Hybrid Modeling and Simulation of Stochastic Effects on Biochemical Regulatory Networks

Mansooreh Ahmadian

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Computer Science & Application

Young Cao, Chair

John J. Tyson, Co-Chair

Lenwood S. Heath

Anuj Karpatne

Jean Peccoud

July 23, 2020

Blacksburg, Virginia

Keywords: Cell Cycle Modeling, Hybrid Stochastic Modeling, Parameter estimation,

Theory-guided machine learning

Hybrid Modeling and Simulation of Stochastic Effects on Biochemical Regulatory Networks

Mansooreh Ahmadian

ABSTRACT

A complex network of genes and proteins governs the robust progression through cell cycles in the presence of inevitable noise. Stochastic modeling is viewed as a key paradigm to study the effects of intrinsic and extrinsic noise on the dynamics of biochemical networks. A detailed quantitative description of such complex and multiscale networks via stochastic modeling poses several challenges. First, stochastic models generally require extensive computations, particularly when applied to large networks. Second, the accuracy of stochastic models is highly dependent on the quality of the parameter estimation based on experimental observations. The goal of this dissertation is to address these problems by developing new efficient methods for modeling and simulation of stochastic effects in biochemical systems. Particularly, a hybrid stochastic model is developed to represent a detailed molecular mechanism of cell cycle control in budding yeast cells. In a single multiscale model, the proposed hybrid approach combines the advantages of two regimes: 1) the computational efficiency of a deterministic approach, and 2) the accuracy of stochastic simulations. The results show that this hybrid stochastic model achieves high computational efficiency while generating simulation results that match very well with published experimental measurements. Furthermore, a new hierarchical deep classification (HDC) algorithm is developed to address the parameter estimation problem in a monomolecular system. The HDC algorithm adopts a neural network that, via multiple hierarchical search steps, finds reasonably accurate ranges for the model parameters. To train the neural network in the presence of experimental data scarcity, the proposed method leverages the domain knowledge from stochastic simulations to generate labeled training data. The results show that the proposed HDC algorithm yields accurate ranges for the model parameters and highlight the potentials of model-free learning for parameter estimation in stochastic modeling of complex biochemical networks.

Hybrid Modeling and Simulation of Stochastic Effects on Biochemical Regulatory Networks

Mansooreh Ahmadian

General Audience Abstract

Cell cycle is a process in which a growing cell replicates its DNA and divides into two cells. Progression through the cell cycle is regulated by complex interactions between networks of genes, transcripts, and proteins. These interactions inside the confined volume of a cell are subject to inherent noise. To provide a quantitative description of the cell cycle, several deterministic and stochastic models have been developed. However, deterministic models cannot capture the intrinsic noise. In addition, stochastic modeling poses the following challenges. First, stochastic models generally require extensive computations, particularly when applied to large networks. Second, the accuracy of stochastic models is highly dependent on the accuracy of the estimated model parameters. The goal of this dissertation is to address these challenges by developing new efficient methods for modeling and simulation of stochastic effects in biochemical networks. The results show that the proposed hybrid model that combines stochastic and deterministic modeling approaches can achieve high computational efficiency while generating accurate simulation results. Moreover, a new machine learning-based method is developed to address the parameter estimation problem in biochemical systems. The results show that the proposed method yields accurate ranges for the model parameters and highlight the potentials of model-free learning for parameter estimation in stochastic modeling of complex biochemical networks.

To my husband, my parents, and my sister.

Acknowledgments

First and foremost, I owe my deepest gratitude to my advisor Dr. Young Cao, and my co-advisor Dr. John J. Tyson, for their continuous support of my Ph.D. study, their patience, motivation, and immense knowledge and experience. I would like to thank them for the amount of time and effort, ideas, and funding they have generously dedicated to making my Ph.D. experience productive and stimulating. I am also grateful to them for believing in me and giving me the freedom to pursue diverse, yet coherent research directions that have made my Ph.D. study a joyful and unforgettable journey. I am thankful for their priceless advice and great supervision that have helped me to grow as a research scientist and find the right path for my future career.

I would like to thank the members of my advisory committee, Dr. Lenwood S. Heath, Dr. Anuj Karpatne, and Dr. Jean Peccoud, for their valuable comments that have helped me to substantially improve the quality of this dissertation. I am grateful for having the opportunity to collaborate with talented researchers during my PhD research. Special thanks to my co-authors, Dr. Shuo Wang and Dr. Minghan Chen for their time and effort.

Last but not the least, words cannot express how grateful I am to my beloved husband, Omid. Without his unparalleled love and support, I could have never been able to finish this dissertation. To my parents and my sister for their continuous and unconditioned love. Thank you for always believing in me and for all of the sacrifices that you have made on my behalf.

Contents

List of Figures

List of Tables

Chapter 1

Overview

Unraveling the physiological characteristics of living cells is contingent upon a fundamental understanding of their underlying regulatory networks. To this end, various mathematical approaches have been proposed to study the regulatory networks of cellular systems capturing different levels of details. Boolean models [1, 2, 3, 4, 5, 6, 7] present an expedient approach to describe the switching networks of a biochemical regulatory network. These parameter-free models are efficient in terms of computations and they are easy to implement. However, due to the qualitative nature of this approach, Boolean models are inadequate to provide a quantitative description of a biological system. Deterministic models [8, 9, 10] that are often described in terms of nonlinear ordinary differential equations (ODEs) and differential-algebraic equations (DAEs), provide a quantitative description of the average dynamics of biological systems. The main assumption of widely-used deterministic approaches is that the concentration of a species represented by these models is a continuously varying quantity that evolves deterministically over time. However, the time-evolution of molecular species within the confined volume of a cell is not deterministic since noise is an inevitable component in cellular systems. Thus, despite being computationally efficient, deterministic models cannot capture the intrinsic and extrinsic noise observed in the wet lab exper-

iments presented in [11, 12]. Therefore, there is a need to develop stochastic models to explain the inherent randomness and the cell-to-cell variability observed in cellular systems.

Most of the stochastic models build on the chemical master equation (CME) that describes the evolution of a biochemical reaction network as a stochastic process [13]. The CME is a system of ordinary differential equations and its solution provides the probability of states of a system over time. Despite providing an exact description, CME is only tractable for simple systems. Moreover, solving the CME directly using numerical methods is computationally expensive and infeasible if a system has a large number of reachable states [14].

To overcome these limitations, the stochastic simulation algorithm (SSA), proposed by Gillespie [15, 16], provides a numerical simulation of the underlying Markov process described by the CME. The SSA is the most common stochastic approach for modeling a well-stirred biochemical regulatory network. It is a Monte Carlo-based simulation that generates a sufficiently large number of state trajectories. More specifically, SSA simulates every single reaction firing in the system, and the time complexity of this algorithm scales with the number of reactions. Thus, the computational cost of SSA for a realistic model of a biological system with fast reactions and high-abundance species is prohibitively high. Different approaches have been proposed to reduce the computational cost of SSA [17, 18, 19, 20, 21, 22, 23, 24, 25, 26].

One of the most promising approaches that leverages the multiscale characteristics of biological systems is a hybrid approach proposed by Haseltine and Rawlings (HR) [24]. With advancements in technology and availability of quantitative experimental data at the level of genes, transcripts, and proteins, the scale of species present in recent mathematical models of biological systems varies in several orders of magnitudes. For instance, while there are a few copies of messenger RNAs in a yeast cell, the abundance of some of the yeast proteins varies between 100 to 10,000. The HR hybrid model takes such scale differences into account and approximates the fast reactions with high abundance using a deterministic formulation (often ODEs) and precisely simulates the

rest of the system using the SSA. This is a partition-based approximation method and if the partitioning strategy is chosen too conservative, i.e., if too many reactions being partitioned in the SSA regime, the computational saving will not be considerable [27]. Thus, it is imperative to find an appropriate partitioning strategy.

Due to the aforementioned limitations of the existing methods (i.e, both deterministic and stochastic models), there is a need to build new efficient stochastic models that can accurately describe the dynamical system of biological networks. Using such accurate quantitative models, one can address many fundamental questions pertaining to the cellular process such as "how does the cell manage to progress through cell cycle robustly in presence of intrinsic and extrinsic noise?", "how does the cell maintain its size from one generation to another?", and "how do mutant cells (i.e., cells with defective or overexpressed genes) respond differently to alternation in environmental conditions such as growth rate?".

The accuracy and reliability of such quantitative models depend on how well their parameters are estimated. The goal of parameter estimation is to find a set of parameter values such that the model can closely reproduce the experimental data. This is referred to as an *inverse problem*. The challenges of solving the inverse problem in systems biology are manifold. Some of the major challenges stem from the experimental data scarcity or having erroneous data in the experimental measurements. Another challenge is that measuring many unknown parameters in large networks of a biological system is not experimentally feasible or even possible. Moreover, in many cases, the inverse problem is not convex due to the nonlinear nature of the models. Therefore, standard optimization methods may not be applied directly [28, 29].

Most of the existing efforts in solving the parameter estimation problem have been focused on deterministic models, particularly, to estimate the parameters of models formulated by nonlinear differential equations [30]. In fact, limited works exist to solve the inverse problem for stochastic models.

In addition, parameter estimation is considerably more difficult for stochastic models compared to deterministic models due to the following reasons: a) stochastic models involve many more parameters compared to their underlying deterministic models, b) the experimental data, required to estimate specific stochastic parameters, are often difficult to measure in wet-lab experiments, and c) the objective function may be a random variable itself, such as in cases where distributions of observables from experimental data are compared with the statistical distributions generated from stochastic simulations. Moreover, most parameter optimization methods return a single best parameter vector, while there might be many parameter vectors that could generate similar system dynamics and characteristics. For instance, the bistable switches are frequently present in the cell cycle system. Most of the parameter values inside the bistability region can be acceptable because, with those parameter values, the model can generate the dynamics of a bistable switch [31]. In contrast, optimization methods only find a single parameter vector as global or local optimal values.

This dissertation is organized in seven chapters, addressing several fundamental problems pertaining to the design and analyses of hybrid models to study the stochastic effects on biochemical networks.

- Chapter 1 presents an overview of the existing literature, a description of key challenges that are addressed in this dissertation, and the summary of contributions.
- Chapter 2 provides preliminary concepts and mathematical background that will be used in the subsequent chapters.
- In Chapter 3, we build a comprehensive hybrid stochastic model based on a detailed molecular mechanism of cell cycle controls in budding yeast [9]. The budding yeast is a powerful, yet economical model organism that provides researchers a tractable system to understand cell functions and design principles. To apply our hybrid scheme to the base deterministic model of protein interactions, we first extend the model to include mRNA species

that are transcribed from cell-cycle genes and translated into proteins. Then, we carry out comprehensive simulations of wild-type yeast cells using both the deterministic and hybrid ODE/SSA models. The results show that our stochastic model can predict the statistical properties of many different cell-cycle variables, including inter-division times, size at birth, and the abundances of specific mRNAs and proteins. We also show that our stochastic simulations are in good agreement with experimental observations. In addition, our results prove that the developed hybrid approach provides a great trade-off between the accuracy and efficiency of numerical simulations. While our hybrid code takes about 15 minutes to simulate 10,000 cell cycles, a similar fully stochastic model takes more than 30 hours to generate results with the same accuracy.

- In Chapter 4, we adjust the parameters of our hybrid stochastic model from Chapter 3 to study the stochastic phenotypic characteristics of more than 100 mutant strains. It is very important to study the non-robust dynamics and stochastic characteristics of mutant cells. Progression through the cell cycle in wild-type eukaryotic cells is significantly robust to various sources of intrinsic and extrinsic noise and variability. However, mutant cells in which critical parts of cell cycle control mechanisms are defective may be prone to more sensitivity in terms of environmental changes such as nutrition condition or intrinsic alternations such as fluctuations of protein levels within the cell [32]. Cancer cells in which genetic mutations are accumulated are good examples of such non-robust dynamics [32]. We propose a scheme to evaluate the viability of a mutant cell that accounts for stochastic phenotypes such as partial viability. Based on that viability rule we show that our model conforms with 104 out of 122 mutant cases.

- In Chapter 5, we study the effect of the size control mechanism on important properties of the budding yeast cell cycle. Cell size is a key characteristic that significantly affects many aspects of cellular physiology. There are specific control mechanisms during cell

cycle that maintain the cell size within a range from generation to generation. Such control mechanisms introduce substantial variabilities to important properties of the cell cycle such as growth and division. To quantitatively study the effect of such variability in progression through the cell cycle, we extend our hybrid stochastic model proposed in Chapter 2 by including the proteins and genes that play roles in the size control mechanism of the budding yeast. The results show that our model 1) confirms that the majority of noise in size control stems from low copy numbers of transcripts in the G1 phase, 2) identifies the size and time regulation modules in the size control mechanism, and 3) conforms with phenotypes of early G1 mutants in exquisite detail.

- In Chapter 6, we propose a theory-guided machine learning-based scheme to estimate the parameters of stochastic models of biochemical networks. To this end, instead of developing a point estimator, our goal is to find an acceptable range for each parameter. We develop an algorithm that starts with a number of candidate ranges for each unknown parameter, and at each iteration, solves the range-selection problem as a classification problem. The classifier that we develop is based on neural networks (NN). To train the NNs, we generate the synthesized training data by using the SSA. The NN, which is the inverse model, learns the relationship between the pseudo-observations and the classes (ranges) of the corresponding parameter vector. The trained NN is then tested using the experimental data that our forward model aims to ultimately match with. We show that our model can successfully find acceptable ranges for the parameter vector of a very simple example.

- In Chapter 7 provides some of the research directions to extend our work and concludes the dissertation.

Chapter 2

Background

2.1 Stochastic Models

2.1.1 Stochastic simulation algorithm

We consider a well-stirred system of N chemical species $\{S_1, \ldots, S_N\}$ and M chemical reactions $\{R_1, \ldots, R_M\}$. We define the state of system at t as a vector $\mathbf{x}(t) = (x_1, \ldots, x_N)$, where x_i is the number of S_i molecules. Each reaction channel R_j is characterized by its propensity function $a_j(\mathbf{x}(t))$ and its state-change vector $\mathbf{v}_j = (v_{1j}, \ldots, v_{Nj})$. The probability of reaction R_j firing in the next infinitesimal time dt is given as $a_j(\mathbf{x}(t))dt$, and v_{ij} S_i molecules will be changed by one R_j reaction. The procedure of the SSA [15, 16] is presented as follows.

1) At time t, compute the propensity values $a_j(\mathbf{x}(t))$, and the sum $a_0(\mathbf{x}(t)) \equiv \sum_{j=1}^{M} a_j(\mathbf{x}(t))$.

2) Generate a time increment τ as a sample of the exponential random variable with mean $1/a_0(\mathbf{x}(t))$.

3) Generate a reaction index j based on the probability function $a_j(\mathbf{x}(t))/a_0(\mathbf{x}(t))(j = 1, \ldots, M)$.

4) Update time $t \leftarrow t + \tau$ and state $\mathbf{x}(t + \tau) \leftarrow \mathbf{x}(t + \tau) + \mathbf{v}_j$.

5) Return to step 1) if stopping condition is not reached.

Although the SSA is a fundamental method for stochastic simulation, its computational cost may become very high when the size of a biochemical system increases, since the SSA simulates every reaction event and the time increment τ often is very small as a consequence of large values of the total propensity $a_0(\mathbf{x}(t))$. Thus it is necessary to develop more efficient simulation methods.

2.1.2 Hybrid stochastic model

The hybrid method presented in this paper was originally proposed by Haseltine and Rawlings [24]. Given a system of N species $\{S_1, \ldots, S_N\}$ and M reactions $\{R_1, \ldots, R_M\}$, these M reactions are partitioned into two subsets. Reactions in subset S_{fast} are formulated by ODEs, and reactions in subset S_{slow} are simulated by the SSA. Let $a_j(\mathbf{x}(t))$ be the propensity function of the j-th reaction in S_{slow}, where $\mathbf{x}(t)$ is the state at time t, and its state-change vector $\mathbf{v}_j = (v_{j1}, \ldots, v_{jN})$. Let τ be the jump interval of the next slow (stochastic) reaction, and μ be its reaction index. Set $t = 0$. The hybrid method simulates the system as follows:

1) Generate two uniform random numbers r_1 and r_2 in $U(0, 1)$.

2) Integrate the ODE system and the integral equation:

$$\int_t^{t+\tau} a_{tot}(\mathbf{x}(s))ds + \log(r_1) = 0, \tag{2.1}$$

where $a_{tot}(\mathbf{x}(t))$ is the sum of propensities of reactions in S_{slow}.

3) Determine μ as the smallest integer satisfying

$$\sum_{i=1}^{\mu} a_i(\mathbf{x}(t)) > r_2 a_{tot}(\mathbf{x}(t)). \tag{2.2}$$

4) Update $\mathbf{x}(t+\tau) \leftarrow \mathbf{x}(t+\tau) + \mathbf{v}_\mu$.

5) Return to step 1) if stopping condition is not reached.

Our implementation is different in step 2). Suppose that the ODE system is given by

$$\mathbf{x}' = f(\mathbf{x}). \tag{2.3}$$

We simply add an integration variable z and an equation

$$z' = a_{tot}(\mathbf{x}), \quad z(t) = \log(r_1), \tag{2.4}$$

where we note that $\log(r_1)$ is negative and a_{tot} is nonnegative. During the simulation, each step starts at time t and numerically integrates ODEs (2.3) and (2.4). When $z(t+\tau) = 0$, the ODE integration stops, where τ is the solution to (2.1). This integration process can be conveniently handled by standard ODE solvers combined with root-finding functions. Note that since z is an integration variable, one may choose to omit it from the error control mechanism [33]. Adding this extra variable will not greatly affect the efficiency.

2.2 Artificial Neural Network

Artificial neural network (ANN) is viewed as an effective function approximator and is adopted in various applications such as natural language processing, computer vision, image analysis, medi-

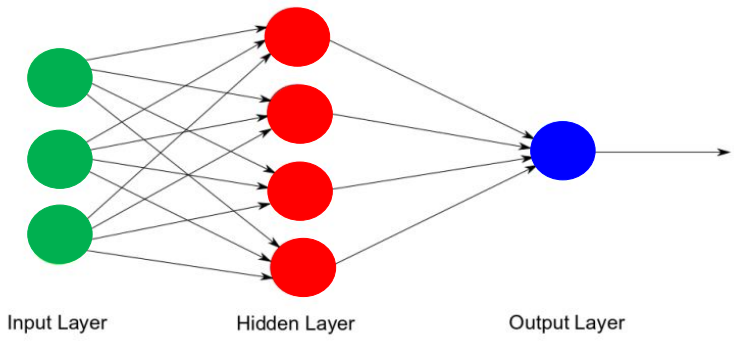

Figure 2.1: The structure of a simple artificial neural network.

cal diagnosis, and many other applications. Figure 2.1 shows a simple example of an ANN. The ANN consists of a set of nodes and a set of links. The nodes correspond to neurons that are the building blocks of the ANN. The links represent the data flow and quantitative connections between neurons. The ANN can be considered as a computing system that learns through processing examples [34]. Each example includes an input along with its associated output. ANNs learn the relationship between the input and output by dynamically adjusting the quantitative connections between the neurons (that are the weights) through the training process. More specifically, during the training process, a set of training inputs that are associated with a set of target outputs are given to the ANN. For each input, the network output will be compared with the target output using a measure of distance, which is known as the loss function. The aim is to minimize the loss function iteratively as we move forward in the training process. That is, through the learning process, the weights of the ANN are adjusted such that the distance between the network output and the target is smaller after each iteration. The training stops when the desired accuracy is reached or when predefined numbers of epochs are trained.

Neurons are the processing units of ANNs. Figure 2.2 shows the general structure of a neuron. Each neuron can have multiple different inputs but it has only one output. As shown in Fig. 2.2,

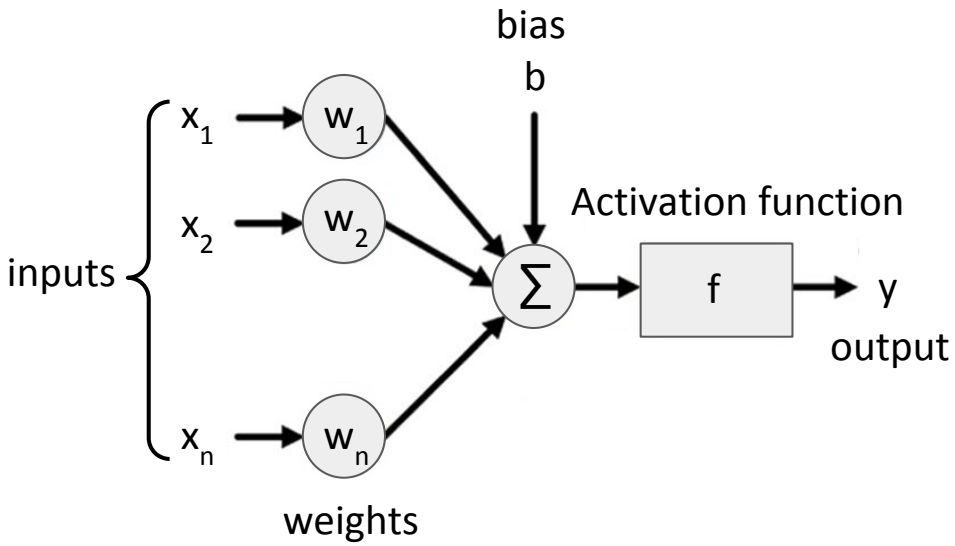

Figure 2.2: A basic neuron in an ANN.

the inputs are combined using a linear combination described by

$$y = f\left(\sum_i w_i x_i + b\right), \tag{2.5}$$

where $\boldsymbol{x}$ and y are, respectively, the input and the output of the neuron. Equation (2.5) describes the output of a neuron that is a linear combination of the inputs $x_i \in \boldsymbol{x}$, weighted by the synaptic weights w_i. Then, a generally non-linear activation function f is applied to generate the final output. The set of inputs to a neuron generally includes a bias b with a constant value. The activation function f can be linear such as the identity activation function or nonlinear such as Sigmoid or the tangent function. Nonlinear activation functions are mostly used for the hidden layers to introduce nonlinearity to the model [35].

2.2.1 Feed-forward neural network.

Feed-forward neural network (FNN) [36, 37] is the simplest and the most common structure of ANNs. In the FNN, starting from the input layer, the information across the network is transferred

in a forward manner. That is, the input data is passed through hidden layers and finally, the output of the network is delivered in the output layer. In a fully-connected FNN, every neuron in a layer is connected with all the neurons in the previous layer. However, the connections may have different weights. In the first iteration of training, weights and biases are assigned randomly. Through the training process, these weights are adjusted such that the difference between the output of the network and the target output (error) is minimized. Among the methods that are used for minimizing the error, backpropagation gradient-based methods are widely used. Backpropagation stands for " backpropagation of error". For an ANN, the gradient-based methods calculate the gradient of the error with respect to the weights. Because the calculation of the gradient starts from the output layer and proceeds backward towards the input layer, the method is call backpropagation that leads to efficient computation of the gradient.

In chapter 6, we use a feedforward neural network with cross-entropy as loss function and scaled conjugate gradient [38] as the training algorithm to solve a classification problem.

Chapter 3

Hybrid Stochastic Model of Budding Yeast Cell Cycle

3.1 Introduction

The eukaryotic cell cycle is a complex process by which a growing cell replicates its DNA and divides into two cells, each capable of repeating the process. Progression through the cycle is controlled by networks of genes, mRNAs, and proteins, with interactions that can be modeled as chemical reaction channels. To unravel the complex dynamics of multiscale reaction networks in higher organisms such as human cells, it is advisable to study single-cell organisms with molecular regulatory networks that are similar yet simpler. For instance, experimental studies and mathematical models of frog eggs [39, 40], fission yeast [41, 42], and budding yeast [43, 44] have shed light on mechanisms of cell cycle regulation in the cells of higher organisms. Extensive experimental studies have been conducted particularly on budding yeast (*Saccharomyces cerevisiae*) to explore gene regulation and signaling pathways of relevance to cell growth and division [45, 46, 47, 48]. Moreover, various modeling approaches, such as deterministic models [8, 9, 10], Boolean net-

works [1, 2, 3, 4, 5, 6, 7, 49], and stochastic models [50, 32, 27, 51, 52, 53, 54], have been adopted to explore the roles of different gene and protein interactions in robust progression through the cell cycle.

Among these models, a deterministic approach is most common. In this approach, the time-dependent variation of each molecular species in the biochemical reaction network is described by a nonlinear ordinary differential equation (ODE), in which the concentration of the substance is considered as a continuous quantity that evolves deterministically over time. However, the time-evolution of molecular species within the confined volume of a budding yeast cell (about 30 fL at birth) is not deterministic. Therefore, despite being able to reproduce certain average characteristics of cell cycle progression in yeast cell populations, a continuous-deterministic model cannot reproduce the cell-to-cell variability observed in wet-lab experiments [11, 12]. For instance, Di Talia *et al.* [12] have reported that the coefficient of variation ($\text{CV} = \frac{\text{standard deviation}}{\text{mean}}$) for G1 time of budding yeast cells (growing on glucose) is 50%.

To capture such high levels of variability, stochastic models have been built using different strategies to incorporate intrinsic and extrinsic sources of noise. In an early stochastic model of the fission yeast cell cycle proposed by Sveiczer *et al.* [55], extrinsic noise was introduced by assuming some sloppiness in the partitioning of cell volume and nucleus volume to daughter cells at the division. A later model by Steuer [56] examined the roles of intrinsic noise in cell cycle progression by adding Gaussian noise to reaction rate equations in a deterministic model. These approaches, however, do not adequately explain the root source of cell-cycle variability in yeast cells, which lies in molecular fluctuations at the level of gene expression [57, 58, 59]. To capture such molecular-level noise, more accurate stochastic methods are required to explicitly model fluctuations in molecular interactions. For this purpose, the best method to implement fluctuating molecular interactions is the *stochastic simulation algorithm* (SSA) proposed by Gillespie [15, 16]. Gillespie's algorithm is a Monte-Carlo approach that numerically simulates the temporal firing of every single reaction in

a chemical reaction network. Gillespie's method assumes that the propensity of every reaction in the model is described by mass-action kinetics. This becomes an issue because most deterministic models of cell cycle regulation, such as those presented in [9, 60, 43, 10], incorporate complex rate laws including Michaelis-Menten kinetics, Hill functions, and ultra-sensitive switches. These complex phenomenological rate laws are used in deterministic models to provide sufficient nonlinearity in reaction kinetics to create bistable switches that flip on and off during progression through the cell cycle. Converting a deterministic model into a stochastic model suitable for Gillespie's SSA by 'unpacking' complex rate laws into elementary reactions is a difficult problem fraught with uncertainties [61].

To address this challenge, several approaches have been tried. The simplest approach, used for example by Mura & Csikasz-Nagy [62], treats all complex rate laws directly as propensity functions of reactions and then applies the SSA. This approximation is subject to considerable errors [61, 63]. For example, Ball *et al.* [32] found that the variability they observed in wet-lab measurements could not be generated by this greatly simplified stochastic approach unless some unrealistic parameter values were chosen. Later Kar *et al.* [50] tried to unpack Michaelis-Menten rate laws in a small (three-variable), deterministic model of the budding yeast cell cycle [43]. Unpacking resulted in a much more complicated system with 19 species and 47 reactions. Although this simple model (with only a few key cell-cycle genes) could generate noise levels that match wet-lab measurements for a few key characteristics of the cell cycle, it is not feasible to apply this approach to more complex models with substantially more genes and proteins. Instead, Barik *et al.* pursued an approach in which the molecular controls of the budding yeast cell cycle are modeled directly in terms of elementary reactions (governed by the law of mass-action) [53, 54]. They included the multisite phosphorylation reactions to introduce the inherent nonlinearity of the cell cycle system into their model. A great advantage of this approach is that the newly designed deterministic model can be converted into its corresponding stochastic version without any approx-

imation. A disadvantage of this approach is that we cannot re-purpose our original deterministic models, which had been carefully designed and parametrized to explain a broad scope of experimental observations. Furthermore, to model the phosphorylation and dephosphorylation reactions that play such important roles in cell cycle progression introduces substantial complexity into the system. Recently, Laomettachit *et al.* proposed a new approach that sidesteps the complications of elementary reactions and mass-action rate laws and that employs a Langevin-type simulation of noisy gene expression [51]. This approach, though promising, also requires an overhaul of the original deterministic models. To take advantage of existing deterministic models in a framework that permits accurate stochastic simulations without 'unpacking', we explore a particular hybrid approach in this chapter.

Gillespie's SSA simulates every single reaction firing. In general, the time complexity of this algorithm scales proportionally with the number of reaction firings. Consequently, SSA-based models involve substantial computational complexity if a reaction network involves many fast reactions. To reduce the high computational cost of the SSA, many optimization methods [17, 18, 19, 20, 21] and approximation methods [22, 23, 24, 25, 26] have been proposed. Among them the *hybrid stochastic* approach, originally proposed by Haseltine and Rawling (HR) [24], performs well because it takes advantage of the multiscale features common in biochemical reaction networks.

The main idea of any hybrid approach is to divide the system into subsystems and solve each subsystem using an appropriate method. The idea of the HR hybrid approach is to partition the dynamical system into *fast* and *slow* reactions, based on the relative time scale of each reaction and the abundances of the reactants. Fast reactions, which fire frequently and often involve high-abundance species, are partitioned into the deterministic (ODE) regime. Meanwhile, slow reactions, which are often found at the gene-expression level, fire much less frequently and are therefore simulated using the SSA. This approach was first applied by Liu *et al.* [27] on the simple three-variable model of the budding yeast cell cycle, originally studied by Kar *et al.* [50]. By partitioning all

gene-expression reactions into the slow (SSA) regime and all protein-level dynamics into the fast (ODE) regime, Liu *et al.* [27] were able to reproduce the noise levels that Kar *et al.* [50] achieved by unpacking the original system into a much more complex one. This success motivated us to apply this approach to the very comprehensive, accurate, and complex deterministic model of yeast cell-cycle controls proposed by Chen *et al.* [9].

3.2 Model

3.2.1 Budding yeast cell cycle

A comprehensive continuous-deterministic model of the budding yeast cell cycle was developed by Chen *et al.* [9] in 2004. By integrating the findings of decades of experimental studies, Chen's model provides an accurate mathematical description of the cell division cycle of budding yeast. The protein regulatory network of Chen's model focuses primarily on the mutual antagonism between mitotic B-type cyclins (Clb1-6) and G1 phase stabilizers (Cdh1, Sic1, and Cdc6). During the growth and division of yeast cells, this antagonism leads to transitions between two coexisting steady states called START (G1→ S) and EXIT (M→ G1). Next, we provide more details on the interactions of the protein-protein regulatory network of this model.

The main regulator of the cell cycle in budding yeast is a cyclin-dependent protein kinase (Cdc28) that is constitutively expressed but periodically activated and inhibited during progression through the cell cycle. Cdc28 forms an active kinase by binding to two families of cyclin partners, Cln1-3 and Clb1-6, each of which is synthesized and degraded in different phases of the cell cycle. In the early G1 phase, Cln3 is the only available partner of Cdc28. As a newborn cell grows, the hetero-dimer of Cln3:Cdc28 accumulates until the cell reaches a critical size, when there is sufficient Cln3:Cdc28 (and a back-up protein, Bck2) to activate the transcription factors SBF and

MBF. These transcription factors are responsible for the production of Cln2 and Clb5, respectively. Cln2 accumulation induces the emergence of the bud and Clb5 initiates DNA synthesis.

In the G1 phase the cyclin-dependent kinase inhibitors (CKI), Sic1 and Cdc6, are in high abundance, and they inhibit the activity of Clb5:Cdc28. However, as the activity of Cln2:Cdc28 increases, these inhibitors get phosphorylated and degraded by the SCF-proteosome pathway. As a result, Clb5 accumulates to initiate DNA synthesis (S phase). Clb2 is the cyclin responsible for driving the cell into mitosis (M phase). As CKIs are removed, Clb2 level rises because Clb2 activates its own transcription factor Mcm1 in an autocatalytic reaction. Rising Clb2:Cdc28 activity phosphorylates and inactivates the transcription factors SBF and MBF. In telophase, Clb2 must be degraded below a threshold so that the cell can exit mitosis and return to the G1 phase. Clb2 degradation is initiated by two proteins, Cdc20 and Cdh1. Cdc20 has been kept inactive in the early stages of mitosis by the 'mitotic checkpoint complex'. When all chromosomes are properly aligned on the mitotic spindle, Cdc20 becomes active and facilitates the degradation of Clb2.

Moreover, as a yeast cell exits mitosis, Cdh1 is activated by a phosphatase, Cdc14, which has been sequestered in the nucleolus by binding to Net1. After full chromosome alignment on the metaphase plate, first Tem1 and then Cdc15 become active. Cdc15 phosphorylates Net1, which leads to the release of Cdc14. Next, Cdc14 activates Cdh1 (which takes over for Cdc20 as the primary initiator of Clb1-6 degradation in G1), and Cdc14 also activates the transcription factor, Swi5, for production of CKIs in G1 phase. In this way, the scene is set for the cell to return to the G1 phase, when the CKIs are abundant and all cyclins (except for Cln3) are out of the picture.

