



# Stochastic Boolean model of normal and aberrant cell cycles in budding yeast



Kittisak Taoma<sup>1,2</sup>, John J. Tyson<sup>3</sup>, Teeraphan Laomettachit<sup>1,2</sup>✉ & Pavel Kraikivski<sup>4,5</sup>✉

The cell cycle of budding yeast is governed by an intricate protein regulatory network whose dysregulation can lead to lethal mistakes or aberrant cell division cycles. In this work, we model this network in a Boolean framework for stochastic simulations. Our model is sufficiently detailed to account for the phenotypes of 40 mutant yeast strains (83% of the experimentally characterized strains that we simulated) and also to simulate an endoreplicating strain (multiple rounds of DNA synthesis without mitosis) and a strain that exhibits ‘Cdc14 endocycles’ (periodic transitions between metaphase and anaphase). Because our model successfully replicates the observed properties of both wild-type yeast cells and many mutant strains, it provides a reasonable, validated starting point for more comprehensive stochastic-Boolean models of cell cycle controls. Such models may provide a better understanding of cell cycle anomalies in budding yeast and ultimately in mammalian cells.

Orderly progression through the eukaryotic cell cycle is governed by molecular circuits that control the timely switching from  $G_1$  into  $S$ - $G_2$ - $M$  and back to  $G_1$ . These transitions typically follow one another in an alternating sequence, but certain disruptions of the control circuits can result in aberrant cell cycles. For example,  $G_1$ - $S$ - $G_1$ - $S$  and  $M$ -( $G_1$ )- $M$ -( $G_1$ ) cycles are observed in some budding yeast mutant strains<sup>1–3</sup>. Moreover, aberrant cell divisions are common occurrences in cancer cells<sup>4,5</sup>.

Ordinary differential equations (ODEs) are often used to model the molecular control circuits governing cell cycle progression and to explain the irreversible transitions from one cell cycle phase to the next. ODEs have been successfully applied to the complex cell cycle regulatory network in budding yeast<sup>6–8</sup>, as well as specific cell cycle transitions controlled by different checkpoints, e.g., the  $G_1/S$  transition<sup>9</sup>, mitotic exit<sup>10,11</sup> and the spindle positioning checkpoint (SPOC)<sup>12</sup>. Although ODE-based approaches can provide comprehensive quantitative details, they require accurate estimation of many kinetic parameters in the equations and substantial computational time to simulate large molecular regulatory networks<sup>13</sup>. Furthermore, accounting for stochastic effects within this framework requires additional quantitative data about cell constituents and significantly greater computational resources<sup>14,15</sup>.

To address these difficulties with ODE modeling, many authors have turned to Boolean methods<sup>16–19</sup>. Recently we have adopted a Boolean Kinetic Monte Carlo (BKMC) approach<sup>20</sup> to explore stochastic Boolean modeling of the budding yeast cell cycle<sup>21</sup>. Although simple (only seven regulatory proteins), the model successfully explained some basic observations of

stochastic cell growth and division in wild-type yeast strains; but it was too simple to account for the phenotypes of any mutant strains. Our goal here is to develop a more comprehensive model that addresses in quantitative detail the phenotypes of certain well-characterized mutant strains, including aberrant cycles such as endoreplication and Cdc14 endocycles<sup>22</sup>. The model’s promising results, on a limited subset of experimental data, suggest that our approach to stochastic Boolean modeling may be worth pursuing in more detail in the future.

The molecular mechanism of our model (Fig. 1) involves 22 cell cycle-related components: fifteen proteins, three checkpoints, three ‘progress’ variables, and one ‘flag’, as defined in Supplementary Table 1.

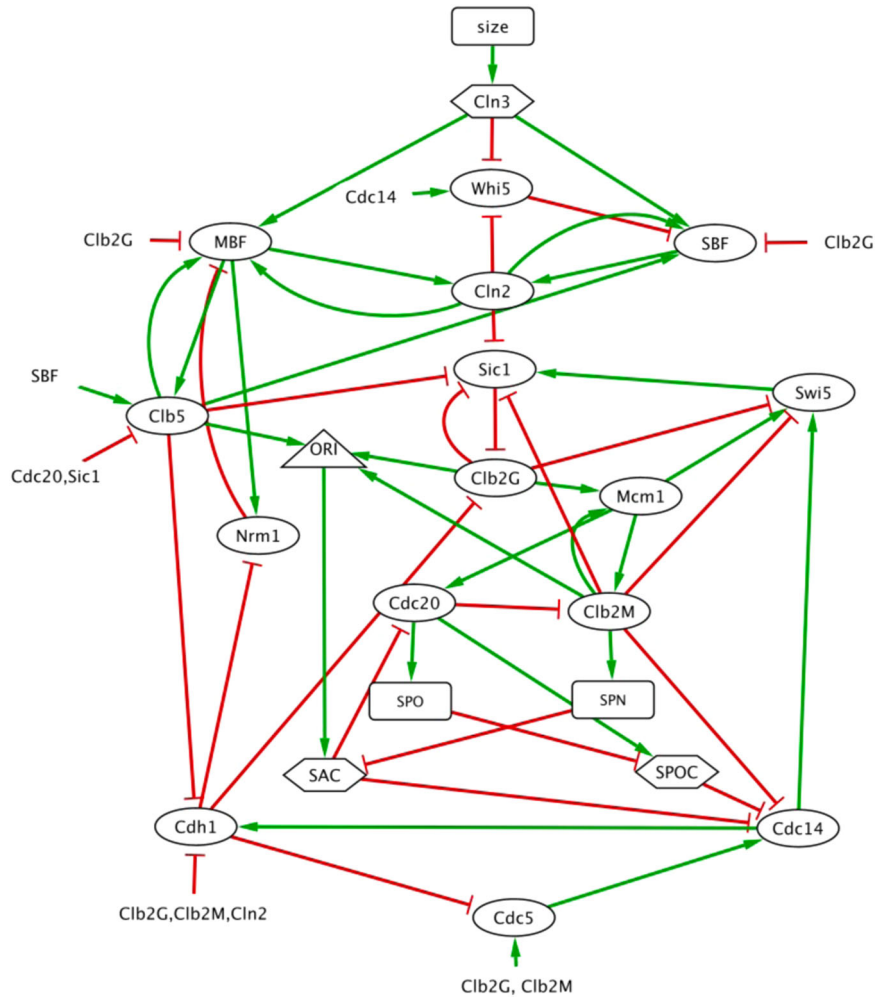
## Methods

### Definition of the model

Our model is based on previous work<sup>20,21</sup> where the Boolean functions and time steps are updated asynchronously and continuously using Gillespie’s stochastic simulation algorithm. We have extended and modified our earlier model in several ways, in order to account for the phenotypes of mutant strains as well as the physiology of wild-type cells. As before, each protein in the model is characterized by a Boolean variable,  $X_{j,b}$ , where  $j = 1, \dots, 15$  indexes the proteins and  $t \geq 0$  is time. Unlike the BKMC of Stoll et al.<sup>20</sup>, where the Boolean variables are updated according to certain specified ‘reaction propensities,’ we update them by a uniform asynchronous scheme in two steps.

<sup>1</sup>Bioinformatics and Systems Biology Program, School of Bioresources and Technology, King Mongkut’s University of Technology Thonburi, Bangkok, 10150, Thailand. <sup>2</sup>Theoretical and Computational Physics Group, Center of Excellence in Theoretical and Computational Science, King Mongkut’s University of Technology Thonburi, Bangkok, 10150, Thailand. <sup>3</sup>Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, 24061, USA. <sup>4</sup>Division of Systems Biology, Academy of Integrated Science, Virginia Polytechnic Institute and State University, Blacksburg, VA, 24061, USA. <sup>5</sup>VT-Center for the Mathematics of Biosystems, Virginia Polytechnic Institute and State University, Blacksburg, VA, 24061, USA. ✉e-mail: [teeraphan.lao@kmutt.ac.th](mailto:teeraphan.lao@kmutt.ac.th); [pavelkr@vt.edu](mailto:pavelkr@vt.edu)

**Fig. 1 | Influence diagram describing our model of cell cycle regulation in budding yeast.** The regulatory network consists of nodes connected by edges representing inhibition (red lines with a blunt end) or activation (green lines with a barbed end). Ovals: proteins; hexagons: checkpoints; rectangles: progress variables; triangle: flag variable.



First, we determine which variables might possibly change in the next time step by the following Boolean functions, where the ‘hat’ indicates the ‘potential’ value of  $X_i$  at time  $t + \Delta t$ :

$$\hat{X}_{i,t+\Delta t} = \text{Heav}(W_{i,t}) = \begin{cases} 1 & \text{if } W_{i,t} > 0 \\ 0 & \text{otherwise} \end{cases}, W_{i,t} = \omega_{i0} + \sum_j \omega_{ij} X_{j,t}, i = 1 \dots 15 \quad (1)$$

Equation 1 takes as input the 15 Boolean variables representing the proteins, and the 3 Boolean variables representing the states of the checkpoints, in order to calculate an intermediate function,  $W_i(X_1, \dots, X_{15}, Cln3, SAC, SPOC)$ , and outputs a Boolean variable,  $\hat{X}_{i,t+\Delta t}$ , the potential update of  $X_i$ . The  $\omega_{ij}$  coefficients define the Boolean function for updating  $X_i$ . For example, the Boolean function  $X_1 = X_2 \text{ AND } X_3$  can be implemented by  $W_1 = -1.5 + X_2 + X_3$ , and  $X_1 = X_2 \text{ OR } X_3$  by  $W_1 = -0.5 + X_2 + X_3$ . Similarly,  $X_1 = X_2 \text{ AND } (\text{NOT } X_3)$  is equivalent to  $W_1 = -0.5 + X_2 - X_3$ ; and  $X_1 = X_2 \text{ OR } (\text{NOT } X_3)$  to  $W_1 = +0.5 + X_2 - X_3$ . In general,  $\omega_{ij} > 0$  if variable  $j$  activates variable  $i$  (a green barbed arrow in Fig. 1), and  $\omega_{ij} < 0$  if variable  $j$  inhibits variable  $i$  (a red blunt connector in Fig. 1). The  $\omega_{ij}$ 's are pure numbers, not rate constants. Their values are chosen once-and-for-all to fix the logical relations in the Boolean dynamics implied by Fig. 1. They are not adjusted to fit quantitative experimental observations (e.g., cycle times, cell size distributions), but they have been adjusted to account for qualitative phenotypes (viable or arrested in a particular phase of the cell cycle) of a panel of mutant strains of budding yeast. In contrast to Boolean modeling with logical functions, we prefer this approach, which is an extension of the method introduced by Li et al.<sup>16</sup>, because the functional form of the  $W_i$ 's

(defined in Eq. 1) is a direct reflection of the network topology in Fig. 1, and the arithmetic calculation is a very compact way to encode a Boolean function. For example, for a node with 4 inputs, there are  $2^{2^4} \approx 65,000$  potential Boolean functions, and the one we choose is specified by the relative values of just five coefficients  $\omega_{i0}, \omega_{i1}, \dots, \omega_{i4}$ . However, this approach is limited in that any linear function  $W_{i,t} = \omega_{i0} + \sum_j \omega_{ij} X_{j,t}$  can be translated into a corresponding logical function for  $\hat{X}_{i,t+\Delta t}$ , but not all possible logical functions can be represented by a linear function. An obvious example is the ‘XOR’ function:  $X_1 = X_2 \text{ XOR } X_3 = (X_2 \text{ OR } X_3) \text{ AND } (\text{NOT } (X_2 \text{ AND } X_3))$ .

