frac{k_a[\text{MPF}](\text{[total Cdc25]} - [\text{Cdc25P}])}{K_a + \text{[total Cdc25]} - [\text{Cdc25P}]} - \frac{k_b[\text{PPase}][\text{Cdc25P}]}{K_b + [\text{Cdc25P}]}$
5. $\frac{d}{dt} [\text{Wee1P}] = \frac{k_e[\text{MPF}](\text{[total Wee1]} - [\text{Wee1P}])}{K_e + \text{[total Wee1]} - [\text{Wee1P}]} - \frac{k_f[\text{PPase}][\text{Wee1P}]}{K_f + [\text{Wee1P}]}$
6. $\frac{d}{dt} [\text{IEP}] = \frac{k_g[\text{MPF}](\text{[total IE]} - [\text{IEP}])}{K_g + \text{[total IE]} - [\text{IEP}]} - \frac{k_h[\text{PPase}][\text{IEP}]}{K_h + [\text{IEP}]}$
7. $\frac{d}{dt} [\text{APC}] = \frac{k_c[\text{IEP}](\text{[total APC]} - [\text{APC}])}{K_c + \text{[total APC]} - [\text{APC}]} - \frac{k_d[\text{PPase}][\text{APC}]}{K_d + [\text{APC}]}$
8. $[\text{Cdk}] = [\text{Total Cdk}] - [\text{MPF}] - [\text{preMPF}]$
9. $k_{25} = V_{25}' ([\text{Total Cdc25}] - [\text{Cdc25P}]) + V_{25}'' [\text{Cdc25P}]$
10. $k_{\text{wee}} = V_{\text{wee}}' [\text{Wee1P}] + V_{\text{wee}}'' ([\text{Total Wee1}] - [\text{Wee1P}])$
11. $k_2 = V_2' ([\text{Total APC}] - [\text{APC}]) + V_2'' [\text{APC}]$

Fig. 2. Mathematical model of the cell cycle control network in frog egg extracts [2]. ODEs describe the rate of change in concentration of every non-constant species in the wiring diagram in Fig. 1.

Although enzyme rate constants are rarely measured or reported directly, reasonable estimates can be derived from time course studies, which abound in the cell cycle literature. For example, the experimental data of Kumagai and Dunphy [17], reproduced in Fig. 3, gives us hints about the rates of phosphorylation of cyclin-Cdk1 by Wee1 in its more active form (V''_{wee}) and its less active form (V'_{wee}). Lanes d–h and i–m show that Cdk1 is rapidly phosphorylated by Wee1 in an interphase extract, when Wee1 is expected to be in its active form. That rapid activation is observed even in the presence of aphidicolin (“+APC;”

not to be confused with the Anaphase Promoting Complex), an inhibitor of DNA replication. Because Cdk1 appears to be completely converted to the tyrosine phosphorylated form within 2 minutes, V''_{wee} must be 1 min^{-1} or larger. On the other hand, in an M-phase arrested extract (lanes a–c), tyrosine phosphorylation of Cdk1 by less active Wee1 is much slower, say $V'_{\text{wee}} \approx 0.01 \text{ min}^{-1}$. Kumagai and Dunphy (1995) [17] also show that the rate of tyrosine phosphorylation is 2.5 times slower when cycloheximide-treated extracts were supplemented with cyclin B monomers rather than with preformed cyclin B-Cdk1 dimers (their Fig. 3A and C), allowing Marlovits et al. (1998) [43] to estimate that $k_3 \approx 0.005 \text{ nM}^{-1} \text{ min}^{-1}$. Although only rough estimates, these values for V'_{wee} , V''_{wee} , and k_3 provide good starting points for more sophisticated parameter optimization procedures. Table 1 provides the ODE files with the Marlovits’ estimates of the rate constants for the ODEs in Fig. 2.

At present, mining the literature to inform parameter estimation remains a laborious task. The body of literature is enormous and even the most carefully constructed database search algorithms will not identify all of the articles that might contain experiments relevant for parameter

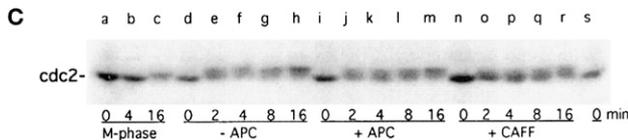


Fig. 3. Experimental data used to estimate the value of V''_{wee} in the Novak-Tyson model of the cell cycle in frog egg extracts. Data are from Fig. 3C of Kumagai and Dunphy, 1995 (17). “Cdc2” refers to Cdk1 (lower band is unphosphorylated on Wee1 sites and upper band is phosphorylated on Wee1 sites). APC = aphidicolin, an inhibitor of DNA replication and CAFF = caffeine, an inhibitor of cell cycle checkpoint kinases.