3.2.2 Building and implementation of hybrid stochastic model

Converting the concentration-based model into population-based model. Chen's mathematical model reproduces the average cell-cycle properties (including cycle time, G1 duration, and cell size

at division) of wild-type budding yeast cells and the variant cell-cycle phenotypes of more than 100 mutant strains. Our goal is to develop a hybrid (stochastic-deterministic) version of this large regulatory network, to quantify the variabilities observed in cell-cycle characteristics and mutant phenotypes within a computationally efficient framework.

Since Chen's model is formulated in terms of normalized (dimensionless) concentrations of proteins, the first step to this goal is to convert the state variables of Chen's model into integer numbers of molecules per cell. This conversion facilitates comparison of our numerical simulation results with observed data from single-cell experiments. Furthermore, it is necessary because, in Gillespie's SSA, state variables are discrete (species populations) rather than continuous (species concentrations). Since a hybrid model involves both SSA and ODEs, it is important that we assure consistency between units of state variables in both the stochastic and deterministic regimes. Therefore, we calculate S_i, the number of molecules of species i in a cell, from the corresponding normalized concentration, $[\mathrm{S}_i]_\mathrm{n}$, by (3.1):

$$S_i(t) = N_\mathrm{A} \cdot [\mathrm{S}_i](t) \cdot V(t) = 0.6 \cdot C_i \cdot [\mathrm{S}_i]_\mathrm{n}(t) \cdot V(t), \tag{3.1}$$

where $[\mathrm{S}_i]$ is the actual concentration of species i (in nanomoles/liter = 10^{-9} mol/L), C_i is the 'characteristic' concentration of species i (used to convert between actual concentration and 'normalized' concentration), $V(t)$ is the volume of the cell (in femtoliters = 10^{-15} L), and $N_\mathrm{A} = 0.6$ is Avogadro's number (when concentration is expressed in nM and volume in fL). One simplifying assumption made in published models [32,51] is to use a constant volume for the size of cell. However, this unrealistic assumption introduces errors into the model because cell size (V) increases exponentially during a cycle ($V(t) = V(0)\mathrm{e}^{k_\mathrm{g}t}$, where k_g is the specific growth rate of yeast cells). We bear in mind that the characteristic concentration C_i, as a scaling factor, cannot be arbitrarily chosen. For instance, if proteins P_1 and P_2 are involved in a reaction to form a complex P_3, then characteristic concentrations of these three proteins have to be the same. This restriction prevents

the possibility to choose different characteristic concentrations to match with experimental data.

Without loss of generality, we use a simple example to explain the procedure of computing the rate constants in our population-based model. Consider the following reactions:

$$\varnothing \xrightarrow{k_{\mathrm{s}}} P \xrightarrow{k_{\mathrm{d}}} \varnothing, \tag{3.2}$$

where k_{s} and k_{d} are constants denoting, respectively, the synthesis and degradation rates of protein P. Given that the ODE is formulated based on $[\mathrm{P}]_{\mathrm{n}}$, the normalized concentration of P, the corresponding ODE for the reactions in (3.2) is given by:

$$\frac{\mathrm{d}[\mathrm{P}]_{\mathrm{n}}}{\mathrm{d}t} = k_{\mathrm{s}} - k_{\mathrm{d}}[\mathrm{P}]_{\mathrm{n}}, \tag{3.3}$$

As mentioned, the cell grows exponentially with a growth rate k_{g}, thus we can write:

$$\frac{\mathrm{d}V}{\mathrm{d}t} = k_{\mathrm{g}}V. \tag{3.4}$$

With $[\mathrm{P}] = \frac{P}{V}$, the differential equation that describes the dynamics of protein P can be formulated by:

$$\frac{\mathrm{d}P}{\mathrm{d}t} = V\frac{\mathrm{d}[\mathrm{P}]}{\mathrm{d}t} + [\mathrm{P}]\frac{\mathrm{d}V}{\mathrm{d}t} = V\frac{\mathrm{d}[\mathrm{P}]}{\mathrm{d}t} + k_{\mathrm{g}}P, \tag{3.5}$$

Using (3.1) and (3.5), we can rewrite (3.3) as:

$$\frac{\mathrm{d}P}{\mathrm{d}t} = k_{\mathrm{s}}^{'}V - (k_{\mathrm{d}} - k_{\mathrm{g}})P, \tag{3.6}$$

where $k_{\mathrm{s}}' = k_{\mathrm{s}} \cdot C_{\mathrm{p}} \cdot N_{\mathrm{A}} \cdot [\mathrm{P}]_{\mathrm{n}} \cdot C_{\mathrm{v}}$. In (3.6), the ODE is based on the population of species P. This conversion procedure is applied to all equations of Chen's model to derive a deterministic model

based on populations of species.

Incorporating the dynamics of mRNAs. We extend the protein regulatory network in Chen's model to include the dynamics of 11 regulated and 8 unregulated mRNAs. This extension is necessary because the major source of intrinsic noise in yeast cells is the small number of mRNA molecules per cell per gene [11].

Experimental observations in yeast cells with increased dosage of genes suggest that the dominant source of variability with respect to cell-cycle time and cell size at division is the low copy number of mRNA and protein molecules in a cell, specifically in G1 phase [12]. Swain *et al.* [58] and Pedraza and Paulsson [64] also have shown that the fluctuations in the level of mRNAs are the primary cause of intrinsic noise. Considering the coefficient of variation (CV[1]) as a measure of variability, they proposed the formula (3.7) where N_{p} and N_{m} are the average molecule numbers of a protein and its corresponding mRNA, τ_{p} and τ_{m} are their half-life times, respectively.

$$\mathrm{CV} = \sqrt{\frac{1}{N_{\mathrm{p}}} + \frac{\tau_{\mathrm{m}}}{\tau_{\mathrm{m}} + \tau_{\mathrm{p}}} \frac{1}{N_{\mathrm{m}}}}, \tag{3.7}$$

Given that population of proteins per budding yeast cell are in the range of 500-5,000 while, there are only 5-10 copies of each mRNAs, the equation (3.7) can be approximated by (3.8):

$$\mathrm{CV} \approx \sqrt{\frac{1}{2N_{\mathrm{m}}}}. \tag{3.8}$$

According to equation (3.8) the major source of the observed variability stems from the mRNAs.

Chen's model, however, did not incorporate the turnover of mRNA molecules, and thus it cannot account for fluctuations stemming from transcriptional noise. For these reasons, Chen's original model must be supplemented with appropriate synthesis and degradation rates for each mRNA,

[1] $\mathrm{CV} = \frac{\text{standard deviation}}{\text{mean}}$

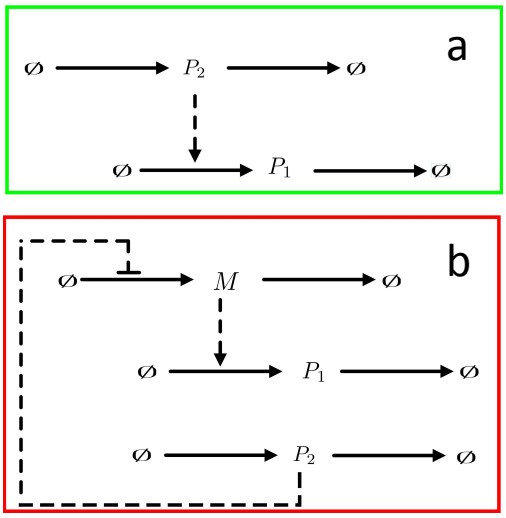

Figure 3.1: **Examples of a biochemical regulatory network. a:** A very simplified bichemical regulatory network in which protein P_1 is regulated by protein P_2. **b:** A more realistic regulatory network in which the transcriptional regulation is included in regulatory network of protein P_1.

as well as realistic rates of translation from mRNA to protein. Next, the procedure of adding transcriptional regulation and computing the corresponding rate laws are explained by a simple example.

Figure 3.1 (a) shows a simple protein regulatory network consisted of proteins P_1 and P_2 that can be represented by the following ODE:

$$\frac{\mathrm{d}P_1}{\mathrm{d}t} = k_{s_{P_1}} f(P_2) - g(P_1, P_2)P_1, \tag{3.9}$$

where $f(P_2)$ and $g(P_1, P_2)$ are, respectively, the synthesis and degradation rate laws of protein P_1. $k_{s_{P_1}}$ represents the background synthesis rate of the P_1. We note that the above rate laws can be either traditional mass-action laws or any complex rate laws such as Michaelis–Menten (MM) or Hill functions. Figure 3.1 (b) shows a simple regulatory network where the protein P_1 is regulated by the mRNA, M. After adding the mRNA, equation 3.9 can be rewritten as:

$$\frac{\mathrm{d}P_1}{\mathrm{d}t} = k_{\mathrm{spm}} M - g(P_1, P_2)P_1, \tag{3.10}$$

where the dynamics of the mRNA is given in (3.11):

$$\frac{\mathrm{d}M}{\mathrm{d}t} = k_{\mathrm{sm}} f(P_2) - k_{\mathrm{dm}} M. \tag{3.11}$$

k_{sm} and k_{dm} denote, respectively, the synthesis and degradation rates of the transcript. In addition, $k_{\mathrm{spm}} = k_{\mathrm{sp}} k_{\mathrm{dm}}/f(P_2)$ is the synthesis rate of protein P_1 after adding the M, which introduces a delay in synthesis of protein P_1. In transcriptional dynamics, mRNAs can be either regulated or unregulated. We note that (3.11) is in general form and describes the dynamics of a regulated mRNA. If $f(P_2) = 1$, then (3.11) describes the dynamics of an unregulated mRNA. Next, we elaborate how these rate lows are calculated.

In our model, based on experimental observations in [65], we assigned half-life times for mRNAs, τ_{m}, in the range of 5-10 min, except for $mCln2$ and $mClb2$, which were assigned shorter half-lives (3 and 2 min, respectively). Given the half-life times for an arbitrary mRNA the $k_{\mathrm{dm}} = \ln(2)/\tau_{\mathrm{m}}$.

The synthesis rate of an mRNA, k_{sm}, is then easy to estimate based on the degradation rate, and the average population of an mRNA molecule, $\langle \mathrm{mRNA} \rangle$. For an unregulated transcript, k_{sm} is chosen such that $\langle \mathrm{mRNA} \rangle = k_{\mathrm{sm}}/k_{\mathrm{dm}}$ is close to the average molecule numbers of the mRNA observed in experiment. For a regulated transcript, the total rate of mRNA synthesis is $\sum(k_{\mathrm{sm},i} \cdot P_i)$ where P_is are the numbers of protein molecules regulating that transcript. The rate constants $k_{\mathrm{sm},i}$ are chosen such that the number of mRNA molecules, averaged over a full cell cycle, is close to the experimentally observed mean number of mRNA molecules in an asynchronous culture. In our model, the synthesis rate of each is estimated to match the mRNA average-abundance measurements in Ball *et al.* [11].

Converting the Algebraic equations into ODEs. We modified Chen's model by introducing ODEs for the concentrations of Cln3 and Bck2 proteins. In Chen's original model, the normalized concentration of Cln3 and Bck2 were assumed to be given by steady-state algebraic equations (3.12) and (3.13),

$$[\text{Cln3}]_n = \frac{C_0 \cdot D_{\text{n3}} \cdot mass}{J_{\text{n3}} + D_{\text{n3}} \cdot mass}, \tag{3.12}$$

$$[\text{Bck2}]_n = B_0 \cdot mass, \tag{3.13}$$

where C_0 determines the maximum concentration of Cln3, D_{n3} is the dosage of the $CLN3$ gene, J_{n3} and B_0 are constants, and *mass* is the 'size' of a cell. We replaced the algebraic equations (3.12) and (3.13) by ODEs in (3.14) and (3.15).

$$\frac{\text{d}Cln3}{\text{d}t} = k_{\text{s,n3}} \cdot V^2(t) - k_{\text{d,n3}} \cdot Cln3, \tag{3.14}$$

$$\frac{\text{d}Bck2}{\text{d}t} = k_{\text{s,k2}} \cdot V^2(t) - k_{\text{d,k2}} \cdot Bck2. \tag{3.15}$$

The synthesis ($k_{\text{s,n3}}$, $k_{\text{s,k2}}$) and degradation ($k_{\text{d,n3}}$, $k_{\text{d,k2}}$) rate constants were estimated so that the half-lives and average abundances of these proteins match with experimental data [66]. The reason for this change is to model the unbalanced partitioning of Cln3 molecules between daughter and mother cells at cell division. According to experimental observations, the concentration of Cln3 in a new-born daughter cell is about 3 times less than its concentration in the mother cell [67, 68], indicating that mother cells get more than their 'fair share' of Cln3 molecules at cell separation. As a consequence of this unequal partitioning of Cln3 between mother and daughter cells at the division, the G1 time of mother cells is much shorter and the G1 time of daughter cells is much longer (on average) than would otherwise be expected. By including Cln3 and Bck2 as

state variables in the model, we can apply an asymmetric partitioning rule with a ratio of 20:80 to daughter and mother cells at cell division. We note that this ratio is set to 40:60 for all other proteins and mRNAs, according to observations in [12].

In addition, we comment that the quadratic dependence of Cln3 and Bck2 synthesis rates on cell size is introduced to account for the major influence that these two proteins have on cell size at the G1-S transition [69,70]. Because the rate of synthesis of these two proteins increases quadratically with cell volume, there is a strong size control on the G1-S transition in our model. Furthermore, we supplemented the model with 7 new ODEs in order to include the synthesis and degradation of the following important proteins: $Cdc15_T$, $Tem1_T$, SBF_T, MBF_T, $Esp1_T$, $Mcm1_T$, and APC_T. The main reason these proteins were not included as a state variable in the original model was that the experimental measurements were not available for many proteins at the time Chen's original model was developed [9].

The variables, variable descriptions, and initial values, as well as equations and parameter values of our model, are provided in Appendix A1-A4.

3.2.3 Hybrid stochastic model

As we mentioned in Section 3.1, the regulatory network of the budding yeast cell cycle is a multiscale system: both the numbers of molecules of mRNAs and proteins and the propensities of individual reactions vary by orders of magnitude. For instance, in budding yeast cells, there are 500-5,000 copies of each protein encoded by only 5-10 copies of the corresponding mRNA. Furthermore, the synthesis and degradation of mRNA species occur much less frequently than the phosphorylation and dephosphorylation of proteins in the cell. A hybrid approach leverages these large scale differences to improve the efficiency of stochastic simulations without sacrificing the accuracy of the computations. A hybrid method divides the system into subsystems, each includ-

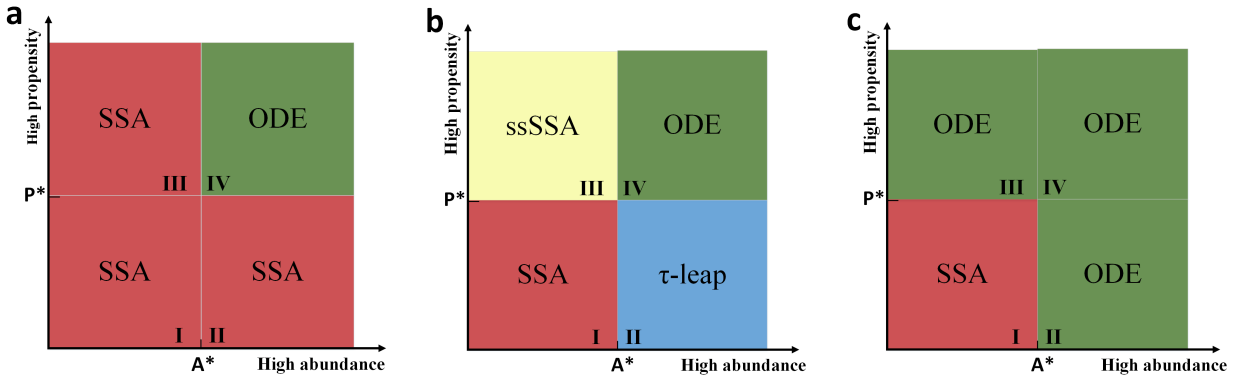

Figure 3.2: **Partitioning strategies for the hybrid model.** Region I : Slow reactions with low-abundance reactants; region II: slow reactions with high-abundance reactants; region III: Fast reactions with low-abundance reactants; region IV: fast reactions with high-abundance reactants. **a:** Conservative partitioning strategy by Haseltine and Rawling [24]. **b:** Possible partitioning strategy based on the definition of reactions in each region. **c:** Partitioning strategy proposed in [27] and used in our model.

ing species and reactions with similar scales, and applies an appropriate simulation method to each subsystem. Figure 3.2 shows various strategies in each of which the partitioning is done by using predefined thresholds for propensities of reactions (P^*) and abundances of reactants (A^*). In this way, the system is divided into four disjoint regions: (I) slow reactions with low-abundance reactants, (II) slow reactions with high-abundance reactants, (III) fast reactions with low-abundance reactants, and (IV) fast reactions with high-abundance reactants. Then an appropriate simulation method is chosen for each region [24, 27, 18].

Reactions in the level of gene expression are examples of reaction channels in region I. Due to low-copy numbers of species in this region, it is unrealistic to assume that the dynamics of the reactants evolve deterministically over time. For this reason reaction channels in region I are simulated using computationally expensive but accurate SSA. Region IV on contrary includes reactions with high frequency and reactants with high abundance. Post-translational reactions are examples of the reactions in region IV. Due to the high abundance of reactants, it is reasonable to approximate the dynamics of state variables in region IV using deterministic methods such as ODEs. Region II and III need more design considerations in order to achieve a desired trade-off

between accuracy and efficiency.

Different strategies have been proposed to choose appropriate simulation methods for reaction channels in regions II and III. Haseltine and Rawling [24] using a conservative strategy (see fig. 3.2 a) employ SSA for all reactions within regions I, II, and III and solves ODEs only in region IV. Figure 3.2 b shows a modification to this strategy, which leads to a more efficient model in terms of computational cost. This less conservative strategy approximates regions II and III, respectively by the τ-leap method [22] and the slow scale Stochastic Simulation Algorithm (ssSSA) [71]. The τ-leap method leaps over reactions with high-abundance reactants, since the inclusion of these reactions may not have a considerable effect on changing the corresponding propensity functions. The ssSSA simulates only slow reactions, assuming that fast reactions are always in partial-equilibrium or steady-state. We follow the strategy proposed by Liu *et al.* [27] where the dynamics of all mRNAs (region I) is placed in the slow subset and simulated by SSA, and the other three regions (II, III, and IV) are placed in the fast subset and modeled with ODEs (see fig. 3.2 c). We shall refer to this partitioning as the 'Liu strategy'. We notice that the partitioning thresholds in Liu strategy are predefined and static. That is, while it is not guaranteed, the fast and slow sets are assumed to remain the same during the simulation. Liu strategy was applied to a three-variable model of cell-cycle and achieved significant efficiency in comparison with other conservative partitioning strategies.

To demonstrate the scale difference in our partitioning strategy, we approximate the propensity function of every reaction by its corresponding rate law function (obtaining a stochastic model with 145 reactions) and track the firing frequency of each of these reactions in a test run of Gillespie's SSA. Figure 3.3 compare the firing frequency of fast and slow reactions. Of 18 million reactions fired in one cell cycle, only about 34,000 (0.2%) involve mRNA turnover, and 99.8% represent fast reactions of protein post-translational modifications. Based on this test run, we estimate that our hybrid scheme will run at least 100 times faster than a brute-force Gillespie simulation of a fully

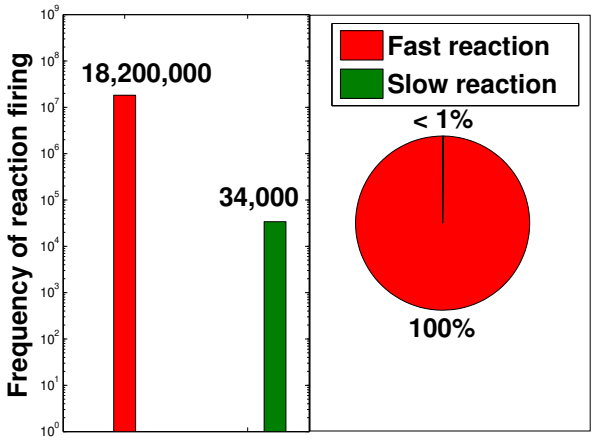

Figure 3.3: **Firing frequency of fast reactions versus slow reactions.** Firing frequency of reactions in the fast subset (fast reactions) within one cell cycle is compared with firing frequency of reactions in the slow subset.

stochastic model. In section 3.3 we show that our hybrid stochastic model, using the Liu strategy, still generates accurate results that agree well with experimental observations.

Algorithm 1 proposed in Liu *et al.* [27] describes the hybrid ODE/SSA algorithm adopted in this chapter, which is a variant of the original HR hybrid method [24]. Consider a well-stirred system with N species in a set $\mathcal{S}$ that interact with each other through M reaction channels in a set $\mathcal{R}$. The reactions in $\mathcal{R}$ are partitioned into two disjoint subsets of fast and slow reactions denoted by $\mathcal{R}_{\text{fast}}$ and $\mathcal{R}_{\text{slow}}$, respectively. The subset $\mathcal{R}_{\text{slow}}$ includes k reactions which are simulated using SSA, while the remaining M-k fast reactions in $\mathcal{R}_{\text{fast}}$ are governed by ODEs. Let $a_j(\boldsymbol{s}, t)$ be the propensity function of the j-th reaction in $\mathcal{R}_{\text{slow}}$, where $\boldsymbol{s} = (S_1(t), \ldots, S_N(t))$ is the state vector with each element $S_i(t)$ representing the number of molecules of species i at time t. In addition, let $\boldsymbol{v}_j = (v_{j1}, \ldots, v_{jN})$ be the state-change vector of the j-th reaction, where v_{ji} denotes the change in the population of species i when reaction j fires. Let τ be the jump interval to the next slow reaction and μ be the index of the reaction that fires. The algorithm only needs to simulate the firings of slow reactions, while integrating the fast subset of ODEs simultaneously in (3.16).

Algorithm 1 Hybrid Stochastic Simulation Algorithm

$HYBRID(\mathcal{R}_{\text{fast}}, \mathcal{R}_{\text{slow}})$

1: $t \leftarrow 0\ t < T$
2: Calculate the propensity function, a_i, for all reactions in slow subset $i = 1, ..., k$.
3: Calculate total propensity function: $a_0(\boldsymbol{s}, t) = \sum_{j=1}^{k} a_j(\boldsymbol{s}, t)$.
4: Generate two uniform random variables r_1 and r_2 in $\mathrm{U}(0, 1)$.
5: Integrate the ODE system until an event occurs at time $t + \tau$ such that

$$\int_{t}^{t+\tau} a_0(\boldsymbol{s}, x)\mathrm{d}x + \ln(r_1) = 0. \tag{3.16}$$

6: Select the smallest μ such that: $\sum_{i=1}^{\mu} a_i(\boldsymbol{s}, t) > r_2 a_0(\boldsymbol{s}, t)$.
7: Update the state variables according to μ_{th} reaction in $\mathcal{R}_{\text{slow}}$.

When a slow reaction fires, the corresponding state variables are updated. In this way the hybrid algorithm generates trajectories of state variables as the system proceeds in time. More details on implementation can be found in [72, 73, 74].

3.2.4 Implementation of stochastic events

When the normalized concentration of Clb2 (denoted as $[\text{Clb2}]_{\text{n}}$) drops below K_{ez}, we reset the auxiliary proteins $[\text{BUD}]_{\text{n}}$ and $[\text{SPN}]_{\text{n}}$ to zero, and divide all species in the cell, except for Cln3 and Bck2, between daughter and mother cells with a 40:60 ratio, according to observations by Di Talia *et al.* [12]. This ratio for Cln3 and Bck2 is set to 20:80 to match with the experimental observations in [68,67]. When [Clb2]+[Clb5] drops below K_{ez2}, $[\text{ORI}]_{\text{n}}$ is reset to zero. These auxiliary proteins are used as flags to specify particular events (check points in the budding yeast cell cycle). That is $[\text{BUD}]_{\text{n}}$=1 indicates bud emergence, $[\text{ORI}]_{\text{n}}$=1 specifies initiation of DNA synthesis and $[\text{SPN}]_{\text{n}}$ =1 signals that the chromosome alignment on spindle is completed.

The reset rules in our hybrid model are similar to the deterministic model by Chen *et al.* [9]. However, to prevent events from misfiring multiple times due to stochastic fluctuations in the hybrid model, we follow specific tactics. In our model, there are two types of events: *First*, events

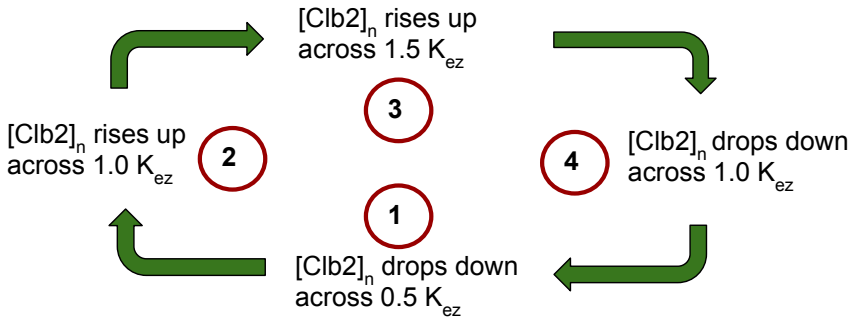

Figure 3.4: **Type-two event monitoring procedure.** Events 2 and 4 are the original events from the deterministic model [9]. Events 1 and 3 are the dummy events added to avoid misfiring of the original events in the presence of stochastic fluctuations. K_{ez} is a dimensionless concentration.

that are associated with the states of certain species that change across predefined thresholds in only one direction (increasing or decreasing), taking BUD as an example. *Second*, events that are associated with the states of certain species that change across predefined thresholds in both directions (increasing and decreasing), taking Clb2 as an instance. For the first type of event, the thresholds are set to the original predefined values at the beginning of the cell cycle. Once an event is triggered, the corresponding threshold is increased by multiplying it with a large number (e.g., 1000) to prevent misfiring.

For the second type of event, we add two dummy events to prevent misfiring. Supplementary figure S3.4 illustrates an example which describes the procedure we use to monitor type-two events for Clb2. The goal is to monitor the occurrence of events 2 and 4. We add events 1 and 3 to prevent misfiring of events 2 and 4 in our hybrid stochastic model. At the beginning of the cell cycle, all thresholds are set to very high values except for the first event which is set to $0.5K_{\text{ez}}$. Once event i occurs, the threshold for i is set to a very large value and the threshold for the event $i+1$ is reset to its original predefined value. Using this procedure, we guarantee that events 2 and 4 correctly take place in the presence of stochastic fluctuations.

3.3 Results

Our hybrid stochastic model of the budding yeast cell cycle, consisting of 45 proteins and 19 mRNAs, is simulated by Algorithm 1 to generate sufficiently large populations of mother and daughter cells to estimate the statistical distributions of various cell-cycle-related properties of wild-type cells. We evaluate our model by comparing numerical simulation results with experimental observations from the published literature.

Figure 3.5 a and b show a deterministic simulation of several protein and mRNA species, respectively, in our model of wild-type budding yeast cells. In the early G1 phase, once the cell grows to a critical size, Cln3 and Bck2 initiate the START event, i.e., the activation of transcription factors for Cln2 and Clb5 production. Cln2 is the cyclin responsible for bud formation. In addition, Cln2 phosphorylates Sic1 and Cdc6, a pair of cyclin-dependent kinase inhibitors (CKIs). Consequently, Clb5-dependent kinase activity rises and initiates DNA replication in the S phase. As CKIs are removed, Clb2 level rises, because Clb2 activates its own transcription factor, Mcm1, in an autocatalytic fashion. Clb2-dependent kinase activity turns off the transcription factors for Cln2 and Clb5 production and is responsible for driving the cell into mitosis (M phase). Clb2 level remains high until metaphase when the proper attachment of chromosomes to the mitotic spindle activates Cdc20. Cdc20 promotes anaphase (the separation of the two strands of replicated chromosomes to opposite poles of the mitotic spindle). At the same time, Cdc20 promotes the degradation of Clb2 and Clb5 and activation of a phosphatase, Cdc14. Cdc14 plays a major role (in budding yeast) in re-establishing the dominance of CKIs in G1 phase, and in replacing Cdc20 by Cdh1 (the protein responsible for Clb2 and Clb5 degradation in G1 phase). Figure 3.5 c and d show the corresponding stochastic trajectories of our hybrid stochastic model. The stochastic trajectories in panel c correctly simulate the sequence of events predicted by the deterministic trajectories in panel a. Panel d shows considerable fluctuations in the numbers of molecules per cell of five mRNA species: three of which ($mClb2$, $mCdc20$, and $mCln2$) undergo periodic transcription during the cell cycle, and

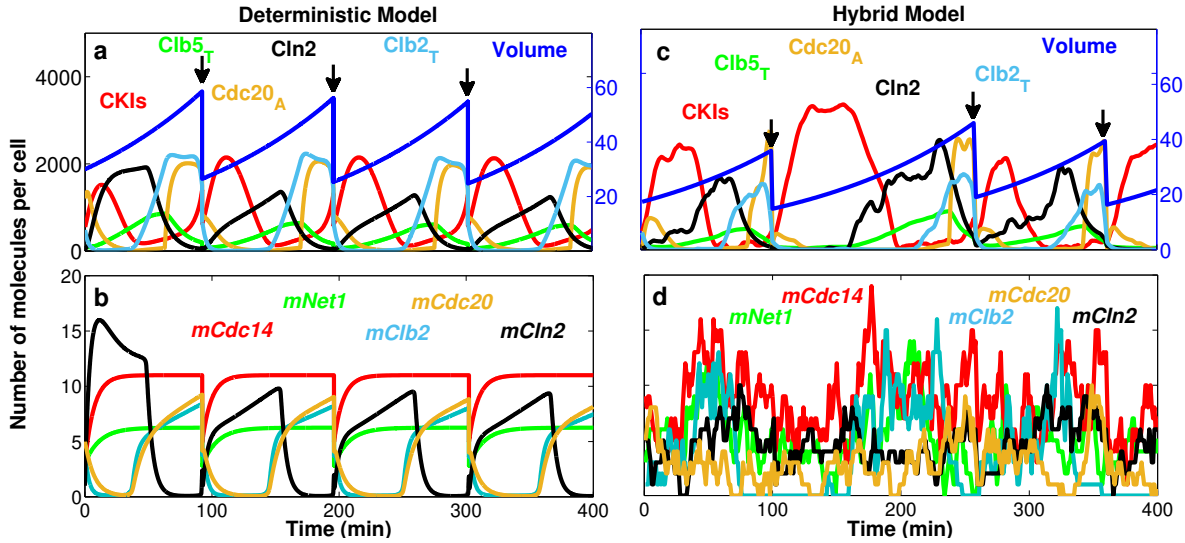

Figure 3.5: **Deterministic and hybrid stochastic simulations of the model. a, b:** The temporal dynamics of representative proteins (a) and mRNAs (b) generated by the deterministic model. The volume of the cell increases exponentially and is divided (at the arrows) asymmetrically between mother (55%) and daughter cell (45%). **c, d**: Stochastic simulation of the same proteins (c) and mRNAs (d) as in panels a and b, generated by a representative run of our hybrid stochastic model. Similar to the deterministic model, the cell grows exponentially; however, at the time of division all species in the cell, except for Cln3 and Bck2, are partitioned between daughter and mother cells with a 40:60 ratio, according to observations by Di Talia *et al.* [12]. Cln3 and Bck2, which are preferentially retained in mother cells [68, 67], are partitioned with a ratio 20:80 between daughter and mother cells. The daughter cell is tracked from division to division in this simulation.

two of which ($mCdc14$ and $mNet1$) are synthesized continuously throughout the cell cycle. Such high variability is expected, due to the stochastic nature of gene expression and the low abundances of mRNA molecules per cell.

We used our hybrid stochastic model to generate more than 20,000 asynchronous wild-type mother and daughter cells growing in glucose medium (mass doubling time about 100 min). These large collections of simulated cells are then used to estimate the distributions of important characteristics of the budding yeast cell cycle, including inter-division time, duration of unbudded phase (G1), duration of budded phase (S-G2-M), and size at birth. In addition we estimate the standard error (SE) of the mean and standard error of coefficient of variation for these characteristics.

Table 3.1 compares the computed summary statistics $\pm$ SE for all cell cycle-related properties

Table 3.1: **Mean and coefficient of variation (CV) for cell cycle properties**. Mean $\pm$ SE and CV $\pm$ SE computed from simulation of the hybrid stochastic model are compared with experimental observations reported by Di Talia *et al.* [12]. The standard errors of the mean are in the same unit of the corresponding characteristic. The number of experimental observations is reported in parenthesis and the number of simulations used to calculate each quantity is at least $10,000$. T_{div}, T_{G1}, T_{SG2M}, and V_{birth} are, respectively, cell cycle duration or the time between two divisions, time from division to next emergence of bud, time from onset of bud to next division, and volume of the cell at birth.