The  $W_i$  functions that define our Boolean model for the 15 cell-cycle regulatory proteins are displayed in Table 1. The values of the 65  $\omega_{i0}$  and  $\omega_{ij}$  coefficients in these functions (for wild-type cells) are specified in Supplementary Table 2A. In Supplementary Table 3 we translate the  $W_i$  functions of our Boolean model into traditional logical functions in terms of the elementary relations: AND, OR, NOT.

The second step in updating our model is to identify the change to be made and the time step  $\Delta t$  to be taken. If more than one protein potentially changes state ( $\hat{X}_{i,t+\Delta t} \neq X_{i,t}$ ), then we choose the one that will actually change at random with equal probabilities, i.e., the Boolean model progresses from one state to the next by uniform asynchronous updating. In the BKMC approach, time is updated by choosing  $\Delta t$  from an exponential distribution, parameterized by the total ‘propensity’ (probability per unit time) for any one of the potential changes to occur. This scheme is based on the assumption that each potential change is an elementary chemical reaction<sup>23</sup>, which certainly doesn’t hold in the case of Boolean modeling. If each change (protein synthesis, degradation, phosphorylation, dephosphorylation, etc.)

**Table 1 | The  $W_i$  functions defining the Boolean network for updating the 15 proteins of the cell-cycle control system**

1	$W_{Whi5} = \omega_{Whi5} + \omega_{Whi5,Cdc14} \cdot Cdc14 - \omega_{Whi5,Cln2} \cdot Cln2 - \omega_{Whi5,Cln3} \cdot Cln3$
2	$W_{SBF} = \omega_{SBF} - \omega_{SBF,Whi5} \cdot Whi5 - \omega_{SBF,Clb2G} \cdot Clb2G + \omega_{SBF,Clb5} \cdot Clb5 + \omega_{SBF,Cln2} \cdot Cln2 + \omega_{SBF,Cln3} \cdot Cln3$
3	$W_{MBF} = \omega_{MBF} - \omega_{MBF,Nrm1} \cdot Nrm1 - \omega_{MBF,Clb2G} \cdot Clb2G + \omega_{MBF,Clb5} \cdot Clb5 + \omega_{MBF,Cln2} \cdot Cln2 + \omega_{MBF,Cln3} \cdot Cln3$
4	$W_{Nrm1} = \omega_{Nrm1} - \omega_{Nrm1,Cdh1} \cdot Cdh1 + \omega_{Nrm1,MBF} \cdot MBF$
5	$W_{Clb5} = \omega_{Clb5} - \omega_{Clb5,Sic1} \cdot Sic1 - \omega_{Clb5,Cdc20} \cdot Cdc20 + \omega_{Clb5,MBF} \cdot MBF + \omega_{Clb5,SBF} \cdot SBF$
6	$W_{Cln2} = \omega_{Cln2} + \omega_{Cln2,SBF} \cdot SBF + \omega_{Cln2,MBF} \cdot MBF$
7	$W_{Sic1} = \omega_{Sic1} + \omega_{Sic1,Swi5} \cdot Swi5 - \omega_{Sic1,Clb2G} \cdot Clb2G - \omega_{Sic1,Clb2M} \cdot Clb2M - \omega_{Sic1,Clb5} \cdot Clb5 - \omega_{Sic1,Cln2} \cdot Cln2$
8	$W_{Cdh1} = \omega_{Cdh1} + \omega_{Cdh1,Cdc14} \cdot Cdc14 - \omega_{Cdh1,Clb2G} \cdot Clb2G - \omega_{Cdh1,Clb2M} \cdot Clb2M - \omega_{Cdh1,Clb5} \cdot Clb5 - \omega_{Cdh1,Cln2} \cdot Cln2$
9	$W_{Clb2G} = \omega_{Clb2G} - \omega_{Clb2G,Cdh1} \cdot Cdh1 - \omega_{Clb2G,Sic1} \cdot Sic1$
10	$W_{Clb2M} = \omega_{Clb2M} + \omega_{Clb2M,Mcm1} \cdot Mcm1 - \omega_{Clb2M,Cdc20} \cdot Cdc20$
11	$W_{Mcm1} = \omega_{Mcm1} + \omega_{Mcm1,Clb2G} \cdot Clb2G + \omega_{Mcm1,Clb2M} \cdot Clb2M$
12	$W_{Cdc5} = \omega_{Cdc5} + \omega_{Cdc5,Clb2G} \cdot Clb2G + \omega_{Cdc5,Clb2M} \cdot Clb2M - \omega_{Cdc5,Cdh1} \cdot Cdh1$
13	$W_{Cdc20} = \omega_{Cdc20} + \omega_{Cdc20,Mcm1} \cdot Mcm1 - \omega_{Cdc20,SAC} \cdot SAC$
14	$W_{Cdc14} = \omega_{Cdc14} + \omega_{Cdc14,Cdc5} \cdot Cdc5 - \omega_{Cdc14,Clb2M} \cdot Clb2M - \omega_{Cdc14,SAC} \cdot SAC - \omega_{Cdc14,SPOC} \cdot SPOC$
15	$W_{Swi5} = \omega_{Swi5} + \omega_{Swi5,Cdc14} \cdot Cdc14 + \omega_{Swi5,Mcm1} \cdot Mcm1 - \omega_{Swi5,Clb2M} \cdot Clb2M - \omega_{Swi5,Clb2G} \cdot Clb2G$

is a series of  $k$  elementary steps that are independent and identically distributed random variables following an exponential distribution with time constant  $\theta$ , then the total time,  $\Delta t$ , for completion of the process is given by a gamma distribution, with density function:

$$\text{gamma}(\Delta t) = \frac{1}{\Gamma(k)\theta^k} (\Delta t)^{k-1} e^{-\frac{\Delta t}{\theta}} \quad (2)$$

The parameters  $k$  and  $\theta$  determine the mean value of  $\Delta t$  between updates (mean =  $k\theta$ ) and its coefficient of variation ( $CV = 1/\sqrt{k}$ ). For wild-type cells, we set  $k = 3$  and  $\theta = 0.3$  min (i.e., mean = 0.9 min and  $CV = 0.58$ ).

In some circumstances, the Boolean model of the protein interactions settles on a steady state ( $\hat{X}_{i,t+\Delta t} = X_{i,t}$  for all  $i = 1, \dots, 15$ ), in which case we allow  $t$  to increase by drawing  $\Delta t$  from a gamma distribution with the same parameter values:  $k = 3$  and  $\theta = 0.3$  min. During the ensuing period, the ‘progress’ variables may change and induce the protein network to leave the steady state and re-enter the cell cycle. For instance, wild-type cells have a  $G_1$  steady state ( $Whi5 = Cdh1 = Sic1 = 1$ , all other Boolean variables = 0). As formalized below, for a cell in this  $G_1$  steady state, the progress variable  $size_t$  steadily increases as the cell grows. When  $size_t > S_0$ , the ‘cell size checkpoint’ is satisfied, and the checkpoint variable  $Cln3_t$  is changed from 0 to 1. This change kicks the protein interaction network out of the  $G_1$  steady state and sets the cell division program in motion. In the simulation of some mutants, the protein interaction network falls into a steady state that it cannot leave (i.e., the cell is arrested at some point in the cell cycle), and we stop the simulation after the arrested state becomes evident.

Finally, after the updated protein (say,  $X_k$ ) is chosen and  $\Delta t$  is determined, all the protein variables are updated as follows:

$$X_{k,t+\Delta t} = \hat{X}_{k,t+\Delta t}, \text{ and } X_{i \neq k,t+\Delta t} = X_{i,t} \quad (3)$$

This completes our description of how the protein variables are updated. The changes we have made to the BKMC scheme of Stoll et al.<sup>20</sup> precludes using their convenient simulation environment MaBoSS (Markovian Boolean Stochastic Simulator, <https://maboss.curie.fr>). Instead, we have implemented our scheme in Python and provided all codes on GitHub.

In addition to the Boolean variables tracking the protein interaction network, the model has a Boolean ‘flag’ called ORI and three Boolean ‘checkpoints’ called Cln3, SAC and SPOC. They are updated as explained in the following paragraphs.

- **ORI** specifies the state of the origins of replication on the chromosomes.  $ORI = 0$  means the chromosomes are unreplicated and the origins are ‘licensed’ to initiate replication.  $ORI = 1$  means that

chromosome replication has been initiated and that the origins are now ‘unlicensed’ (i.e., unable to initiate a new round of DNA replication). The value of ORI at any time  $t$  is determined simply by the presence of Clb-dependent kinase activity:

$$ORI_t = Clb5_t \text{ OR } Clb2G_t \text{ OR } Clb2M_t \quad (4)$$

- **Cln3** is an indicator of cell growth.  $Cln3 = 0$  indicates that a cell is too small to start S phase;  $Cln3 = 1$  means that it has grown large enough to warrant a new round of DNA replication and cell division:

$$Cln3_t = \text{Heav}(size_t - S_0) \quad (5)$$

‘Size’ is the progress variable that controls the size checkpoint:  $size_t > S_0 > 0$ , where  $S_0$  is the minimum size necessary start the S-G<sub>2</sub>-M sequence.  $S_0$  is a positive random variable assigned to a cell at birth from a lognormal distribution,

$$\text{lognormal}(S_0) = \frac{1}{S_0 \sigma \sqrt{2\pi}} \exp\left(-\frac{(\ln S_0 - \mu)^2}{2\sigma^2}\right) \quad (6)$$

For a lognormal distribution,  $S_{\text{mean}} = e^\mu(1 + \frac{1}{2}\sigma^2 + \dots)$  and  $S_{\text{CV}} = \sigma(1 + \frac{1}{4}\sigma^2 + \dots)$ . For simplicity, we encode the lognormal distribution with  $\mu = \ln(S_{0,\text{mean}})$  and  $\sigma = S_{0,\text{CV}}$ , which are suitable approximations for our purposes. We chose  $S_{0,\text{mean}} = 0.6$  so that the size of dividing cells is roughly 1 (arbitrary unit), and  $S_{0,\text{CV}} = 0.1$  to fit the observed variability (typically ~10%) of size at division. Note: the CV of cell size at bud emergence is ~5%, according to Table S11 of Di Talia et al.<sup>24</sup>.

During every time step  $\Delta t$ ,  $size_t$  increases according to:

$$size_{t+\Delta t} = size_t \cdot e^{r \cdot \Delta t} \quad (7)$$

where  $r$  is the specific growth rate of cells. In glucose medium, the mass-doubling time is ~90 min, so we choose  $r = 0.0077 \text{ min}^{-1}$ . At cell division (an event to be defined later), the size of the dividing cell,  $size_{\text{@div}}$ , is distributed asymmetrically to the progeny:

$$size_{\text{newborn\_mother}} = f \cdot size_{\text{@div}}, \text{ } size_{\text{newborn\_daughter}} = (1 - f) \cdot size_{\text{@div}} \quad (8)$$

We draw  $f$  from a lognormal distribution with  $f_{\text{mean}} = 0.55$  and  $f_{\text{CV}} = 0.1$ ; these values give a good agreement with the mean and CV of size at birth for both mother and daughter cells growing in glucose medium.