Table 1

Input file for WinPP to reproduce Fig. 4

```

# A model of MPF regulation in frog egg extracts
# after (Novak & Tyson, J. Cell Sci., 1993)

# Warning: WinPP and XPP are case insensitive.

# Differential Equations (Figure 2)

dCyclin/dt = k1- k2*Cyclin - k3*Cyclin*Cdk
dMPF/dt = k3*cyclin*Cdk- k2*MPF - kwee*MPF + k25*preMPF
dpreMPF/dt = kwee*MPF- k25*preMPF - k2*preMPF
dCdc25P/dt = F(TotCdc25-Cdc25P,MPF,ka,KKa) - F(Cdc25P,PPase,kb,KKb)
dWee1P/dt = F(TotWee1-Wee1P,MPF,ke,KKe) - F(Wee1P,PPase,kf,KKf)
dIEP/dt = F(TotIE-IEP,MPF,kg,KKg) - F(IEP,PPase,kh,KKh)
dAPC/dt = F(TotAPC-APC,IEP,kc,KKc) - F(APC,PPase,kd,KKd)

# Conservation of Cdk subunits

Cdk = TotCdk - MPF - preMPF

# Rate functions embodying the regulatory signals

k25 = V25*(TotCdc25-Cdc25P) + V25*Cdc25P
kwee = Vwee*Wee1P + Vwee*(TotWee1-Wee1P)
k2 = V2*(TotAPC-APC) + V2*APC

# Michaelis-Menten Rate Law

F(S,E,k,Km) = k*E*S/(Km+S)

# Parameter values (from Marlovits et al. [43])

par k1=1, V2'=.005, V2"=.25, k3=.005
par V25'=.017, V25"=.17, Vwee'=.01, Vwee"=1
par ka=.02, KKa=.1, kb=.1, KKb=1
par kc=.13, KKc=.01, kd=.13, KKd=1
par ke=.02, KKe=.1, kf=.1, KKf=1
par kg=.02, KKg=.01, kh=.15, KKh=.01
par TotCdk=100

# These parameters are arbitrarily set to 1.

par TotCdc25=1, TotWee1=1, TotIE=1, TotAPC=1, PPase=1

# Initial conditions

init Cyclin=0, MPF=0, preMPF=0, Cdc25P=0, Wee1P=0, IEP=1, APC=1

# Define "auxiliary" variables for plotting purposes

aux TotCyclin=cyclin+MPF+preMPF
aux Wee1=10*(TotWee1-Wee1P)

# Simulator settings

@ meth=stiff, Total=200, Bounds=1000

# Plotter settings

@ xhi=200, ylo=0, yhi=25
@ nplot=2, yp2=MPF

done