		Mother Cell		Daughter Cell	
		Hybrid model	Experiment	Hybrid model	Experiment
T_{div} (min)	mean $\pm$ SE	87 ± 0.22	87 ± 1 (116)	111 ± 0.36	112 ± 3 (97)
T_{G1}(min)		18 ± 0.06	16 ± 0.59 (158)	37 ± 0.21	37 ± 2 (202)
T_{SG2M} (min)		69 ± 0.20	72 ± 1 (116)	73 ± 0.26	76 ± 2 (97)
V_{birth} (fL)		41 ± 0.08	40	27 ± 0.05	28
T_{div} (min)	CV $\pm$ SE	0.26 ± 0.0017	0.14 ± 0.01 (116)	0.33 ± 0.0023	0.22 ± 0.02 (97)
T_{G1}(min)		0.35 ± 0.0024	0.50 ± 0.05 (158)	0.60 ± 0.0041	0.50 ± 0.05 (202)
T_{SG2M} (min)		0.30 ± 0.0020	0.17 ± 0.02 (116)	0.36 ± 0.0017	0.2 ± 0.06 (97)
V_{birth} (fL)		0.28 ± 0.0014	0.18	0.28 ± 0.0014	0.20

with experimental data reported by Di Talia *et al.* [12]. The results in Table 3.1 show that the model accurately reproduces the mean of these important properties of the wild-type budding yeast cell cycle. Despite the fact that the coefficients of variation reproduced by our model are generally larger than what is observed in the experiment, they are in a comparable range. In accordance with experimental observations, the G1 phase is the noisiest phase in the cell cycle, the variability in daughter cells is more than mother cells. The estimated standard errors are significantly smaller than the experimental observations. In fact, we expect such low standard errors due to the large number of simulations. We note that the standard error for the volume of a cell at birth is not

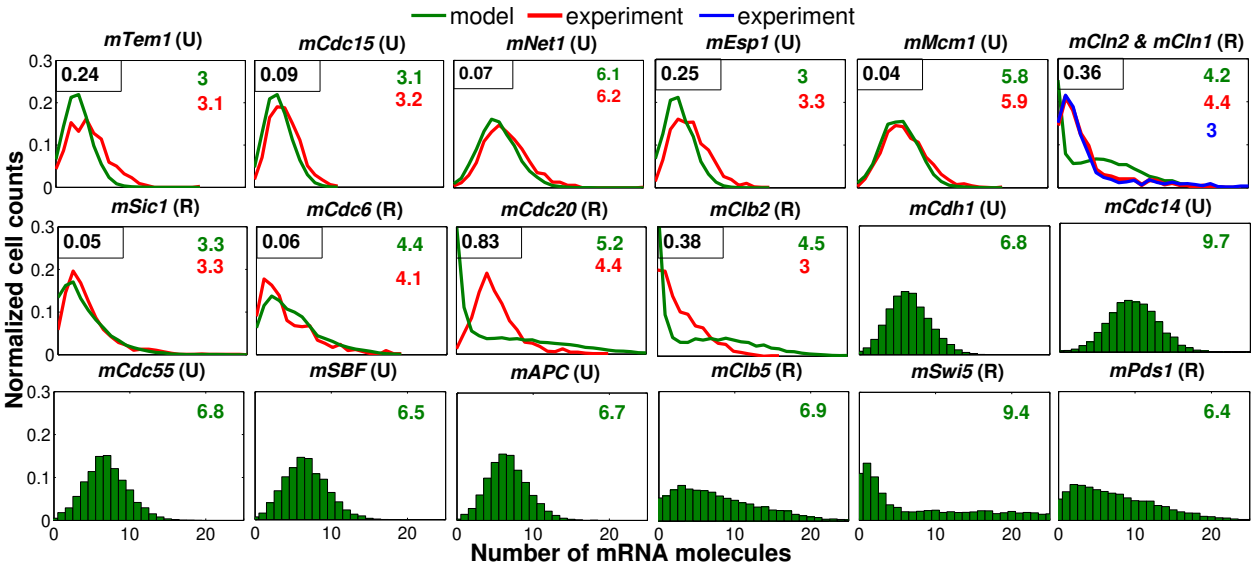

Figure 3.6: **Histograms of mRNAs for a population of wild-type cells growing in glucose medium.** The histograms of mRNA molecules generated from a stochastic run of the hybrid model (in green) are compared with experimental observations [11] (in red and blue colors) for a population of wild-type cells growing in glucose. (In the simulation the growth rate is set to 0.0072 min^{-1} to reproduce the 96 min mass-doubling time of wild-type cells growing in glucose culture medium.) U and R in parenthesis indicate, respectively, unregulated and transcriptionally regulated mRNAs. The histograms in red are reproduced from the experimental data reported by Ball *et al.* [11]. For the last 8 transcripts, experimental data are not available. On the top-right corner the average number of mRNA molecules is compared with experiment where available. On the top-left corner the Kullback-Leibler divergence (D_{KL}) is reported to quantify the difference between the two distributions. $D_{\text{KL}} = 0$ indicates that the two distributions in question are identical. In our model $mCln2$ stands for $mCln1 + mCln2$. In experiment, however, they are measured separately. Here, the histograms in red and blue are, respectively, $mCln1$ and $mCln2$. Similarly, in our model $mClb2$ describes the abundance of both $mClb1$ and $mClb2$; however, the histogram reproduced from the experimental data refers only to $mClb2$.

reported in column 4 and 6, because cell volume is not measured directly by Di Talia *et al.* [12], but rather $V(t)$ is estimated as a function of time.

Next, we compare our simulations to the observed distributions of mRNA molecules in wild-type yeast cells. We have 11 unregulated mRNAs ($mCdh1$, $mTem1$, $mCdc15$, $mCdc14$, $mNet1$, $mCdc55$, $mEsp1$, $mSBF$, $mMBF$, $mMcm1$, $mAPC$) and 8 transcriptionally regulated mRNAs ($mClb5$, $mCln2$, $mClb2$, $mSic1$, $mCdc6$, $mSwi5$, $mCdc20$, $mPds1$) in our model. Figure 3.6 compares the histograms of these mRNAs with the distributions measured by Ball *et al.* [11]. In

the original deterministic model, MBF and SBF are described by the same algebraic equation since they were supposed to have the same dynamics [9]. In adding $mSBF$ and $mMBF$ to the model, we kept the same assumption and therefore, the histograms of the two unregulated mRNAs ($mSBF$ and $mMBF$) are very similar. For this reason we just include one of them ($mSBF$) in Figure 2. To quantify the difference between a distribution generated by our model and the corresponding experimental observations, we use the Kullback-Leibler divergence ($D_{\text{KL}} = \sum_{x \in X} R(x) log(\frac{Q(x)}{R(x)})$) where R is the distribution from simulation and Q from experiment. The computed value of the KL divergence is reported on the top-left corner of each subplot. The smaller is D_{KL}, the more similar are the two distributions.

Generally, there is a good agreement between observed and simulated mRNA distributions, except for $mCln1 + mCln2$ and $mCdc20$. As expected, the unregulated transcripts follow Poisson distributions, which are consistent with experimental measurements. The value of D_{KL} computed for these distribution is small. The cell-cycle regulated transcripts, which follow long-tailed, non-Poisson distributions, are well-fit by two-component Poisson distributions as reported by [11, 54]. (We note that in our model $mClb2$ represents both $mClb1$ and $mClb2$, and $mCln2 = mCln1 + mCln2$, whereas in the experiment these cyclin mRNAs are tracked independently. Therefore, we do not expect a particularly good match between the computed and observed distributions for these transcripts. As expected, the values of D_{KL} computed for these distribution are large.)

Table 3.2 compares the average abundances of proteins as observed in [66] and simulated by our model. We use a sufficiently large population of cells from at least $10,000$ simulations to calculate the average abundance (number of molecules per cell) and the standard error of the mean for each protein. Note that, for the proteins listed in Table 3.2, only a single measurement has been made experimentally, so the standard error is not available for comparison. Except for the Sic1, the agreement between simulation results and experimental observations is quite reasonable.

Table 3.2: **Average abundances of protein molecules per cell.** The average abundance $\pm$ standard error of proteins in molecules per cell, computed by the hybrid stochastic model, are compared with experimental observations reported in [66]. In our model Clb5 stands for Clb5 and Clb6, Clb2 stands for Clb1 and Clb2, and Cln2 stands for Cln1 and Cln2. We are reporting the total abundance of each protein, which includes protein molecules that are either phosphorylated or unphosphorylated, and that are bound in complexes or free. That is, Clb5_T = Clb5 + C5 + C5P + F5 + F5P, Clb2_T = Clb2 + C2 + C2P + F2 + F2P, and Sic1_T = Sic1 + Sic1P + C2 + C2P + C5 + C5P.

Protein	Average abundance ± standard error		Protein	Average abundance ± standard error	
	Experiment	Hybrid model		Experiment	Hybrid model
Cln3	108	109 ± 2	Swi5_T	688	658 ± 6
Cln2_T = Cln1 + Cln2	1589 = 319 + 1270	1647 ± 18	Tem1	573	544 ± 11
Clb5_T	420	516 ± 5	Cdc15	238	257 ± 4
Clb2_T	693	736 ± 12	Net1_T	1590	1579 ± 8
Sic1_T	768	511 ± 7	Cdc55	3170	3357 ± 25

3.4 Summary

In this chapter, we have presented a hybrid stochastic model of the molecular mechanism controlling progression through the budding yeast cell cycle. Our model provides a good match with experimental observations of many important characteristics of the budding yeast cell cycle, including inter-division time, cell size. Compared with other approaches to stochastic modeling, our hybrid stochastic approach has several advantages. In a multiscale regulatory network such as cell cycle controls, the major source of intrinsic noise can be attributed to low copy numbers of mRNA species in the gene-protein regulatory network. In fact, in budding yeast cells, there are only 5-10 copies of each mRNA species encoding the production of corresponding proteins at levels of 500-5,000 molecules per cell. In such circumstances, small fluctuations in the population of mRNAs will result in substantial fluctuations in the corresponding protein levels. With this in mind, the key idea of the hybrid scheme is to partition the dynamics of mRNA species into the stochastic

regime, in order to capture the major effects of random fluctuations in mRNA numbers, and to keep the protein dynamics in the deterministic framework, to achieve greater simulation efficiency. In addition, in this scheme, it is not necessary to reformulate the complex rate laws governing protein interactions as elementary mass-action rate laws, which is a great advantage from a modeling standpoint.

We have applied our hybrid stochastic method to a detailed molecular mechanism of cell cycle controls in budding yeast [9]. To apply our scheme to Chen's model, which is a deterministic model of protein interactions, we first had to extend the model to include mRNA species that are transcribed from cell-cycle genes and translated into proteins. Then we carried out comprehensive simulations of wild-type yeast cells using both the deterministic and hybrid ODE/SSA models. Our stochastic model predicts the statistical properties of many different cell-cycle variables, including inter-division times, size at birth, and the abundances of specific mRNAs and proteins, and our stochastic simulations are in accord with most experimental observations. In addition, our results prove that our hybrid approach to stochastic/deterministic simulations can achieve a good trade-off between the accuracy and efficiency of numerical simulations. FORTRAN code takes about 15 min to simulate 10,000 cell cycles on an Intel i7-3770 processor with 16G memory running a Linux environment. A similar system using a fully stochastic model may take more than one day (for example, when the FORTRAN code of Barik *et al.* [54] is run using the same work station, it takes more than 30 hours to generate a population of 10,000 yeast cells).

3.5 Appendix A

A.1 Variables, description of variables, and initial values

Table 3.3: **Variables and initial values.** Except for mass which is a normalized (dimensionless) variable, all other variables are based on number of molecules per cell. Initial values are given for all variables that are governed by differential equations. Some other variables are expressed in algebraic functions of the time-dependent variables, so their initial values are not listed here.

Variable	Description	Initial value
mass	Indicator of cell size (dimensionless)	1.04
Cln2	Total cyclins Cln1 and Cln2 (represented as Cln2 in the model)	21
Clb5	Active forms of cyclins Clb5 and Clb6 (represented as Clb5 in the model)	48
Clb2	Active forms of Clb1 and Clb2 (represented as Clb2 in the model)	233
Sic1	A stoichiometric inhibitor of Cdc28/Clb2 and Cdc28/Clb5	8
Sic1P	Phophorylated form of Sic1	6
C2	A complex formed by Clb2 and Sic1	149
C5	A complex formed by Clb5 and Sic1	32
C2P	Phophorylated form of C2	31
C5P	Phophorylated form of C5	6
Cdc6	A stoichiometric inhibitor of Cdc28/Clb2	22
Cdc6P	Phophorylated form of Cdc6	11
F2	A complex formed by Clb2 and Cdc6	66
F5	A complex formed by Clb5 and Cdc6	0
F2P	Phophorylated form of F2	26
F5P	Phophorylated form of F5	0
Swi5	Transcription factor for Sic1 and Cdc6	613
$Swi5_T$	Total Swi5	634

APC	Anaphase Promoting Complex	285
APCP	Active (phosphorylated) form of APC	16
$Cdc20_T$	Total Cdc20	4059
Cdc20A	Active form of Cdc20 (an APC partner)	752
Cdh1T	Total Cdh1	1273
Cdh1	Active Cdh1 (an APC partner)	895
Tem1	A GTP-binding protein	1140
$Tem1_T$	Total Tem1	1166
Cdc15	A kinase required for nuclear division	342
$Cdc15_T$	Total Cdc15	518
Cdc14	A phosphatase required for mitosis exit	121
$Cdc14_T$	Total of Cdc14	607
Net1	A stoichiometric inhibitor of Cdc14	7
$Net1_T$	Total Net1	838
Net1P	Phosphorylated form of Net1	-
RENT	A protein complex of Net1 and Cdc14 that regulates nucleolar silencing and telophase	324
RENTP	Phosphorylated form of RENT	-
Cdc55	Active form of CDC55 phosphatase	1161
Esp1	A protein required for sister chromatid separation	9
$Esp1_T$	Total Esp1	42
Pds1	A stoichiometric inhibitor of Esp1	2
ORI	An auxiliary variable representing proteins that signal the onset of DNA synthesis	2374

BUD	An auxiliary variable representing proteins that initiate a new bud	0
SPN	An auxiliary variable representing proteins that signal the alignment of all chromosomes	0
SBF	Transcription factor for Cln2	-
SBF_T	Total SBF	527
MBF	Transcription factor for Clb5	-
MBF_T	Total MBF	442
Mcm1	Transcription factor for Clb2, Cdc20 and Swi5	-
$Mcm1_T$	Total Mcm1	10363
Cln3	A G1-cyclin initiating START events	50
Bck2	A back-up protein initiating START events	25
$Clb5_T$	Total Clb5	-
$Clb2_T$	Total Clb2	-
$Sic1_T$	Total Sic1	-
$Cdc6_T$	Total Cdc6	-
CKI_T	Total cyclin inhibitors Sic1 and Cdc6	-
$mCdh1$	Cdh1 messenger RNA	4
$mTem1$	Tem1 messenger RNA	2
$mCdc15$	Cdc15 messenger RNA	2
$mCdc14$	Cdc14 messenger RNA	6
$mNet1$	Net1 messenger RNA	3
$mCdc55$	Cdc55 messenger RNA	4
$mEsp1$	Esp1 messenger RNA	2

$mSBF$	SBF messenger RNA	4
$mMBF$	MBF messenger RNA	4
$mMcm1$	Mcm1 messenger RNA	3
$mAPC$	APC messenger RNA	4
$mCln2$	Cln2 messenger RNA	1
$mClb5$	Clb5 messenger RNA	1
$mClb2$	Clb2 messenger RNA	4
$mSic1$	Sic1 messenger RNA	3
$mCdc6$	Cdc6 messenger RNA	4
$mSwi5$	Swi5 messenger RNA	9
$mCdc20$	Cdc20 messenger RNA	5
$mPds1$	Pds1 messenger RNA	4

A.2 Equations

Table 3.4: **Equations.**

$\frac{\mathrm{d}mass}{\mathrm{d}t} = k_{\mathrm{g}} \cdot mass$
$\frac{\mathrm{d}Cln2}{\mathrm{d}t} = (k'_{\mathrm{s,n2}} \cdot c_{\mathrm{cln2}} \cdot mass + k''_{\mathrm{s,n2}} \cdot \frac{c_{\mathrm{cln2}}}{c_{\mathrm{sbf}}} \cdot SBF) \cdot mass \cdot mCln2 - k_{\mathrm{d,n2}} \cdot Cln2 + k_{\mathrm{g}} \cdot Cln2$
$\frac{\mathrm{d}Clb5}{\mathrm{d}t} = (k'_{\mathrm{s,b5}} \cdot c_{\mathrm{clb5}} \cdot mass + k''_{\mathrm{s,b5}} \cdot \frac{c_{\mathrm{clb5}}}{c_{\mathrm{mbf}}} \cdot MBF) \cdot mass \cdot mClb5 + (k_{\mathrm{d3,c1}} \cdot C5\mathrm{P} + k_{\mathrm{di,b5}} \cdot C5) +$ $(k_{\mathrm{d3,f6}} \cdot F5\mathrm{P} + k_{\mathrm{di,f5}} \cdot F5) - (V_{\mathrm{d,b5}} + \frac{k_{\mathrm{as,b5}}}{c_{\mathrm{sic1}} \cdot mass} \cdot Sic1 + \frac{k_{\mathrm{as,f5}}}{c_{\mathrm{cdc6}} \cdot mass} \cdot Cdc6) \cdot Clb5 + k_{\mathrm{g}} \cdot Clb5$
$\frac{\mathrm{d}Clb2}{\mathrm{d}t} = (k'_{\mathrm{s,b2}} \cdot c_{\mathrm{clb2}} \cdot mass + k''_{\mathrm{s,b2}} \cdot \frac{c_{\mathrm{clb2}}}{c_{\mathrm{mcm1}}} \cdot Mcm1) \cdot mass \cdot mClb2 + (k_{\mathrm{d3,c1}} \cdot C2\mathrm{P} + k_{\mathrm{di,b2}} \cdot C2) +$ $(k_{\mathrm{d3,f6}} \cdot F2\mathrm{P} + k_{\mathrm{di,f2}} \cdot F2) - (V_{\mathrm{d,b2}} + \frac{k_{\mathrm{as,b2}}}{c_{\mathrm{sic1}} \cdot mass} \cdot Sic1 + \frac{k_{\mathrm{as,f2}}}{c_{\mathrm{cdc6}} \cdot mass} \cdot Cdc6) \cdot Clb2 + k_{\mathrm{g}} \cdot Clb2$
$\frac{\mathrm{d}Sic1}{\mathrm{d}t} = (k'_{\mathrm{s,c1}} \cdot c_{\mathrm{sic1}} \cdot mass + k''_{\mathrm{s,c1}} \frac{c_{\mathrm{sic1}}}{c_{\mathrm{swi5}}} \cdot Swi5) \cdot mSic1 + (V_{\mathrm{d,b2}} + k_{\mathrm{di,b2}}) \cdot C2\mathrm{P} + (V_{\mathrm{d,b5}} + k_{\mathrm{di,b5}}) \cdot$ $C5\mathrm{P} + \frac{k_{\mathrm{pp,c1}}}{c_{\mathrm{cdc14}} \cdot mass} \cdot Cdc14 \cdot Sic1\mathrm{P} - (\frac{k_{\mathrm{as,b2}}}{c_{\mathrm{clb2}} \cdot mass} \cdot Clb2 + \frac{k_{\mathrm{as,b5}}}{c_{\mathrm{clb5}} \cdot mass} \cdot Clb5 + V_{\mathrm{kp,c1}}) \cdot Sic1 + k_{\mathrm{g}} \cdot Sic1$
$\frac{\mathrm{d}Sic1P}{\mathrm{d}t} = V_{\mathrm{kp,c1}} \cdot Sic1 - (\frac{k_{\mathrm{pp,c1}}}{c_{\mathrm{cdc14}} \cdot mass} \cdot Cdc14 + k_{\mathrm{d3,c1}}) \cdot Sic1\mathrm{P} + V_{\mathrm{d,b2}} \cdot C2\mathrm{P} + V_{\mathrm{d,b5}} \cdot C5\mathrm{P} + k_{\mathrm{g}} \cdot Sic1\mathrm{P}$

$$\frac{\mathrm{d}C2}{\mathrm{d}t} = \frac{k_{\mathrm{as,b2}}}{c_{\mathrm{clb2}} \cdot mass} \cdot \mathrm{Clb2} \cdot Sic1 + \frac{k_{\mathrm{pp,c1}}}{c_{\mathrm{cdc14}} \cdot mass} \cdot Cdc14 \cdot C2\mathrm{P} - (k_{\mathrm{di,b2}} + V_{\mathrm{d,b2}} + V_{\mathrm{kp,c1}}) \cdot C2 + k_{\mathrm{g}} \cdot C2$$

$$\frac{\mathrm{d}C5}{\mathrm{d}t} = \frac{k_{\mathrm{as,b5}}}{c_{\mathrm{clb5}} \cdot mass} \cdot \mathrm{Clb5} \cdot Sic1 + \frac{k_{\mathrm{pp,c1}}}{c_{\mathrm{cdc14}} \cdot mass} \cdot Cdc14 \cdot C5\mathrm{P} - (k_{\mathrm{di,b5}} + V_{\mathrm{d,b5}} + V_{\mathrm{kp,c1}}) \cdot C5 + k_{\mathrm{g}} \cdot C5$$

$$\frac{\mathrm{d}C2\mathrm{P}}{\mathrm{d}t} = V_{\mathrm{kp,c1}} \cdot C2 - (\frac{k_{\mathrm{pp,c1}}}{c_{\mathrm{cdc14}} \cdot mass} \cdot Cdc14 + k_{\mathrm{d3,c1}} + V_{\mathrm{d,b2}}) \cdot C2\mathrm{P} + k_{\mathrm{g}} \cdot C2\mathrm{P}$$

$$\frac{\mathrm{d}C5\mathrm{P}}{\mathrm{d}t} = V_{\mathrm{kp,c1}} \cdot C5 - (\frac{k_{\mathrm{pp,c1}}}{c_{\mathrm{cdc14}} \cdot mass} \cdot Cdc14 + k_{\mathrm{d3,c1}} + V_{\mathrm{d,b5}}) \cdot C5\mathrm{P} + k_{\mathrm{g}} \cdot C5\mathrm{P}$$

$$\frac{\mathrm{d}Cdc6}{\mathrm{d}t} = (k'_{\mathrm{s,f6}} \cdot c_{\mathrm{cdc6}} \cdot mass + k''_{\mathrm{s,f6}} \cdot \frac{c_{\mathrm{cdc6}}}{c_{\mathrm{swi5}}} \cdot Swi5 + k'''_{\mathrm{s,f6}} \cdot \frac{c_{\mathrm{cdc6}}}{c_{\mathrm{sbf}}} \cdot SBF) \cdot mCdc6 + (V_{\mathrm{d,b2}} + k_{\mathrm{di,f2}}) \cdot F2 + (V_{\mathrm{d,b5}} + k_{\mathrm{di,f5}}) \cdot F5 + \frac{k_{\mathrm{pp,f6}}}{c_{\mathrm{cdc14}} \cdot mass} \cdot Cdc14 \cdot Cdc6P - (\frac{k_{\mathrm{as,f2}}}{c_{\mathrm{clb2}} \cdot mass} \cdot Clb2 + \frac{k_{\mathrm{as,f5}}}{c_{\mathrm{clb5}} \cdot mass} \cdot Clb5 + V_{\mathrm{kp,f6}}) \cdot Cdc6 + k_{\mathrm{g}} \cdot Cdc6$$

$$\frac{\mathrm{d}Cdc6\mathrm{P}}{\mathrm{d}t} = V_{\mathrm{kp,f6}} \cdot Cdc6 - (\frac{k_{\mathrm{pp,f6}}}{c_{\mathrm{cdc14}} \cdot mass} \cdot Cdc14 + k_{\mathrm{d3,f6}}) \cdot Cdc6\mathrm{P} + V_{\mathrm{d,b2}} \cdot F2\mathrm{P} + V_{\mathrm{d,b5}} \cdot F5\mathrm{P} + k_{\mathrm{g}} \cdot Cdc6\mathrm{P}$$

$$\frac{\mathrm{d}F2}{\mathrm{d}t} = \frac{k_{\mathrm{as,f2}}}{c_{\mathrm{clb2}} \cdot mass} \cdot \mathrm{Clb2} \cdot Cdc6 + \frac{k_{\mathrm{pp,f6}}}{c_{\mathrm{cdc14}} \cdot mass} \cdot Cdc14 \cdot F2\mathrm{P} - (k_{\mathrm{di,f2}} + V_{\mathrm{d,b2}} + V_{\mathrm{kp,f6}}) \cdot F2 + k_{\mathrm{g}} \cdot F2$$

$$\frac{\mathrm{d}F5}{\mathrm{d}t} = \frac{k_{\mathrm{as,f5}}}{c_{\mathrm{clb5}} \cdot mass} \cdot Clb5 \cdot Cdc6 + \frac{k_{\mathrm{pp,f6}}}{c_{\mathrm{cdc14}} \cdot mass} \cdot Cdc14 \cdot F5\mathrm{P} - (k_{\mathrm{di,f5}} + V_{\mathrm{d,b5}} + V_{\mathrm{kp,f6}}) \cdot F5 + k_{\mathrm{g}} \cdot F5$$

$$\frac{\mathrm{d}F2\mathrm{P}}{\mathrm{d}t} = V_{\mathrm{kp,f6}} \cdot F2 - (\frac{k_{\mathrm{pp,f6}}}{c_{\mathrm{cdc14}} \cdot mass} \cdot Cdc14 + k_{\mathrm{d3,f6}} + V_{\mathrm{d,b2}}) \cdot F2\mathrm{P} + k_{\mathrm{g}} \cdot F2\mathrm{P}$$

$$\frac{\mathrm{d}F5\mathrm{P}}{\mathrm{d}t} = V_{\mathrm{kp,f6}} \cdot F5 - (\frac{k_{\mathrm{pp,f6}}}{c_{\mathrm{cdc14}} \cdot mass} \cdot Cdc14 + k_{\mathrm{d3,f6}} + V_{\mathrm{d,b5}}) \cdot F5\mathrm{P} + k_{\mathrm{g}} \cdot F5\mathrm{P}$$

$$\frac{\mathrm{d}Swi5_{\mathrm{T}}}{\mathrm{d}t} = (k'_{\mathrm{s,swi}} \cdot c_{\mathrm{swi5}} \cdot mass + k''_{\mathrm{s,swi}} \cdot \frac{c_{\mathrm{swi5}}}{c_{\mathrm{mcm1}}} \cdot Mcm1) \cdot mSwi5 - k_{\mathrm{d,swi}} \cdot Swi5T + k_{\mathrm{g}} \cdot Swi5_{\mathrm{T}}$$

$$\frac{\mathrm{d}Swi5}{\mathrm{d}t} = (k'_{\mathrm{s,swi}} \cdot c_{\mathrm{swi5}} \cdot mass + k''_{\mathrm{s,swi}} \cdot \frac{c_{\mathrm{swi5}}}{c_{\mathrm{mcm1}}} \cdot Mcm1) \cdot mSwi5 + \frac{k_{\mathrm{a,swi}}}{c_{\mathrm{cdc14}} \cdot mass} \cdot Cdc14 \cdot (Swi5_{\mathrm{T}} - Swi5) - (k_{\mathrm{d,swi}} + \frac{k_{\mathrm{i,swi}}}{c_{\mathrm{clb2}} \cdot mass} \cdot Clb2) \cdot Swi5 + k_{\mathrm{g}} \cdot Swi5$$

$$\frac{\mathrm{d}APC\mathrm{P}}{\mathrm{d}t} = \frac{k_{\mathrm{a,apc}} \cdot \frac{c_{\mathrm{apc}}}{c_{\mathrm{clb2}}} \cdot Clb2 \cdot (APC - APC\mathrm{P})}{J_{\mathrm{a,apc}} \cdot c_{\mathrm{apc}} \cdot mass + APC - APC\mathrm{P}} - \frac{k_{\mathrm{i,apc}} \cdot c_{\mathrm{apc}} \cdot mass \cdot APC\mathrm{P}}{J_{\mathrm{i,apc}} \cdot c_{\mathrm{apc}} \cdot mass + APC\mathrm{P}} + k_{\mathrm{g}} \cdot APC\mathrm{P}$$

$$\frac{\mathrm{d}Cdc20_{\mathrm{T}}}{\mathrm{d}t} = (k'_{\mathrm{s,20}} \cdot c_{\mathrm{cdc20}} \cdot mass + k''_{\mathrm{s,20}} \cdot \frac{c_{\mathrm{cdc20}}}{c_{\mathrm{mcm1}}} \cdot Mcm1) \cdot mCdc20 - k_{\mathrm{d,20}} \cdot Cdc20_{\mathrm{T}} + k_{\mathrm{g}} \cdot Cdc20_{\mathrm{T}}$$

$$\frac{\mathrm{d}Cdc20\mathrm{A}}{\mathrm{d}t} = (k'_{\mathrm{a,20}} + \frac{k''_{\mathrm{a,20}}}{c_{\mathrm{cdc20}} \cdot mass} \cdot APC\mathrm{P}) \cdot (Cdc20_{\mathrm{T}} - Cdc20\mathrm{A}) - (k_{\mathrm{mad2}} + k_{\mathrm{d,20}}) \cdot Cdc20\mathrm{A} + k_{\mathrm{g}} \cdot Cdc20\mathrm{A}$$

$$\frac{\mathrm{d}Cdh1_{\mathrm{T}}}{\mathrm{d}t} = k_{\mathrm{s,cdh}} \cdot c_{\mathrm{cdh1}} \cdot mass \cdot mCdh1 - k_{\mathrm{d,cdh}} \cdot Cdh1_{\mathrm{T}} + k_{\mathrm{g}} \cdot Cdh1_{\mathrm{T}}$$

$$\frac{\mathrm{d}Cdh1}{\mathrm{d}t} = k_{\mathrm{s,cdh}} \cdot c_{\mathrm{cdh1}} \cdot mass \cdot mCdh1 - k_{\mathrm{d,cdh}} \cdot Cdh1 + \frac{V_{\mathrm{a,cdh}} \cdot c_{\mathrm{cdh1}} \cdot mass \cdot (Cdh1_{\mathrm{T}} - Cdh1)}{J_{\mathrm{a,cdh}} \cdot c_{\mathrm{cdh1}} \cdot mass + Cdh1_{\mathrm{T}} - Cdh1} - \frac{V_{\mathrm{i,cdh}} \cdot c_{\mathrm{cdh1}} \cdot mass \cdot Cdh1}{J_{\mathrm{i,cdh}} \cdot c_{\mathrm{cdh1}} \cdot mass + Cdh1} + k_{\mathrm{g}} \cdot Cdh1$$

$$\frac{\mathrm{d}Tem1}{\mathrm{d}t} = \frac{k_{\mathrm{lte1}} \cdot c_{\mathrm{tem1}} \cdot mass \cdot (Tem1_{\mathrm{T}} - Tem1)}{J_{\mathrm{a,tem}} \cdot c_{\mathrm{tem1}} \cdot mass + Tem1_{\mathrm{T}} - Tem1} - \frac{k_{\mathrm{bub2}} \cdot c_{\mathrm{tem1}} \cdot mass \cdot Tem1}{J_{\mathrm{i,tem}} \cdot c_{\mathrm{tem1}} \cdot mass + Tem1} + k_{\mathrm{g}} \cdot Tem1$$

$$\frac{\mathrm{d}Cdc15}{\mathrm{d}t} = (\frac{k'_{\mathrm{a,15}}}{c_{\mathrm{tem1}} \cdot mass} \cdot (Tem1_{\mathrm{T}} - Tem1) + \frac{k''_{\mathrm{a,15}}}{c_{\mathrm{tem1}} \cdot mass} \cdot \mathrm{Tem1} + \frac{k'''_{\mathrm{a,15}}}{c_{\mathrm{cdc14}} \cdot mass} \cdot Cdc14) \cdot (Cdc15T - Cdc15) - k_{\mathrm{i,15}} \cdot Cdc15 + k_{\mathrm{g}} \cdot Cdc15$$

$$\frac{\mathrm{d}Cdc14_{\mathrm{T}}}{\mathrm{d}t} = k_{\mathrm{s,14}} \cdot c_{\mathrm{cdc14}} \cdot mass \cdot mCdc14 - k_{\mathrm{d,14}} \cdot Cdc14_{\mathrm{T}} + k_{\mathrm{g}} \cdot Cdc14_{\mathrm{T}}$$

$$\frac{\mathrm{d}Cdc14}{\mathrm{d}t} = k_{\mathrm{s,14}} \cdot c_{\mathrm{cdc14}} \cdot mass \cdot mCdc14 - k_{\mathrm{d,14}} \cdot Cdc14 + k_{\mathrm{d,net}} \cdot (RENT + RENT\mathrm{P}) + k_{\mathrm{di,rent}} \cdot RENT + k_{\mathrm{di,rentp}} \cdot RENT\mathrm{P} + k_{\mathrm{g}} \cdot Cdc14 - (\frac{k_{\mathrm{as,rent}}}{c_{\mathrm{net1}} \cdot mass} \cdot Net1 + \frac{k_{\mathrm{as,rentp}}}{c_{\mathrm{net1}} \cdot mass} \cdot Net1\mathrm{P}) \cdot Cdc14$$

$$\frac{\mathrm{d}Net1_{\mathrm{T}}}{\mathrm{d}t} = k_{\mathrm{s,net}} \cdot c_{\mathrm{net1}} \cdot mass \cdot mNet1 - k_{\mathrm{d,net}} \cdot Net1_{\mathrm{T}} + k_{\mathrm{g}} \cdot Net1_{\mathrm{T}}$$