- **SAC**, the ‘spindle assembly checkpoint’, is a Boolean variable indicating the state of alignment of replicated chromosomes on the mitotic spindle:

$$SAC_t = \text{Heav}(1 - SPN_t) \text{ AND } ORI_t \quad (9)$$

where  $ORI_t = 1$  indicates that DNA replication has been initiated, and  $SPN_t$  tracks the progression of the replicated chromosomes on the spindle;  $SPN_t = 1$  indicating complete alignment.  $SPN_t$  is initialized at 0 when the cell enters mitosis, i.e., when  $Clb2M$  turns on, and  $SPN_t$  increases in each time step thereafter, according to:

$$SPN_{t+\Delta t} = SPN_t + \Delta SPN \quad (10)$$

where  $\Delta SPN$  is chosen (in each time step) from a lognormal distribution with parameters  $SPN_{\text{mean}} = 0.07$  and  $SPN_{\text{CV}} = 0.03$ . Hence, it takes  $\sim 14$  steps ( $14k\theta = 13$  min) from entry into mitosis until all chromosomes are aligned on the metaphase plate.

- **SPOC**, the ‘spindle position checkpoint’, is a Boolean variable indicating that the fully aligned mitotic spindle is properly positioned in the neck between mother and bud:

$$SPOC_t = \text{Heav}(1 - SPO_t) \text{ AND } Cdc20_t \quad (11)$$

where  $Cdc20_t = 1$  indicates that anaphase has been initiated, and  $SPO_t$  tracks the movement of the two incipient nuclei during anaphase and telophase.  $SPO_t = 1$  indicates that the bud has received its nucleus. When SPOC turns off,  $Cdc14$  is activated and the cell completes the transition from telophase to  $G_1$ .  $SPO_t$  is initialized at 0 when the cell enters anaphase, i.e., when  $Cdc20$  turns on, and  $SPO_t$  increases in each time step thereafter, according to:

$$SPO_{t+\Delta t} = SPO_t + \Delta SPO \quad (12)$$

where  $\Delta SPO$  is chosen (in each time step) from a lognormal distribution with parameters  $SPO_{\text{mean}} = 0.07$  and  $SPO_{\text{CV}} = 0.03$ , so it takes  $\sim 13$  min for the daughter chromosomes to be properly partitioned to the bud.

In addition to the 65  $\omega_i$  and  $\omega_j$  coefficients defined in Supplementary Table 2A, the equations defining our model involve 14 adjustable parameters: 4 parameters for updating the SPN and SPO variables, 2 parameters for the specific growth rate (mass doubling time) in glucose and galactose media, 4 parameters to determine the fraction  $f$  of a dividing cell that is apportioned to the mother cell in glucose and galactose, 2 parameters to determine the critical cell size  $S_0$ , and 2 parameters to define  $\Delta t$  for updating the Boolean model. These 14 parameters are manually adjusted to fit the model to experimental observations in both wild-type and mutant strains. See Supplementary Tables 2B-D for their values in wild-type cells.

### Experimental evidence for the Boolean model

Cell cycle progression through  $G_1$  phase is inhibited by  $Whi5$ , which is inactivated (phosphorylated) by  $Cln3$ ,  $Bck2$ ,  $Cln1$ , and  $Cln2$ <sup>25,26</sup> and activated (dephosphorylated) by  $Cdc14$  phosphatase. In our model (Fig. 1), the ‘ $Cln3$ ’ variable accounts for both  $Cln3$  and  $Bck2$  proteins, and the ‘ $Cln2$ ’ variable represents both  $Cln1$  and  $Cln2$ . In  $G_1$  phase ( $Cln3 = Cln2 = Cdc14 = 0$ ),  $Whi5$  is active, and it can be inactivated by either  $Cln3 = 1$  or  $Cln2 = 1$ .  $Whi5$  stays inactive throughout S- $G_2$ -M and is activated by  $Cdc14$  as the cell exits mitosis, provided either  $Cln3 = 0$  or  $Cln2 = 0$ . These interactions imply the logical function on Row 1 of Supplementary Table 3:

$Whi5 = (\text{NOT}Cdc14 \text{ AND NOT}(Cln2 \text{ OR } Cln3)) \text{ OR } (Cdc14 \text{ AND NOT}(Cln2 \text{ AND } Cln3))$ .

This logical function is implemented in our model by the ‘ $W$ ’ function in Row 1 of Table 1.

In late  $G_1$ ,  $Cln3$  is activated by cell growth, causing inactivation of  $Whi5$  and subsequent activation of the SBF and MBF transcription factors (Table 1, Rows 2 & 3). Activation of SBF and MBF defines the ‘START’ transition in the budding yeast cell cycle, after which yeast cells set off on an irreversible path to DNA synthesis, mitosis and cell division. At first, the transcription factors are kept active by  $Cln3$ -,  $Cln2$ - and  $Clb5$ -dependent kinase activities, but later they are inactivated by  $Clb2$ -dependent kinase<sup>27-30</sup>. In addition, MBF is regulated by a negative feedback loop with  $Nrm1$  (Table 1, Rows 3 & 4)<sup>31</sup>.

After the START transition, MBF and SBF activate the synthesis of  $Clb5$ ,  $Clb6$ , and  $Cln1$ ,  $Cln2$  cyclins, which are responsible for DNA replication (Row 5) and budding (Row 6), respectively<sup>32</sup>. (In our notation, ‘ $Clb5$ ’ represents both  $Clb5$  and  $Clb6$  cyclins, and ‘ $Clb2$ ’ represents both  $Clb1$  and  $Clb2$  cyclins.) The origin licensing variable  $ORI$ , is set to 0 (origins licensed) when both  $Clb5$  and  $Clb2$  are inactivated as a mother cell exits mitosis and divides, then  $ORI$  is flipped to 1 (DNA replication begins) when either  $Clb5$  or  $Clb2$  is activated in the next cell cycle (Eq. 4).

Once DNA replication is initiated,  $Clb5$  activity promotes the accumulation of active  $Clb1$  and  $Clb2$  cyclins by suppressing  $Sic1$  (Row 7) and  $Cdh1$  (Row 8) in late S phase<sup>33</sup>. Subsequently, a positive feedback loop with  $Mcm1$  transcription factor sets off rapid accumulation of  $Clb1$  and  $Clb2$  cyclins, which drive the cell into M phase<sup>34</sup> (Rows 9-11). The  $Clb2G$  variable represents  $Clb1$  and  $Clb2$  cyclin-dependent kinase activities in late S- and  $G_2$  phases, and  $Clb2M$  represents their higher activities in M phase. When  $Clb2M = 1$ , the cell enters M phase and  $Cdc5$  is activated (Row 12).

In M phase, the spindle assembly checkpoint (SAC) prevents the metaphase-to-anaphase transition until all sister chromatids achieve bipolar alignment on the mitotic spindle. The SAC turns ON ( $SAC = 0 \rightarrow 1$ ) when DNA replication begins ( $ORI = 1$ ) (Eq. 9). Progress in aligning the replicated chromosomes on the mitotic spindle is tracked by the  $SPN$  variable. When the cell enters M phase ( $Clb2M = 1$ ), the (continuous) variable  $SPN_t$  starts to increase (Eq. 10). When  $SPN_t$  reaches 1 (i.e., all chromosomes are aligned on the metaphase plate),  $SAC$  is set to zero.

Once the cell passes the SAC ( $SAC = 1 \rightarrow 0$ ),  $Cdc20$  is activated ( $Cdc20 = 0 \rightarrow 1$ ) and it promotes the metaphase-anaphase transition.  $SPN_t$  is reset to zero, and the Spindle Position Checkpoint (SPOC) is activated ( $SPOC = 0 \rightarrow 1$ ) to ensure that both mother and daughter cells receive a full set of chromosomes before cytokinesis<sup>35</sup>. Spindle positioning is monitored by the (continuous)  $SPO$  variable (Eq. 12). When  $SPO_t = 1$ , the SPOC is satisfied ( $SPOC = 1 \rightarrow 0$ ), and  $SPO$  is reset to 0.

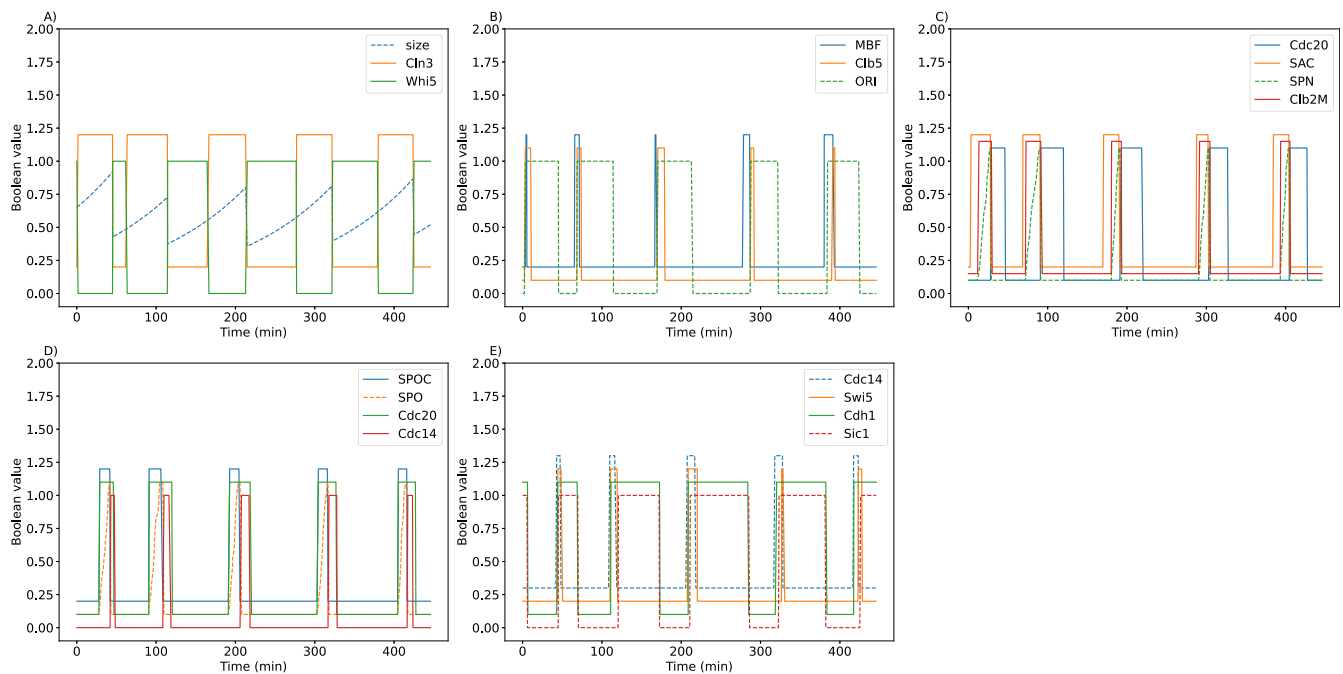
To exit mitosis,  $Cdc14$  must be fully activated (i.e., released from the RENT complex in the nucleolus), which is a consequence of both the FEAR and MEN pathways<sup>36</sup>. FEAR is activated when  $SAC \rightarrow 0$  and MEN when  $SPOC \rightarrow 0$  (Row 14). Finally,  $Cdc14$  activates  $Cdh1$  and  $Swi5$  (Rows 8 & 15), and  $Swi5$  (a transcription factor) activates  $Sic1$  (Row 7). Together  $Sic1$  and  $Cdh1$  reverse the activities of all cyclins. When  $Clb5$ ,  $Clb2M$  and  $Clb2G$  activities are destroyed by the combined actions of  $Cdc20$ ,  $Cdh1$  and  $Sic1$ , the mother cell divides according to the rule in Eq. 8, and the newborn cells re-enter  $G_1$ <sup>37</sup>.