```

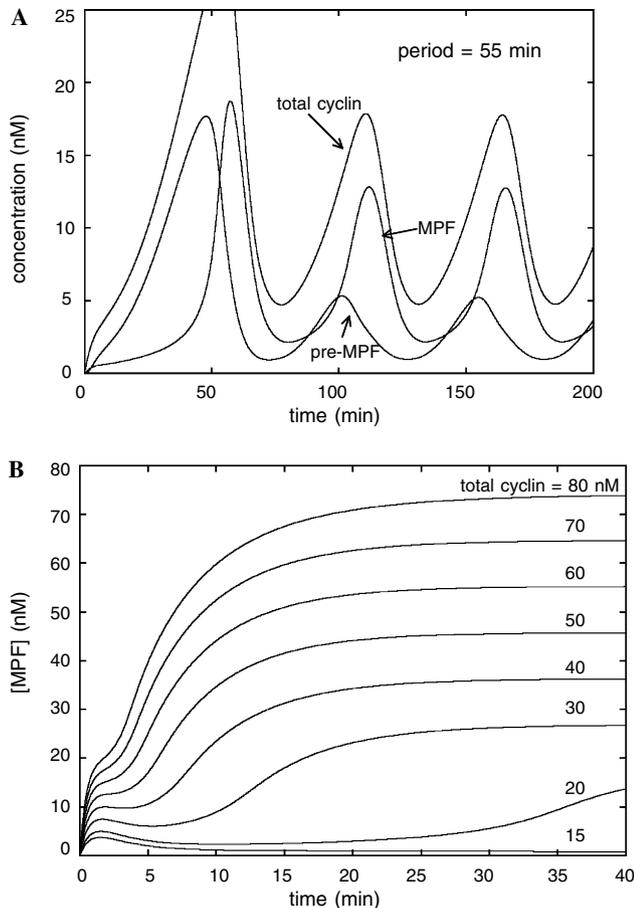


Fig. 4. Simulations of the mathematical model of the cell cycle in frog egg extracts describe known and previously unobserved behaviors of the control system. (A) Simulation of the changes in cyclin, preMPF and MPF levels over time in a cycling frog egg extract [2]. Computed from the ‘ode’ file in Table 1. The simulations are consistent with a body of experimental data. (B) Simulations of the lag time into mitosis after addition of various fixed concentrations of nondegradable cyclin B. Computed from the ‘ode’ file in Table 1, with $k_1 = V'_2 = V''_2 = 0$. The lengthening lag times at progressively lower concentrations of cyclin B was not validated experimentally until 2003, when sufficiently small increments in cyclin B were used for the experiments [8].

estimation. As modeling efforts become more widely utilized and organized, links between specific components of the model (parameters or details of the wiring diagram) and the data upon which they are based (a specific figure or page in a research article) need to be built. Meanwhile, even a limited amount of quantitative information can be useful in constructing the initial set of parameters, constraining portions of the model and allowing the governing ODEs to be numerically simulated to see how close or far the model lies from describing observed behaviors.

Once the process of mining the literature has been exhausted, then the remaining parameters must be estimated by other means. Even in cases where the majority of the parameter values are unknown or where it would be feasible to measure parameter values directly, we recommend that the initial parameter set be derived with a minimal investment in new experimentation. Through computer

simulations (described below), parameters can be twiddled with efficiency so that the modeler can develop a sense of which variables under which conditions are sensitive to which parameters. Such simulation will help design experiments that will yield the most quantitative “bang” for one’s funding “buck.” If during this twiddling phase, no reasonable parameter set can be developed to approximate the behavior of the system, then the wiring diagram may need to be reconsidered, and experiments carried out to test fundamental qualitative assumptions of the model rather than generating specific quantitative data.

2.4. Running simulations of the model

With an initial parameter set in hand, the next step is to “run” the model, that is, to use software that solves ODEs numerically in order to simulate the behavior of the system over time. WinPP (for Windows) and XPP (for Unix) are two simulation programs freely available from Dr. G. Bard Ermentrout, University of Pittsburgh. Go to <http://www.math.pitt.edu/~phase> for links to software downloads and tutorials. The basic output of a simulation is a graph of the concentration of each component over time.

To run a simulation one needs, in addition to differential equations (Fig. 2) and rate constants (Table 1), “initial conditions” for all the variables (i.e., concentrations of each variable at $t = 0$). Initial conditions can usually be estimated readily from the experimental protocol of a particular experiment. For example, a “cycling” extract is prepared by activating the APC, which degrades most of the cyclin in the extract, so it is reasonable to set $[IEP] = [APC] = 1$ and $[cyclin] = [MPF] = [preMPF] = 0$ at $t = 0$. Furthermore, we assume that $[Cdc25P] = [Wee1P] = 0$ at $t = 0$, because MPF is inactive. These common sense considerations fix reasonable initial conditions for a simulation of the temporal evolution of a cycling extract.

The ODEs, parameter values, and initial conditions are communicated to WinPP in a text file called an ‘.ode’ file (see Table 1). By opening this .ode file in WinPP.exe and clicking ‘Run’ ‘Go,’ one will generate Fig. 4A, which simulates the temporal evolution of cyclin B, preMPF and active MPF in a cycling frog-egg extract. A comparison of the simulation to experimental observations shows that the model successfully reproduces the basic features of the egg extract: Cyclin accumulates during interphase then is rapidly degraded at mitosis, and a time lag exists in which tyrosine-phosphorylated cyclin:Cdk dimers form, then these dimers are converted to the active form of MPF.