$$\frac{\mathrm{d}Net1}{\mathrm{d}t} = k_{\mathrm{s,net}} \cdot c_{\mathrm{cdc14}} \cdot mass \cdot mNet1 - k_{\mathrm{d,net}} \cdot Net1 + k_{\mathrm{d,14}} \cdot RENT + k_{\mathrm{di,rent}} \cdot RENT - \frac{k_{\mathrm{as,rent}}}{c_{\mathrm{cdc14}} \cdot mass} \cdot Cdc14 \cdot Net1 + V_{\mathrm{pp,net}} \cdot Net1\mathrm{P} - V_{\mathrm{kp,net}} \cdot Net1 + k_{\mathrm{g}} \cdot Net1$$

$$\frac{\mathrm{d}RENT}{\mathrm{d}t} = -(k_{\mathrm{d,14}} + k_{\mathrm{d,net}}) \cdot RENT - k_{\mathrm{di,rent}} \cdot RENT + \frac{k_{\mathrm{as,rent}}}{c_{\mathrm{cdc14}} \cdot mass} \cdot Cdc14 \cdot Net1 + V_{\mathrm{pp,net}} \cdot RENT\mathrm{P} - V_{\mathrm{kp,net}} \cdot RENT + k_{\mathrm{g}} \cdot RENT$$

$$\frac{\mathrm{d}Cdc55}{\mathrm{d}t} = k_{\mathrm{s,55}} \cdot c_{\mathrm{cdc55}} \cdot mass \cdot mCdc55 - V_{\mathrm{d,55}} \cdot Cdc55 + k_{\mathrm{g}} \cdot Cdc55$$

$$\frac{\mathrm{d}Pds1}{\mathrm{d}t} = (k'_{\mathrm{s,pds}} \cdot c_{\mathrm{pds1}} \cdot mass + k''_{\mathrm{s1,pds}} \cdot \frac{c_{\mathrm{pds1}}}{c_{\mathrm{sbf}}} \cdot SBF + k''_{\mathrm{s2,pds}} \cdot \frac{c_{\mathrm{pds1}}}{c_{\mathrm{mcm1}}} \cdot Mcm1) \cdot mPds1 + k_{\mathrm{di,esp}} \cdot PE - (V_{\mathrm{d,pds}} + \frac{k_{\mathrm{as,esp}}}{c_{\mathrm{esp1}} \cdot mass} \cdot Esp1) \cdot Pds1 + k_{\mathrm{g}} \cdot Pds1$$

$$\frac{\mathrm{d}Esp1}{\mathrm{d}t} = -\frac{k_{\mathrm{as,esp}}}{c_{\mathrm{pds1}} \cdot mass} \cdot Pds1 \cdot Esp1 + (k_{\mathrm{di,esp}} + V_{\mathrm{d,pds}}) \cdot PE + k_{\mathrm{g}} \cdot Esp1$$

$$\frac{\mathrm{d}ORI}{\mathrm{d}t} = k_{\mathrm{s,ori}} \cdot (e_{\mathrm{ori,b5}} \cdot \frac{c_{\mathrm{ori}}}{c_{\mathrm{clb5}}} \cdot Clb5 + e_{\mathrm{ori,b2}} \cdot \frac{c_{\mathrm{ori}}}{c_{\mathrm{clb2}}} \cdot Clb2) - k_{\mathrm{d,ori}} \cdot ORI + k_{\mathrm{g}} \cdot ORI$$

$$\frac{\mathrm{d}BUD}{\mathrm{d}t} = k_{\mathrm{s,bud}} \cdot (e_{\mathrm{bud,n2}} \cdot \frac{c_{\mathrm{bud}}}{c_{\mathrm{cln2}}} \cdot Cln2 + e_{\mathrm{bud,n3}} \cdot \frac{c_{\mathrm{bud}}}{c_{\mathrm{cln3}}} \cdot Cln3 + e_{\mathrm{bud,b5}} \cdot \frac{c_{\mathrm{bud}}}{c_{\mathrm{clb5}}} \cdot Clb5) - k_{\mathrm{d,bud}} \cdot BUD + k_{\mathrm{g}} \cdot BUD$$

$$\frac{\mathrm{d}SPN}{\mathrm{d}t} = \frac{k_{\mathrm{s,spn}} \cdot c_{\mathrm{spn}} \cdot mass \cdot Clb2}{J_{\mathrm{spn}} \cdot c_{\mathrm{clb2}} \cdot mass + Clb2} - k_{\mathrm{d,spn}} \cdot SPN + k_{\mathrm{g}} \cdot SPN$$

$$\frac{\mathrm{d}Tem1_{\mathrm{T}}}{\mathrm{d}t} = k_{\mathrm{s,tem1t}} \cdot c_{\mathrm{tem1}} \cdot mass \cdot mTem1 - k_{\mathrm{d,tem1t}} \cdot Tem1_{\mathrm{T}} + k_{\mathrm{g}} \cdot Tem1_{\mathrm{T}}$$

$$\frac{\mathrm{d}Cdc15_{\mathrm{T}}}{\mathrm{d}t} = k_{\mathrm{s,cdc15t}} \cdot c_{\mathrm{cdc15}} \cdot mass \cdot mCdc15 - k_{\mathrm{d,cdc15t}} \cdot Cdc15_{\mathrm{T}} + k_{\mathrm{g}} \cdot Cdc15_{\mathrm{T}}$$

$$\frac{\mathrm{d}Esp1_{\mathrm{T}}}{\mathrm{d}t} = k_{\mathrm{s,esp1t}} \cdot c_{\mathrm{esp1}} \cdot mass \cdot mEsp1 - k_{\mathrm{d,esp1t}} \cdot Esp1_{\mathrm{T}} + k_{\mathrm{g}} \cdot Esp1_{\mathrm{T}}$$

$$\frac{\mathrm{d}SBF_{\mathrm{T}}}{\mathrm{d}t} = k_{\mathrm{s,sbft}} \cdot c_{\mathrm{sbf}} \cdot mass \cdot mSBF - k_{\mathrm{d,sbft}} \cdot SBF_{\mathrm{T}} + k_{\mathrm{g}} \cdot SBF_{\mathrm{T}}$$

$$\frac{\mathrm{d}MBF_{\mathrm{T}}}{\mathrm{d}t} = k_{\mathrm{s,mbft}} \cdot c_{\mathrm{mbf}} \cdot mass \cdot mMBF - k_{\mathrm{d,mbft}} \cdot MBF_{\mathrm{T}} + k_{\mathrm{g}} \cdot MBF_{\mathrm{T}}$$

$$\frac{\mathrm{d}Mcm1_{\mathrm{T}}}{\mathrm{d}t} = k_{\mathrm{s,mcm1t}} \cdot c_{\mathrm{mcm1}} \cdot mass \cdot mMcm1 - k_{\mathrm{d,mcm1t}} \cdot Mcm1_{\mathrm{T}} + k_{\mathrm{g}} \cdot Mcm1_{\mathrm{T}}$$

$$\frac{\mathrm{d}APC}{\mathrm{d}t} = k_{\mathrm{s,apct}} \cdot c_{\mathrm{apc}} \cdot mass \cdot mAPC - k_{\mathrm{d,apct}} \cdot APC + k_{\mathrm{g}} \cdot APC$$

$$\frac{\mathrm{d}Cln3}{\mathrm{d}t} = k_{\mathrm{s,n3}} \cdot c_{\mathrm{cln3}} \cdot mass^2 - k_{\mathrm{d,n3}} \cdot Cln3 + k_{\mathrm{g}} \cdot Cln3$$

$$\frac{\mathrm{d}Bck2}{\mathrm{d}t} = k_{\mathrm{s,k2}} \cdot c_{\mathrm{bck2}} \cdot mass^2 - k_{\mathrm{d,k2}} \cdot Bck2 + k_{\mathrm{g}} \cdot Bck2$$

$$G(V_{\mathrm{a}}, V_{\mathrm{i}}, J_{\mathrm{a}}, J_{\mathrm{i}}) = \frac{2J_{\mathrm{i}}V_{\mathrm{a}}}{V_{\mathrm{i}}-V_{\mathrm{a}}+J_{\mathrm{a}}V_{\mathrm{i}}+J_{\mathrm{i}}V_{\mathrm{a}}+\sqrt{(V_{\mathrm{i}}-V_{\mathrm{a}}+J_{\mathrm{a}}V_{\mathrm{i}}+J_{\mathrm{i}}V_{\mathrm{a}})^2-4(V_{\mathrm{i}}-V_{\mathrm{a}})J_{\mathrm{i}}V_{\mathrm{a}}}}$$

$$SBF = G(V_{\mathrm{a,sbf}}, V_{\mathrm{i,sbf}}, J_{\mathrm{a,sbf}}, J_{\mathrm{i,sbf}}) \cdot SBF_{\mathrm{T}}$$

$$MBF = G(V_{\mathrm{a,sbf}}, V_{\mathrm{i,sbf}}, J_{\mathrm{a,sbf}}, J_{\mathrm{i,sbf}}) \cdot MBF_{\mathrm{T}}$$

$$Mcm1 = G(\frac{k_{\mathrm{a,mcm}}}{c_{\mathrm{clb2}} \cdot mass} \cdot \mathrm{Clb2}, k_{\mathrm{i,mcm}}, J_{\mathrm{a,mcm}}, J_{\mathrm{i,mcm}}) \cdot Mcm1_{\mathrm{T}}$$

$$Clb5_{\mathrm{T}} = Clb5 + C5 + C5\mathrm{P} + \mathrm{F5} + F5\mathrm{P}$$

$$Clb2_{\mathrm{T}} = Clb2 + C2 + C2\mathrm{P} + F2 + F2\mathrm{P}$$

$$Sic1_{\mathrm{T}} = Sic1 + Sic1\mathrm{P} + C2 + C2\mathrm{P} + C5 + C5\mathrm{P}$$

$$Cdc6_{\mathrm{T}} = Cdc6 + Cdc6\mathrm{P} + F2 + F2\mathrm{P} + F5 + F5\mathrm{P}$$

$$CKI_{\mathrm{T}} = Sic1_{\mathrm{T}} + Cdc6_{\mathrm{T}}$$

$$RENT\mathrm{P} = Cdc14_{\mathrm{T}} - RENT - Cdc14$$

$$Net1\mathrm{P} = Net1_{\mathrm{T}} - Net1 - Cdc14_{\mathrm{T}} + Cdc14$$

$$\mathrm{PE} = Esp1_{\mathrm{T}} - Esp1$$

$$V_{\mathrm{d,b5}} = k'_{\mathrm{d,b5}} + k''_{\mathrm{d,b5}} \cdot Cdc20\mathrm{A}$$

$$V_{\mathrm{d,b2}} = k'_{\mathrm{d,b2}} + k''_{\mathrm{d,b2}} \cdot Cdh1 + k_{\mathrm{d,b2p}} \cdot Cdc20\mathrm{A}$$

$$V_{\mathrm{a,sbf}} = k_{\mathrm{a,sbf}} \cdot (e_{\mathrm{sbf,n2}} \cdot Cln2 + e_{\mathrm{sbf,n3}} \cdot (Cln3 + Bck2) + e_{\mathrm{sbf,b5}} \cdot Clb5)$$

$$V_{\mathrm{i,sbf}} = k'_{\mathrm{i,sbf}} + k''_{\mathrm{i,sbf}} Clb2$$

$$V_{\mathrm{kp,c1}} = k_{\mathrm{d1,c1}} + \frac{k_{\mathrm{d2,c1}}}{J_{\mathrm{d2,c1}}+\mathrm{Sic1T}} \cdot (e_{\mathrm{c1,n3}} \cdot Cln3 + e_{\mathrm{c1,k2}} \cdot Bck2 + e_{\mathrm{c1,n2}} \cdot Cln2 + e_{\mathrm{c1,b5}} \cdot Clb5 + e_{\mathrm{c1,b2}} \cdot Clb2)$$

$$V_{\mathrm{kp,f6}} = k_{\mathrm{d1,f6}} + \frac{k_{\mathrm{d2,f6}}}{J_{\mathrm{d2,f6}}+Cdc6} \cdot (e_{\mathrm{f6,n3}} \cdot Cln3 + e_{\mathrm{f6,k2}} \cdot Bck2 + e_{\mathrm{f6,n2}} \cdot Cln2 + e_{\mathrm{f6,b5}} \cdot Clb5 + e_{\mathrm{f6,b2}} \cdot Clb2)$$

$$V_{\mathrm{a,cdh}} = k'_{\mathrm{a,cdh}} + k''_{\mathrm{a,cdh}} \cdot Cdc14$$

$$V_{\mathrm{i,cdh}} = k'_{\mathrm{i,cdh}} + k''_{\mathrm{i,cdh}} \cdot (e_{\mathrm{cdh,n3}} \cdot Cln3 + e_{\mathrm{cdh,n2}} \cdot Cln2 + e_{\mathrm{cdh,b2}} \cdot Clb2 + e_{\mathrm{cdh,b5}} \cdot Clb5)$$

$V_{\text{pp,net}} = k'_{\text{pp,net}} + k''_{\text{pp,net}} \cdot Cdc55$
$V_{\text{kp,net}} = (k'_{\text{kp,net}} + k''_{\text{kp,net}} \cdot Cdc55) \cdot mass$
$V_{\text{d,55}} = k'_{\text{d,55}} + k''_{\text{d,55}}(J_{20,55} + Cdc20\text{A}) \cdot \frac{J_{\text{pds}}}{J_{\text{pds}} + Pds1}$
$V_{\text{d,pds}} = k'_{\text{d1,pds}} + k''_{\text{d2,pds}} \cdot Cdc20\text{A} + k''_{\text{d3,pds}} \cdot Cdh1$

A.3 Reactions and propensities

Table 3.5: Reaction channels and corresponding propensity functions for the slow subset in the hybrid model.

Reaction	Propensity Function	Reaction	Propensity Function
$\phi \rightarrow mCdh1$	$k_{\text{s,mcdh1}}$	$mCdh1 \rightarrow \phi$	$k_{\text{d,mcdh1}} \cdot mCdh1$
$\phi \rightarrow mTem1$	$k_{\text{s,mtem1}}$	$mTem1 \rightarrow \phi$	$k_{\text{d,mtem1}} \cdot mTem1$
$\phi \rightarrow mCdc15$	$k_{\text{s,mcdc15}}$	$mCdc15 \rightarrow \phi$	$k_{\text{d,mcdc15}} \cdot mCdc15$
$\phi \rightarrow mCdc14$	$k_{\text{s,mcdc14}}$	$mCdc14 \rightarrow \phi$	$k_{\text{d,mcdc14}} \cdot mCdc14$
$\phi \rightarrow mNet1$	$k_{\text{s,mnet1}}$	$mNet1 \rightarrow \phi$	$k_{\text{d,mnet1}} \cdot mNet1$
$\phi \rightarrow mCdc55$	$k_{\text{s,mcdc55}}$	$mCdc55 \rightarrow \phi$	$k_{\text{d,mcdc55}} \cdot mCdc55$
$\phi \rightarrow mEsp1$	$k_{\text{s,mesp1}}$	$mEsp1 \rightarrow \phi$	$k_{\text{d,mesp1}} \cdot mEsp1$
$\phi \rightarrow mSBF$	$k_{\text{s,msbf}}$	$mSBF \rightarrow \phi$	$k_{\text{d,msbf}} \cdot mSBF$
$\phi \rightarrow mMBF$	$k_{\text{s,mmbf}}$	$mMBF \rightarrow \phi$	$k_{\text{d,mmbf}} \cdot mMBF$
$\phi \rightarrow mMcm1$	$k_{\text{s,mmcm1}}$	$mMcm1 \rightarrow \phi$	$k_{\text{d,mmcm1}} \cdot mMcm1$
$\phi \rightarrow mAPC$	$k_{\text{s,mapc}}$	$mAPC \rightarrow \phi$	$k_{\text{d,mapc}} \cdot mAPC$
$\phi \rightarrow mCln2$	$k_{\text{s,mcln2}} \cdot (k'_{\text{s,n2}} \cdot c_{\text{cln2}} \cdot mass + k''_{\text{s,n2}} \cdot \frac{c_{\text{cln2}}}{c_{\text{sbf}}} \cdot SBF)$	$mCln2 \rightarrow \phi$	$k_{\text{d,mcln2}} \cdot mCln2$

$\phi \rightarrow mClb5$	$k_{\text{s,mclb5}} \cdot (k'_{\text{s,b5}} \cdot c_{\text{clb5}} \cdot mass + k''_{\text{s,b5}} \cdot \frac{c_{\text{clb5}}}{c_{\text{mbf}}} \cdot MBF)$	$mClb5 \rightarrow \phi$	$k_{\text{d,mclb5}} \cdot mClb5$
$\phi \rightarrow mClb2$	$k_{\text{s,mclb2}} \cdot (k'_{\text{s,b2}} \cdot c_{\text{clb2}} \cdot mass + k''_{\text{s,b2}} \cdot \frac{c_{\text{clb2}}}{c_{\text{mcm1}}} \cdot Mcm1)$	$mClb2 \rightarrow \phi$	$k_{\text{d,mclb2}} \cdot mClb2$
$\phi \rightarrow mSic1$	$k_{\text{s,msic1}} \cdot (k'_{\text{s,c1}} \cdot c_{\text{sic1}} \cdot mass + k''_{\text{s,c1}} \frac{c_{\text{sic1}}}{c_{\text{swi5}}} \cdot Swi5)$	$mSic1 \rightarrow \phi$	$k_{\text{d,msic1}} \cdot mSic1$
$\phi \rightarrow mCdc6$	$k_{\text{s,mcdc6}} \cdot (k'_{\text{s,f6}} \cdot c_{\text{cdc6}} \cdot mass + k''_{\text{s,f6}} \cdot \frac{c_{\text{cdc6}}}{c_{\text{swi5}}} \cdot Swi5 + k'''_{\text{s,f6}} \cdot \frac{c_{\text{cdc6}}}{c_{\text{sbf}}} \cdot SBF)$	$mCdc6 \rightarrow \phi$	$k_{\text{d,mcdc6}} \cdot mCdc6$
$\phi \rightarrow mSwi5$	$k_{\text{s,mswi5}} \cdot (k'_{\text{s,swi}} \cdot c_{\text{swi5}} \cdot mass + k''_{\text{s,swi}} \cdot \frac{c_{\text{swi5}}}{c_{\text{mcm1}}} \cdot Mcm1)$	$mSwi5 \rightarrow \phi$	$k_{\text{d,mswi5}} \cdot mSwi5$
$\phi \rightarrow mCdc20$	$k_{\text{s,mcdc20}} \cdot (k'_{\text{s,20}} \cdot c_{\text{cdc20}} \cdot mass + k''_{\text{s,20}} \cdot \frac{c_{\text{cdc20}}}{c_{\text{mcm1}}} \cdot Mcm1)$	$mCdc20 \rightarrow \phi$	$k_{\text{d,mcdc20}} \cdot mCdc20$
$\phi \rightarrow mPds1$	$k_{\text{s,mpds1}} \cdot (k'_{\text{s,pds}} \cdot c_{\text{pds1}} \cdot mass + k''_{\text{s1,pds}} \cdot \frac{c_{\text{pds1}}}{c_{\text{sbf}}} \cdot SBF + k''_{\text{s2,pds}} \cdot \frac{c_{\text{pds1}}}{c_{\text{mcm1}}} \cdot Mcm1)$	$mPds1 \rightarrow \phi$	$k_{\text{d,mpsd1}} \cdot mPds1$

A.4 Parameters

Table 3.6: Basal parameter values for the wild-type cell cycle. Parameters that start with lower case *k* are rate constants (min^{-1}). All other parameter are dimensionless.

$k_{\text{g}} = 0.0072$	$k'_{\text{s,n2}} = 0$	$k''_{\text{s,n2}} = 0.03$	$k_{\text{d,n2}} = 0.12$
$k'_{\text{s,b5}} = 0.0001$	$k''_{\text{s,b5}} = 0.0007$	$k'_{\text{d3,c1}} = 1$	$k_{\text{di,b5}} = 0.06$
$k_{\text{as,f5}} = 0.01$	$k_{\text{d3,f6}} = 1$	$k_{\text{di,f5}} = 0.01$	$k_{\text{as,b5}} = 50$

$k'_{s,b2} = 0.0003$	$k''_{s,b2} = 0.0114$	$k_{di,b2} = 0.05$	$k_{di,f2} = 0.5$
$k_{as,b2} = 50$	$k_{as,f2} = 15$	$k'_{s,c1} = 0.0036$	$k''_{s,c1} = 0.0359$
$k_{pp,c1} = 4$	$k'_{s,f6} = 0.0059$	$k''_{s,f6} = 0.0295$	$k'''_{s,f6} = 0.0001$
$k_{pp,f6} = 4$	$k'_{s,swi} = 0.0007$	$k''_{s,swi} = 0.0114$	$k_{d,swi} = 0.08$
$k_{a,swi} = 2$	$k_{i,swi} = 0.05$	$k_{a,apc} = 0.1$	$k_{i,apc} = 0.15$
$k'_{s,20} = 0.0014$	$k''_{s,20} = 0.1364$	$k'_{a,20} = 0.05$	$k''_{a,20} = 0.2$
$k_{d,20} = 0.3$	$k_{s,cdh} = 0.0014$	$k_{d,cdh} = 0.01$	$k'_{a,15} = 0.002$
$k''_{a,15} = 1$	$k'''_{a,15} = 0.001$	$k_{i,15} = 0.5$	$k_{s,14} = 0.02$
$k_{d,14} = 0.1$	$k_{d,net} = 0.03$	$k_{di,rent} = 1$	$k_{di,rentp} = 2$
$k_{as,rent} = 200$	$k_{as,rentp} = 1$	$k_{s,net} = 0.0134$	$k_{s,55} = 0.0143$
$k'_{s,pds} = 0$	$k''_{s1,pds} = 0.0043$	$k''_{s2,pds} = 0.0079$	$k_{di,esp} = 0.5$
$k_{as,esp} = 50$	$k_{s,ori} = 2$	$k_{d,ori} = 0.06$	$k_{s,bud} = 0.2$
$k_{d,bud} = 0.06$	$k_{s,spn} = 0.1$	$k_{d,spn} = 0.06$	$k_{s,tem1t} = 0.0150$
$k_{d,tem1t} = 0.046$	$k_{s,cdc15t} = 0.0142$	$k_{d,cdc15t} = 0.046$	$k_{s,esp1t} = 0.0139$
$k_{d,esp1t} = 0.046$	$k_{s,sbft} = 0.007$	$k_{d,sbft} = 0.0462$	$k_{a,mcm} = 5$
$k_{s,mbft} = 0.0066$	$k_{d,mbft} = 0.046$	$k_{s,apct} = 0.0065$	$k_{d,apct} = 0.0462$
$k_{s,mcm1t} = 0.0078$	$k_{d,mcm1t} = 0.046$	$k_{i,mcm} = 0.15$	$k''_{d3,pds} = 0.04$
$k_{a,sbf} = 0.38$	$k'_{i,sbf} = 0.6$	$k''_{i,sbf} = 8$	$k'_{d,b5} = 0.01$
$k'_{d,b2} = 0.003$	$k''_{d,b2} = 0.4$	$k_{d,b2p} = 0.15$	$k_{d1,c1} = 0.01$
$k_{d2,c1} = 1$	$k''_{d,b5} = 0.16$	$k_{d1,f6} = 0.01$	$k_{d2,f6} = 1$
$k'_{a,cdh} = 0.01$	$k''_{a,cdh} = 0.8$	$k'_{i,cdh} = 0.001$	$k''_{i,cdh} = 0.08$
$k'_{pp,net} = 0.05$	$k''_{pp,net} = 3$	$k'_{kp,net} = 0.01$	$k''_{kp,net} = 0.6$
$k'_{d,55} = 0.17$	$k''_{d,55} = 2$	$k'_{d1,pds} = 0.01$	$k''_{d2,pds} = 0.2$
$k_{s,n3} = 0.0080$	$k_{d,n3} = 0.12$	$k_{s,k2} = 0.0048$	$k_{d,k2} = 0.12$

$k_{\text{mad2}} = 8$ for $[\text{ORI}]_{\text{n}} > 1$and $[\text{SPN}]_{\text{n}} < 1;$ otherwise 0.01			
$k_{\text{bud2}} = 1$ for $[\text{ORI}]_{\text{n}} > 1$ and $[\text{SPN}]_{\text{n}} < 1;$ otherwise 0.2			
$k_{\text{lte1}} = 1$ for $[\text{SPN}]_{\text{n}} > 1$ and $[\text{Clb2}]_{\text{n}} > 1;$ otherwise 0.1			
$K_{\text{ez}} = 0.3$	$K_{\text{ez2}} = 0.2$	$J_{\text{i,cdh}} = 0.03$	$J_{\text{a,apc}} = 0.1$
$J_{\text{i,apc}} = 0.1$	$J_{\text{a,cdh}} = 0.03$	$J_{\text{pds}} = 0.04$	$J_{\text{a,sbf}} = 0.01$
$J_{\text{n3}} = 6$	$J_{\text{a,tem}} = 0.1$	$J_{\text{i,tem}} = 0.1$	$J_{\text{spn}} = 0.14$
$J_{\text{d2,f6}} = 0.05$	$J_{\text{a,mcm}} = 0.1$	$J_{\text{i,mcm}} = 0.1$	$J_{\text{d2,c1}} = 0.05$
$J_{\text{i,sbf}} = 0.01$	$J_{20,55} = 0.15$	$e_{\text{cdh,b2}} = 1.2$	$e_{\text{cdh,b5}} = 8$
$e_{\text{bud,n3}} = 0.05$	$e_{\text{ori,b5}} = 0.9$	$e_{\text{ori,b2}} = 0.45$	$e_{\text{bud,n2}} = 0.12$
$e_{\text{sbf,b5}} = 2$	$e_{\text{bud,b5}} = 0.25$	$e_{\text{sbf,n2}} = 2$	$e_{\text{sbf,n3}} = 10$
$e_{\text{c1,b5}} = 0.1$	$e_{\text{c1,n3}} = 0.3$	$e_{\text{c1,k2}} = 0.03$	$e_{\text{c1,n2}} = 0.06$
$e_{\text{f6,n2}} = 0.06$	$e_{\text{c1,b2}} = 0.45$	$e_{\text{f6,n3}} = 0.3$	$e_{\text{f6,k2}} = 0.03$
$e_{\text{cdh,n2}} = 0.4$	$e_{\text{f6,b5}} = 0.1$	$e_{\text{f6,b2}} = 0.55$	$e_{\text{cdh,n3}} = 0.25$
$c_{\text{clb2}} = 479.2$	$c_{\text{cln2}} = 239.6$	$c_{\text{cln3}} = 599$	$c_{\text{bck2}} = 599$
$c_{\text{cdh1}} = 1198$	$c_{\text{ori}} = 119.8$	$c_{\text{bud}} = 119.8$	$c_{\text{spn}} = 119.8$
$c_{\text{tem1}} = 178.2$	$c_{\text{clb5}} = 479.2$	$c_{\text{cdc20}} = 1797$	$c_{\text{sic1}} = 479.2$
$c_{\text{cdc6}} = 479.2$	$c_{\text{cdc14}} = 275.54$	$c_{\text{net1}} = 275.54$	$c_{\text{cdc15}} = 479.2$
$c_{\text{pds1}} = 39.5$	$c_{\text{esp1}} = 39.5$	$c_{\text{sbf}} = 479.2$	$c_{\text{mbf}} = 479.2$
$c_{\text{apc}} = 263.56$	$c_{\text{cdc55}} = 4193$	$c_{\text{rent}} = 4193$	$c_{\text{mcm1}} = 4193$
$k_{\text{s,mCdc15}} = 0.45$	$k_{\text{d,mCdc15}} = 0.14$	$k_{\text{s,mTem1}} = 0.43$	$k_{\text{d,mTem1}} = 0.14$
$k_{\text{s,mNet1}} = 0.5$	$k_{\text{d,mNet1}} = 0.1$	$k_{\text{s,mCdc14}} = 1.5$	$k_{\text{d,mCdc14}} = 0.14$
$k_{\text{s,mEsp1}} = 0.3$	$k_{\text{d,mEsp1}} = 0.08$	$k_{\text{s,mCdc55}} = 0.97$	$k_{\text{d,mCdc55}} = 0.14$
$k_{\text{s,mMBF}} = 0.97$	$k_{\text{d,mMBF}} = 0.14$	$k_{\text{s,mSBF}} = 0.94$	$k_{\text{d,mSBF}} = 0.14$
$k_{\text{s,mPds1}} = 0.24$	$k_{\text{d,mPds1}} = 0.14$	$k_{\text{s,mMcm1}} = 0.8$	$k_{\text{d,mMcm1}} = 0.14$

$k_{\text{s,mAPC}} = 0.97$	$k_{\text{d,mAPC}} = 0.14$	$k_{\text{s,mCln2}} = 0.0117$	$k_{\text{d,mCln2}} = 0.23$
$k_{\text{s,mClb5}} = 0.3916$	$k_{\text{d,mClb5}} = 0.14$	$k_{\text{s,mClb2}} = 0.0034$	$k_{\text{d,mClb2}} = 0.35$
$k_{\text{s,mSic1}} = 0.0067$	$k_{\text{d,mSic1}} = 0.14$	$k_{\text{s,mCdc6}} = 0.0083$	$k_{\text{d,mCdc6}} = 0.14$
$k_{\text{s,mSwi5}} = 0.0016$	$k_{\text{d,mSwi5}} = 0.14$	$k_{\text{s,mCdc20}} = 0.0001$	$k_{\text{d,mCdc20}} = 0.14$
$k_{\text{s,mCdh1}} = 0.97$	$k_{\text{d,mCdh1}} = 0.14$		

Chapter 4

Stochastic Phenotypes of Budding Yeast Mutant Strains

4.1 Introduction

Cell cycle is a sequence of events through which a living cell replicates its DNA and other necessary machinery and divides into two newborn cells. Such development for wild-type cells is significantly robust to intrinsic and extrinsic alternations in growth conditions. For instance, the fate of a wild-type cell is very robust against changes in the growth rate of the media (an example of extrinsic alternation) it grows in. Similarly, stochastic fluctuations of protein levels within the cell (an example of intrinsic alternations) has very limited effects on the progression through the cell cycle. However, mutant cells in which critical parts of cell cycle control mechanisms are altered may be prone to much more sensitivity in terms of environmental changes [32]. Cancer cells in which genetic mutations are accumulated are good examples of such non-robust dynamics [75]. For instance, the size and shape of the cells and the proliferation rate in cancer cells are severely abnormal [75]. Therefore, it is imperative to study the non-robust dynamics and stochastic

characteristics of mutant cells.

The phenotypes of more than 100 mutant strains of the budding yeast cell cycle have been studied by Chen *et al.* [60,9] using a deterministic model. More recently, Kerakofski *et al.* used a more comprehensive deterministic model to predict the viability of 256 mutant strains of the budding yeast [10]. In both deterministic models, the fate of the cells is decided deterministically as either *viable* (able to divide and commence a new cycle) or *inviable* (unable to exit mitosis or failure in a successful division). However, unlike many mutants that are completely viable or inviable, there are specific mutant strains of Saccharomyces cerevisiae which are sensitive to the growth rate of their nutrition environment. For instance $CLB2-db\Delta clb5\Delta$ mutant is inviable in glucose but partially viable on slower growth media such as raffinose. This stochastic phenotype cannot be explained by a deterministic model. Building stochastic models for model organisms such as budding yeast that can accurately account for such stochastic phenotypes can shed light on understanding such non-robust dynamics in higher eukaryotes and ultimately may lead to understanding the disease and therapeutic interventions [32].

In this chapter, we extend our hybrid stochastic model of the budding yeast cell cycle to study the phenotypes of more than 100 mutant strains. Moreover, we are interested to explore a quantitative method to evaluate the viability of a mutant strain given that stochasticity is included in the model.

4.2 Quantitative viability criteria

In the original deterministic model by Chen *et al.* [60], a cell is considered viable if the following conditions are met:

1. certain events, listed in [60], take place in a proper sequence,

2. in particular, cell division occurs after budding, and

3. cell mass does not exceed a predetermined threshold ($mass$ at division < 10).

The same set of rules should also apply to the stochastic model. A division is considered successful if the aforementioned viability criteria are met; otherwise, the cell is considered to be inviable. From our numerical simulations, the probability of successful division p is estimated by

$$p = \frac{\text{number of successful divisions}}{\text{number of successful divisions} + \text{number of failures}}.$$

Using this metric, we consider a mutant strain viable if $p > 0.75$, inviable if $p < 0.65$, and partially viable otherwise. This viability criterion is based on the following considerations.

Let N_0 be the initial population of cells in an experiment. After one cycle, the average number of cells that divide is pN_0, while $(1-p)N_0$ cells exit the cycle and stop dividing. Thus, after one cycle, the total population of cells is $(1+p)N_0$, of which $2pN_0$ cells completed the previous cell cycle and $(1-p)N_0$ cells have ceased to divide (we call them dead cells). In our simulations, we disregard the $(1-p)N_0$ dead cells; hence, the number of actively dividing cells in the second cycle is $N_1 = 2pN_0$, and the expected number of actively dividing cells after k cycles will be $N_k = (2p)^k N_0$.