### Description of mutant simulations

A mutant strain in which ‘gene K’ is deleted is modeled by setting the Boolean variable for ‘protein K’ = 0 for all time  $t > 0$ . A mutant strain overexpressing gene L from a GAL promoter is modeled by adding a constant  $\gamma_L$  to the Boolean variable for protein L; i.e., if protein L is chosen to be updated at time  $t$ , then Eq. 3 is modified to  $X_{L,t+\Delta t} = \hat{X}_{L,t+\Delta t} + \gamma_L$ . Note that this modification may change the logical Boolean function specified by any  $W_{i,t}$  for which the sum includes  $X_{L,t}$  (which now has the value  $\gamma_L$  or  $1 + \gamma_L$ ). Also, for all mutant strains with the GAL promoter, we set  $r = 0.0046 \text{ min}^{-1}$  and  $f_{\text{mean}} = 0.58$ .

For mutant strains that exhibit endoreplication and  $Cdc14$  endocycles, we also change the parameters of the gamma distribution (Eq. 2) in order to





**Fig. 2 | Simulation of wild-type budding yeast cell cycles.** Each subplot shows the dynamics of key cell cycle components for five division cycles. **A** Cell size and molecular components that regulate the progression through  $G_1$ . **B** The initiation of DNA replication. **C** Spindle assembly progress in response to the activation of Clb2M. **D** Spindle orientation progress after anaphase. **E** Cdc14 activation and resetting into  $G_1$ . All simulations follow the lineage of mother cells from initial

conditions corresponding to stationary  $G_1$  phase:  $Whi5=Sic1=Cdh1 = 1$ , all other variables = 0, and  $size=0.65$  (the average size of a mother cell just after division). Some of the variables are offset for clearer visualization:  $Clb5, SPN, Clb2M, SPO, Cdc20,$  and  $Cdh1$  are offset by 0.1;  $Cln3, MBF, SAC, SPOC$  and  $Swi5$  are offset by 0.2; and  $Cdc14$  is offset by 0.3.

match the period of oscillation to experimental observations. For example, for the *GAL-CLB2-dbΔ* mutant strain, Clb2 (which inhibits Cdh1) is non-degradable and its level is high; therefore, the activation of Cdh1 is delayed compared to wild-type cells with normal levels of Clb2. Also, because Clb2 activates Cdc5, the inactivation of Cdc5 is delayed when the level of Clb2 is high. Therefore, for this mutant strain, the gamma-distribution parameters are set to  $k = 15, \theta = 1.5$  min (mean = 22.5 min and CV = 0.25) for calculating  $\Delta t$  for Cdc5 inactivation and Cdh1 activation. Similar reasoning applies to the *clb1-5Δ* mutant strain that exhibits endoreplication cycles. In the absence of most Clbs, Cdh1 inactivation and Clb6 activation are delayed, and therefore, we set mean = 22.5 min and CV = 0.25 for the timing of these events in this mutant strain.

All parameter changes that are made to model mutant strains are summarized in Supplementary Tables 4 and 5.

## Results

To assess the potential of our method, we present simulations of wild-type cell cycles, of population-level properties of budding yeast cultures, and of mutant strains that exhibit aberrant cycles. We also use the model to predict phenotypes of mutant strains that have not yet been characterized experimentally.

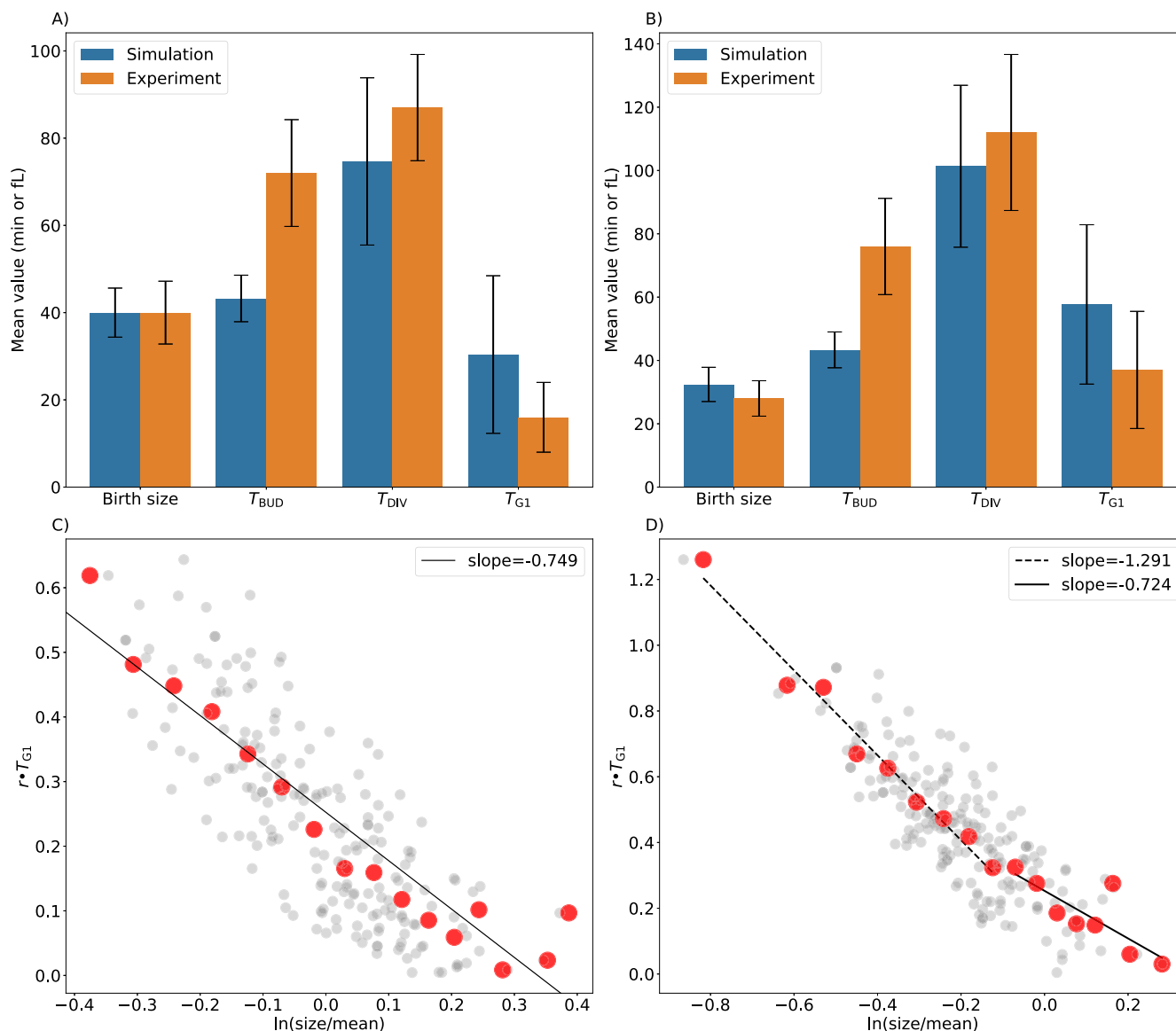
### Simulation of cell cycle progression in wild-type cells

Figure 2 shows simulations of key components regulating cell cycle events in wild-type budding yeast. Newborn yeast cells must grow to a ‘critical size’ in order to activate Cln3 and subsequently to inactivate Whi5, which then permits activation of SBF and MBF transcription factors (Fig. 2A). MBF induces the synthesis of Clb5, which induces DNA replication ( $ORI = 1$  identifies the onset of S phase, Fig. 2B). The spindle assembly progress variable ( $SPN$ ) indicates progression through  $G_2/M$  into metaphase. When  $SPN \geq 1$ , the spindle assembly checkpoint variable ( $SAC$ ) is set to zero, which allows the activation of Cdc20 and the cell to progress into anaphase (see Fig. 2C).

The spindle-orientation progress variable ( $SPO$ ) accounts for proper segregation of chromosomes into mother and daughter cell compartments during anaphase. When  $SPO \geq 1$ , the spindle orientation checkpoint ( $SPOC$ ) is set to zero, which allows Cdc14 to be fully released from the nucleolus (Fig. 2D). Once Cdc14 is released, it activates Cdh1 and Swi5 (which initiates synthesis of Sic1), thereby resetting the cell back to  $G_1$  (Fig. 2E).

Model simulations of wild-type cells are very robust with respect to perturbations of the 65  $\omega_{ij}$  coefficients in Supplementary Table 2A. Each one of them can be perturbed by at least 40% and the model predicts ‘viable’ wild-type cell cycles (see Supplementary Fig. 2A). When three specific coefficients ( $\omega_{Mcm1,Clb2G}, \omega_{Clb2M,Mcm1}$  and  $\omega_{Cdc20,Mcm1}$ ) are reduced by 50% (Supplementary Fig. 2B), the model fails to yield successful division cycles because, in each case, a crucial component (Mcm1, Clb2M and Cdc20, respectively) fails to activate.

Next, the model is used to simulate the exponential expansion of a population of budding yeast cells, in order to estimate the means and standard deviations of observable cell-cycle measures: the period from birth to division ( $T_c$ ), the duration of  $G_1$  phase from mitotic exit to S phase ( $T_{G1}$ ), the period from budding to division ( $T_{bud}$ ), and cell size at birth. We compare these simulation results with corresponding experimental data from ref.<sup>24</sup> in Fig. 3. Overall, the model accurately simulates these population-level properties in both mother cells (Fig. 3A) and daughter cells (Fig. 3B), although the model overestimates  $T_{G1}$  variability in both mother and daughter cells. Similar discrepancies were observed in simulations based on previous models<sup>21,38</sup>. Figure 3C and D show the joint distributions of size-at-birth and  $T_{G1}$  for mother and daughter cells. As in ref.<sup>24</sup>, we plot  $r \cdot T_{G1}$  vs.  $\ln(size/mean)$ , where (in glucose medium)  $r = 0.0077 \text{ min}^{-1}$  and ‘mean’ = mean size of mother cells at birth = 40 fL = 0.55 in units of dimensionless size. The simulated data points are fitted with trendlines, as was done to analyze the experimental data<sup>24</sup>. Although the estimated slopes of the trendlines for mother cells (−0.74) and daughter cells (large −0.72 & small −1.29) show similar trends to the experimental data (slope = −0.1 for



**Fig. 3 | Population-level statistics from model simulations and experimental observations.** The mean values and coefficients of variation for four cell-cycle properties in populations of mother cells (A) and daughter cells (B). To transform dimensionless cell size into volume in fL, we used a conversion factor of 75 fL, which was derived by equating the mean size of mother cells to the experimental mean

volume of ~40 fL. To visualize the joint distributions of size at birth and  $G_1$  duration ( $T_{G1}$ ) in mother cells (C) and daughter cells (D), we plot (grey dots) 200 simulated cells sampled from the whole population. To estimate the trends in the data, we plot (red dots) the average value of the grey dots in bins of size 2 fL, exactly as implemented by the authors of the experimental data<sup>24</sup>.

mother cells;  $-0.3$  &  $-0.7$  for large & small daughter cells, respectively), the theoretical slopes are considerably more negative than observations warrant, suggesting that size-control in the model is considerably stronger than in reality.