The simulation represents the behavior of the cell cycle engine in an unperturbed extract. One could also vary parameters to represent experiments in which one or more components of the system is altered. For example, in a classic set of experiments by Solomon et al. [18], synthesis of endogenous cyclin was blocked by addition of cycloheximide, and extracts were supplemented with fixed amounts of recombinant, non-degradable cyclin B (Δ cyclin B).

As done originally by Novak and Tyson (1993) [2], we can simulate this experiment by setting $k_1 = V'_2 = V''_2 = 0$ and setting the initial concentration of cyclin to a fixed value. By varying the initial concentration of cyclin (Fig. 4B), we can draw two conclusions from the model. First of all, the graph of active MPF over time indicates that the initial cyclin concentration must exceed a threshold (between 15 and 20 nM in Fig. 4B) in order to activate MPF. This prediction agrees well with the experimental data [18]. The simulation also suggests that as cyclin concentration approaches this activation threshold from larger values, the lag time to activate MPF becomes longer and longer. This prediction was first made by Novak and Tyson in 1993. Although no published studies showed evidence of this “critical slowing down,” subsequent experiments confirmed this prediction [8]. This example illustrates that mathematical models of the cell cycle can both be informed by and can inform experimental studies.

2.5. Parameter optimization

Initial simulations of the model, based on the original estimates of parameter values, may or may not approximate the behavior of the system to the modeler’s satisfaction. The more qualitative data available to build the wiring diagram and quantitative data available to constrain the parameter values, the closer the simulations are likely to be to the biological system. However, rarely have all rate constants been measured, and estimates of initial conditions typically rely on semi-quantitative and relative data (such as that obtained by Western blotting). Therefore, most parameter sets will need to be optimized to bring the model and the biological system into better quantitative agreement. For relatively simple models, such as our example here, consisting of 7 ODEs and 24 kinetic constants, it may be possible to derive a satisfying parameter set “by hand,” varying one parameter at a time and determining how well simulations match known behaviors of the system. However, this approach is inefficient and tedious, requires considerable advance knowledge of the system in order to make a good “starting guess” of the parameters, and is virtually impossible for more complex models consisting of many variables and many more parameters.

A collection of global parameter optimization tools is publicly available in the Gepasi modeling platform (discussed below) [19]. The choice of which tool to apply remains somewhat empirical [19,20] and largely the realm of the theoretician.

In our example of the cell cycle engine in frog egg extracts, a deterministic global optimizer (called VTDIRRECT) followed by local optimization (using ODRPACK, an implementation of the Levenberg–Marquardt method) was applied to derive an optimized set of rate constants [21]. Overall, simulations using the optimized parameter set give a better fit to the experimental data than simulations using the Marlovits parameter values (Table 1).