In cell-viability experiments, colony formation is typically assessed after 24 h growth of a series of ten-fold diluted inocula. For wild-type yeast cells ($p \approx 1$) growing on rich glucose medium (cycle time $\approx$ 12 h), each inoculum should increase by a factor of about $2^{12} = 4096$. The colony sizes after 24 h growth of ten-fold serial dilutions would be (4000 N_0, 400 N_0, 40 N_0, 4 N_0), of which the first would be too dense to quantify, the last would be too sparse to see, and the middle two would be used to assess the viability of mutant strains. For a mutant cell with $p < 0.5$, no visible colony will grow from the initial inoculum, and the mutant will be scored 'inviable'. For mutant strains with $0.5 < p < 1$, we must consider how the colony growth assay compares to wild-type cells. For $p = 0.8$ the initial inoculum grows to 280 N_0, which is comparable to the first

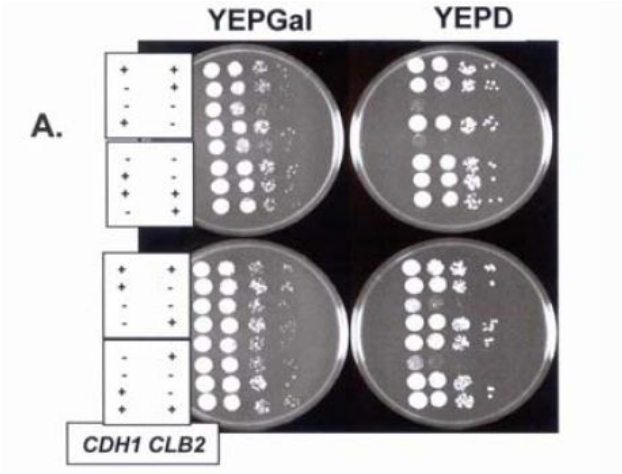

Figure 4.1: **The experimental observation of $cdh1\Delta$ $clb2\Delta$, an inviable strain.** This figure is copied from Supplementary Text of the paper by Cross *et al.* [76]. Left column, “YEPGal" is growth on galactose, and right colmn “YEPD" is growh on glucose.

dilution of the wild-type cells, and we would score this mutant strain as ‘viable’. For $p = 0.75$ the initial inoculum grows to 130 N_0, which is denser than the second dilution of the wild-type cells, and we would score this mutant strain as ‘probably viable’. For $p = 0.65$ the initial inoculum grows to 23 N_0, which is less dense than the second dilution of the wild-type cells, and we would score this mutant strain as ‘hardly viable’. These calculations suggest that a mutant strain be considered viable if $p \geqslant 0.75$ and nonviable if $p \leqslant 0.65$. For $0.65 < p < 0.75$, the strain is identified as partially viable.

Given this viability criteria, we can use our hybrid stochastic model to explain many of the stochastic observations that could not be explained by any deterministic approach. In a deterministic model, a viable strain always divides and commences a new cycle while an inviable strain never divides. However, in experimental observations, some “viable" cells might exit the cell cycle and an inviable cell might divide for a few cycles. Figure 4.1 shows an image of plates from an experimental study of $cdh1$ $clb2$ double mutant cells. The + and - on the far left indicates if the gene is deleted or not. For instance, + + means both $CDH1$ and $CLB2$ are intact and + - means $CDH1$ is intact and $CLB2$ is deleted. For the $cdh1\Delta$ $clb2\Delta$ strain (see - - experiments) after four serial

dilutions, still a few cells are visible on the YEPGal plate but not on the YEPD plate (in Fig.4.1, see the fourth row of the corresponding experiment in each plate). This means in spite of the fact that this double mutant is concluded to be inviable in both glucose and galactose media [9], there are quite a few cells that are dividing and commencing new cycles on galactose growth medium. A deterministic model cannot describe such detailed observations, while in section 4.3, we will show that our hybrid stochastic model accounts for such stochastic phenotypes in exquisite detail.

4.3 Results

Based on these criteria developed in section 4.2, we assess the viability of 122 mutant strains of budding yeast that were studied in the modeling paper of Chen *et al.* [9].

To demonstrate the significant roles of noise in some of these mutants, we discuss two multiple-mutant strains, $cln1\Delta\ cln2\Delta\ bck2\Delta$ and $cln3\Delta\ bck2\Delta$ multi-copy $CLN2$ (Fig. 4.2), in some detail. According to experimental observations, the $cln1\Delta\ cln2\Delta\ bck2\Delta$ strain [77] is viable. However, due to deletion of START cyclins Cln1 and Cln2, the cell requires a longer time than normal to form a bud and hence grows to a larger size at division, in comparison with wild-type cells. Figure 4.2a shows that in the deterministic model the cell consistently exits mitosis and divides successfully with size larger than normal, as observed experimentally [77]. In the hybrid stochastic simulation, however, due to the stochastic nature of the process, there is a finite probability that a cell may exit the cycle and become arrested in some phase of the cell cycle. In Fig. 4.2b, for instance, the cell grows too large in the G1 phase and never divides again, while in Fig. 4.2c, it exits mitosis and divides successfully. The probability of successful completion of the cell division cycle, in this case, is computed to be $p \approx 0.84$. As shown in Fig. 4.2d, the total number of cells in our computational culture increases exponentially, with a number-doubling time (NDT) of 140 min, which is slower than the NDT of a fully viable wild-type culture (approx 100 min). Therefore,

we conclude that the hybrid stochastic simulation correctly confirms the viability, but the reduced growth rate, of the $cln1\Delta\ cln2\Delta\ bck2\Delta$ strain.

Next we consider the inviable mutant strain $cln3\Delta\ bck2\Delta$ multi-copy $CLN2$ [78]. Figure 4.2e shows that in the deterministic model the mutant cell is arrested in the G1 phase and grows without dividing until it dies. In the hybrid stochastic simulation, although many of the cells become arrested in G1 (see Fig. 4.2f), some cells manage to exit the G1 phase, complete the cell cycle, commence a new cycle and divide a few times (see Fig. 4.2g). Nonetheless, according to Fig. 4.2h the total number of cells in our computational culture declines with time because the probability of cell division is only $p \approx 0.40$. Therefore, we conclude that the hybrid stochastic simulation correctly confirms the inviability of $cln3\Delta\ bck2\Delta$ multi-copy $CLN2$ strain.

Based on our hybrid stochastic simulations of all 122 mutant strains in Chen's database, we find that the model successfully reproduces the phenotypes of 103 of these strains. Our results for all mutant strains are reported in Table4.1 in Appendix B.

$CLB2db\Delta\ clb5\Delta$ is a mutant with an interesting stochastic phenotype: it is inviable when grown on glucose medium but 'partially viable' when grown on raffinose (a sugar that supports a slower growth rate than glucose) [76]. Due to the deletion of the destruction box of $CLB2$, Clb2 protein is in excess at telophase and the cell is unable to exit mitosis and divide, even in the absence of Clb5 protein (due to deletion of the $CLB5$ gene). When growing on raffinose, however, many of these mutant cells (approximately 60%-75%) can exit mitosis and commence a new cycle, whereas the remaining cells (25%-40%) are arrested in telophase and never re-enter the cell cycle [32]. The NDT of the double-mutant cells (250 - 300 min) is observed to be much longer than the NDT of wild-type cells (160 min) growing in raffinose medium [32].

Simulation results of Chen's deterministic model predict that $CLB2db\Delta\ clb5\Delta$ cells are inviable on glucose and viable on galactose and raffinose media. Clearly, we cannot expect a deterministic

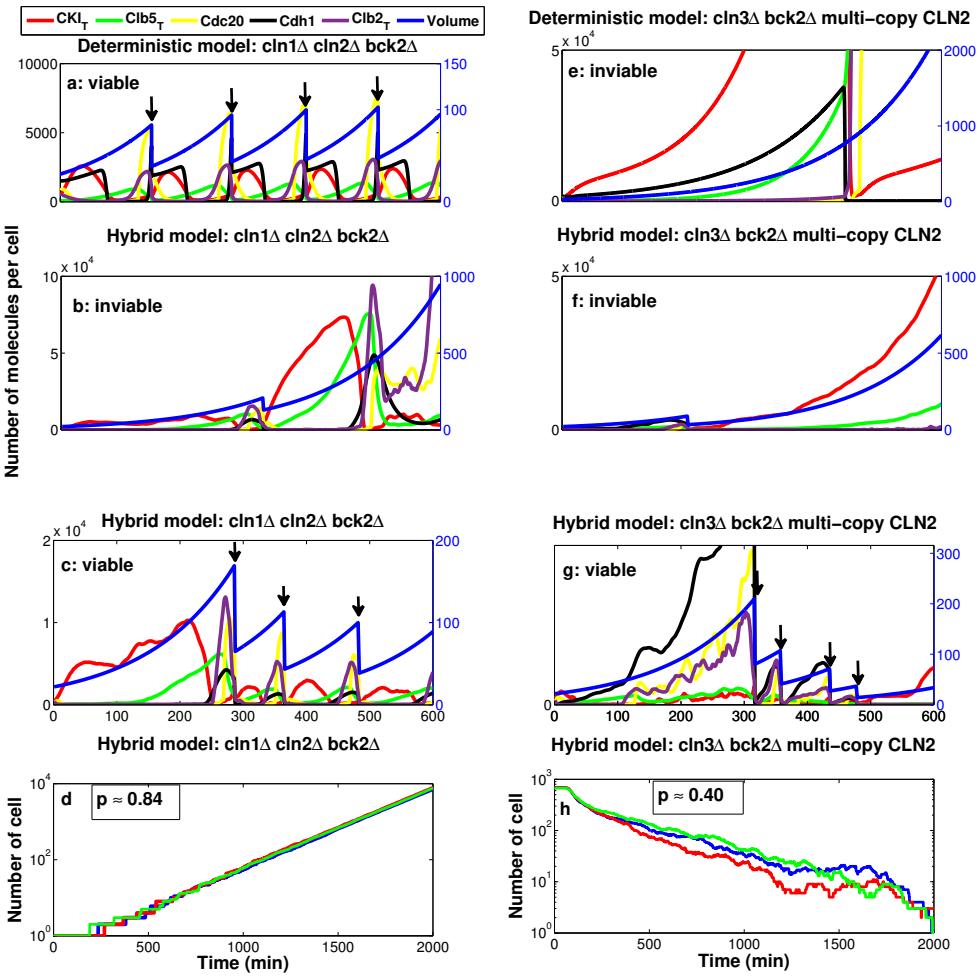

Figure 4.2: **Comparison of deterministic and stochastic trajectories of two different multiple-mutant strains**. **a:** Deterministic trajectories of $cln1\Delta\ cln2\Delta\ bck2\Delta$; the cell consistently exits mitosis and divides (the divisions are indicated by arrows). **b-c:** Stochastic trajectories of $cln1\Delta\ cln2\Delta\ bck2\Delta$ from two independent runs. In panel b the cell becomes arrested in the G1 phase while in panel c the cell divides successfully. **d:** The total number of cells as a function of time; we start each simulation with one cell and count the total number of cells over time for 2000 min. The probability of division is calculated as $p \approx 0.84$ which indicates that the $cln1\Delta\ cln2\Delta\ bck2\Delta$ strain is viable according to our definition. The semilog plot in panel d shows that the number of cells increases exponentially (NDT $\approx$ 140 min) in our computational culture. **e:** Deterministic trajectories of $cln3\Delta\ bck2\Delta$ multi-copy $CLN2$; the cell arrests permanently in the G1 phase. **f-g:** Stochastic trajectories of $cln3\Delta\ bck2\Delta$ multi-copy $CLN2$ from two independent runs. In panel f the cell becomes arrested in the G1 phase after one cycle, while in panel g the cell exits mitosis and divides successfully several times. **h:** The total number of cells as a function of time; we start the simulation with 1000 cells and count the total number of viable cells over time for 2000 min. The probability of division is calculated as $p \approx 0.40$ which indicates that the $cln3\Delta\ bck2\Delta$ multi-copy $CLN2$ strain is inviable. The semilog plot in panel h shows that the total number of cells decreases exponentially in our computational culture.

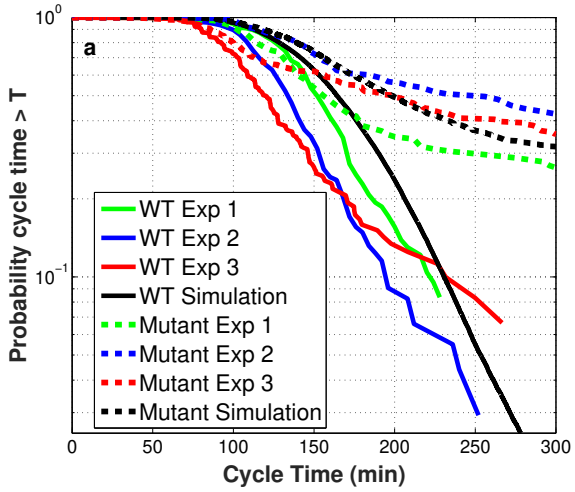

Figure 4.3: **Stochastic phenotype of $CLB2db\Delta$ $clb5\Delta$ mutant strain.** Comparison of wild-type and $CLB2db\Delta$ $clb5\Delta$ mutant cells growing in raffinose. The probability that a cell divides with a cycle time longer than a specific time T is plotted for wild-type cells (solid lines) and mutant cells (dotted lines). The black lines are generated by our hybrid stochastic model and the red-blue-green lines are the results of three independent experimental runs by Ball *et al.* [32]. To model growth on raffinose medium in our simulation, the specific growth rate of cells is set to 0.00433 min^{-1} (MDT = 160 min).

model to capture the stochastic properties of such a 'partially viable' mutant strain. Our hybrid stochastic model, however, describes the phenotype in exquisite detail. The probability of division for $CLB2db\Delta$ $clb5\Delta$ strain is $p \approx 0.68$ confirming the partial viability of the mutant according to our viability criterion. Figure 4.3a shows the cumulative probability, $P(T)$, of cycle times for wild-type and $CLB2db\Delta$ $clb5\Delta$ mutant cells growing in raffinose. ($P(T)$ is the probability that the cycle time of a randomly chosen cell is longer than a specified time, T.) As shown in Fig. 4.3a, $P(T)$ for the mutant cells levels off at approximately 35% as T increases, whereas, for wild-type cells, $P(T)$ drops steadily (below 5%) as T increases beyond 250 min. Cumulative distributions of cycle times computed by our hybrid stochastic model (black lines) are in excellent agreement with the experimental distributions (red-blue-green lines) for both wild-type and the double-mutant cells.

Another interesting mutant strain is $CLB1$ $clb2\Delta$ $cdh1\Delta$, for which the $CLB1$ gene is intact and $CLB2$ and $CDH1$ genes are deleted. Due to the mutual antagonism between Clb2 and

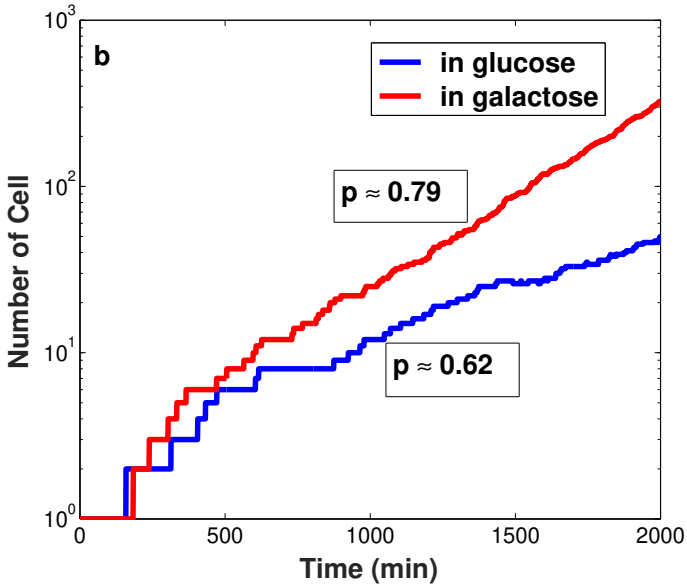

Figure 4.4: **Stochastic phenotype of $CLB1$ $clb2\Delta$ $cdh1\Delta$ mutant strain.** Comparison of cell proliferation for colonies of $CLB1$ $clb2\Delta$ $cdh1\Delta$ cells growing in glucose (blue) or galactose (red). The probability of division in our computational culture is given in the boxes next to each simulation. To mimic growth in glucose and galactose media, respectively, the specific growth rates are set to 0.0072 and 0.004621 min^{-1}, i.e., MDT = 96 and 150 min, respectively.

Cdh1, the deletion of both genes, $CDH1$ and $CLB2$ might be consistent with the viability of the double-mutant strain, provided Clb1 is still functional. Indeed, experimental observations show that $CLB1$ $clb2\Delta$ $cdh1\Delta$ cells are *poorly viable* in glucose medium and *viable* when growing on galactose [76]. Chen's deterministic model does not capture this phenotype; the model predicts the mutant cells to be viable in both media. However, simulation results of our hybrid stochastic model (Figure 4.4b) can reproduce the observed phenotype. The probability of division computed for a population of cells growing in glucose is ≈ 0.62, which suggests *poor* viability. In galactose, the corresponding probability of division is ≈ 0.79, which indicates that the mutant grows well in this medium. Figure 4.4b confirms a faster increase in cell number in the slower growth medium (galactose) which is in agreement with experimental observations.

4.4 Summary

In this chapter, we present a hybrid stochastic model of the molecular mechanism controlling progression through the budding yeast cell cycle. Our model provides a good match with experimental observations of many important characteristics of the budding yeast cell cycle, including inter-division time, cell size, and the phenotypes of more than 100 mutant strains. Compared with other approaches to stochastic modeling, our hybrid stochastic approach has several advantages. In a multiscale regulatory network such as cell cycle controls, the major source of intrinsic noise can be attributed to low copy numbers of mRNA species in the gene-protein regulatory network. In fact, in budding yeast cells, there are only 5-10 copies of each mRNA species encoding the production of corresponding proteins at levels of 500-5,000 molecules per cell. In such circumstances, small fluctuations in the population of mRNAs will result in substantial fluctuations in the corresponding protein levels. With this in mind, the key idea of the hybrid scheme is to partition the dynamics of mRNA species into the stochastic regime, to capture the major effects of random fluctuations in mRNA numbers, and to keep the protein dynamics in the deterministic framework, to achieve greater simulation efficiency. In addition, in this scheme, it is not necessary to reformulate the complex rate laws governing protein interactions as elementary mass-action rate laws, which is a great advantage from a modeling standpoint.

In this chapter, we have extended our hybrid stochastic method to a detailed molecular mechanism of cell cycle controls in budding yeast [9]. To apply our scheme to Chen's model, which is a deterministic model of protein interactions, we first had to extend the model to include mRNA species that are transcribed from cell-cycle genes and translated into proteins. Then we carried out comprehensive simulations of wild-type yeast cells and more than 100 mutant strains, using both the deterministic and hybrid ODE/SSA models. Results have shown that our stochastic simulations are in accord with most experimental observations, including detailed phenotypic characteristics of 103 out of 122 mutant cases. Although Chen's deterministic model may seem to 'score' bet-

ter on the 'viability' of mutant strains, it is not so highly constrained as our stochastic model by consideration of the statistical properties of these mutant cells, especially the characteristics of 'partially viable' mutants. In addition, our results prove that our hybrid approach to stochastic/deterministic simulations can achieve a good trade-off between the accuracy and efficiency of numerical simulations. FORTRAN code takes about 15 min to simulate 10,000 cell cycles on an Intel i7-3770 processor with 16G memory running a Linux environment. A similar system using a fully stochastic model may take more than one day. For example, when the FORTRAN code of Barik *et al.* [54] is run using the same work station, it takes more than 30 hours to generate a population of 10,000 yeast cells. Therefore, simulation of 122 mutants using a fully stochastic model may require months of computations while our hybrid stochastic model performs such a task in less than 30 hours.

4.5 Appendix B

Table 4.1: **List of mutant strains.** Simulation results of our hybrid stochastic model (in column 6) are compared with experimental data (in column 4) and simulation results of the deterministic model by Chen *et al.* [9] (in column 5). The parameter modification column reports the changes made to the parameter values to simulate each strain. 'V' means 'viable', 'In' means inviable, and 'pV' partially viable. We generate a sufficiently large population of mother and daughter cells, expanding over time, to compute the probability of division. In column 6, we report the probability of division as well as the number of simulations used to calculate this probability. Orange typeface in column 5 indicates disagreement between the Chen deterministic model and the experimental observations. The red typeface in column 6 indicates disagreement between the hybrid model and the experimental observations.

No	Mutant strain	Parameter modification	Exp	Chen model	Hybrid model
1	Wild-type (WT) on glucose (Glc)	-	V	V	V : 0.97 (11,000)

2.1	WT on galactose (Gal)	k_g=0.004621 (MDT=150)	V	V	V : 0.93 (22,762)
2.2	WT on raffinose (Raff)	k_g=0.00433 (MDT=160)	V	V	V : 0.93 (14,198)
3	$cln1\Delta\ cln2\Delta$	$k''_{s,n2}$=0	V	V	V : 0.92 (13,710)
4	GAL-$CLN2\ cln1\Delta\ cln2\Delta$	$k'_{s,n2}$=0.12, $k''_{s,n2}$=0, k_g=0.004621	V	V	V : 0.90 (24,500)
5	$cln1\Delta\ cln2\Delta\ sic1\Delta$	$k''_{s,n2}$=0, $k'_{s,c1} = k''_{s,c1}$=0	V	V	V : 0.75 (14,010)
6	$cln1\Delta\ cln2\Delta\ cdh1\Delta$	$k''_{s,n2}$=0, $k_{s,cdh}$=0, init CDH1T=CDH1=0	V	V	In : 0.30 (9,373)
7	GAL-$CLN2\ cln1\Delta\ cln2\Delta\ cdh1\Delta$	$k'_{s,n2}$=0.12, $k''_{s,n2}$=0, $k_{s,cdh}$=0, k_g=0.004621, init CDH1T=CDH1=0	V	V	V : 0.76 (11,594)
8	$cln3\Delta$	$k_{s,n3}$=0	V	V	V : 0.98 (9,990)
9	GAL-$CLN3$	$k_{s,n3}$=5×$k_{s,n3}$, k_g=0.004621	V	V	V : 0.94 (9,610)
10	$bck2\Delta$	$k_{s,k2}$=0	V	V	V : 0.97 (11,326)
11	multi-copy $BCK2$	$k_{s,k2}$=5×$k_{s,k2}$	V	V	V : 0.96 (10,032)

12	$cln1\Delta\ cln2\Delta\ bck2\Delta$	$k''_{s,n2}$=0, $k_{s,k2}$=0	V	V	V : 0.84 (11,294)
13	$cln3\Delta\ bck2\Delta$	$k_{s,n3}$=0, $k_{s,k2}$=0	In	In	In : 0.10 (9,310)
14	$cln3\Delta\ bck2\Delta\ GAL\text{-}CLN2\ cln1\Delta\ cln2\Delta$	$k_{s,n3}$=0, $k_{s,k2}$=0, $k'_{s,n2}$=0.12, $k''_{s,n2}$=0, k_g=0.004621	V	V	V : 0.95 (17,452)
15	$cln3\Delta\ bck2\Delta$ multi-copy $CLN2$	$k_{s,n3}$=0, $k_{s,k2}$=0 $k''_{s,n2}$=5×$k''_{s,n2}$	In	In	In : 0.40 (16,523)
16	$cln3\Delta\ bck2\Delta\ sic1\Delta$	$k_{s,n3}$=0, $k_{s,k2}$=0, $k'_{s,c1} = k''_{s,c1}$=0	In	In	In : 0.44 (9,614)
17	$cln1\Delta\ cln2\Delta\ cln3\Delta$	$k'_{s,n2}$=$k''_{s,n2}$=0, $k_{s,n3}$=0	In	In	In : 0.04 (9,700)
18	$cln1\Delta\ cln2\Delta\ cln3\Delta$ $GAL\text{-}CLN2$	$k''_{s,n2}$=0, $k_{s,n3}$=0, $k'_{s,n2}$=0.12, k_g=0.004621	V	V	V : 0.95 (17,640)
19	$cln1\Delta\ cln2\Delta\ cln3\Delta$ $GAL\text{-}CLN3$	$k'_{s,n2}$=$k''_{s,n2}$=0, $k_{s,n3}$=5×$k_{s,n3}$, k_g=0.004621	V	V	V : 0.84 (12,598)
20	$cln1\Delta\ cln2\Delta\ cln3\Delta$ $sic1\Delta$	$k'_{s,n2}$=0, $k''_{s,n2}$=0, $k_{s,n3}$=0, $k'_{s,c1} = k''_{s,c1}$= 0	V	V	V : 0.76 (12,904)
21	$cln1\Delta\ cln2\Delta\ cln3\Delta$ $cdh1\Delta$	$k'_{s,n2}$=$k''_{s,n2}$=0, $k_{s,n3}$=0, $k_{s,cdh}$=0, init CDH1=CDH1T=0	In	In	In : 0.03 (9,800)

22	$cln1\Delta\ cln2\Delta\ cln3\Delta$ multi-copy $CLB5$	$k'_{s,n2}$=$k''_{s,n2}$=0, $k_{s,n3}$=0, $k'_{s,b5}$=4×$k'_{s,b5}$, $k''_{s,b5}$=4×$k''_{s,b5}$	V	V	V : 0.86 (20,376)
23	$cln1\Delta\ cln2\Delta\ cln3\Delta$ $GAL\text{-}CLB5$	$k'_{s,n2}$=0, $k''_{s,n2}$=0, $k_{s,n3}$=0, $k'_{s,b5}$=15×$k'_{s,b5}$, k_g=0.004621	V	V	V : 0.99 (17,856)
24	$cln1\Delta\ cln2\Delta\ cln3\Delta$ multi-copy $BCK2$	$k'_{s,n2}$=$k''_{s,n2}$=0, $k_{s,n3}$=0, $k_{s,k2}$=10×$k_{s,k2}$	V	V	V : 0.89 (11,486)
25	$cln1\Delta\ cln2\Delta\ cln3\Delta$ $GAL\text{-}CLB2$	$k'_{s,n2}$=$k''_{s,n2}$=0, $k_{s,n3}$=0, $k'_{s,b2}$=2×$k'_{s,b2}$, k_g=0.004621	In	In	In : 0.01 (10,030)
26	$cln1\Delta\ cln2\Delta\ cln3\Delta$ $apc-ts$	$k'_{s,n2}$=$k''_{s,n2}$=0, $k_{s,n3}$=0, $k_{s,cdh}$=0, $k'_{s,20}$=$k''_{s,20}$=0, init CDH1=CDH1T=0	In	In	In : 0 (10,000)
27	$sic1\Delta$	$k'_{s,c1}$ = $k''_{s,c1}$=0	V	V	V : 0.87 (16,198)
28	$GAL\text{-}SIC1$	$k'_{s,c1}$=10×$k'_{s,c1}$, k_g=0.004621	V	V	V : 0.90 (9,848)
29	$GAL\text{-}SIC1db\Delta$	$k'_{s,c1}$=10×$k'_{s,c1}$, $k_{d3,c1}$=0, k_g=0.004621	In	In	In : 0 (10,000)
30	$GAL\text{-}SIC1\ cln1\Delta\ cln2\Delta$	$k'_{s,c1}$=10×$k'_{s,c1}$, $k''_{s,n2}$=0, k_g=0.004621	In	In	In : 0 (1,000)

31	*GAL-SIC1 GAL-CLN2 cln1Δ cln2Δ*	$k'_{s,c1}=10\times k'_{s,c1}$, $k'_{s,n2}=0.12$, $k''_{s,n2}=0$, $k_g=0.004621$	V	V	V : 0.90 (11,180)
32	*GAL-SIC1 cln1Δ cln2Δ cdh1Δ*	$k'_{s,c1}=k''_{s,c1}=0.12$, $k''_{s,n2}=k_{s,cdh}=0$, $k_g=0.004621$, init CDH1T=CDH1=0	In	In	In : 0 (1,000)
33	*GAL-SIC1 GAL-CLN2 cln1Δ cln2Δ cdh1Δ*	$k'_{s,c1}=0.12$, $k'_{s,n2}=0.12$, $k''_{s,n2}=0$, $k_{s,cdh}=0$, $k_g=0.004621$, init CDH1T=CDH1=0	V	V	V : 0.80 (14,376)
34	*cdh1Δ*	$k_{s,cdh}=0$, init CDH1T=CDH1=0	V	V	In : 0.49 (15,078)
35	*Cdh1* constitutively active	$k''_{i,cdh}=0$, $k_g=0.004621$, $k_{s,cdh}=3\times k_{s,cdh}$	In	In	In : 0.32 (9,238)
36	*sic1Δ cdh1Δ*	$k'_{s,c1}=k''_{s,c1}=0$, $k_{s,cdh}=0$, init CDH1T=CDH1=0	In	In	In : 0 (10,000)
37	*sic1Δ cdh1Δ GAL-CDC20*	$k'_{s,c1}=k''_{s,c1}=0$, $k_{s,cdh}=0$, $k'_{s,20}=10$, $k_g=0.004621$	V	V	Gal : In : 0.28 (9,640) Glc : pV : 0.71 (10,708)
38	*cdc6Δ2-49*	$k'_{s,f6}=k''_{s,f6}=k'''_{s,f6}=0$	V	V	V : 0.97 (11,482)

39	$sic1\Delta$ $cdc6\Delta$2-49	$k'_{s,c1}=k''_{s,c1}=0$, $k'_{s,f6}=k''_{s,f6}=k'''_{s,f6}=0$	V	V	In : 0.01 (9,016)
40	$cdh1\Delta$ $cdc6\Delta$2-49	$k'_{s,f6}=k''_{s,f6}=k'''_{s,f6}=0$, $k_{s,cdh}=0$, init CDH1T=CDH1=0	V	V	In : 0.16 (10,331)
41[1]	$sic1\Delta$ $cdc6\Delta$2-49 $cdh1\Delta$	$k'_{s,c1}= k''_{s,c1}= 0$, $k_{s,cdh}=0$, $k'_{s,f6}= k''_{s,f6}=k'''_{s,f6}=0$, init CDH1T=CDH1=0	In	In	In
42	$sic1\Delta$ $cdc6\Delta$2-49 $cdh1\Delta$ GAL-$CDC20$	$k'_{s,c1}=k''_{s,c1}= 0$, $k'_{s,f6}=k''_{s,f6}=k'''_{s,f6}=0$, $k_{s,cdh}= 0$, $k'_{s,20}=4$, k_g=0.004621, init CDH1T=CDH1=0	V	V	In : 0.02 (10,088)
43	$swi5\Delta$	$k'_{s,swi}=k''_{s,swi}=0$	V	V	In : 0.53 (23,878)
44	$swi5\Delta$ GAL-$CLB2$	$k'_{s,swi}=k''_{s,swi}=0$, $k_{s,b2}$=0.12, k_g=0.004621	In	In	In : 0.34 (9,790)

[1]Experimental observations show that $sic1\Delta$ $cdc6\Delta$2-49 $cdh1\Delta$ (mutant # 41) and $swi5\Delta$ $cdh1\Delta$ (mutant # 45) strains are not able to undergo the cell division and thus are inviable [79]; both strains have similar phenotype and become arrested as binucleate cells with 4C DNA content. Our simulation (similar to Chen's deterministic simulation [9]) shows that the cells are inviable; however, they become arrested in telophase with 2C DNA content. We consider this simulation result as an inconsistency with experimental observations.

45[1]	$swi5\Delta\ cdh1\Delta$	$k'_{s,swi}$= $k''_{s,swi}$=0, $k_{s,cdh}$=0, init CDH1T=CDH1=0	In	In	In
46	$swi5\Delta\ cdh1\Delta\ GAL\text{-}SIC1$	$k'_{s,swi}$= $k''_{s,swi}$=0, $k_{s,cdh}$=0, $k'_{s,c1}$=10$\times k'_{s,c1}$, k_g=0.004621, init CDH1T=CDH1=0	V	V	V : 0.83 (10,636)
47	$clb1\Delta\ clb2\Delta$	$k'_{s,b2}$=$k''_{s,b2}$=0	In	In	In : 0 (10,000)
48	$CLB1\ clb2\Delta$	$k'_{s,b2}$=0.0003, $k''_{s,b2}$=0.013 (33% WT)	V	In	V : 0.82 (26,466)
49	$GAL\text{-}CLB2$	$k'_{s,b2}$=0.12, k_g=0.004621	V	V	V : 0.86 (17,714)
50	multi-copy $GAL\text{-}CLB2$	$k'_{s,b2}$=0.96 (8 copies of $GAL\text{-}CLB2$), k_g=0.004621	In	In	In : 0 (10,000)
51[2]	$CLB1\ clb2\Delta\ cdh1\Delta$	$k'_{s,b2}$=0.0003, $k''_{s,b2}$=0.013 (33% WT), $k_{s,cdh}$=0, init CDH1T=CDH1=0	Gal :V Glc : pV	Gal : V Glc : V	Gal : V : 0.79 (17,608), Glc : In : 0.62 (14,719)

[2]Experimental observations (see supplementary figure 1, panel A in [80]) show that $CLB1\ clb2\Delta\ cdh1\Delta$ cells grow well on galactose; however, they exhibit poor viability on glucose. That is why in column 4 we report this mutant strain as partially viable (pV). 'Glc' refers to growth in glucose medium; 'Gal' to galactose.