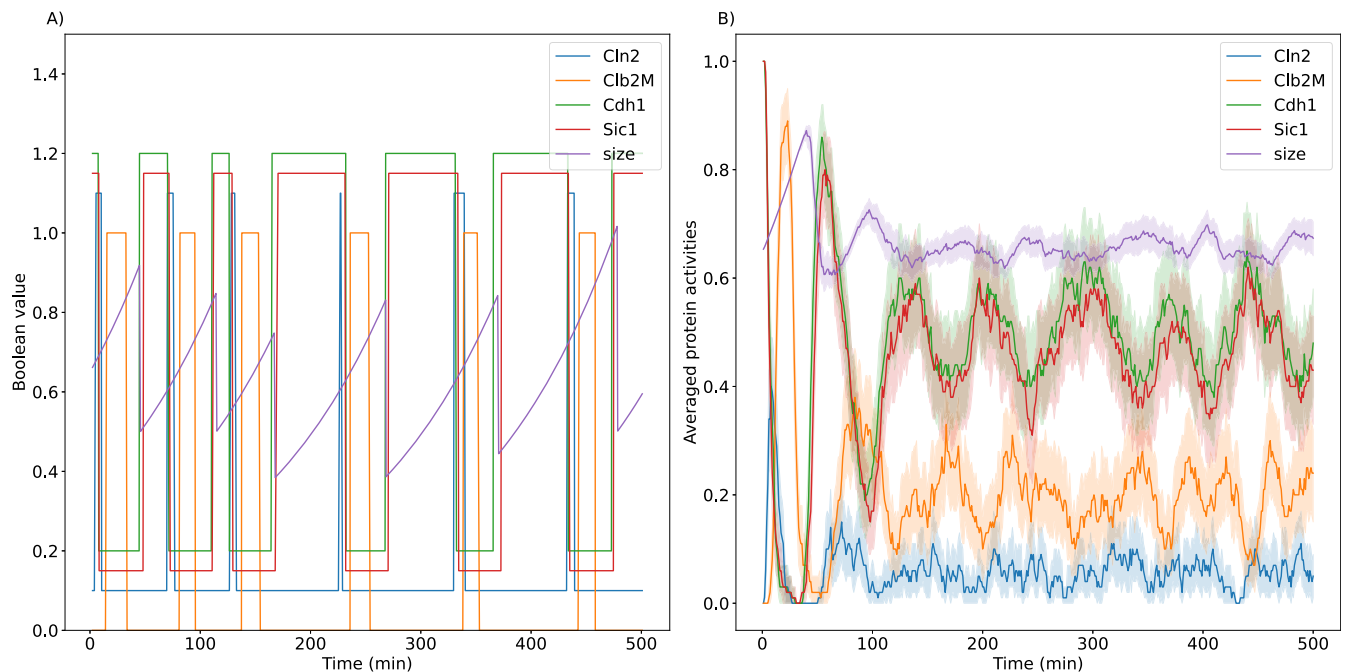
To estimate how fast a simulated population of cells loses synchrony over time, 100 cells with average size of 0.65 were initiated in the  $G_1$  state and tracked (both mother and daughter cells) for 500 min (Fig. 4A). Cell size and protein activities of all cells extant at time  $t$  were averaged and the results plotted as functions of  $t$  (Fig. 4B). The population-average results clearly show a loss of synchrony that agrees well with observations<sup>39</sup>. Protein activities and cell size quickly lose synchrony due to the unequal division of material between daughter and mother cells.

Some of the mutant strains discussed in later sections are grown in galactose medium, rather than glucose. To simulate this change of growth medium, we change the specific growth rate  $r$  to  $0.0046 \text{ min}^{-1}$  (mass doubling time = 150 min, typical for growth on galactose) and the division fraction  $f_{\text{mean}}$  to 0.58 (cell division is more asymmetric in poorer growth

medium). With these changes, the interdivision times in galactose medium for mother and daughter cells are computed to be 113 min and 184 min, respectively, in good agreement with the observations of Lord and Wheals (1980).

### Simulation of known mutant phenotypes

To test the accuracy of our model in accounting for mutant phenotypes, we simulated cell cycle progression in 48 experimentally characterized mutant strains, including gene deletion and overexpression mutants. Of these strains, our model agrees with 40 observed phenotypes; see Supplementary Table 4. The model's success rate for predicting phenotypes of mutant strains depends on our choice of the 65  $\omega_{ij}$  parameters in Supplementary Table 2A, because these parameters determine the precise logical functions at play in wild type and mutant cells. We have obtained a success rate of 83% by manual adjustment of the parameters. Presumably a higher success rate could be achieved by an automated parameter estimation procedure, but we defer this step to later developments of the modeling approach.



**Fig. 4 | Loss of synchrony of protein activities from model simulations.** **A** A single cell is tracked over time from one division to the next. *Cdh1*, *Sic1* and *Cln2* are offset by 0.2, 0.15 and 0.1, respectively. **B** Many such simulations are combined to predict

the loss of synchrony in a population of cells. Solid line = average activity, vertical bar indicates  $\pm$  95% Confidence Interval.

In this subsection, we focus on mutant strains exhibiting aberrant cycles. For the *clb1-5Δ* strain, in which all Clbs—except Clb6—are deleted, cells replicate the genome multiple times without mitosis<sup>3</sup>, a phenotype called endoreplication. Because the Clb5 variable represents both Clb5 and Clb6, the action of Clb6 in *clb1-5Δ* mutant is simulated by reducing the basal parameter  $\omega_{Clb5}$  to  $-1.1$ , and the values of parameters  $\omega_{i,Clb5}$  (which describe the influence of Clb5 and Clb6 on target protein *i*) were reduced by eight-fold (see Supplementary Table 4). Because MBF-induced transcription of the *CLB6* gene is slower than the *CLB5* gene, the accumulation of Clb6 protein is delayed by 22.5 min compared to 0.9 min in wild type cells. Also, due to absence of most Clbs in the *clb1-5Δ* mutant, Cdh1 inactivation is delayed by 22.5 minutes compared to 0.9 minutes in wild type. As Fig. 5A shows, after  $t = 250$  min when all Clbs except Clb6 are deleted, the *clb1-5Δ* mutant fails to enter mitosis and to divide, and cell size becomes extremely large. However, MBF, Clb6, Cdh1 and Nrm1 continue to oscillate, driven by the negative feedback loop MBF  $\rightarrow$  Clb6  $\neg$  Cdh1  $\neg$  Nrm1  $\neg$  MBF. Figure 5B shows the distribution of endoreplication periods. The estimated period of Clb6 oscillations,  $57.9 \pm 14.3$  min, is in good agreement with experimental observations<sup>3</sup>.

The mean delays we have introduced (22.5 min for the accumulation of Clb6 and the inactivation of Cdh1) are somewhat arbitrary. The experimental observation (that the period of endoreplication is  $\sim 60$  min) only fixes the total delay to be  $\sim 45$  min (see Supplementary Fig. 3), which we split evenly between the two processes.

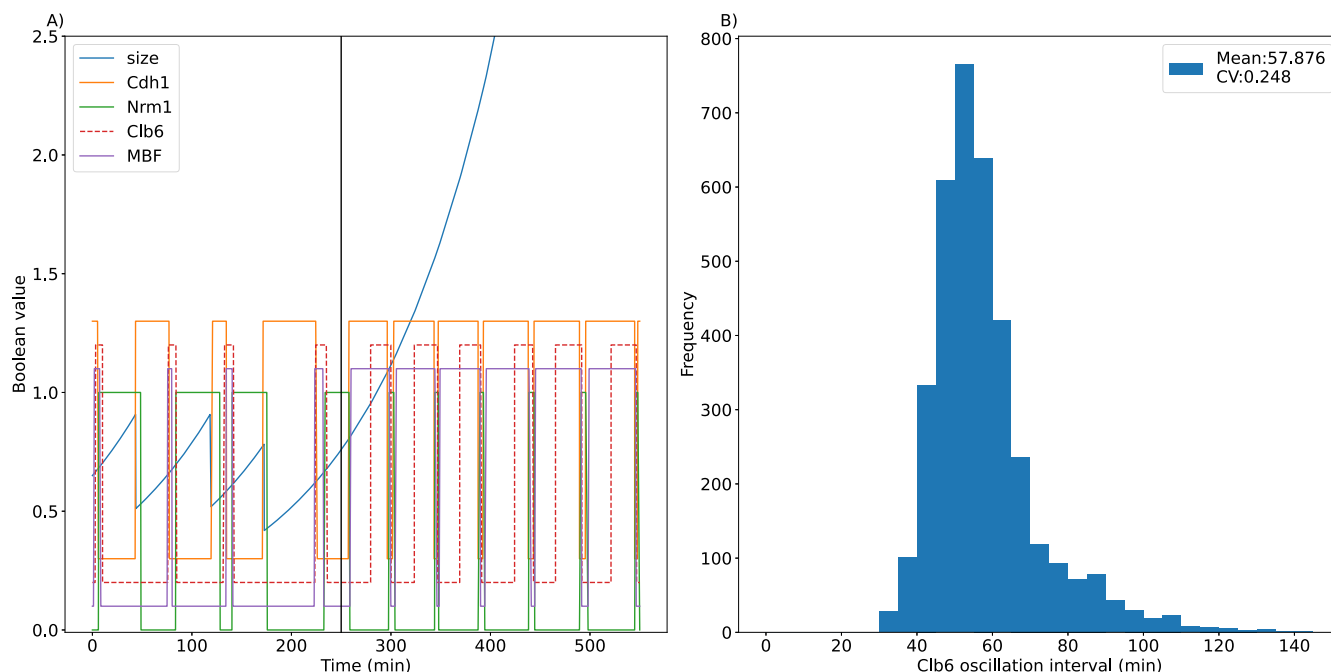
Endoreplication cycles in our model disappear when any component (i.e., Cdh1, Nrm1, MBF or Clb6) in the negative feedback loop is deleted. In addition, Cln3, Cln2, and SBF are essential for endoreplication cycles. Further, we tested all pairwise deletions of model components in *clb1-5Δ* mutant strain to identify other mutant strains that might exhibit endoreplication cycles. Mutant strains that lose or retain endoreplication cycles are shown in Supplementary Fig. 1A.