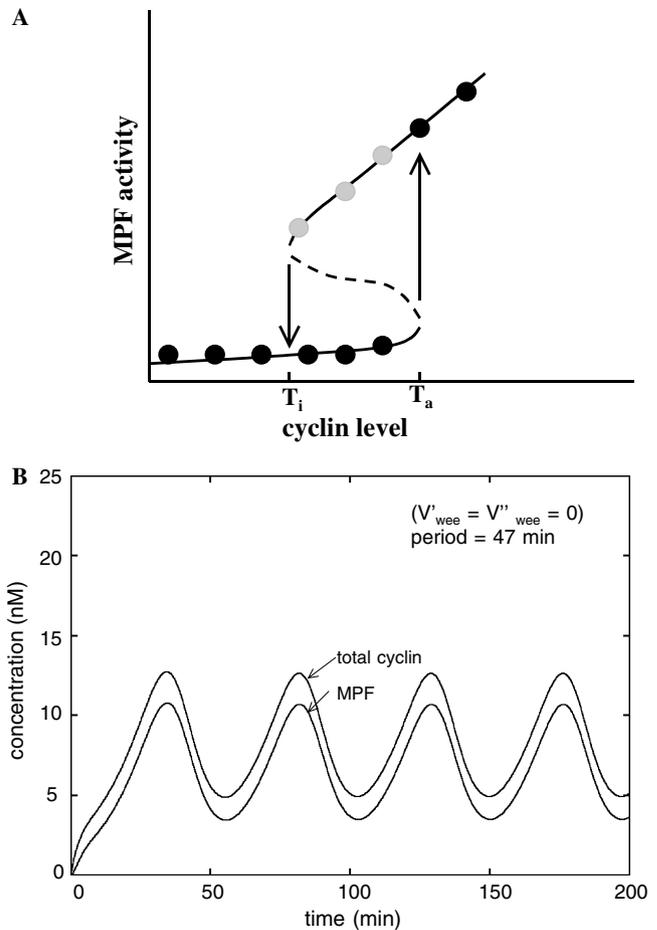


Fig. 5. Theoretical predictions from Novak and Tyson [2]. (A) Hysteresis drives mitotic transitions in frog egg extracts. The simulation indicates that the concentration of cyclin required to drive high MPF activity (and thereby enter mitosis) is higher than the concentration required to maintain high MPF activity (and thereby stay in mitosis). For the Marlovits parameter set, Table 1, $T_i = 8$ nM and $T_a = 17$ nM. The theoretical prediction of two distinct thresholds was confirmed experimentally [8, 9]. (B) MPF oscillations are faster and smaller if Cdk1 is not phosphorylatable by Wee1. Computed from the ‘ode’ file in Table 1, with $V'_{wee} = V''_{wee} = 0$. This prediction was confirmed experimentally by Pomerening et al. [11].

2.6. Experimental testing of the model

Once parameter optimization is used to bring a mathematical model into reasonable quantitative and qualitative agreement with existing experimental data, the next step is to test predictions made by the model experimentally. The rationale for doing so is twofold: the predictive ability of a model is a rigorous test of its accuracy, and more importantly, a major reason for building a model in the first place is to make novel predictions that inform subsequent rounds of experimentation. Virtually any experiment can be simulated in silico by varying parameter values in logical ways: (e.g., decreasing an enzyme’s rate constant to represent the effect of a specific inhibitor, setting the rates of synthesis of all proteins to zero in the presence of cycloheximide, setting the rate of synthesis of a specific protein to zero for a gene

deletion mutant, or setting the initial concentration of protein to zero for an immunodepletion). Conditions that produce the most interesting and measurable cell behaviors are good candidates to test experimentally. Other considerations are whether or not the experiment is feasible (e.g., depleting a Cdk without removing associated cyclins could be problematic) and whether a particular experiment will help define ambiguous parts of the model (e.g., resolve uncertainties in the wiring diagram or constrain a parameter value).

The interplay between modeling and experimentation is at its most valuable when a model can lead to novel, non-intuitive predictions about the behavior of the system. For example, Novak and Tyson's prediction that a hysteresis loop governs entry into and exit from mitosis in frog egg extracts (Fig. 5A), was tested and validated experimentally [8,9]. Using a modified protocol of Solomon et al. [18], in which cyclin levels were controlled in cycloheximide-treated extracts by the addition of exogenous nondegradable cyclin B, a variety of experiments indicated that higher levels of cyclin are required to enter mitosis than to stay in mitosis. Another key prediction of the Novak–Tyson model is that, if the control of MPF activity by Wee1-dependent phosphorylation of Cdk1 is eliminated, then MPF activity continues to oscillate but at a higher frequency and smaller amplitude (compare Figs. 4A and 5B). This prediction was validated experimentally in a series of clever experiments in which Wee1- and Cdc25-control loops were eliminated by adding a nonphosphorylatable Cdk1 (called the Cdc2AF mutant) and inhibiting endogenous Cdk1 [11]. These elegant yet straightforward experiments were inspired by systems-level predictions derived from mathematical models rather than the published experimental studies that preceded them.

Because the frog egg extract is biochemically tractable, experimental testing of the model was performed by biochemical manipulations, such as altering the concentration of cyclin in the system. Genetic approaches can be used to test both quantitative and qualitative predictions of the models that describe cell cycle networks in genetically tractable organisms. For example, a comprehensive mathematical model of the cell cycle control network in *Saccharomyces cerevisiae* describes the behavior of more than 100 genetic mutants [4,22]. Rigorous tests of the model have been performed by creating new mutants predicted by the model to give the most informative phenotypes [23] and by challenging the model to predict phenotypes of yet unpublished mutants [24]. Complete details of the model are available at an easily navigable web site: http://mpf.biol.vt.edu/research/budding_yeast_model/pp/.