52[3]	$CLB1\ clb2\Delta\ pds1\Delta$	$k'_{s,b2}$=0.0003, $k''_{s,b2}$=0.013 (33% WT), $k''_{s1,pds}$=$k''_{s2,pds}$=0	In	V	In
53	$GAL\text{-}CLB2\ sic1\Delta$	$k'_{s,b2}$=0.12, $k'_{s,c1}$=$k''_{s,c1}$=0, k_g=0.004621	In	In	In : 0.20 (10,090)
54	$GAL\text{-}CLB2\ cdh1\Delta$	$k'_{s,b2}$=0.12, $k_{s,cdh}$=0, k_g=0.004621, init CDH1T=CDH1=0	In	In	In : 0 (10,000)
55	$CLB2\text{-}db\Delta$	$k_{d,b2p}$=0, $k''_{d,b2}$=0.03 (8.5% activity left due to KEN box).	In	In	In : 0.27 (9,892)
56	$CLB2\text{-}db\Delta$ on galactose	$k_{d,b2p}$=0, $k''_{d,b2}$=0.03, k_g=0.004621	In	In	In : 0.59 (13,848)
57	$CLB2\text{-}db\Delta$ multi-copy $SIC1$	$k_{d,b2p}$=0, $k''_{d,b2}$=0.03, $k'_{s,c1}$=10×$k'_{s,c1}$, $k''_{s,c1}$=10×$k''_{s,c1}$	V	V	V : 0.79 (11,900)
58	$CLB2\text{-}db\Delta\ GAL\text{-}SIC1$	$k_{d,b2p}$=0, $k''_{d,b2}$=0.03, $k'_{s,c1}$=10×$k'_{s,c1}$, k_g=0.004621	V	V	V : 0.79 (9,044)

[3]Simulation results show that $CLB1\ clb2\Delta\ pds1\Delta$ (mutant # 52), and $pds1\Delta$ (mutant # 80) strains are able to exit mitosis; however, Esp1 is active throughout the cell cycle (even before spindle alignment). Thus, according to our viability rules, these mutants are inviable.

59	*CLB2-dbΔ* multi-copy *CDC6*	$k_{d,b2p}$=0, $k''_{d,b2}$=0.03, $k'_{s,f6}$=5×$k'_{s,f6}$, $k''_{s,f6}$=5×$k''_{s,f6}$, $k'''_{s,f6}$=5×$k'''_{s,f6}$	V	V	In : 0.57 (13,484)
60	*CLB2-dbΔ clb5Δ*	$k_{d,b2p}$=0, $k''_{d,b2}$=0.03, $k'_{s,b5}$=$k''_{s,b5}$=0	In	In	In : 0.24 (9,672)
61	*CLB2-dbΔ clb5Δ* on galactose and raffinose	$k_{d,b2p}$=0, $k''_{d,b2}$=0.03, $k'_{s,b5}$=$k''_{s,b5}$=0, k_g=0.004621 (Gal), k_g=0.00433 (Raff)	pV	V	Gal : In : 0.62 (17,338) Raff : pV : 0.68 (13,364)
62	*GAL-CLB2-dbΔ*	$k_{d,b2p}$=0, $k''_{d,b2}$=0.03, $k''_{s,b2}$=0.12, k_g=0.004621	In	In	In : 0.03 (10,118)
63	*clb5Δ clb6Δ*	$k'_{s,b5}$=$k''_{s,b5}$=0	V	V	V : 0.98 (16,653)
64	*cln1Δ cln2Δ clb5Δ clb6Δ*	$k'_{s,n2}$=$k''_{s,n2}$=0, $k'_{s,b5}$=$k''_{s,b5}$=0	In	In	In : 0 (10,000)
65	*GAL-CLB5*	$k'_{s,b5}$=15×$k'_{s,b5}$, k_g=0.004621	V	V	V : 0.94 (9,612)
66[4]	*GAL-CLB5 sic1Δ*	$k'_{s,b5}$=15×$k'_{s,b5}$, $k'_{s,c1}$=$k''_{s,c1}$=0, k_g=0.004621	In	In	In

[4]Simulation results show that mutant strains *GAL-CLB5 sic1Δ* (mutant # 66), *CLB5-dbΔ sic1Δ* (mutant # 69), and *GAL-CLB5-dbΔ* (mutant # 72) are able to exit mitosis; however, [ORI] is not relicensed (due to high activity of Clb5 in the G1 phase). Thus, according to our viability rules, these mutants are inviable.

67	GAL-$CLB5$ $cdh1\Delta$	$k'_{s,b5}=10\times k'_{s,b5}$, $k_{s,cdh}=0$, $k_g=0.004621$, init CDH1T=CDH1=0	In	V	In : 0.60 (20,782)
68	$CLB5$-$db\Delta$	$k''_{d,b5}=0$	V	V	V : 0.95 (11,656)
69[4]	$CLB5$-$db\Delta$ $sic1\Delta$	$k''_{d,b5}=0$, $k'_{s,c1}=k''_{s,c1}=0$	In	In	In
70	$CLB5$-$db\Delta$ $pds1\Delta$	$k''_{d,b5}=0$, $k''_{s1,pds}=k''_{s2,pds}=0$	V	V	V : 0.97 (12,539)
71	$CLB5$-$db\Delta$ $pds1\Delta$ $cdc20\Delta$	$k''_{d,b5}=0$, $k''_{s1,pds}=k''_{s2,pds}=0$, $k'_{s,20}=k''_{s,20}=0$	In	In	pV : 0.65, (12,846)
72[4]	GAL-$CLB5$-$db\Delta$	$k'_{s,b5}=15\times k'_{s,b5}$, $k''_{d,b5}=0$, $k_g=0.004621$	In	In	In
73	$cdc20$-ts	$k'_{s,20}=k''_{s,20}=0$	In	In	In : 0.31 (9,710)
74	$cdc20\Delta$ $clb5\Delta$	$k'_{s,20}=k''_{s,20}=0$, $k'_{s,b5}=k''_{d,b5}=0$	In	In	In : 0.58 (9,944)
75	$cdc20\Delta$ $pds1\Delta$	$k'_{s,20}=k''_{s,20}=0$, $k''_{s1,pds}=k''_{s2,pds}=0$	In	In	pV : 0.66 (16,274)
76	$cdc20\Delta$ $pds1\Delta$ $clb5\Delta$	$k'_{s,20}=k''_{s,20}=0$, $k'_{s,b5}=k''_{d,b5}=0$, $k''_{s1,pds}=k''_{s2,pds}=0$	V	V	V : 0.95 (10,920)
77	GAL-$CDC20$	$k'_{s,20}=6$, $k_g=0.004621$	In	In	V : 0.99 (11,355)

78	$cdc20$-ts $mad2\Delta$	$k'_{s,20}$=$k''_{s,20}$=0, k_{mad2}=0.01	In	In	In : 0.31 (9,720)
79	$cdc20$-ts $bub2\Delta$	$k'_{s,20}$=$k''_{s,20}$=0, k_{bub2}=0, init BUB2=0	In	In	In : 0.25 (9,586)
80[3]	$pds1\Delta$	$k''_{s1,pds}$=$k''_{s2,pds}$=0	V	In	In
81[5]	$esp1$-ts	$k_{as,esp}$= 0.1, $k_{di,esp}$= 0.002, (1/500 × WT)	In	In	In
82[6]	$PDS1$-$db\Delta$	$k''_{d2,pds}$=$k''_{d3,pds}$=0	In	In	In
83[6]	GAL-$PDS1$-$db\Delta$	$k'_{s,pds}$= 0.1, k_g=0.004621 $k''_{d2,pds}$=$k''_{d3,pds}$=0	In	In	In
84[6]	GAL-$PDS1$-$db\Delta$ $esp1$-ts	$k'_{s,pds}$= 0.1, $k''_{d2,pds}$= $k''_{d3,pds}$=0, $k_{as,esp}$=0.1, $k_{di,esp}$=0.002, (1/500× WT), k_g=0.004621	In	In	In
85	GAL-$ESP1$ $cdc20$-ts	$k'_{s,20}$=$k''_{s,20}$=0, $k'_{s,esp1t}$=3×$k'_{s,esp1t}$, k_g=0.004621	In	In	In : 0.47 (10,057)

[5]Simulation results show that $esp1$-ts strains are able to exit mitosis while the sister chromatids are not yet separated. Therefore, according to our viability rule this mutant is inviable.

[6]Simulation results show that $PDS1$-$db\Delta$ (mutant # 82), GAL-$PDS1$-$db\Delta$ (mutant # 83), and GAL-$PDS1$-$db\Delta$ $esp1$-ts (mutant # 84) strains are able to exit mitosis; however, Esp1 is inactive throughout the cell cycle which means sister chromatids are still attached. Thus, according to our viability rules, these mutants are inviable.

86	$tem1\Delta$	$k''_{a,15}$=0.002 (= $k'_{a,15}$), $k_{s,mTem1}$=0	In	In	pV : 0.73 (14,096)
87	$GAL\text{-}TEM1$	$k_{s,tem1t}$=5×$k_{s,tem1t}$, k_g=0.004621	V	V	V : 0.93 (15,072)
88	$tem1\text{-}ts$ multi-copy $CDC15$	$k''_{a,15}$=0.002, $k_{s,mTem1}$=0, $k_{s,tem1t}$=0, $k_{s,cdc15t}$=5×$k_{s,cdc15t}$	V	V	V : 0.77 (17,050)
89	$tem1\text{-}ts$ $GAL\text{-}CDC15$	$k''_{a,15}$=0.002, $k_{s,mTem1}$=0, $k_{s,tem1t}$=0, $k_{s,cdc15t}$=15×$k_{s,cdc15t}$, k_g=0.004621	V	V	V : 0.81 (15,616)
90	$tem1\Delta$ $net1\text{-}ts$	$k''_{a,15}$=0.002, $k_{s,mTem1}$=0, $k_{s,tem1t}$=0, $k_{as,rent}$=10, $k_{as,rentp}$=0.05, (5% WT)	V	V	V : 0.99 (16,436)
91	$tem1\text{-}ts$ multi-copy $CDC14$	$k''_{a,15}$=0.002, $k_{s,14}$=0.4 (2 copies), $k_{s,mTem1}$=$k_{s,tem1t}$=0	V	V	V : 0.96 (10,470)
92	$cdc15\Delta$	$k''_{kp,net}$=0, $k_{s,cdc15t}$=0	In	In	pV : 0.73 (13,862)
93	multi-copy $CDC15$	$k_{s,cdc15t}$=5×$k_{s,cdc15t}$	V	V	V : 0.97 (15,020)
94	$cdc15\text{-}ts$ multi-copy $TEM1$	$k''_{kp,net}$=0, $k_{s,cdc15t}$=0, $k_{s,tem1t}$=5×$k_{s,tem1t}$	In	In	pV : 0.73 (14,076)

95	$cdc15\Delta\ net1\text{-}ts$	$k''_{\text{kp,net}}$=0, $k_{\text{as,rentp}}$=0.05, $k_{\text{as,rent}}$=10, (5% WT)	V	V	V : 0.99 (27,612)
96	$cdc15\text{-}ts$ multi-copy $CDC14$	$k''_{\text{kp,net}}$=0, $k_{\text{s,14}}$=0.4 (2 copies)	V	V	V : 0.96 (24,884)
97	$net1\text{-}ts$	$k_{\text{as,rentp}}$=0.05, $k_{\text{as,rent}}$=10, (5% WT)	V	V	V : 0.99 (30,3160)
98	$GAL\text{-}NET1$	$k_{\text{s,net}}$=5×$k_{\text{s,net}}$, k_{g}=0.004621	In	In	In : 0.50 (10,966)
99	$cdc14\text{-}ts$	$k_{\text{s,14}}$=0	In	In	In : 0.08 (9,600)
100	$GAL\text{-}CDC14$	$k_{\text{s,14}}$=4×$k_{\text{s,14}}$, k_{g}=0.004621	In	In	In : 0.04 (10,004)
101	$GAL\text{-}NET1$ $GAL\text{-}CDC14$	$k_{\text{s,net}}$=3×$k_{\text{s,net}}$, $k_{\text{s,14}}$=3×$k_{\text{s,14}}$, k_{g}=0.004621	V	V	V : 0.94 (10,651)
102	$net1\Delta$ $cdc20\text{-}ts$	$k_{\text{as,rentp}}$=0.05, $k_{\text{as,rent}}$=10, (5% WT) $k'_{\text{s,20}}$=$k''_{\text{s,20}}$=0	In	In	In : 0.55 (10,974)
103	$cdc14\text{-}ts$ $GAL\text{-}SIC1$	$k_{\text{s,14}}$=0, $k'_{\text{s,c1}}$=10×$k'_{\text{s,c1}}$, k_{g}=0.004621	In	In	In : 0.49 (10,482)

104	*cdc*14-*ts* then *GAL-SIC*1	$k_{s,14}$=0 for 180 min, then 3 copies of GAL-SIC1 ($k'_{s,c1}$=3×$k'_{s,c1}$, k_g=0.004621)	able to exit mito-sis	exits	not able to exit mitosis
105	*cdc*14-*ts sic*1*Δ*	$k'_{s,c1}$=$k''_{s,c1}$=0, (16%Cdc14 activity) $k_{pp,c1}$=$k_{pp,f6}$=0.64, $k_{a,swi}$=0.32, $k''_{a,cdh}$=0.128	In	In	In : 0.27 (9,725)
106	*cdc*14-*ts cdh*1*Δ*	$k_{s,cdh}$=0, init CDH1=CDH1T=0 (30%Cdc14 activity) $k_{pp,c1}$=$k_{pp,f6}$=1.2, $k_{a,swi}$=0.6, $k''_{a,cdh}$=0.13	In	In	In : 0.16 (9,200)
107	*cdc*14-*ts GAL-CLN*2	$k'_{s,n2}$=0.12, $k''_{s,n2}$=0, k_g=0.004621, (10%Cdc14 activity) $k_{pp,c1}$=$k_{pp,f6}$=0.8, $k_{a,swi}$=0.4, $k''_{a,cdh}$=0.16	In	In	In : 0.61 (11,612)
108	*TAB*6-1	$k_{as,rentp}$=0.05, $k_{as,rent}$=10,(5% WT)	V	V	V : 0.99 (13,010)
109	*TAB*6-1 *cdc*15*Δ*	$k_{as,rentp}$=0.05, $k_{as,rent}$=10 (5% WT), $k''_{kp,net}$=0	V	V	V : 0.99 (12,960)

110	*TAB6-1 clb5Δ clb6Δ*	$k_{as,rentp}$=0.05, $k_{as,rent}$=10 (5% WT), $k'_{s,b5}$=$k''_{s,b5}$=0	In	In	V : 0.99 (11,638)
111	*TAB6-1 CLB1 clb2Δ*	$k_{as,rentp}$= 0.05, $k_{as,rent}$=10 (5% WT), $k'_{s,b2}$=0.0003, $k''_{s,b2}$=0.013, (33% WT)	V	V	V : 0.98 (11,314)
112	*mad2Δ*	k_{mad2}=0.01	V	V	V : 0.98 (16,510)
113	*bub2Δ*	k_{bub2}=0, init BUB2=0	V	V	V : 0.98 (12,488)
114	*mad2Δ bub2Δ*	k_{bub2}=0, k_{mad2}=0.01, init BUB2=0	V	V	V : 0.99 (11,659)
115	*APC-A*	$k''_{a,20}$=0	V	V	V : 0.97 (12,113)
116	*APC-A cdh1Δ*	$k''_{a,20}$=0, $k_{s,cdh}$=0, init CDH1T=CDH1=0	In	In	In : 0.18 (9,911)
117	*APC-A cdh1Δ* on galactose	$k''_{a,20}$=0, $k_{s,cdh}$=0, init CDH1T=CDH1=0, k_g=0.004621	In	In	In : 0.47 (10,796)
118	*APC-A cdh1Δ* multi-copy *SIC1*	$k''_{a,20}$=0, $k_{s,cdh}$=0, $k'_{s,c1}$=15×$k'_{s,c1}$, $k''_{s,c1}$=15×$k''_{s,c1}$, init CDH1T=CDH1=0	V	V	V : 0.83 (16,949)

119	*APC-A cdh1Δ GAL-SIC1*	$k''_{a,20}$=0, $k_{s,cdh}$=0, $k'_{s,c1}$=20 $\times k'_{s,c1}$, k_g=0.004621, init CDH1T=CDH1=0	V	V	V : 0.75 (9,934)
120	*APC-A cdh1Δ* multi-copy *CDC6*	$k''_{a,20}$=0, $k_{s,cdh}$=0, $k'_{s,f6}$=10$\times k'_{s,f6}$, $k''_{s,f6}$=10$\times k''_{s,f6}$, $k'''_{s,f6}$=10$\times k'''_{s,f6}$, init CDH1T=CDH1=0	V	V	V : 0.78 (12,700)
121	*APC-A cdh1Δ GAL-CDC6*	$k''_{a,20}$=0, $k_{s,cdh}$=0, $k'_{s,f6}$=15$\times k'_{s,f6}$, k_g=0.004621, init CDH1T=CDH1=0	V	V	V : 0.82 (11,321)
122	*APC-A cdh1Δ* multi-copy *CDC20*	$k''_{a,20}$=0, $k_{s,cdh}$=0, $k'_{s,20}$=25$\times k'_{s,20}$, $k''_{s,20}$=25$\times k''_{s,20}$, init CDH1T=CDH1=0	V	V	V : 0.79 (26,557)
123	*APC-A sic1Δ*	$k''_{a,20}$=0, $k'_{s,c1}$=$k''_{s,c1}$=0	V	V	V : 0.83 (14,922)
124	*APC-A GAL-CLB2*	$k''_{a,20}$=0, $k'_{s,b2}$=0.48, k_g=0.004621	In	In	In : 0 (10,000)

Chapter 5

Stochastic Size Control Mechanism in the Budding Yeast Cell Cycle

5.1 Introduction

Cell size is a key characteristic that significantly affects many aspects of cellular physiology. There are specific control mechanisms during cell cycle that maintain the cell size within a range from generation after generation. Such complex control mechanisms are regulated by positive and negative feedbacks. The feedbacks create bistable switches to make the progression through different phases (G1 $\rightarrow$ S $\rightarrow$ G2 $\rightarrow$ M) irreversible. The specification of these biochemical feedbacks varies between different organisms. Extensive experimental studies have been carried out to identify the underlying molecular mechanisms that regulate the cell cycle [81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 69, 65, 66, 91, 92]. Many signaling pathways in regulatory networks in addition to a myriad of mutant phenotypes have been explored. Along with experimental studies, various mathematical models, including deterministic models, boolean networks, stochastic models, and hybrid approaches have been developed to quantitatively describe the cell cycle as a dynamical system.

Despite the large body of experimental and mathematical works, there are still important aspects of the cell cycle control mechanism that require further studies. For instance, mechanisms that control the cell size increase the survival chance for a cell under different circumstances, e.g., in the presence of the molecular noise. However, despite their importance, it is still not clear what underlying control mechanisms maintain the robustness of cell size. Experimental data shows that in most types of living cells, size control is often regulated through feedbacks between the time a cell spends in each phase of cell cycle and the cell volume [93, 94, 95, 96, 97, 98, 99]. Nonetheless, the specification of those feedbacks may vary across different living cells, such as fission yeast, mammalian eukaryotes, epithelial cells, and budding yeast.

Since yeasts are genetically tractable in experimental studies, most studies are focused on these unicellular organisms, in particular the budding yeast. Prior studies show that the asymmetric division of budding yeast cells produces smaller daughter cells and larger mother cells. Those small daughter cells spend a longer time in the G1 phase before they go through the START transition when a cell makes an irreversible commitment to commence a new cycle. However, those mother cells of the same size commence a new cycle soon after division [12]. This observation suggests that the size control is not strong in mother cells, even if their size is as small as a daughter cell.

Di Talia *et al.* [12] has studied the effect of molecular noise and size control in the variability of the budding yeast cell cycle. More specifically, they used single-cell imaging of fluorescent-labeled budding yeast cells to measure the G1 time. According to their observations: 1) the hypothesis that a substantial portion of intrinsic noise stems from the noise in gene expression level is confirmed. More specifically, they showed that the variabilities in G1 time decrease with the square root of the ploidy and increased dosage of G1 cyclins, 2) size control plays important role in G1 time variability of daughter cells, but not on that of mother cells, and 3) the G1 cyclin genes *CLN2* and *CLN3* dominantly control the regulatory dynamics of the START transition by size control and time control modules. In fact, the size control module enforces longer G1 times to small size daughter

cells, but not mother cells. However, the time control module is the same in both daughter and mother cells. The study by Di Talia provides a comprehensive data set that can be used to construct new informative models to quantitatively study the effect of size control module on the variability of the cell cycle.

To quantitatively study the effect of such variability in progression through the cell cycle, detailed stochastic models are required. The multiscale feature is inherent in reaction rates and reactant populations inside living cells. For instance, the post-translational reactions (such as phosphorylation/dephosphorylation) during the budding yeast cell cycle are several orders of magnitude more frequent than transcriptional reactions. Moreover, species in a system may also exhibit different scales of populations. For example, mRNAs with average abundance of 5-10 molecules per cell are translated into proteins with the average abundance of 1,000-10,000 molecules per cell. The HR hybrid method leverages the efficiency of solving ordinary differential equations (ODEs) and the accuracy of SSA by integrating both deterministic and stochastic approaches in a single model.

The main contribution of this chapter is a new hybrid model that quantitatively describes key characteristics of the cell cycle, such as inter-division times and cell sizes, distribution of mRNAs, as well as the partial viability of specific mutant strains. Building on our previous work in chapter!3, our new model includes the transcripts of the early G1 phase. This feature is in a direct contrast with existing works, such as [9, 100, 101], that disregard the dynamics of early G1 proteins (Cln3 and Bck2) and do not include the G1 cyclin transcripts (*mCln3* and *mBck2*). The proposed model enables studying the effect of noise on G1 cyclins and size control mechanism in the budding yeast cell cycle. In fact, the developed model partitions all 45 proteins and 21 mRNAs, respectively, in fast and slow subsets. Using this partitioning, we obtain the dynamics of the system by solving ODEs for the fast subset and applying SSA to the slow subset. In addition, we use our model to predict variabilities in mutant strains of the G1 phase. Finally, we present comprehensive

results for the performance comparison between our proposed model and the experimental observations reported by Di Talia *et al.* [12]. In this chapter, we modify our hybrid stochastic model introduced in chapter 2 to study the effect of molecular noise and size control mechanism on the variabilities in the cell cycle of the budding yeast *Saccharomyces cerevisiae*. The proposed model provides an accurate, yet computationally efficient approach for simulation of an intricate system by integrating the deterministic and stochastic simulation schemes. The developed hybrid stochastic model can successfully capture several key features of the cell cycle observed in experimental data. In particular, the proposed model: 1) confirms that the majority of noise in size control stems from low copy numbers of transcripts in the G1 phase, 2) identifies the size and time regulation modules in the size control mechanism, and 3) conforms with phenotypes of early G1 mutants in exquisite detail.

5.2 Model and simulation

The model we explore in this chapter is based on the molecular regulatory network originally proposed by Chen *et al.* [9]. Chen's model is a comprehensive deterministic model that accounts for average properties of wild-type budding yeast cells, in addition to the phenotypes of more than 100 mutant strains. Next, we briefly summarize the regulatory network of Chen's model.

Cdc28 is the main regulator of the budding yeast cell cycle which is assumed to be constitutively expressed. Cdc28 forms an active kinase by binding to two families of cyclin partners, Cln1-3 and Clb1-6. In the early G1 phase, Cln3 and Bck2 (a backup protein) are the main partners of Cdc28. As a newborn cell grows, the Cln3 and Bck2 proteins accumulate in the nucleus to activate the transcription factors SBF and MBF. These two transcription factors are responsible for the production of Cln2 and Clb5. Cln2 accumulation induces the emergence of bud and Clb5 initiates DNA synthesis. The activation of SBF corresponds to a transition called the START. Once the cell

goes through the START transition, it commits to a new cycle. Shortly after the emergence of bud and initiation of DNA synthesis in the S phase, the level of Clb2 rises, and the spindle assembly starts to form. Later, the cell goes through another transition called the FINISH, during which a pair of proteins, Cdc20 and Cdh1, become active and facilitate the degradation of Clb2 and Clb5. Through this transition, the level of Clb2 drops, the cell divides, and returns to the G1 phase.

Chen's model describes the protein-protein regulatory network but does not include the dynamics of mRNAs. There is strong evidence suggesting that the noise in the cell cycle mainly results from the low copy number of mRNAs [12, 32, 11]. Thus, to construct a model that accounts for the variabilities in the cell cycle, it is crucial to incorporate the dynamics of mRNAs along with protein regulators in a stochastic model. Therefore, in our earlier work in [101], we substantially extended Chen's model by adding the dynamics of 19 important mRNAs. Moreover, we constructed a hybrid stochastic model that not only explains the average properties of the budding yeast cell cycle, it can also reproduce the variability in critical characteristics of the cell cycle, in addition to stochastic phenotypes of specific mutants. Next, we briefly discuss the HR hybrid model. Fully stochastic simulation becomes substantially slow when reactants with high abundance or reactions with high frequency are involved [17]. That is because the time complexity of SSA scales with the number of reactions [15]. The HR hybrid method integrates stochastic and deterministic approaches to construct a stochastic model that achieves a good trade-off between accuracy and efficiency. The deterministic method, typically formulated by nonlinear ODEs, is a computationally efficient approach to describe the average properties of a system. Meanwhile, the stochastic counterpart implements the computationally expensive SSA to generate stochastic time-evolution trajectories of the state variables (here the molecule counts of mRNAs). The main idea of the HR hybrid stochastic model is to partition the system of reactions into *fast* and *slow* subsets, apply the computationally expensive SSA only to the slow reactions, and then solve ODEs for the fast subset.

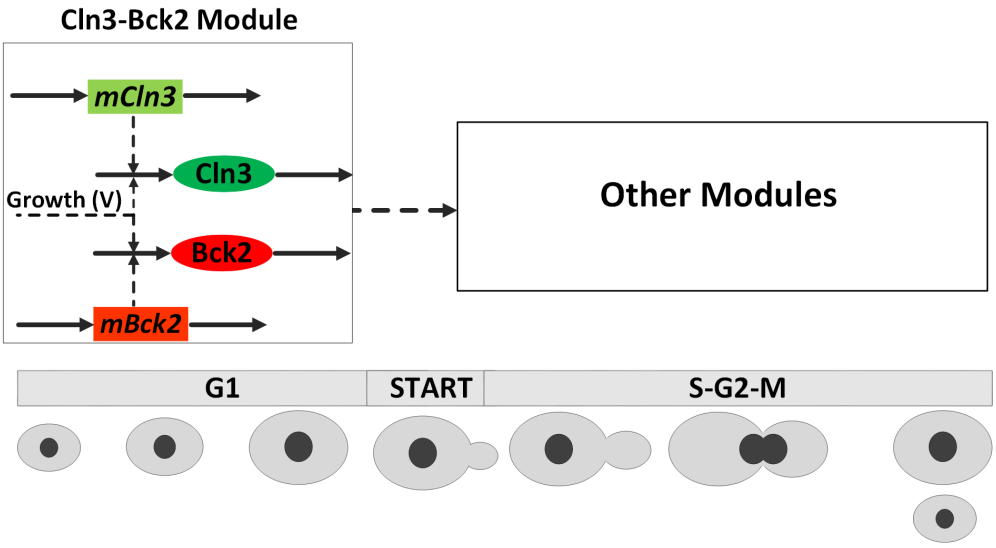

Figure 5.1: **Modified Cln3-Bck2 module for budding yeast cell cycle model.** Solid arrows represent synthesis/degradation reactions and dashed arrows represent activation. The *mCln3* and *mBck2* are unregulated mRNAs constitutively transcribed. As cell grows in the G1 phase, Cln3 and Bck2 proteins accumulate and the cell goes through the START transition, during which the cell commits to a new round of division. The START transition is identified by the activation of two important transcription factors (SBF and MBF), which are included in *Other Modules* and not presented here.

In our model, the reaction channels in the region I includes the synthesis and degradation of transcripts. After such partitioning, we apply the HR hybrid algorithm to simulate the trajectories of state variables. The step by step algorithm is provided in [102]. In those works, we have shown that placing the dynamics of mRNAs into SSA regime and solving ODEs for protein regulatory network leads to sufficiently accurate results, and significantly reduces the computational cost. In this chapter, we substantially improve our extended model in 3 as follows:

- First, we incorporate the dynamics of two early G1-phase proteins, Cln3 and Bck2 as depicted in Fig 5.1. In fact, Cln3 and Bck2 play important roles in the START transition and thus are necessary to be included into the model. In Chen's original model [9] the activities of

Cln3 and Bck2 were formulated by algebraic equations (5.1) and (5.2), respectively.

$$[\text{Cln3}] = \frac{C_0 \cdot D_{\text{n3}} \cdot [\text{mass}]}{(J_{\text{n3}} + D_{\text{n3}} \cdot [\text{mass}])}, \tag{5.1}$$

$$[\text{Bck2}] = [\text{mass}] \cdot B_0 \tag{5.2}$$

Here C_0 determines the maximum activity of Cln3, D_{n3} is the dosage of *CLN3* gene, J_{n3} is the Michaelis Menton constant, and B_0 is the Bck2 constant. We note that $[X]$ denotes the concentration of species X. These algebraic equations present an underlying assumption that the aforementioned proteins are always in steady states. We relax this assumption and modify these two algebraic equations into corresponding ODEs in (5.3) and (5.4):

$$\frac{\text{dCln3}}{\text{d}t} = k_{\text{s,n3}} \cdot V^2 \cdot \text{mCln3} - (k_{\text{d,n3}} - k_{\text{g}}) \cdot \text{Cln3}, \tag{5.3}$$

$$\frac{\text{dBck2}}{\text{d}t} = k_{\text{s,k2}} \cdot V \cdot \text{mBck2} - (k_{\text{d,k2}} - k_{\text{g}}) \cdot \text{Bck2}, \tag{5.4}$$

where k_{g} is the growth rate of the cell, V indicates the volume of the cell, $k_{\text{s,n3}}$ and $k_{\text{d,n3}}$ are the synthesis and degradation rates of Cln3 and $k_{\text{s,k2}}$ and $k_{\text{d,k2}}$ are the synthesis and degradation rates of Bck2. The growth rate is defined as $k_{\text{g}} = \ln 2/\text{MDT}$, where MDT is the mass doubling time of the culture. MDT is measured experimentally for different nutrient cultures. For instance, the MDT for glucose is observed to be approximately 90 minutes and thus, the growth rate of glucose culture is estimated as $\ln 2/90 \approx 0.0077$ min^{-1}. The synthesis and degradation rates are estimated as follow: we first estimate the degradation rate defined by $k_{\text{d,p}} = \ln 2/\tau_{\text{p}}$, where τ_{p} is the half-life time of the protein P. Then, we estimate the synthesis rate such that the average abundance of the protein matches with experimental observation reported in [66]. Table 5.1 lists the estimated parameters.

The reason we modified the algebraic equations in (5.1) and (5.2) into ODEs in (5.3) and

Table 5.1: Estimated parameters in fast subset of the extended model.

Parameter	Value	Parameter	Value
$k_{\text{s,n3}}$	0.0015	$k_{\text{d,n3}}$	0.12
$k_{\text{s,k2}}$	0.024	$k_{\text{d,k2}}$	0.14
k_{g}	0.0072		

(5.4) is to follow the experimental observations in [67, 68]. These observations show that *CLN3* gene is down-regulated in a newborn daughter cell. More specifically, Di Talia *et al.* [67] observed that the *CLN3* gene is about 3 times less expressed in daughter cells in comparison with a newly divided mother cell. This can be modeled by choosing a partitioning ratio of 25:75 for the distribution of Cln3 and Bck2 between daughter and mother cells at the division. To apply this partitioning ratio, the dynamics of Cln3 and Bck2 should be formulated by ODEs rather than algebraic equations, since this particular partition rule can be only applied to state variables, not those species in algebraic equations. That is because the abundance of species formulated by algebraic equations completely depends on other species and thus, they cannot be partitioned with individual rules.

- The second modification is that we coupled the synthesis rate of Cln3 quadratically to growth as (V^2). The reason is that to have the abundance of Cln3 to be indicative of cell growth, this protein must be synthesized at an accelerated rate in comparison with other proteins such as Cln2 or Clb5 [53]. Through many different sorts of experimental measurements [91, 103, 104, 105], it has been observed that the size control at START transition is primarily through the action of Cln3 proteins. For instance, it has been shown that the translation rate of *mCln3* with a low intrinsic initiation rate, enhances more significantly by increasing the availability of ribosomal precursors [105]. Moreover, the translation of *mCln3* is affected by specific sequence that is sensitive to cell growth rate [103]. Additionally, the *mCln3* level rises more than other mRNAs such as *mCln2* in the presence of glucose [104]. These observations support the idea that Cln3 has stronger dependencies with cell size in comparison with other

Table 5.2: Modified reactions and propensities in slow subset of the extended model.