At this point we note that, despite the fact that MBF is a transcription factor for Nrm1 (Fig. 1), our logical function for updating Nrm1 (Supplementary Table 3, row 4) is  $Nrm1 = NOTCdh1$ , independent of whether MBF is active or not. In other words, with our choice of  $\omega_{Nrm1} = 0.5$ , even if  $MBF = 1$ , Nrm1 cannot accumulate if  $Cdh1 = 1$ . If we were to choose

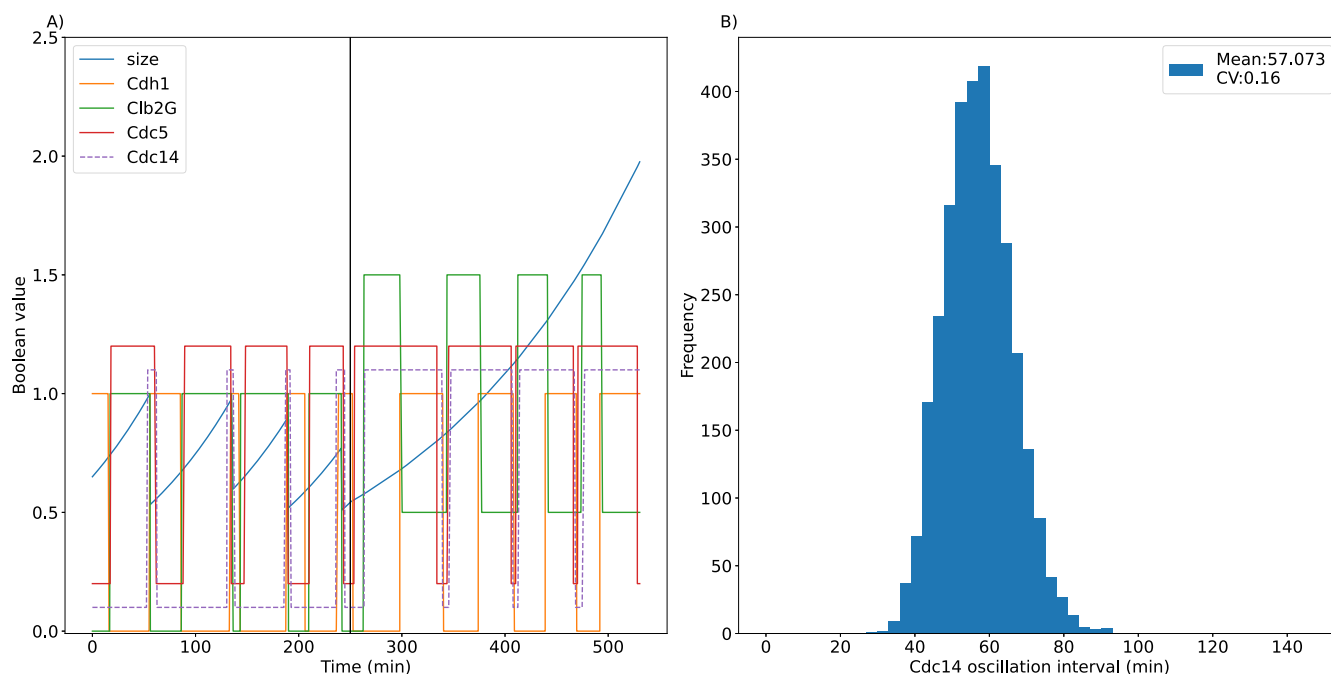
$\omega_{Nrm1} = 1.5$ , then the logical function would be  $Nrm1 = MBF \text{ OR } NOTCdh1$ , because MBF-induced expression of Nrm1 would overwhelm its degradation initiated by Cdh1. In this case ( $\omega_{Nrm1} = 1.5$ ), the *clb1-5Δ* mutant strain still exhibits endoreplication, but the distribution of Clb6 oscillation periods is bimodal (Supplementary Fig. 4). In these mutant cells,  $Cln3 = 1$ ,  $Clb2G = Clb2M = 0$ , and  $\omega_{MBF,Clb5} = 3/8$ ; hence,  $W_{MBF} > 0$  if  $Nrm1 = 0$  and  $< 0$  if  $Nrm1 = 1$ , so  $MBF = NOTNrm1$ . In this case, when  $Cdh1 = 1$ , the regulatory network has a short, negative feedback loop ( $Nrm1 = MBF$ ,  $MBF = NOTNrm1$ ) that, under asynchronous updating, can oscillate a few times ( $Nrm1 \text{ ON}$ ,  $MBF \text{ OFF}$ ,  $Nrm1 \text{ OFF}$ ,  $MBF \text{ ON}$ , ...) before MBF is selected to turn on Clb6 and the cell proceeds to endoreplicate. This effect is evident in the temporal simulations of MBF and Nrm1 in Supplementary Fig. 5. With our preferred choice of  $\omega_{Nrm1} = 0.5$ , the short, negative feedback loop is broken and the endoreplication cycles are more uniform (see Supplementary Fig. 6). Furthermore, with the alternative parameter choice ( $\omega_{Nrm1} = 1.5$ ), in the strain *clb1-5Δ cln2Δ*, Cdh1 is constitutively active and the short NFL is persistently in play; hence, the cells exhibit rapid endoreplication cycles with a period of 6 updates. To avoid these unlikely results, we prefer the choice  $\omega_{Nrm1} = 0.5$  and the logical function  $Nrm1 = NOTCdh1$ .

Another mutant strain that exhibits aberrant cycles is *GAL-CLB2-dbΔ*. The high level of non-degradable Clb2 causes cell cycle arrest in mitosis. Although the cells cannot exit mitosis, the high level of Clb2 supports the activation of Cdc5 which promotes Cdc14 release, Cdc14 then activates Cdh1 which degrades Cdc5. This negative loop ( $Cdc5 \rightarrow Cdc14 \rightarrow Cdh1 \neg Cdc5$ ) results in Cdc14 endocycles (Fig. 6A). Because Clb2 activity is high in this mutant strain, the time delays for inactivation of Cdc5 and for activation of Cdh1 are increased to 22.5 min. Figure 6B shows the distribution of endocycle periods. The averaged period of Cdc14 oscillations is  $57.1 \pm 9.1$  min, in agreement with experimental observations<sup>12</sup>. Supplementary Figure 1B shows that Cdc14 endocycles depend not only on *CDC5*, *CDH1* and *CDC14* genes, as expected, but also on *CLB2*, *MCM1* and *CDC20* genes, and, furthermore, that the double-deletion strain, *cdc20Δ mcm1Δ*, restores Cdc14 endocycles. The genetic dependencies of Cdc14 endocycles are described as follows.

1. *CDC20*. Clb2M activity oscillates between 0 and 1 during normal cell cycles.



**Fig. 5 | Simulation of endoreplication cycles in the *clb1-5Δ* mutant strain. A** Wildtype cell cycles for  $t < 250$  min, and endoreplication cycles for  $t > 250$  min. *Cdh1*, *Clb6* and *MBF* are offset by 0.30, 0.20, and 0.10, respectively. **B** The distribution *Clb6* oscillation periods.



**Fig. 6 | The simulation of *Cdc14* endocycles in the *GAL-CLB2-dbΔ* mutant strain. A** Wildtype cell cycles for  $t < 250$  min, and *Cdc14* endocycles for  $t > 250$  min. As before, *Cdc5* and *Cdc14* are offset by 0.20 and 0.10, respectively, for clearer visualization. On the other hand, *Clb2G* (the green curve) is not offset for plotting purposes but rather because *Clb2* protein is overproduced in this mutant strain. **B** The distribution of *Cdc14* oscillation periods.

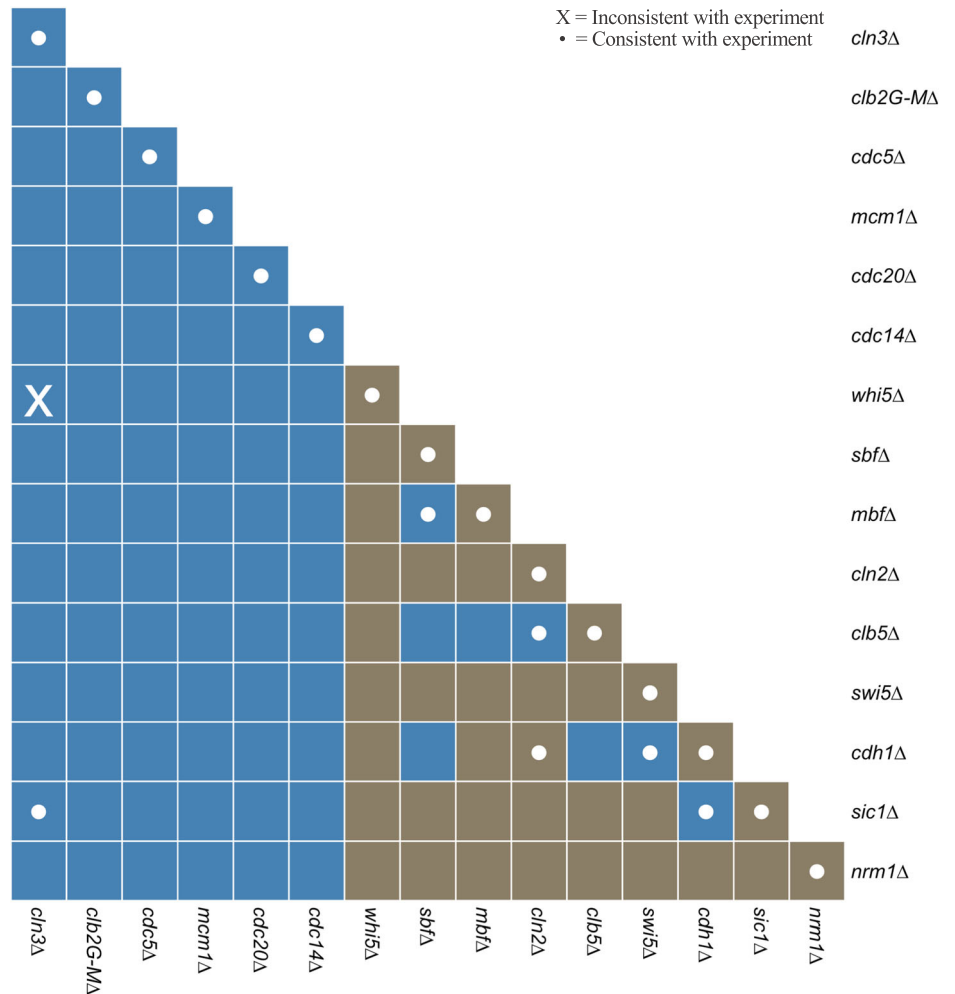
2. *cdc20Δ*. *Clb2M* = 1, and the cell cannot exit mitosis because *Cdh1* and *Sic1* are inhibited.
3. *GAL-CLB2-dbΔ*. *Clb2M* activity oscillates between 0.5 and 1.5, which supports *Cdc14* endocycles but blocks return to  $G_1$ .
4. *GAL-CLB2-dbΔ cdc20Δ*. *Clb2M* activity is constitutively high in this strain because *Mcm1* is constitutively active, and *Clb2M* = 1.5 blocks *Cdc14* endocycles by inhibiting *Cdh1* and activating *Cdc5*.
5. *GAL-CLB2-dbΔ cdc20Δ mcm1Δ*. In the absence of both *Mcm1* and *Cdc20*, *Clb2M* = 0.5, which is a ‘sweet spot’ for *Cdc14* endocycles.