3. Modeling tools and environments

The development of modeling “environments” that contain suites of tools necessary for model building, simulations, data fitting, and data management is an advancement that will benefit both experts and novices

building models of molecular regulatory networks. The modeling environments described here are publicly available, relatively user friendly, and operate on personal computers running Windows, Macintosh, and/or Linux operating systems. Furthermore, each environment contains tools to translate models into the Systems Biology Markup Language (SBML), a grammar that is becoming widely adopted by the biochemical network modeling community to exchange models [25].

3.1. Gepasi (<http://www.gepasi.org>)

Gepasi, the first package for modeling biochemical networks, was originally released in 1993 [26] and has been enhanced by regular additions and revisions [27]. The current version (3.3) runs on Windows and Linux operating systems. Gepasi supports modeling of time course data in an environment that includes a graphical user interface (GUI) for model building, tools for simulations, and a collection of parameter estimation algorithms. Gepasi allows the definition of spatial compartments and includes a tool called MEG that allows modeling of spatially heterogeneous systems such as cell cultures or tissues [28]. The environment also includes a number of tutorials to guide novice users through key aspects of the modeling environment. Although the Gepasi 3.3 is reported to be the last iteration, COPASI (complex pathway simulator), a new modeling environment based on Gepasi, is being developed. Test versions of COPASI are available at <http://www.copasi.org>. COPASI is designed to be more user-friendly, compatible with the latest versions of SBML, and functional in Windows, Macintosh and Linux operating systems.

3.2. Virtual cell (<http://www.vcell.org>)

Virtual Cell, a modeling environment created by the National Resource for Cell Analysis and Modeling, may be the most user-friendly for the non-expert [29]. The current version, 4.0, can be run as a Java application or applet link on Windows, Macintosh OSX 10.1 and higher, and Linux operating systems. User support is available by e-mail. Programs for building models, running simulations, and managing documents are included, with plans for parameter estimation tools in Version 4.2. Virtual Cell places particular emphasis on modeling spatially distributed systems, providing tools to formulate and simulate partial differential equations (PDEs), which are better suited to represent spatiotemporal dynamics in cell signaling networks [30]. Virtual Cell 4.1, available as a beta version, allows representation of two- and three-dimensional membrane diffusion and visualization of three-dimensional surfaces. Virtual Cell utilizes its own language, VCML, which can be translated to SBML. A host of known incompatibilities between the two languages is documented on the Virtual Cell website.

3.3. Systems biology workbench (<http://sbw.kgi.edu>)

The Systems Biology Workbench (SBW) provides a suite of tools generically suited for modeling biochemical networks [31]. The current version 2.5.0 operates in Windows, Linux and Macintosh OSX. SBW provides a GUI for model building (JDesigner) and a simulation tool (Jarnac) [32]. Most importantly, SBW enables modeling applications written in diverse languages to communicate and operate together. The SBW group has worked for interoperability with the BioSPICE modeling projects, described below. SBW requires some background in dynamical systems and computer science to utilize the full suite of programs, but the basic operation of JDesigner/Jarnac is designed to be user-friendly.

3.4. JigCell (<http://jigcell.biol.vt.edu>)

JigCell is a modeling environment for Windows being developed as part of the larger DARPA BioSPICE effort [33,34]. Although still a work-in-progress, JigCell includes a convenient spreadsheet-style Model Builder, a Run Manager for organizing related parameter sets and running multiple simulations [35], a Comparator for comparing simulations to experimental data [36], and a Parameter Estimator [21,37]. Although JigCell is suited for modeling any biochemical network, it was developed with a particular emphasis on modeling the cell cycle. A model of the cell cycles of frog egg extracts (described above) and a model of the more complex budding yeast cell cycle have been used as test cases for developing the various components of JigCell, and these models are available on the website. JigCell is designed for interoperability with other modeling environments under the BioSPICE umbrella, and thus, is compatible with SBW as well.