Reaction	Reaction rate	Parameter value	Propensity
$\phi \rightarrow$ mCln3	$k_{\text{s,mcln3}}$	0.5	$k_{\text{s,mcln3}} \cdot V$
mCln3 $\rightarrow \phi$	$k_{\text{d,mcln3}}$	0.10	$k_{\text{d,mcln3}} \cdot mCln3$
$\phi \rightarrow$ mBck2	$k_{\text{s,mbck2}}$	0.7	$k_{\text{s,mbck2}} \cdot V$
mBck2 $\rightarrow \phi$	$k_{\text{d,mbck2}}$	0.14	$k_{\text{d,mBck2}} \cdot mbck2$

proteins such as Cln2. For this reason, we believe that a quadratic coupling of size and Cln3 better reflects the significant role of Cln3 in START transition. We note that $mass$ is an indicator of the cell size in Chen's model. We also alter this variable $mass$ by replacing it with a volume variable (V), which is a more sensible metric of size and can be directly compared with experimental data.

- The last modification includes the dynamics of early G1 transcripts in the subset of slow reactions. Table 2 lists the synthesis and degradation rates of the transcripts, *mCln3* and *mBck2*. To estimate the corresponding reaction rates, the degradation rate is defined first. For an mRNA denoted by *m*, the degradation rate is $k_{\text{dm}} = \ln 2/\tau_{\text{m}}$, where τ_{m} is the half-life time of transcript m. In our model, the half-life time values are in the range of 5–10 minutes, consistent with the experimental measurements of Miller *et al.* reported in [65]. The synthesis rate, k_{sm}, is also estimated such that the average abundance of the mRNA is consistent with experimental measurements: $\langle m \rangle = k_{\text{sm}}/k_{\text{dm}}$, where $\langle m \rangle$ denotes the average abundance of the corresponding mRNA, m.

We apply the HR hybrid stochastic algorithm [101] to the modified model with the listed parameters in Tables 5.1 and 5.2.

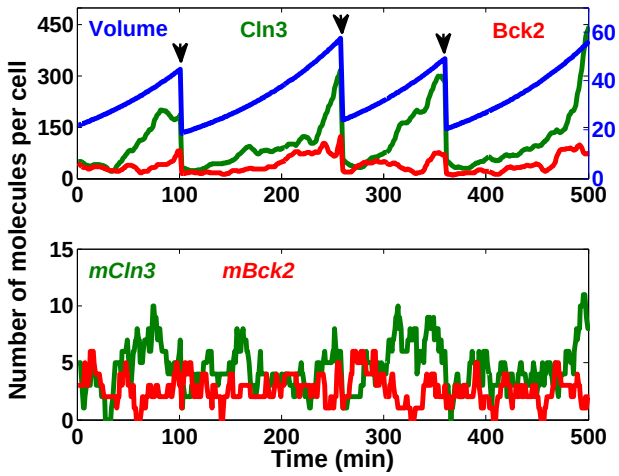

Figure 5.2: **Oscillatory dynamics of proteins and mRNAs in the Cln3-Bck2 module.** The temporal dynamics of specific proteins (A) and mRNAs (B) generated by an independent run of the hybrid stochastic model are demonstrated base on the numbers of molecules per cell. The volume of the cell (solid blue line) grows exponentially and divides asymmetrically (the division is indicated by arrows) between daughter and mother cells. This asymmetric division follows the experimental observations by Di Talia *et al.* [67, 12]. Except for Cln3 that divides with a ratio of $25:75$ between daughter and mother, the rest of proteins along with the volume are partitioned with a ratio of $40:60$. In simulation this is implemented by setting the initial values of the progenies for the next simulation run according to the aforementioned ratios.

5.3 Results

The modified hybrid stochastic model is used to generate sufficiently large populations of daughter and mother cells (at least $5,000$ cells in each independent simulation run) starting from one cell at $t = 0$. Figure 5.2 shows the oscillatory dynamics of the protein and mRNA molecules in the modified Cln3-Bck2 module. The cell volume grows exponentially with time and divides asymmetrically between the daughter and mother cells. As the cell grows in size, the levels of Cln3 and Bck2 proteins rise. The average abundance of the proteins and mRNAs are computed from the simulation results. The Cln3 protein, with an average abundance of 186 molecules per cell, matches very well with experimental observations (216 molecules per cell) [66]. The unregulated mRNAs, *mCln3* and *mBck2*, are constitutively expressed (see Fig. 5.2 bottom panel) with an average abundance of about 5, which is within the experimental range of 5-10 copies of unique

mRNAs per cell. First, we show that gene expression noise introduces significant variability to the cell cycle. One way to look into this is by increasing the ploidy, because ploidy increases the average number of transcripts. Thus, if the source of noise stems from transcripts, the ploidy will reduce the variability. Experimental studies show that when the ploidy level is doubled, the abundance of all cellular components, as well as the volume of the cell, are doubled while the concentration of species remains the same. In our simulations, to generate populations of diploid and tetraploid wild-types, the synthesis rates of all transcripts are estimated such that the average numbers of transcripts are, respectively, twice and four times larger than haploid wild-type cells. Moreover, the average volume of the cell is increased accordingly. To quantify the variability the coefficient of variation (CV = standard deviation/mean) is computed for each population.

Figure 5.3(A-L) shows the histogram of the G1 time, generated from simulations of sufficiently large populations of haploids, diploids, and tetraploids for both daughter and mother cells. Comparing the covariances (CVs) of daughter cells (on left) and mother cells (on right), we notice that the G1 variability is reduced from top to bottom in both daughter and mother cells. Thus, we can infer that the noise in gene expression level induces substantial variability to the cell cycle system.

Next, we investigate the effect of the noise introduced by G1 cyclin genes *CLN3* and *CLN2* along with *BCK2* by increasing the dosage of the corresponding transcripts. To this end, in simulations, the synthesis rates of the transcripts are estimated such that their average abundance is enlarged by 4 folds. Figure 5.3(G-L) demonstrates the histogram of the G1 time for the cell population with an increased dosage of genes. The decreasing pattern of CVs in comparison with haploid wild-type cells is evident for both daughter and mother cells. Thus, the results of our model are consistent with the hypothesis that the low copy number of transcripts results in substantial variability in the cell cycle. In particular, in the G1 phase, this variability is more considerable due to the limited number of species.

This analysis shows how the noise reduction in level of gene expression results in reduction

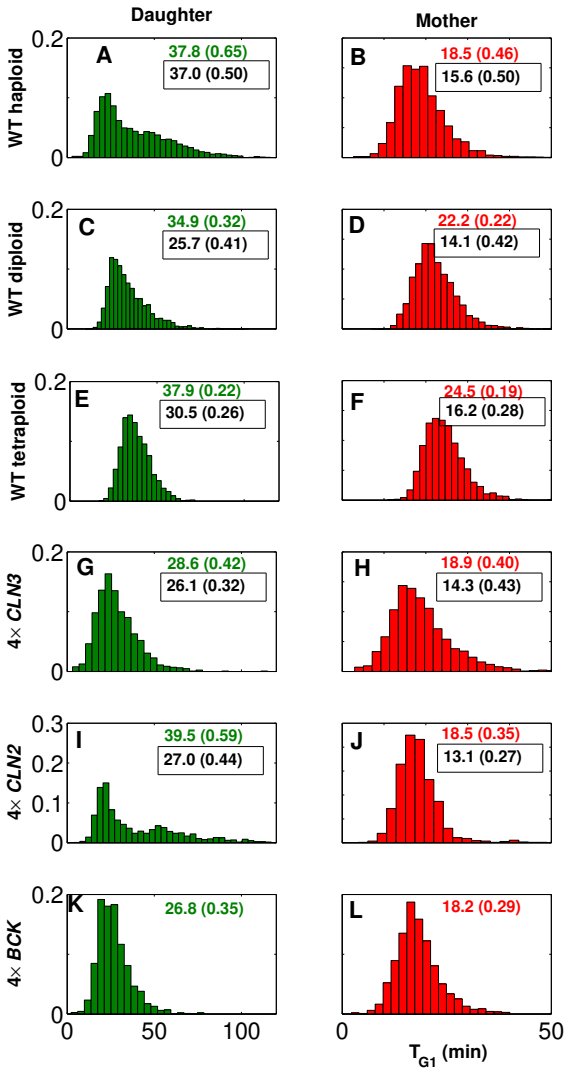

Figure 5.3: **Increasing the ploidy or the G1 cyclin transcripts decreases the variabilities in G1 time.** The histograms of G1 time duration for daughter and mother cells are generated by simulation of a sufficiently large population of cells in different simulation scenarios. The summary statistics are given for quantitative comparison. The mean (CV) of simulation is compared with experimental data (in box). (A , B): WT haploid; (B , C): WT Diploids; (E , F): WT tetraploid, (G , H): 4 $\times$*CLN3*; (I , J): 4 $\times$*CLN2*; (K , L): 4 $\times$*Bck2*; left panels: daughters; right panels: mothers.

in variability of the G1 time. Next, we investigate the effect of size control on such variability. Deterministic size control has long been proposed as a mechanism that regulates the G1 time [93, 95]. This means that cells stay in the G1 phase until they reach a critical size. Such deterministic size control ensures that all cells bud at the same size. Since the cell size at birth is variable, the size control would assure smaller cells stay long enough to reach the critical size. This introduces variability in G1 duration. To quantify such variability in experiment, instead of geometric volume estimation which is less reliable, Di Talia *et al.* related size at bud, V_{bud}, to size at birth, V_{birth}, through the amount of time the cell spends in the G1 phase (T_{G1}) by:

$$V_{\text{bud}} = V_{\text{birth}} \mathrm{e}^{k_{\text{g}} T_{\text{G1}}}, \tag{5.5}$$

where k_{g} is the growth rate of the media. We notice that the underlying assumption of (5.5) is that the population growth is exponential, which is consistent with experimental observations for the budding yeast [12]. Equation (5.5) leads to $k_{\text{g}} T_{\text{G1}} = \ln(V_{\text{bud}}) - \ln(V_{\text{birth}})$, where $k_{\text{g}} T_{\text{G1}}$ can be used as an indicator of the volume change during the G1 phase. In simulations, however, change in volume of cells from birth to bud can be directly recorded by defining specific flags. Therefore, we use the volume change in our simulations instead of $k_{\text{g}} T_{\text{G1}}$.

Figure 5.4(A-B) shows the correlation between the relative change in volume during the G1 phase and the scaled volume at birth for mother and daughter cells in glucose. The growth of mother cells in the G1 phase is almost independent of the size at birth. There is a very good match between the experimental observations (slope = -0.1) and our model (slope = -0.13). For daughter cells, however, strong size control is evident (see Fig. 5.4B). Small daughter cells need to stay longer and grow larger in the G1 phase before they go through START transition and commit to a new cycle. The size control is stronger in smaller daughters (slope = -0.97) compared with the larger ones (slope = -0.2). The quantified correlations are, respectively, comparable with (slope = -0.7) and (slope = -0.3) from experiments [12].

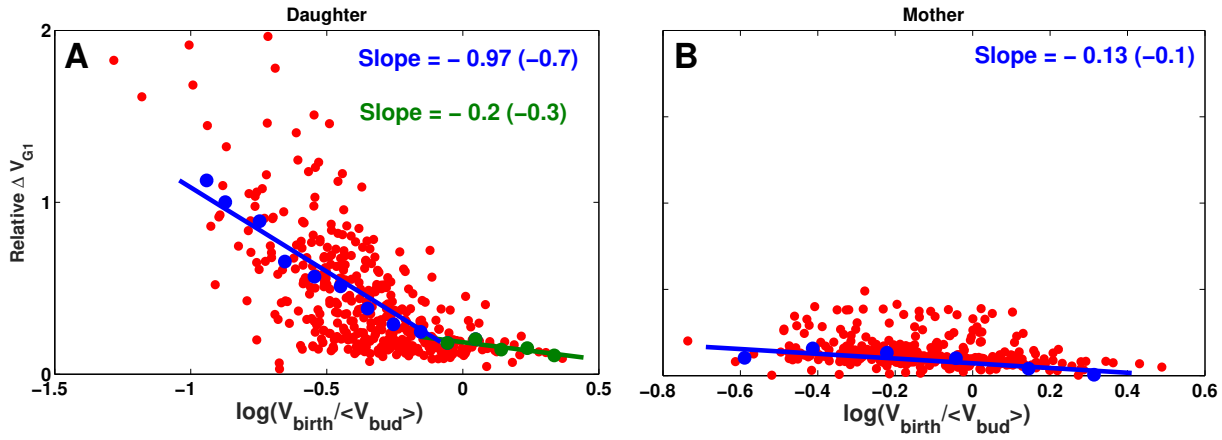

Figure 5.4: **Cell cycle variability and size control for both daughter and mother cell.** Size control leads to a negative correlation between cell size at birth and growth in the G1 phase. Plotting the logarithm of size at birth, V_{birth}, scaled with the average volume at bud, $\langle V_{\text{bud}} \rangle$, with respect to relative growth in G1 $((V_{\text{bud}} - V_{\text{birth}})/V_{\text{birth}})$ quantifies how strong the G1 size control is in daughter **(A)** and mother **(B)** cells. A slope of -1 would indicate a perfect size control, whereas a slope of 0 indicates a lack of size control. Wild-type budding yeast shows imperfect size control in glucose.

This analysis quantitatively describes two causes for G1 variability: 1) variability that is introduced by size control and 2) variability that is independent of size and stems from molecular noise [12]. The simulation results of cells with increased copies of genes provide more quantitative evidence in this regard. The efficient size control observed in wild-type daughters (see fig.5.4A) is reduced by ploidy and increasing the dosage of G1 cyclins. Figure 5.5 shows the correlation between relative volume growth during the G1 phase and the scaled volume at birth. The linear fit to the binned data demonstrates a smaller slop in diploid and tetraploid daughters in comparison with haploid wild-type in fig.5.4A. Similarly, the less efficient size control is evident in daughter cells with increased dosage of *CLN3*, *BCK2*, and *CLN2*. In addition, panels E, G, and I of fig. 5.5 show that increasing the number of *CLN3* and *BCK2* copies substantially alters the size control while the increased dosage of *CLN2* does not. The lack of size control in mother cells is also evident in right-hand side panels in fig. 5.5. Our model describes the stochastic size control of the budding yeast cell cycle very well and is quantitatively consistent with experimental observations by Di Talia *et al.* [12]

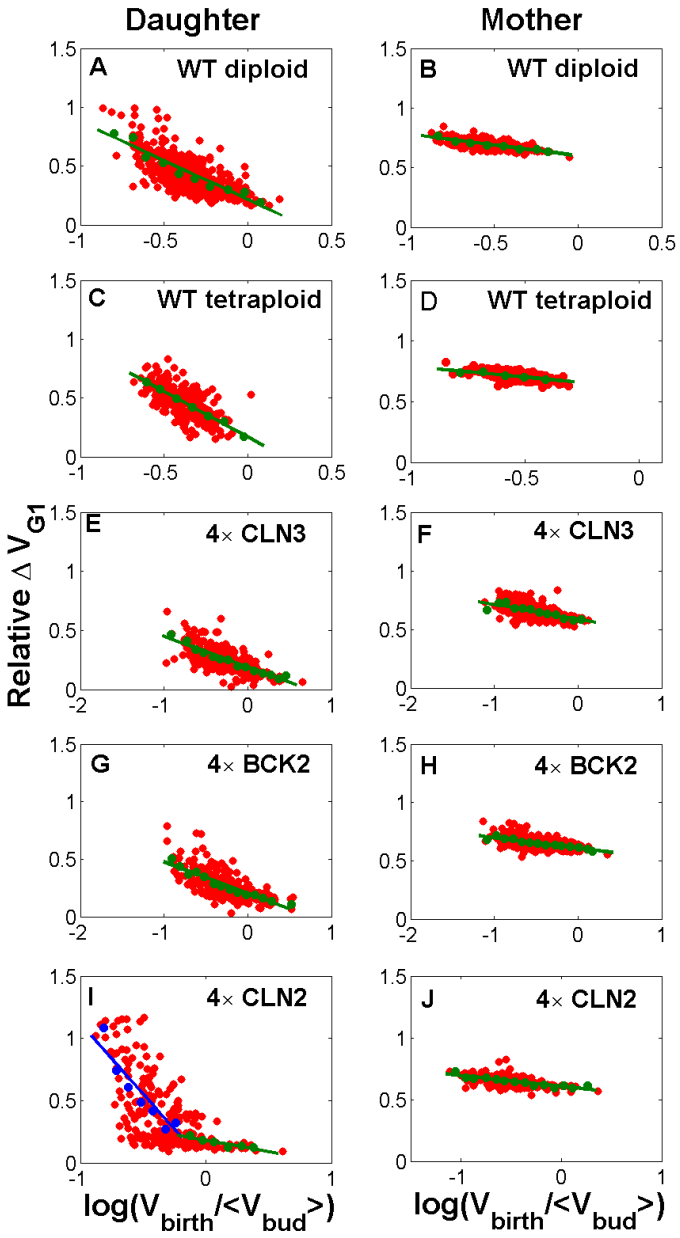

Figure 5.5: **Size control variability is reduced by ploidy and increasing the copy numbers of G1 cyclins in both mother and daughter cells.** The correlation between the relative change in volume of a cell during the G1 phase (RelativeΔV_{G1}) and the normalized logarithm of cell size at birth ($\log(V_{\mathrm{birth}}/ < V_{\mathrm{bud}} >)$ shows that the cell size independent noise is decreased by ploidy and by increasing the number of copies of G1 cyclin as well as BCK2.

Thus far, using different analysis we have shown that our hybrid stochastic model captures many properties of the cell cycle and matches very well with experimental data. Next, we use the proposed model to predict the variability in phenotypes of mutant strains in START network, more specifically the genes that play roles in size control. The average phenotypes of mutants have been modeled deterministically first by Chen *et al.* in [60, 9] and recently by Kraikivski *et al.* in [10]. However, the variability of cell cycle properties in different mutant strains needs stochastic modeling and yet is not addressed by prior works.

Figure 5.6(A-B) shows the results from hybrid stochastic simulation for some of the viable mutants of START network. Important properties of the cell cycle including division time (T_{div}), G1 time (T_{G1}), volume at birth (V_{birth}), and S/G2/M duration ($T_{\text{S/G2/M}}$) are computed for the population of the mutants. According to results in Figs. 5.6A and 5.6B for both daughter and mother cells, the average properties of the cell cycle predicted by our proposed model is consistent with the experimental observations reported in [9]. For instance, for those strains with deleted G1 cyclin genes such as $cln3\Delta$ and $cln2\Delta$, the G1 time is significantly longer than the wild-type cell. That is because the budding is delayed due to the deletion of these genes. For a similar reason, the G1 phase is prolonged in $bck2\Delta$ and $cln2\Delta bck2\Delta$. Moreover, since these strains stay longer in the G1 phase, they grow larger before they commence a new cycle. For this reason, the size of these mutants after division (at birth) is larger on average in comparison with wild-type cells [9, 60]. $GAL-CLN3$ and multicopy$BCK2$ with increased activity of *CLN3* and *BCK2* produce cells with smaller size. Such phenotype of these mutant strains is precisely described by our model.

Figure 5.6(C-D) shows the percent change in variability of cell cycle properties under perturbation of *CLN3*, *CLN2*, and *BCK2* genes compared with wild-type cell. According to the prediction of the model, the deletion of *CLN3* gene induces the highest amount of noise into almost all properties of the cell cycle, particularly to the G1 phase. The G1 time is highly variable because Cln3 protein is the cyclin that contributes the most in the G1 phase (by activation of SBF which is the

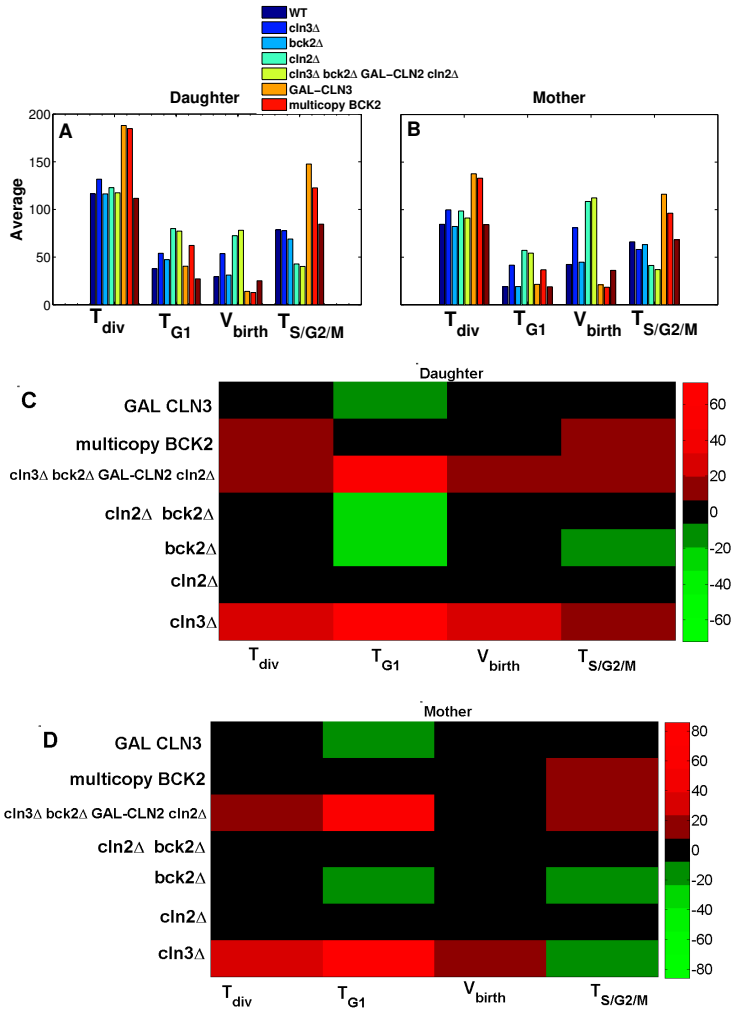

Figure 5.6: **Mutant phenotypes of selected strains in START network.** The average (A-B) and CV (C-D) of critical properties of cell cycle (including cycle time (T_{div}), G1 time T_{G1}, duration of S/G2/M phase, ($T_{\text{S/G2/M}}$) and size at birth (V_{birth}) are computed for simulated population of selected mutant strains for both daughter and mother cell. Time is in min and volume in fL.

transcription factor of Cln2). Hence, in the absence of *CLN3* gene, the START transition becomes sloppy. Another observation is that *cln3* deletion induces more variability to the mother cell than the daughter cell. This can be explained by the asymmetric division of Cln3 and *mCln3* in budding yeast. Considering the experimental observations in [67], we partition both Cln3 and *mCln3* with ratio of 25:75 between daughter and mother cells. Thus, the absence of Cln3 protein has a more considerable effect on the variability of mother cells.

Bck2 is a back-up protein in the G1 phase. In the absence of this protein, Cln3 is able to drive the cell to START transition and hence, the deletion of the *bck2* does not greatly affect the properties of the cell cycle. However, increasing the dosage of this gene in multicopy$BCK2$ significantly decreases the variability of the G1 phase. That is because multiple copies of the gene decrease the fluctuation in the level of gene expression, and consequently, in G1 duration.

Deleting *CLN2* gene in $cln2\Delta$ strain does not considerably change the variability in cell size or other phases of the cell cycle, except for the G1 phase of the daughter cell. This phenotype is not consistent with our expectation and under current parameterization, this observation requires more investigation. *CLN1* and *CLN2* genes have overlapping functions in formation of bud and spindle pole body duplication [60]. Due to the similarity of Cln1 and Cln2 proteins, the combined activity of these proteins is typically presented by only Cln2 to simplify the model. One possible solution to study the dynamics of this family of cyclins in more detail is to include the dynamics of Cln1 and *mCln1* in the model.

Experimental studies show that the noisiest phase of the budding yeast cell cycle is G1 with a coefficient of variation equal to 50% for both daughter and mother cells. Thus, one may infer that the variability in other phases (S/G2/M) is dictated by the G1 variability. However, Debashis *et al.* in [54] showed that the variability in cell cycle times and volume at birth are driven by the variability of the S/G2/M phase (bud phase), rather than the G1 (unbud) phase. Next, we study the correlation between variability in division time and size at birth with respect to variability in the

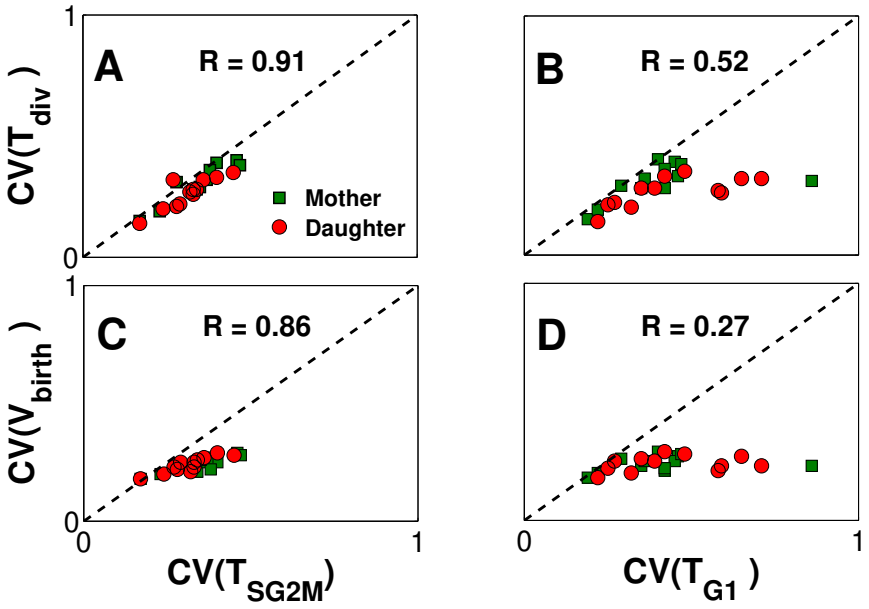

Figure 5.7: **Correlation between the Variability of size and duration of different phases of cell cycle.** CVs of T_{div} **(a and b)** and V_{birth} **(c and d)** are correlated with CVs of $T_{\text{S/G2/M}}$ and T_{G1} for simulated population of haploids, diploids, tetraploids as well as several viable mutant strains. Correlation coefficients (R) is computed and given on each panel.

duration of the G1 and S/G2/M phases. To this end, the CVs of G1, S/G2/M, and division time along with the cell volume at birth are computed from simulation of the following populations: wild-type haploids, diploids, and tetraploids in glucose, mutant strains presented in Fig. 5.6, and the cells with increased dosage of genes presented in Fig. 5.7. Moreover, the product-moment correlation coefficient is computed to quantify such correlations.

Figure 5.7(A-D) demonstrates that variability in cycle time and volume with respect to variability in the duration of the G1 and S/G2/M phases. There is a strong correlation (R = 0.91) between the CVs of division time and S/G2/M duration (see Fig. 5.7A), while such correlation is less evident for G1 duration with a coefficient of correlation equal to 0.52 (see Fig. 5.7B). Similarly, Fig. 5.7(C-D) shows that CVs of birth size are more strongly correlated to the less noisy phase of the cell cycle (S/G2/M), rather than the G1 phase. The reason for such non-intuitive observation lies in the size control mechanism. According to the results in Fig. 5.4, size control imposes a negative

correlation between the size at birth and G1 duration to make the cells bud at the same size such that the variability in size is minimized before START transition [54]. Hence, major variability of the cell size from birth to division takes place after START and during the S/G2/M phase.

5.4 Summary

In this chapter, we have presented a new hybrid stochastic (ODE/SSA) model to address the variability in the budding yeast cell cycle introduced by noise in the level of gene expression and stochastic size control. We have significantly modified existing works by adding a modified Cln3-Bck2 module. This module is required to study the G1 variability due to the importance of Cln3 and Bck2 proteins at the early G1 phase. We have applied the HR hybrid method in our developed model to maintain a good trade-off between accuracy and efficiency by integrating deterministic and stochastic simulations. We note that the full stochastic simulation is computationally expensive to be used in our large model that includes 66 proteins and mRNAs. To incorporate the inherent molecular fluctuations of the cell cycle in our model, we have enhanced the protein-protein regulatory module of Chen's deterministic model to a gene-protein regulatory network, where *mCln3* and *mBck2* are incorporated.

We validate our model though comprehensive numerical simulations and present several comparisons with wet-lab experimental data. With manageable computational complexity, our approach has successfully accounted for a broad range of single-cell experimental observations on wild-type and mutant cells. In particular, our model has confirmed that a substantial amount of deleterious noise in the cell cycle stems from the low copy number of mRNAs in the early G1 phase. The imperfect sizer of the budding yeast control mechanism has also been predicted by the proposed model. Moreover, the effect of size control on variability of certain mutant phenotypes in Cln3-Bck2 module have been quantitatively described. The proposed model has yielded promising

preliminary results that can be used to build more comprehensive models of size control regulatory network.

Chapter 6

Parameter Estimation for Stochastic Models of Biochemical Systems; A Machine Learning Approach

6.1 Introduction

Quantitative computational models are increasingly used in systems biology to provide a system-level understanding of the complex dynamics of cellular systems. The accuracy and reliability of such quantitative models depend on how well their parameters are determined. Therefore, parameter identification is an important task towards the development of accurate models that can be reliably used for developing and testing hypotheses and ultimately for prediction and refinement of experimental design. A quantitative model at the molecular level consists of many model variables and parameters. Typically the model variables represent either concentration or population of macromolecules (i.e., mRNAs and proteins). The parameters can be categorized into the following two classes. The first class includes parameters that have physical meaning such as reaction rates,

and synthesis and degradation constants. The second class that does not directly have a biochemical interpretation includes the parameters that are added to the model as a result of approximations or reductions [106]. The exponents and constant rates in Hill equation [107] are good examples of parameters in the second class. The Hill equation is used to approximate the quantification of ultrasensitivity in cellular systems [107].

Most parameters in both classes are unknown. Although some parameters in the first class can be measured in wet-lab experiments, such measurements are costly. Moreover, the parameters in the second class are typically unmeasurable through experiments. Therefore, instead of measuring the parameters through experiments, it is more feasible to measure the model variables (i.e., observables such as the abundance of mRNAs and proteins) and estimate the parameters such that the model variables fit the observables measured in experiments. Such indirect determination of unknown parameters has been widely used in systems biology [108, 109, 110, 111, 30].

Various approaches have been adopted to estimate the parameters of quantitative models in systems biology. One common approach is based on numerical optimization methods [112, 113, 114, 115, 29]. Linear and non-linear least squares fitting [112], evolutionary algorithms [113, 114], simulated annealing [115] are examples of optimization methods. Mendes *et al.* [29] surveys a number of these optimization-based methods for estimating the parameters of biochemical pathways. Most of these optimization algorithms start from an initial random point and aim to iteratively minimize an objective function that represents the difference between the output of the model and the experimental measurements. The main drawback of these methods is that they are mostly based on extensive search over a large parameter space, and hence, they are computationally expensive. Moreover, optimization methods often result in a single best parameter vector that corresponds to a (sometimes local) minimum of the objective function obtained in search (that is not necessarily the global minimum that exists), and thus, they may not perform well if measurements obtained from wet-lab experiments are substantially noisy [106]. Another alternative approach is

the probabilistic Bayesian inference [116] that quantifies the uncertainties in both experimental data and the intrinsic noise that rises from the low copy number of transcripts. The main advantage of Bayesian inference is that this method results in a probability distribution for the parameters rather than a single best parameter vector. The main issue of Bayesian inference is that the computational cost of calculating the multidimensional integration problem is high [106]. Several Bayesian-based methods have been proposed that do not require the computation of the likelihood probability [117, 118, 119, 120, 121]. Kalman filter and its variants which have originally been proposed for state estimation have also been extended to solve the problem of parameter estimation [122, 123, 124].

Most of the existing efforts to solve the problem of parameter estimation have been focused on deterministic models. Due to the recent advancements in experimental techniques, cell biologists can perform wet-lab measurements at a single-cell level, which shows substantial cell-to-cell variability between individual cells. To capture such variability, stochastic models are increasingly developed [32, 53, 54, 51, 102, 125]. Parameter estimation for such stochastic models is considerably more difficult, because the relationship between the model output and the parameters that produce those outputs is highly stochastic. Thus far, limited work exists, such as the approximate maximum likelihood method [126] and Quasi-Newton stochastic optimization [127, 128, 129], to solve the problem of parameter estimation in a stochastic setting.

In this study, we propose a novel approach, referred to as *hierarchical deep classification (HDC)*, to estimate the parameters of stochastic models of biochemical systems. In contrast with conventional optimization-based methods that focus on the point estimation of the unknown parameters, the proposed HDC scheme aims to find an acceptable range for each parameter. The HDC develops an ML-based model-free approach that can capture the highly stochastic and complex relationship between the input trajectories and parameter ranges. Therefore, the proposed method provides an acceptable parameter region rather than returning a single set of best parameters. Since we do not

have sufficiently large data from the experiment to train such an ML-based model, we use the forward model (i.e., the stochastic model) to generate sufficient synthesized training data. This data will be used together with the experimental data to train and test the ML model. We note that our proposed approach falls within the category of *theory-guided machine learning* [130] since we are using the theory-based forward model to train the ML model to solve the inverse problem.