**Model predictions**

Next, we used the model to predict the viability of mutant strains with single and double deletions of model variables. Because some model variables already represent two genes ( $MBF = SWI6 + MBP1$ ,  $SBF = SWI6 + SWI4$ ,  $CLN3 = CLN3 + BCK2$ ,  $CLN2 = CLN1 + CLN2$ ,  $CLB5 = CLB5 + CLB6$ , and  $CLB2 = CLB1 + CLB2$ ), combining their deletions with the deletion of other model components generates triple- or quadruple-deletion strains of budding yeast. The model, with 15 genetic components, correctly predicts the viability/inviability of the 15 ‘single’ deletion strains (Fig. 7). Of the



**Fig. 7 | Predicted phenotypes of double-deletion strains.** Each rectangular cell corresponds to a combination of two deleted components listed along the vertical and horizontal axes. Blue = inviable, brown = viable phenotype. A white dot indicates that the prediction is consistent with experimental observations, and X indicates a difference between the simulated and observed viability. The elements along the diagonal correspond to single-deletion strains.



15×14/2 = 105 ‘double’ deletion strains, only seven combinations have been characterized experimentally, and the model correctly predicts six of them; only the *cln3Δ bck2Δ whi5Δ* strain is incorrectly predicted to be inviable (see also, row 26 of Supplementary Table 4). The viability of this strain indicates that the only essential function of Cln3+Bck2 is to inactivate Whi5, but in our model, even when *Whi5* = 0, Cln3 must be activated (by cell growth) in order to activate SBF and MBF at START. In the cell (but not in our model), this role of Cln3 is most likely backed up by Cln2.

The model predicts that eight double-deletion strains are synthetically lethal (the eight blue squares in the brown ‘background’ in Fig. 7). Four of these strains are already known to be synthetic lethal; the other four combinations, which are genuine predictions of the model, are described below.

1. *cdh1Δ clb5Δ*. In *cdh1Δ*, SBF compensates for the inhibition of MBF by Nrm1. In *clb5Δ*, Cln2 is available to inhibit Cdh1 and Sic1. In *cdh1Δ clb5Δ*, Nrm1 and Sic1 are permanently active, and Cln2 and Clb2 are permanently repressed.
2. *cdh1Δ sbfΔ*. In *cdh1Δ*, SBF compensates for the inhibition of MBF by Nrm1. In *sbfΔ*, MBF drives Clb5 synthesis and cell cycle progression. In *cdh1Δ sbfΔ*, SBF is absent and MBF is inhibited by Nrm1, so the cell arrests in G<sub>1</sub>.
3. *clb5Δ sbfΔ*. In *clb5Δ*, Cln2 is available to inhibit Cdh1 and Sic1. In *sbfΔ*, MBF drives Clb5 synthesis and cell cycle progression. In *clb5Δ sbfΔ*, neither Clb5 nor Cln2 are available to inactivate Cdh1 and Sic1, so the cell is arrested in G<sub>1</sub>.
4. *clb5Δ mbfΔ*. In *clb5Δ*, Cln2 is available to inhibit Cdh1 and Sic1. In *mbfΔ*, SBF drives Clb5 synthesis and cell cycle progression. Because Cln2 synthesis requires both SBF and MBF activities in our model, the

*clb5Δ mbfΔ* strain, lacking both Clb5 and Cln2, cannot inactivate Cdh1 and Sic1, so cells arrest in G<sub>1</sub>.

These explanations suggest that the synthetic lethal strains, *cdh1Δ clb5Δ* and *cdh1Δ sbfΔ*, should be rescued by further deletion of *NRM1*. The triple mutants, *cdh1Δ clb5Δ nrm1Δ* and *cdh1Δ sbfΔ nrm1Δ*, are indeed viable in our model simulations (Supplementary Fig. 7).

## Discussion

In this study, we developed a stochastic Boolean model of the budding yeast cell cycle that can correctly explain both normal cell cycle progression and aberrant cycles (endoreplication and Cdc14 endocycles). Understanding the mechanism of aberrant cycles and identifying controls that suppress these cycles in wild-type cells is important because these cycles are found in many types of cancer<sup>40</sup>. Our model suggests that endoreplication cycles are driven by the negative feedback loop MBF → Clb6 -| Cdh1 -| Nrm1 -| MBF, and that Cdc14 endocycles are driven by the negative loop Cdc5 → Cdc14 → Cdh1 -| Cdc5. Although our model has not yet been optimized to account for mutant phenotypes, it already accounts for the viability/inviability of 40 gene deletion and overexpression mutant strains out of 48 strains tested. Furthermore, we used the model to predict the phenotypes of 120 mutant strains carrying one-, two-, three- or four gene deletions.

The stochastic Boolean model proposed here is closely related to the stochastic differential equation (SDE) approach proposed in Laomettachtit et al.<sup>38</sup> and applied in great detail to the budding yeast cell cycle in Kraikivski et al.<sup>8</sup>. At the core of the SDE approach is a nonlinear ODE model of a protein

signaling network:

$$\frac{1}{\beta_i} \frac{dY_i}{dt} = H(\sigma_i W_i) - Y_i, \text{ where } W_i = \omega_{i0} + \sum_j \omega_{ij} Y_j, \text{ and } H(\sigma x) = \frac{1}{1 + e^{-\sigma x}} \quad (13)$$

where  $\beta_i$ ,  $\sigma_i$ , and  $\omega_{ij}$ 's are parameters. Equations 1 and 13 are closely related because  $H(\sigma x) \rightarrow \text{Heav}(x)$  as  $\sigma \rightarrow \infty$ . With this identification in mind, it would be (relatively) easy to translate a stochastic Boolean model (in the format proposed here) to a continuous, nonlinear SDE model of the type proposed by Laomettachtit et al.<sup>38</sup>. Indeed, some predictions by Kraikivski et al.<sup>8</sup> are consistent with our stochastic Boolean model: e.g., the *cln1Δ cln2Δ swi6Δ mbp1Δ* strain is predicted to be viable, whereas the *cln3Δ bck2Δ cdh1Δ* strain is inviable.

It is instructive to compare our model to the Boolean models of Li et al.<sup>16</sup>, Irons<sup>19</sup> and Fauré et al.<sup>18</sup>, see Supplementary Table 6. Li et al.<sup>16</sup> were primarily concerned with demonstrating the 'robustness' of cell cycle progression: *START* → DNA synthesis & budding → prophase-metaphase-anaphase-telophase → early G<sub>1</sub> arrest. In addition to emphasizing the robustness of the cell-cycle control system, Irons focused on 'sub-network analysis,' identifying a 'core network' (his Fig. 6C) that is very similar to the four-variable ODE model of Battogtokh and Tyson<sup>41</sup> (their Fig. 5, where their 'Hct1' plays the same role as Iron's 'CKI'). Fauré et al.<sup>18</sup> were concerned to show how simple logical models could be 'composed' into a single 'comprehensive' model that rivals a detailed ODE model<sup>6</sup> in terms of successfully simulating mutant cell phenotypes. We are primarily concerned with modifying the Boolean approach to allow for stochastic modeling in terms of real time and real cell-size variables, in order to compare model simulations with experimental data, as well as to simulate the phenotypes of ~150 mutant strains (40% known experimentally and 60% novel predictions).

All four models follow the time evolution of 10 dynamical variables representing the core cell-cycle regulatory proteins Cdh1, Sic1, Cln3/Bck2, SBF/MBF, Cln1/2, Clb5/6, Mcm1, Swi5, Clb1/2 and Cdc20. Irons adds Cdc14, Yhp1/Yox1 (a transcriptional repressor of Cln3), and two 'pathways' FEAR and MEN. Fauré et al.<sup>18</sup> ignore Yhp1/Yox1 and unpack the FEAR and MEN pathways. We add Whi5, Nrm1 and Cdc5. Like Li et al.<sup>16</sup>, we define the protein interaction network through  $\text{Heav}(W_i)$  functions, whereas Irons and Fauré et al.<sup>18</sup> use logical functions. Li/Irons/Fauré update their models synchronously and deterministically in discrete time steps, whereas we update our model asynchronously and stochastically in real time (min). Irons introduces 'dummy' variables to simulate time delays in protein synthesis and degradation and in cell cycle events; Fauré et al.<sup>18</sup> achieve similar effects by assuming that ~40% of their proteins (e.g., Clb2, Clb5 and Cdc20) are multi-state variables (2 or 3 'active' states). We have only one multi-state variable: Clb2G (low activity of Clb2 in G<sub>2</sub> phase) and Clb2M (higher activity form in M phase). We account for time delays by choosing  $\Delta t$  (updating intervals) from a gamma distribution with mean and CV chosen to fit observed temporal progression through the cell cycle in wild-type and endocycling mutants. Only our model has a real variable 'size(t)' to track cell growth. To track cell growth and division, Fauré introduced two multistate variables, called 'MASS' and 'CYTOKINESIS,' which were updated separately from the other variables by a set of 'priority rules.' Li's model goes through a sequence of 13 states from *START* (activation of Cln3) to 'Stationary G<sub>1</sub>' (*Cdh1 = Sic1 = 1*, all other variables = 0), and it 'cycles' only if *Cln3 = 0* in Stationary G<sub>1</sub> is flipped to 1 'externally' (say, by cell growth). Iron's model goes through a repetitive sequence of 19 states, driven by a negative feedback loop, *Cln3* → *SBF/MBF* → *Yhp1/Yox1* —| *Cln3*, with a lengthy time delay. Fauré's model has a 22-state cycle because the activation of MASS in early G<sub>1</sub> phase drives the *START* transition. Li et al.<sup>16</sup> discussed checkpoints briefly but did not consider mutant phenotypes. Irons' model is consistent with the phenotypes (viable/inviable) of 13 deletion mutants and with 4 types of checkpoints ('*START*,' 'morphogenesis,' 'spindle assembly,' and 'DNA damage'). Fauré et al.<sup>18</sup> successfully simulated 135 mutant strains by defining alternative logical rules for particular components of the network, to account for the genetic changes in each mutant strain. Using our model, we analyzed both

deletion and overproduction mutants (correctly predicting the phenotypes of 40/48 strains) and simulated mutant strains exhibiting endoreplication (multiple rounds of DNA replication) and Cdc14 endocycles (multiple attempts at mitotic exit). We also predicted the phenotypes of 'double deletion' mutants: of 105 such strains, 9 phenotypes are known experimentally, and our model agrees with 8 of them. In addition, our model is stochastic, so it can be used to predict cell-cycle time distributions, cell size distributions, and correlations between birth size and time spent in G<sub>1</sub> phase. Our model simulations are in good agreement with most of these statistical measures.