4. Conclusions

Because of the enormous body of data, the ability to test quantitative theoretical predictions in powerful biochemical and genetic systems, and the obvious value in deciphering this regulatory network, rigorous mathematical models of the eukaryotic cell cycle have been under development for over a decade, long before systems biology became fashionable. Early models were largely theoretical, and parameter values were based on scanty experimental measurements, back-of-envelope calculations, and a good deal of guesswork. However, in more recent years, serious collaborations between experimentalists and modelers have begun, and efforts to make modeling tools accessible to the non-expert have emerged. At present, we cannot provide a “recipe” or “kit” for building and using a mathematical model of your favorite molecular regulatory network. Serious work will probably require collaboration with professional modelers for the near future. However, we are confident that as computer tools get better and theoretical concepts seem more familiar, mathematical model-

ing will take its place in the molecular cell biologist’s toolkit alongside SDS–PAGE and PCR!

5. For further study

This review was written for the non-expert. We recommend as a first foray into modeling, that newcomers familiarize themselves with the example provided here and/or tutorial models found on the websites described in Section 3. For those interested in delving deeper into the theory and practice of modeling complex biochemical networks, we suggest literatures [38–42].

Acknowledgments

We thank the NIH (GM59688 to J.C.S) and DARPA/ARFL (F30602-01-2-0572 to J.J.T) for funding. Thanks to Dr. Brian Wroble for critical reading of this manuscript.

References

- [1] B. Aguda, *Proc. Natl. Acad. Sci. USA* 96 (1999) 11352–11357.
- [2] B. Novak, J.J. Tyson, *J. Cell Sci.* 106 (1993) 1153–1168.
- [3] A. Goldbeter, *Proc. Natl. Acad. Sci. USA* 88 (1991) 9107–9111.
- [4] K.C. Chen, A. Csikasz-Nagy, B. Gyorffy, J. Val, B. Novak, J.J. Tyson, *Mol. Biol. Cell* 11 (2000) 369–391.
- [5] Z. Qu, J.N. Weiss, W.R. MacLellan, *J. Cell Sci.* 117 (2004) 4199–4207.
- [6] Z. Qu, J.N. Weiss, W.R. MacLellan, *Am. J. Physiol. Cell Physiol.* 284 (2003) C349–C364.
- [7] C.D. Thron, *Biophys. Chem.* 57 (1996) 239–251.
- [8] W. Sha, J. Moore, K. Chen, A. Lassaletta, C.-S. Yi, J. Tyson, J. Sible, *Proc. Natl. Acad. Sci. USA* 100 (2003) 975–980.
- [9] J.R. Pomeroy, E.D. Sontag, J.E.J. Ferrell, *Nat. Cell Biol.* 5 (2003) 346–351.
- [10] W. Xiong, J.E.J. Ferrell, *Nature* 426 (2003) 460–465.
- [11] J.R. Pomeroy, S.Y. Kim, J.E.J. Ferrell, *Cell* 122 (2005) 565–578.
- [12] S. Bai, D. Goodrich, C.D. Thron, E. Tecarro, M. Obeyesekere, *Cell Cycle* 2 (2003) 46–52.
- [13] M.A. Felix, J.C. Labbe, M. Doree, T. Hunt, E. Karsenti, *Nature* 346 (1990) 379–382.
- [14] E.R. Kramer, C. Gieffers, G. Holzl, M. Hengstschlager, J.M. Peters, *Curr. Biol.* 8 (1998) 1207–1210.
- [15] A. Hershko, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 354 (1999) 1571–1575.
- [16] T. Lorca, A. Castro, A.M. Martinez, S. Vigneron, N. Morin, S. Sigrist, C. Lehner, M. Doree, J.C. Labbe, *EMBO J.* 17 (1998) 3565–3575.
- [17] A. Kumagai, W.G. Dunphy, *Mol. Biol. Cell* 6 (1995) 199–213.
- [18] M.J. Solomon, M. Glotzer, T.H. Lee, M. Phillippe, M.W. Kirschner, *Cell* 63 (1990) 1013–1024.
- [19] P. Mendes, D.B. Kell, *Bioinformatics* 14 (1998) 869–883.
- [20] C.G. Moles, P. Mendes, J.R. Banga, *Genome Res.* 13 (2003) 2467–2474.
- [21] J.W. Zwolak, J.J. Tyson, L.T. Watson, *J. Comput. Biol.* 12 (2005) 48–63.
- [22] K.C. Chen, L. Calzone, A. Csikasz-Nagy, F.R. Cross, B. Novak, J.J. Tyson, *Mol. Biol. Cell* 15 (2004) 3841–3862.
- [23] F.R. Cross, V. Archambault, M. Miller, M. Klovstad, *Mol. Biol. Cell* 13 (2002) 52–70.
- [24] F.R. Cross, L. Schroeder, M. Kruse, K.C. Chen, *Mol. Biol. Cell* 16 (2005) 2129–3138.
- [25] M. Hucka, A. Finney, H.M. Sauro, H. Bolouri, J.C. Doyle, H. Kitano, A.P. Arkin, B.J. Bornstein, D. Bray, A. Cornish-Bowden,

- A.A. Guellar, S. Dronov, E.D. Gilles, M. Ginkey, V. Gor, I.I. Goryanin, W.J. Hedley, T.C. Hodgman, J.H. Hofmeyr, P.J. Hunter, N.S. Juty, J.L. Kasberger, A. Kremling, U. Kummer, N.L. Novere, L.M. Loew, D. Lucio, P. Mendes, E. Minch, E.J. Mjolsness, Y. Nakayama, M.R. Nelson, P.F. Nielsoen, T. Sakurada, J.C. Schaff, B.E. Shapiro, T.S. Shimizu, H.D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, J. Wang, *Bioinformatics* 19 (2003) 524–531.
- [26] P. Mendes, *Comput. Appl. Biosci.* 9 (1993) 563–571.
- [27] P. Mendes, *Trends Biochem. Sci.* 22 (1997) 361–363.
- [28] P. Mendes, D.B. Kell, *Bioinformatics* 17 (2001) 288–289.
- [29] I.I. Moraru, J.C. Schaff, B.M. Slepchenko, L.M. Loew, *Ann., N.Y. Acad. Sci.* 971 (2002) 595–596.
- [30] N.J. Eungdamrong, R. Iyengar, *Biol. Cell* 96 (2004) 355–362.
- [31] H.M. Sauro, M. Hucka, A. Finney, C. Wellock, H. Bolouri, J. Doyle, H. Kitano, *OMICS* 7 (2003) 355–372.
- [32] H.M. Sauro, in: J.-H.S. Hofmeyr, J.M. Rohwer, J.L. Snoep (Eds.), *Animating the Cellular Map: Proceedings of the 9th International Meeting on BioThermoKinetics*, Stellenbosch University Press, 2000, pp. 221–228.
- [33] N.A. Allen, L. Calzone, K.C. Chen, A. Ciliberto, N. Ramakrishnan, C.A. Shaffer, J.C. Sible, J.J. Tyson, M.T. Vass, L.T. Watson, J.W. Zwolak, *OMICS* 7 (2003) 285–299.
- [34] N.A. Allen, C.A. Shaffer, N. Ramakrishnan, M.T. Vass, L.T. Watson, *Simulation* 79 (2003) 674–688.
- [35] M.T. Vass, N.A. Allen, C.A. Shaffer, N. Ramakrishnan, L.T. Watson, J.J. Tyson, *Bioinformatics* 20 (2004) 3680–3681.
- [36] N.A. Allen, C.A. Shaffer, N. Ramakrishnan, M.T. Vass, L.T. Watson, *Simulation* 79 (2005) 674–688.
- [37] J.W. Zwolak, J.J. Tyson, L.T. Watson, *IEE Proc.-Syst. Biol.* 152 (2005) 81–92.
- [38] C.P. Fall, E.S. Marland, J.M. Wagner, J.J. Tyson (Eds.), *Computational Cell Biology*, Springer, Springer, New York, 2000.
- [39] J.J. Tyson, K.C. Chen, B. Novak, *Curr. Opin. Cell Biol.* 15 (2003) 221–231.
- [40] D.A. Lauffenburger, J.J. Linderman, *Receptors: models for binding, trafficking, and signaling*, Oxford University Press, New York, 1993.
- [41] L.A. Segel, *Modeling dynamic phenomena in molecular and cellular biology*, Cambridge University Press, New York, 1989.
- [42] A. Goldbeter, *Biochemical Oscillations and Cellular Rhythms: The molecular bases of periodic and chaotic behaviour*, Cambridge University Press, New York, 1996.
- [43] G. Marlovits, C.T. Tyson, B. Novak, J.J. Tyson, *Modeling M-phase control in Xenopus oocyte extracts: the surveillance mechanism for unreplicated DNA*, *Biophys. Chem.* 72 (1998) 169–184.