When updating Boolean models asynchronously, there are subtle timing issues that must be recognized. In our approach (which was motivated by Gillespie's stochastic simulation algorithm), we first identify those variables that potentially change in the next iteration and then choose one of them at random to actually change. Alternatively, in each iteration one might choose at random any variable to be updated, whether or not it changes state. The latter method avoids certain unrealistic consequences of the former method; e.g., the period of a negative feedback oscillator (in terms of number of iterations) may depend on how other variables could potentially change state, whether or not the 'other variables' are causally connected to the oscillator variables. For example, in our model, Cdc5, Cdc14 and Cdh1 are engaged in a negative feedback loop, but potential oscillations of this loop are suppressed by the overpowering cycle of ClbM activation during normal cell division cycles. However, in the *GAL-CLB2-dbΔ* mutant strain, the constitutively high activity of ClbM allows the Cdc5-Cdc14-Cdh1 NFL to oscillate freely, albeit at an unrealistically high frequency. To match the Cdc14 oscillation period to observations, we 'renormalize' the transition times by increasing the mean delays of the gamma distributions. This parameter change is, in a sense, the price we pay to fit a simple, stochastic Boolean model to experimental data in real time.

Our modeling approach supplements the simplicity of Boolean models with quantitative details, such as real continuous time, and with easily interpretable, adjustable parameters that are helpful in accounting for mutant phenotypes. Therefore, our approach lies somewhere between Boolean models that lack quantitative details necessary to explain experimental observations and ODE models that provide all these details at the expense of estimating many obscure kinetic rate constants. In our approach, only the parameters that determine the delays must be estimated by fitting model simulations to experimental data. Also, our approach incorporates stochastic effects at minimal computational cost.

## Data availability

The generated data for reproducing all figures in this study are available on GitHub: <https://github.com/Ktaoma/A-continuous-time-Boolean-model-of-the-endocycle-events-in-budding-yeast>.

## Code availability

The model is available in BioModels<sup>42</sup> with the identifier MODEL2405030001 (<https://www.ebi.ac.uk/biomodels/MODEL2405030001>). Computer codes to simulate all figures are available on GitHub: <https://github.com/Ktaoma/A-continuous-time-Boolean-model-of-the-endocycle-events-in-budding-yeast>.

Received: 12 January 2024; Accepted: 6 October 2024;

Published online: 18 October 2024

## References

1. Lu, Y. & Cross, F. R. Periodic cyclin-Cdk activity entrains an autonomous Cdc14 release oscillator. *Cell* **141**, 268–279 (2010).
2. Manzoni, R. et al. Oscillations in Cdc14 release and sequestration reveal a circuit underlying mitotic exit. *J. Cell Biol.* **190**, 209–222 (2010).
3. Simmons Kovacs, L. A. et al. Cyclin-dependent kinases are regulators and effectors of oscillations driven by a transcription factor network. *Mol. Cell* **45**, 669–679 (2012).
4. Shu, Z., Row, S. & Deng, W. M. Endoreplication: the good, the bad, and the ugly. *Trends Cell Biol.* **28**, 465–474 (2018).

5. Edgar, B. A., Zielke, N. & Gutierrez, C. Endocycles: a recurrent evolutionary innovation for post-mitotic cell growth. *Nat. Rev. Mol. Cell Biol.* **15**, 197–210 (2014).
6. Chen, K. C. et al. Integrative analysis of cell cycle control in budding yeast. *Mol. Biol. Cell* **15**, 3841–3862 (2004).
7. Chen, K. C. et al. Kinetic analysis of a molecular model of the budding yeast cell cycle. *Mol. Biol. Cell* **11**, 369–391 (2000).
8. Kraikivski, P., Chen, K. C., Laomettachit, T., Murali, T. M. & Tyson, J. J. From START to FINISH: computational analysis of cell cycle control in budding yeast. *NPJ Syst. Biol. Appl* **1**, 15016 (2015).
9. Barberis, M., Klipp, E., Vanoni, M. & Alberghina, L. Cell size at S phase initiation: an emergent property of the G1/S network. *PLoS Comput Biol.* **3**, e64 (2007).
10. Hancioglu, B. & Tyson, J. J. A mathematical model of mitotic exit in budding yeast: the role of Polo kinase. *PLoS One* **7**, e30810 (2012).
11. Vinod, P. K. et al. Computational modelling of mitotic exit in budding yeast: the role of separase and Cdc14 endocycles. *J. R. Soc. Interface* **8**, 1128–1141 (2011).
12. Howell, R. S. M., Klemm, C., Thorpe, P. H. & Csikász-Nagy, A. Unifying the mechanism of mitotic exit control in a spatiotemporal logical model. *PLoS Biol.* **18**, e3000917 (2020).
13. Karlebach, G. & Shamir, R. Modelling and analysis of gene regulatory networks. *Nat. Rev. Mol. Cell Biol.* **9**, 770–780 (2008).
14. Tyson, J. J., Laomettachit, T. & Kraikivski, P. Modeling the dynamic behavior of biochemical regulatory networks. *J. Theor. Biol.* **462**, 514–527 (2019).
15. Barik, D., Ball, D. A., Peccoud, J. & Tyson, J. J. A stochastic model of the yeast cell cycle reveals roles for feedback regulation in limiting cellular variability. *PLoS Comput Biol.* **12**, e1005230 (2016).
16. Li, F., Long, T., Lu, Y., Ouyang, Q. & Tang, C. The yeast cell-cycle network is robustly designed. *Proc. Natl Acad. Sci. USA* **101**, 4781–4786 (2004).
17. Davidich, M. I. & Bornholdt, S. Boolean network model predicts knockout mutant phenotypes of Fission yeast. *PLoS One* **8**, e71786 (2013).
18. Fauré, A. et al. Modular logical modelling of the budding yeast cell cycle. *Mol. Biosyst.* **5**, 1787 (2009).
19. Irons, D. J. Logical analysis of the budding yeast cell cycle. *J. Theor. Biol.* **257**, 543–559 (2009).
20. Stoll, G., Viara, E., Barillot, E. & Calzone, L. Continuous time boolean modeling for biological signaling: application of Gillespie algorithm. *BMC Syst. Biol.* **6**, 1–18 (2012).
21. Laomettachit, T., Kraikivski, P. & Tyson, J. J. A continuous-time stochastic Boolean model provides a quantitative description of the budding yeast cell cycle. *Sci. Rep.* **12**, 20302 (2022).
22. Novak, B. & Tyson, J. J. Mitotic kinase oscillation governs the latching of cell cycle switches. *Curr. Biol.* **32**, 2780–2785.e2 (2022).
23. Gillespie, D. T. Stochastic simulation of chemical kinetics. *Annu Rev. Phys. Chem.* **58**, 35–55 (2007).
24. Di Talia, S., Skotheim, J. M., Bean, J. M., Siggia, E. D. & Cross, F. R. The effects of molecular noise and size control on variability in the budding yeast cell cycle. *Nature* **448**, 947–951 (2007).
25. Skotheim, J. M., Di Talia, S., Siggia, E. D. & Cross, F. R. Positive feedback of G1 cyclins ensures coherent cell cycle entry. *Nature* **454**, 291–296 (2008).
26. de Bruin, R. A. M., McDonald, W. H., Kalashnikova, T. I., Yates, J. & Wittenberg, C. Cln3 activates G1-specific transcription via phosphorylation of the SBF bound repressor Whi5. *Cell* **117**, 887–898 (2004).
27. Charvin, G., Oikonomou, C., Siggia, E. D. & Cross, F. R. Origin of irreversibility of cell cycle start in budding yeast. *PLoS Biol.* **8**, e1000284 (2010).
28. Dirick, L. & Nasmyth, K. Positive feedback in the activation of G1 cyclins in yeast. *Nature* **351**, 754–757 (1991).
29. Koch, C., Schleiffer, A., Ammerer, G. & Nasmyth, K. Switching transcription on and off during the yeast cell cycle: Cln/Cdc28 kinases activate bound transcription factor SBF (Swi4/Swi6) at start, whereas Clb/Cdc28 kinases displace it from the promoter in G2. *Genes Dev.* **10**, 129–141 (1996).
30. Nasmyth, K. Control of the yeast cell cycle by the Cdc28 protein kinase. *Curr. Opin. Cell Biol.* **5**, 166–179 (1993).
31. Ostapenko, D. & Solomon, M. J. Anaphase promoting complex-dependent degradation of transcriptional repressors Nrm1 and Yhp1 in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **22**, 2175–2184 (2011).
32. Wittenberg, C. & Reed, S. I. Cell cycle-dependent transcription in yeast: promoters, transcription factors, and transcriptomes. *Oncogene* **24**, 2746–2755 (2005).
33. Yeong, F. M., Lim, H. H., Wang, Y. & Surana, U. Early expressed Clb proteins allow accumulation of mitotic cyclin by inactivating proteolytic machinery during S phase. *Mol. Cell Biol.* **21**, 5071–5081 (2001).
34. Maher, M., Cong, F., Kindelberger, D., Nasmyth, K. & Dalton, S. Cell cycle-regulated transcription of the CLB2 gene is dependent on Mcm1 and a ternary complex factor. *Mol. Cell Biol.* **15**, 3129–3137 (1995).
35. Caydasi, A. K., Ibrahim, B. & Pereira, G. Monitoring spindle orientation: spindle position checkpoint in charge. *Cell Div.* **5**, 28 (2010).
36. Queralt, E. & Uhlmann, F. Cdk-counteracting phosphatases unlock mitotic exit. *Curr. Opin. Cell Biol.* **20**, 661–668 (2008).
37. Visintin, R. et al. The phosphatase Cdc14 triggers mitotic exit by reversal of Cdk-dependent phosphorylation. *Mol. Cell* **2**, 709–718 (1998).
38. Laomettachit, T., Chen, K. C., Baumann, W. T. & Tyson, J. J. A model of yeast cell-cycle regulation based on a standard component modeling strategy for protein regulatory networks. *PLoS One* **11**, e0153738 (2016).
39. Woldringh, C. L., Huls, P. G. & Vischer, N. O. Volume growth of daughter and parent cells during the cell cycle of *Saccharomyces cerevisiae* a/alpha as determined by image cytometry. *J. Bacteriol.* **175**, 3174–3181 (1993).
40. Zhang, J., Qiao, Q., Xu, H., Zhou, R. & Liu, X. Human cell polyploidization: the good and the evil. *Semin Cancer Biol.* **81**, 54–63 (2022).
41. Battogtokh, D. & Tyson, J. J. Bifurcation analysis of a model of the budding yeast cell cycle. *Chaos* **14**, 653–661 (2004).
42. Malik-Sheriff, R. S. et al. BioModels-15 years of sharing computational models in life science. *Nucleic Acids Res.* **48**, D407–D415 (2020).

**Acknowledgements**

K.T. acknowledges the Petchra Pra Jom Klao Ph.D. Research Scholarship (KMUTT – NSTDA) from King Mongkut’s University of Technology Thonburi (No: 103/2563).

**Author contributions**

J.J.T., T.L., and P.K. did the conceptualization of the work. K.T. developed the Python code, did the simulations, and prepared the figures. K.T., J.J.T., T.L., and P.K. analyzed, and interpreted the model simulation results. All authors participated in writing the manuscript.

**Competing interests**

The authors declare no competing interests.

**Additional information**

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41540-024-00452-3>.

**Correspondence** and requests for materials should be addressed to Teeraphan Laomettachit or Pavel Kraikivski.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024