This chapter is organized as follows. Section 6.2 presents the proposed hierarchical classification method and illustrates how to use this approach to find an acceptable parameter ranges for a stochastic model of a biochemical system. Section 6.3 presents the proposed HDC algorithm. Section 6.4 illustrates the implementation of the HDC algorithm through a simple example. Section 6.5 concludes Chapter 6 with a discussion on the capabilities of the proposed algorithm in the field of systems biology.

6.2 Method

So far, in Chapters 3-5, we have focused on the problem of developing and analyzing the hybrid stochastic models that generate the trajectories of state variables (i.e., proteins and mRNAs) from a known parameter vector. We refer to this problem as the *forward problem*. In this chapter, our goal is to solve the *inverse problem* for identifying a parameter vector that could drive the forward model to reproduce a given set of experimental observations. Figure 6.1 shows the block diagram of the proposed approach. The output of the proposed scheme is a *range* for model parameters rather than a single set of best values. That is particularly desirable because "sloppiness" is a known characteristic of the parameters in systems biology [131]; meaning that many quantitative models in systems biology have parameter values that are distributed possibly over several orders of magnitudes [106]. Hence, given the intrinsic fluctuations in systems, uncertainty in experimental measurements, and the sloppy model parameters, reaching an ensemble of parameter sets or ranges

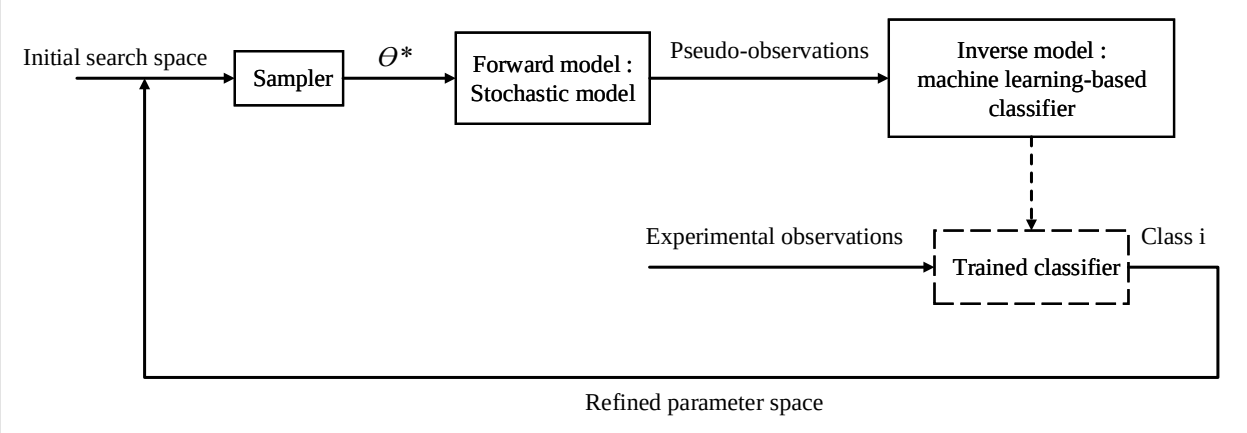

Figure 6.1: **Block diagram of the proposed hierarchical deep classification method.** Given $\boldsymbol{\theta}^*$, the vector of parameters we aim to estimate, we generate pseudo-observations using the forward model. This is done by uniformly sampling from an initial search space $\mathcal{S}$. Dividing the search space of each parameter by two, we define each subspace as a class and label the pseudo-observations by the class of the corresponding parameter. The labeled data (the pairs of pseudo-observations and the class of their corresponding $\boldsymbol{\theta}^*$) are then used to train the inverse model. We design the inverse model as a classifier with the goal of finding the class of experimental data to which our forward model is desired to fit. Once the class of the experimental data is determined (if needed) we further divide the selected class into new classes, and repeat the process until we reach the desired resolution for the range of model parameters.

for the parameter vector results in a more robust predictive model [131].

As shown in Fig. 6.1, the forward model is in fact the stochastic model with a parameter vector $\boldsymbol{\theta}$. Our goal is to estimate acceptable ranges for the parameter vector $\boldsymbol{\theta}^* \subseteq \boldsymbol{\theta}$. In our proposed method, the search starts with an arguably large search space $\mathcal{S} \subset \mathbb{R}^n$, where $n = |\boldsymbol{\theta}^*|$ is the cardinality of the set $\boldsymbol{\theta}^*$. We divide the search space $\mathcal{S}$ into 2^n subspaces $\mathcal{S}_i, i = 1, \cdots, n$, where $\bigcup_i \mathcal{S}_i = \mathcal{S}$ and $\mathcal{S}_i \cap \mathcal{S}_j = \varnothing$ for any $i, j = 1, \cdots, n$ and $i \neq j$. Figure 6.2 shows the subspaces for a forward model with $n = 2$. Hereinafter, we refer to each subspace as a class. In fact, we treat the parameter identification (i.e., the inverse problem) as a *classification problem* in which we aim to map the input trajectories (or any of their statistical representation) into a specific class. In Sec.6.3, we will develop a new hierarchical algorithm that solves the inverse problem iteratively to reduce the search space and achieve a refined range for the parameter vector.

To solve the inverse problem, we repeatedly take K samples $\boldsymbol{\theta}^*_{k,i}, k = 1, \cdots, K$, from each class

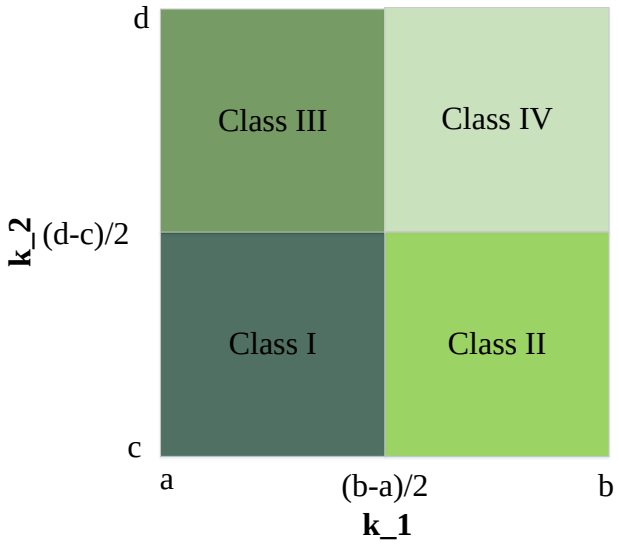

Figure 6.2: **Defining the classes for a given search space.** Four classes are defined, given $\boldsymbol{\theta}^* = [k_1, k_2]$, $k_1 \in [a, b]$, and $k_2 \in [c, d]$.

$\mathcal{S}_i$ and use the sampled parameters to generate pseudo-observations via the forward (stochastic) model. As shown in Fig. 6.1, the generated pseudo-observations along with the corresponding class labels will then be fed to the ML-based inverse model as the synthetic training data. To test the trained classifier, we use the experimental data that the forward aims to closely reproduce. The outcome of the test (i.e., the class that the experimental data is classified into) will serve as the new search space in the next round of the iterative algorithm (see Sec. 6.3 for further details). This process is repeated until a stopping condition is met. The class of experimental data in the final iteration will be reported as the acceptable parameter range for the parameter vector.

6.3 Hierarchical deep classification algorithm

To implement the systematic approach proposed in the previous section, we develop a novel procedure in Algorithm 2 which we refer to as HDC. While the proposed approach in Sec. 6.2 is not limited to a specific type of classifier, here, we focus on employing an artificial neural net-

Algorithm 2 Hierarchical Deep Classification

Input: Labeled training data, test data (experimental data), initial classes: $\mathcal{S}_i \subset \mathbb{R}^n$, initial distance matrix $\mathbf{D}_{\text{JS}}$.

1: **while** $\exists (i,j) \in \mathbf{D}_{\text{JS}}$ such that $d_{\text{JS}}^{ij} > \alpha$ **do**
2: Train the ANN model using the labeled training data.
3: Test the trained ANN using the experimental data to find the corresponding class $\mathcal{S}^*$.
4: Update the subspaces by decomposing the search space of the selected class $\mathcal{S}^*$ into n new subspaces (as described in Sec. 6.2).
5: Generate new labeled data by repeatedly sampling parameters from the new search space.
6: Recalculate the distance matrix D_{JS}.
7: **end while**

Output: Selected class $\mathcal{S}^*$ of the experimental data.

work (ANN) to perform the classification. One of the key reasons for using ANN is that this model-free approach is widely used as a universal function approximator [132]. In fact, ANN is a powerful method to establish the nonlinear complex relationship between the input trajectories (or their statistical representations) and the output parameter ranges. Next, we explain the steps of the proposed HDC method in Algorithm 2.

The input of the HDC algorithm is the synthetic training data generated by the forward stochastic model, as elaborated in the previous section. Here, we use the SSA as the forward model. We note that other forward models can also be used in the HDC algorithm. The generated trajectories are then sampled by time window, τ. In Sec. 6.4, we will discuss how we can select this time window. Since SSA is a Monte-Carlo based method, one trajectory is not sufficient to represent the system. Thus, we generate a large set of trajectories, or one extended trajectory for a long period of time, and extract the probability transition matrix for the state variables (see Sec.6.4 for further details). We label each transition probability matrix with the class of its corresponding parameters. With these labeled data, we train the ANN that can learn the relationship between the transition matrix (as the input) and the class of parameters (as the output). Then we test the trained ANN with the experimental observations we are trying to fit to, in our forward SSA model. The output of this test is a class that will be used as the refined search space in the next iteration of the HDC

algorithm. To determine the stopping condition, we calculate the $n \times n$ distance matrix $\mathbf{D}_{\text{JS}}$ with (i, j)-th element, d_{JS}^{ij}, equal to the Jensen–Shannon divergence [133] between classes i and j, for $i, j = 1, \cdots, n$. We note that other forms of distances can also be used. If such distance is greater than a predefined threshold α, we can infer that different classes are generating indistinguishable transition probability matrices and the algorithm must continue. While the choice of α may seem to be arbitrary, we will show that it is rather a simple task.

The next section utilizes the HDC algorithm in order to estimate the parameter vectors of two chemically reacting systems taken from the field of systems biology.

6.4 Results

To demonstrate the merits of the proposed HDC algorithm, we apply it to an example of a mono-molecular reaction system. We select this simple reaction system since we are able to analytically derive the possible states and the relationship between the parameters and the transition probability matrix. This way, we can have a baseline to compare with the outcome of the proposed HDC algorithm. Next, we first derive an analytical representation of the parameter vector for the mono-molecular reaction system. Using the results of these analyses, synthetic experimental data will be generated and will be used as test data in the HDC scheme.

6.4.1 Analytical derivation of the parameter vector

Consider the following simple mono-molecular reaction system described by

$$\mathrm{X} \underset{k_{\mathrm{b}}}{\overset{k_{\mathrm{f}}}{\rightleftharpoons}} \mathrm{Y}, \tag{6.1}$$

where X and Y are the reacting molecules. Moreover, k_{f} and k_{b} are, respectively, the forward and backward reaction rates. Our goal is to estimate $\boldsymbol{\theta}^* = [k_{\mathrm{f}}, k_{\mathrm{b}}]$ by using the data that comes from multiple measurements of X. We further assume that the collected data is just the population of X measured at fixed time intervals: $t_i = t_0 + i\tau$, where τ is a fixed time window that is used to conduct the measurements.

We begin with a deterministic system described by the following ODE for the population (or concentration) of X:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = -k_{\mathrm{f}} \cdot X + k_{\mathrm{b}} \cdot Y. \tag{6.2}$$

For this system, there is a conservation law such that $X + Y = c$, where c is a constant and does not change under the reversible reactions. Thus, (6.2) can be rewritten as

$$\frac{\mathrm{d}X}{\mathrm{d}t} = -k_{\mathrm{b}} \cdot c - (k_{\mathrm{f}} + k_{\mathrm{b}}) \cdot X. \tag{6.3}$$

Solving (6.3) results in a deterministic trajectory of species X. In addition, for this simple example, k_{f} and k_{b} can be estimated using approaches such as the least squares method. However, our goal is to estimate the parameter vector $\boldsymbol{\theta}^* = [k_{\mathrm{f}}, k_{\mathrm{b}}]$ given that the system in (6.1) is stochastic and formulated under stochastic frameworks such as the SSA or its derivatives. We know that X generated by the SSA is a random realization of possible trajectories of X. That is, we aim to estimate $\boldsymbol{\theta}^*$ while taking into account that the trajectory of X is random. To simplify the analytical derivations, we consider a specific case where $c = 1$, i.e., we have only two states. In other words, there is only one particle that switches its states between X and Y.

Considering a probability representation for the system, let

$$Z_i = \mathrm{P}(\mathrm{S}_i), \quad i = 1, 2, \tag{6.4}$$

where Z_1 and Z_2 are, respectively, the probability of the states $\mathrm{S}_1 = \{X = 1, Y = 0\}$ and $\mathrm{S}_2 = \{X = 0, Y = 1\}$. Moreover, given that $X + Y = 1$, it is easy to follow that $Z_1 + Z_2 = 1$ and

$$\frac{\mathrm{d}Z_1}{\mathrm{d}t} = k_\mathrm{f} Z_1 + k_\mathrm{b} Z_2 = k_\mathrm{b} - (k_\mathrm{f} + k_\mathrm{b}) Z_1. \tag{6.5}$$

Solving (6.5), we have:

$$Z_1(\tau) = \frac{k_\mathrm{b}}{k_\mathrm{f} + k_\mathrm{b}} + e^{-(k_\mathrm{f}+k_\mathrm{b})\tau} \left[Z_1(0) - \frac{k_\mathrm{b}}{k_\mathrm{f} + k_\mathrm{b}} \right], \tag{6.6}$$

where $Z_1(0)$ is the initial probability and is either 1 or 0 depending on the initial state. Suppose that initially the system is at state S_1, that is $Z_1(0) = 1$. Then we have:

$$Z_1(\tau) = \gamma + (1 - \gamma) e^{-\lambda\tau}, \quad \text{and} \quad Z_2(\tau) = (1 - \gamma) \left(1 - e^{-\lambda\tau}\right), \tag{6.7}$$

where $\lambda = k_\mathrm{f} + k_\mathrm{b}$, $\gamma = \frac{k_\mathrm{b}}{\lambda}$. Similarly, if $Z_1(0) = 0$, we have

$$Z_1(\tau) = \gamma \left(1 - e^{-\lambda\tau}\right), \quad \text{and} \quad Z_2(\tau) = (1 - \gamma) + \gamma e^{-\lambda\tau}. \tag{6.8}$$

Moreover, given an initial state of S_1 (i.e., $Z_1(0) = 1$), one can find the probability of transitioning to the state S_2 by calculating $Z_2(\tau)$. More generally, we define a *transition probabilities* p_{ij} as $p_{ij} = \mathrm{P}(\mathrm{S}_j(\tau)|\mathrm{S}_i(0))$, i.e, the conditional probability of the state of the system be S_j after time τ, given that the initial state of the system is S_i. Based on (6.7) and (6.8), the transition matrix $\mathbf{T}$, including the transition probabilities, can be calculated as follow:

$$\mathbf{T} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} \gamma + (1 - \gamma) e^{-\lambda\tau} & (1 - \gamma)(1 - e^{-\lambda\tau}) \\ \gamma(1 - e^{-\lambda\tau}) & (1 - \gamma) + \gamma e^{-\lambda\tau} \end{bmatrix}. \tag{6.9}$$

We notice that the transition matrix can also be directly estimated from the measurement data.

Among all the measurement data, we can count the total number of cases N_1 where the system starts at state S_1. Then, after time τ, the system either stays at S_1 (we count the total number to be m_{11}) or jumps to S_2 (we count the total number to be m_{12}). We thus have $m_{11} + m_{12} = N_1$. Similarly, we can count the total number of cases where the system starts at state S_2 and denote it as N_2. Then, after time τ, the system either jumps to S_1 (we count the total number to be m_{21}) or stays at S_2 (we count the total number to be m_{22}). We also have $m_{21} + m_{22} = N_2$. All these numbers are directly collected from measurement data. Therefore, using the empirical measurements, the transition matrix can be approximated as follow:

$$\mathbf{T} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \approx \begin{bmatrix} \frac{m_{11}}{N_1} & \frac{m_{12}}{N_1} \\ \frac{m_{21}}{N_2} & \frac{m_{22}}{N_2} \end{bmatrix}. \tag{6.10}$$

Furthermore, since $p_{11} + p_{12} = p_{21} + p_{22} = 1$, we can calculate p_{12} and p_{21}, respectively, from p_{11} and p_{22}. From equations (6.9) we have

$$p_{11} + p_{22} = 1 + e^{-\lambda\tau}, \tag{6.11}$$

which provides a way to estimate λ as

$$\lambda = -\frac{1}{\tau}\ln(p_{11} + p_{22} - 1) = -\frac{1}{\tau}\ln(p_{11} - p_{21}). \tag{6.12}$$

Note that if $\lambda\tau >> 1$, then $e^{-\lambda\tau} \approx 0$. In that case, $p_{11} + p_{22} \approx 1$ or $p_{11} \approx p_{21}$. This indicates that the system is in a stochastic quasi-steady state, i.e., the state at time τ is not affected by its initial state at time 0. In that case, we cannot estimate λ, and there will be just one independent value available. Hereby, we can only estimate γ from

$$\frac{p_{12}}{p_{21}} = \frac{1}{\gamma} - 1, \tag{6.13}$$

which results in

$$\gamma = \frac{1}{1 + \frac{p_{12}}{p_{21}}} = \frac{p_{21}}{p_{12} + p_{21}}. \tag{6.14}$$

The above analysis demonstrates a way to use the measurement data for estimating system parameters of a very simple system. When a system is subject to randomness, we may use the transition probabilities and the distribution of states to estimate the parameters. In the simplest case, the solution exists and is unique as given in equations (6.12) and (6.14).

6.4.2 Simulation results

Validation of the theoretical derivations: We use the SSA for the reaction system in (6.1) to calculate the transition probability matrix $\mathbf{T}$ for the sampled trajectory of X. The calculated transition probability matrix is then plugged into (6.12) and (6.14) to calculate $\boldsymbol{\theta}^* = [k_{\mathrm{f}}, k_{\mathrm{b}}]$. To validate our analytical derivations for parameter estimation, let the actual parameter vector be $[k_{\mathrm{f}}, k_{\mathrm{b}}] = [2, 1]$ and $\tau = 0.1$. Then, using the SSA and from (6.10), the transition probability matrix can be calculated as

$$\mathbf{T} = \begin{bmatrix} 0.83 & 0.17 \\ 0.09 & 0.91 \end{bmatrix}. \tag{6.15}$$

Using the calculated transition probabilities in (6.15), and the equations in (6.12) and (6.14), the parameter vector is estimated to be $[k_{\mathrm{f}}, k_{\mathrm{b}}] = [1.94, 1.03]$. This result is quite comparable with the actual parameters that were used in the SSA.

It is worth noting that the time window τ plays a crucial role in accuracy of the result. We need to choose a sufficiently small time window so that characteristics of the transition between states are accurately captured. However, a very small time window can increase the computational or measurement costs. Therefore, there is a trade-off between the cost and accuracy that must be taken

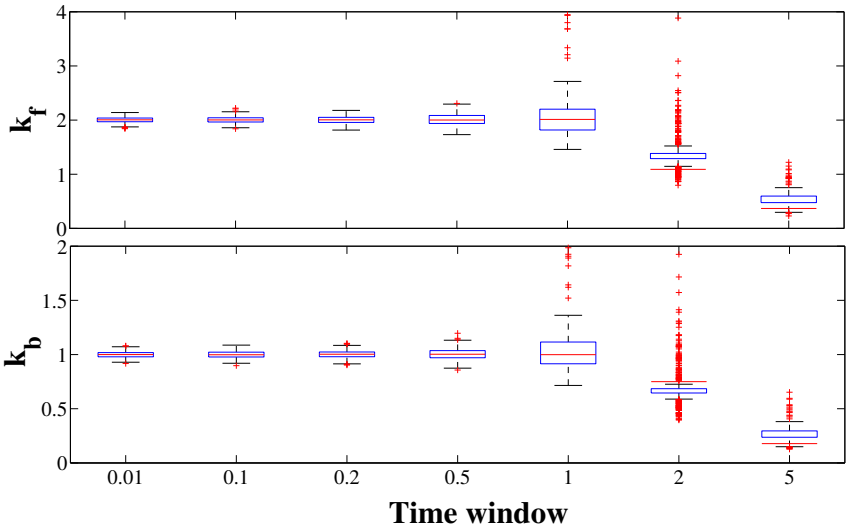

Figure 6.3: **Variability of estimated parameters for different time windows.** Given $[k_{\mathrm{f}}, k_{\mathrm{b}}] = [2, 1]$, we run 500 simulations for time windows $\tau = 0.01, 0.1, 0.2, 0.5, 1, 2, 5$ and calculate the parameters using the equation in (6.12) and (6.14) and the transition probability matrix from each SSA simulation.

into account when selecting the time window. Figure 6.3 shows the box plot of the estimated values for k_{f} and k_{b} versus different time windows τ, where the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Moreover, estimated values that are considered outliers are plotted individually using the "+" symbol. As we expect, the smallest time window $\tau = 0.01$ results in the most accurate parameter vector and the very large time windows such as $\tau = 2$ and $\tau = 5$ result in estimated values that are far from the actual parameters. To obtain a trade-off between the accuracy and cost, $\tau = 0.1$ or $\tau = 0.2$ seems to be reasonable options. We let $\tau = 0.1$ in our simulations.

Constructing, training, and testing the inverse model: Next, we use the proposed HDC algorithm to estimate the parameters of the mono-molecular system described in (6.1). To produce the inputs of the HDC algorithm, we begin with a sufficiently wide search space, let $k_{\mathrm{f}} \in [0.001, 10]$ and $k_{\mathrm{b}} \in [0.001, 10]$. Then the classes are defined according to the demonstration in Fig. 6.2 where k_{f} and k_{b} are, respectively, map to k_1 and k_2 and the SSA is used to generate 1,000 labeled data for

Table 6.1: Summary of the parameter estimation of the monomolecular system using the proposed HDC algorithm.

Iteration	**Initial search space**	**Selected class**	**Refined search space**	**Accuracy of NN**
1	k_f = [0.001 , 10] k_b = [0.001 , 10]	I	k_f = [0.001 , 5] k_b = [0.001 , 5]	97%
2	k_f = [0.001 , 5] k_b = [0.001 , 5]	I	k_f = [0.001 , 2.5] k_b = [0.001 , 2.5]	95.7%
3	k_f = [0.001 , 2.5] k_b = [0.001 , 2.5]	II	k_f = [1.25 , 2.5] k_b = [0.001 , 1.25]	93.8%
4	k_f = [1.25 , 2.5] k_b = [0.001 , 1.25]	IV	k_f = [1.89 , 2.5] k_b = [0.63 , 1.25]	80.2%
5	k_f = [1.89 , 2.5] k_b = [0.63 , 1.25]	III	k_f = [1.89 , 2.2] k_b = [0.94 , 1.25]	62.9%
6	k_f = [1.89 , 2.2] k_b = [0.94 , 1.25]	I	k_f = [1.89 , 2.04] k_b = [0.94 , 1.09]	46.2%

each class. It is worth noting that the labeled data are randomized before being used for training the ANN. Due to the lack of experimental data for the considered system, we generate a separate set of synthetic data that will be used as the experimental data to test the trained ANN. In particular, the test data is generated by randomly selecting k_{f} from [1.9,2.1] and k_{b} from [0.9,1.1] and plugging them into equations (6.12) and (6.14). The calculated experimental transition probability matrices are fed to the trained ANN as the test data. The next step is to design a neural network as the inverse model. We choose a simple structure that is a fully connected feed-forward network with 4 inputs (corresponding to the 4 elements of the transition matrix), 4 output outputs (corresponding to the labels of the 4 classes), and 25 neurons in the hidden layers as the base structure of the inverse model. The initial weights and biases are assigned randomly. We use the cross-entropy [35] as the loss function and scaled conjugate gradient backpropagation [134] as the training algorithm.

Table 6.1 summarizes the simulation results of the proposed HDC method for the first 6 itera-

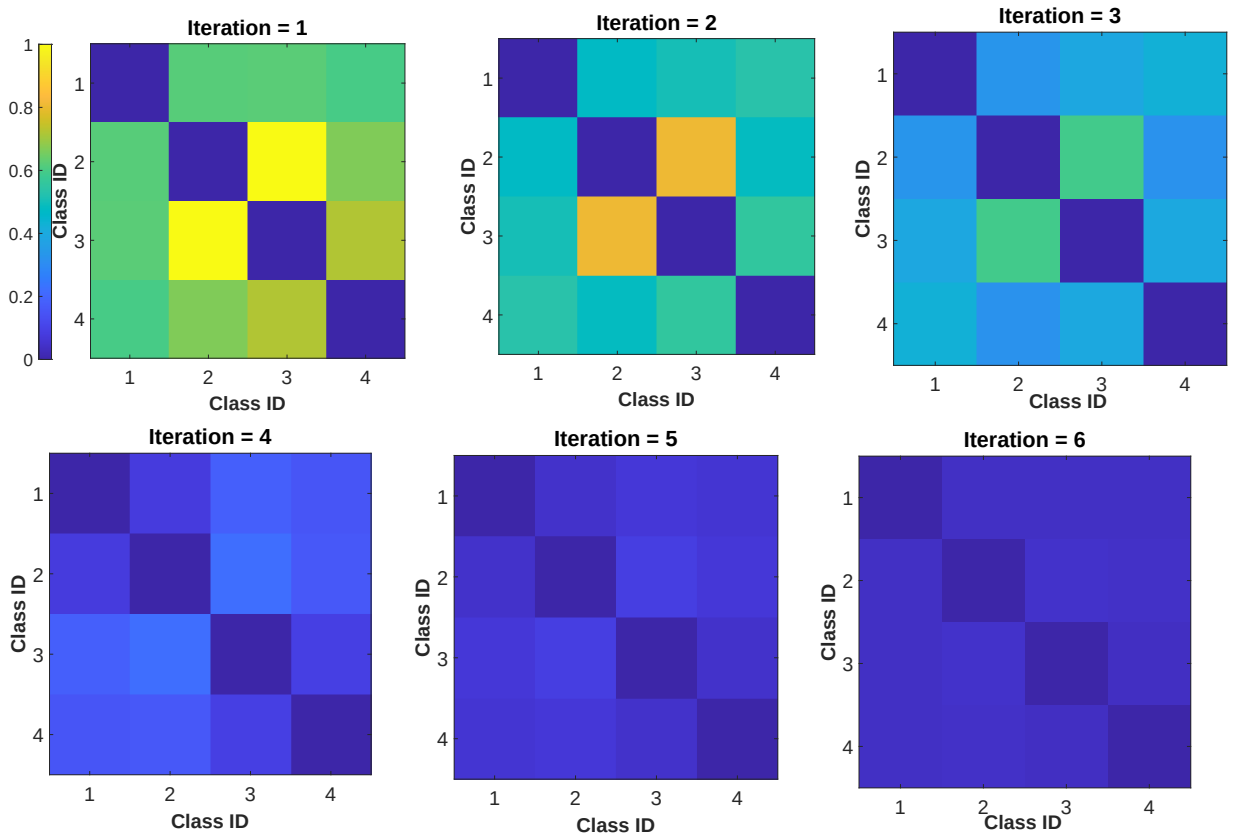

Figure 6.4: **Average distance between classes in monomolecular system.** The average of Jensen-Shannon distance between the transition martices corresponding to classes i and j (where $i, j = 1, \cdots, 4$) is computed and normalized for 1,000 simulations.

tions. In the first iteration, the trained network with 97% accuracy classifies the experimental data to class 1. Here, the accuracy is defined as the number of correctly classified data points divided by the total number of data points (including both correctly and incorrectly classified data points). In the second iteration, we reduce the search space to the selected class 1 from the first iteration. Then, the classes are redefined based on the new search space. We synthesize sufficiently large sets of training data for the new classes and retrain the ANN model. Table 6.1 shows that in the second iteration, class 1 is selected and the acceptable parameter region is updated to [0.001 , 2.5] for both k_{f} and k_{b}. According to the proposed HDC algorithm in Algorithm 2, the stopping condition is met when the normalized average distances between the input data are less than the

threshold α. We reach this condition at iteration 4 if $\alpha = 0.2$, and therefore, [1.89 , 2.5] and [0.63 , 1.25] are the estimated parameter ranges for k_{f} and k_{b}, respectively. The results show that the proposed HDC method can successfully provide parameter ranges (for both parameters) that are very close to the actual parameters ([1.9,2.1] for k_{f} and [0.9,1.1] for k_{b}) used to synthesize the so-called experimental data.

Figure 6.4 shows the normalized average distance matrix $\bar{\mathbf{D}}_{\mathrm{JS}}$ where each element (i, j) is the normalized average distance $\bar{d}^{ij}_{\mathrm{JS}}$ between the transition matrices of the classes i and j, for $i, j = 1, \cdots, 4$, in mono-molecular system. More specifically, $\bar{d}^{ij}_{\mathrm{JS}}$ is computed by averaging over 1,000 measurements of d^{ij}_{JS} where d^{ij}_{JS} is the Jensen-Shannon distance between the transition matrices constructed by the parameters that are drawn from classes i and j. To calculate the Jensen-Shannon distance between two transition matrices, $n \times n$ matrices are reshaped into $1 \times n^2$ vectors by concatenating the rows. We note that there is no distance between the two same classes and thus, the diagonal elements of $\bar{\mathbf{D}}_{\mathrm{JS}}$ are zero. In addition, $\bar{\mathbf{D}}_{\mathrm{JS}}$ is symmetric since the distance between class i and class j is the same as the distance between class j and class i. According to Fig. 6.4, in the first iteration with a large search space, the distance between different classes is large. The distance is still significant in the second iteration where the search space is reduced. However, starting from the third iteration, the distance between the transition matrices decreases. In the fourth iteration, all distances are less than $\alpha = 0.2$ which means we have reached the stopping condition. If we continue refining the search space (i.e., selecting a smaller value for α), then the synthesized data will not present any recognizable pattern to the ANN model. This means that we are labeling the same (or very similar) input data with different classes. Nonetheless, it is known that the performance of ANN models drastically declines once the same (or almost the same) input data are labeled with different classes [135]. This can also be observed from the results in Table 6.1. According to Table 6.1, the accuracy of the ANN model declines drastically after the fourth iteration.

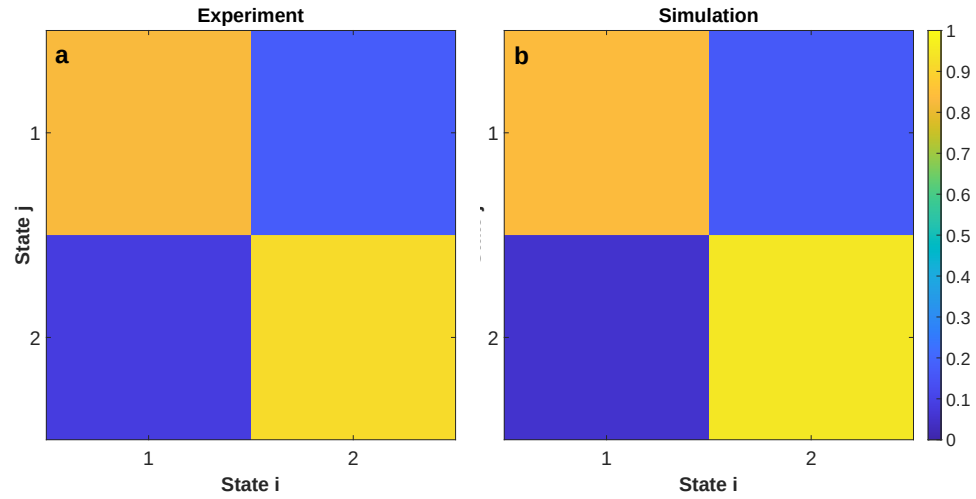

Figure 6.5: **Transition probability matrix, experiment versus simulation.** Transition probability matrix used as experimental data (on left hand side) is compared with the average transition probability matrix computed by simulation (on the right hand side). For the computed transition probability matrix we sample 1,000 sets of parameters from the estimated range for each parameter ($k_{\mathrm{f}} \in [1.89, 2.5]$ and $k_{\mathrm{b}} \in [0.63, 1.25]$). The plot on the right hand side shows the average of the 1,000 simulated matrices.

To further validate the accuracy of the proposed HDC algorithm, in Fig. 6.5, we compare the transition matrices from the experiment and simulation. To compute the experimental transition matrix, we draw 10 parameter values for k_{f}, from [1.9,2.1], and k_{b}, from [0.9,1.1]. Then, we calculate the 10 transition matrices using equations (6.12) and (6.14). The transition probability matrix on the left hand side in Fig. 6.5a is the average of the 10 computed transition matrices. Figure 6.5b shows the average transition matrix calculated based on randomly drawing 1,000 pairs of parameters from the selected parameter range $\mathcal{S}^* \subset \mathbb{R}^2$. Figure 6.5b is the average over those 1,000 computed values. The results show no significant distance between the experimental transition matrix and the one generated by the forward model using parameters from the selected parameter range by the HDC algorithm. In fact, The result shows that the HDC algorithm can successfully estimate an acceptable parameter range such that we can reconstruct the experimental data by the forward model using any combination of parameters within the estimated region.

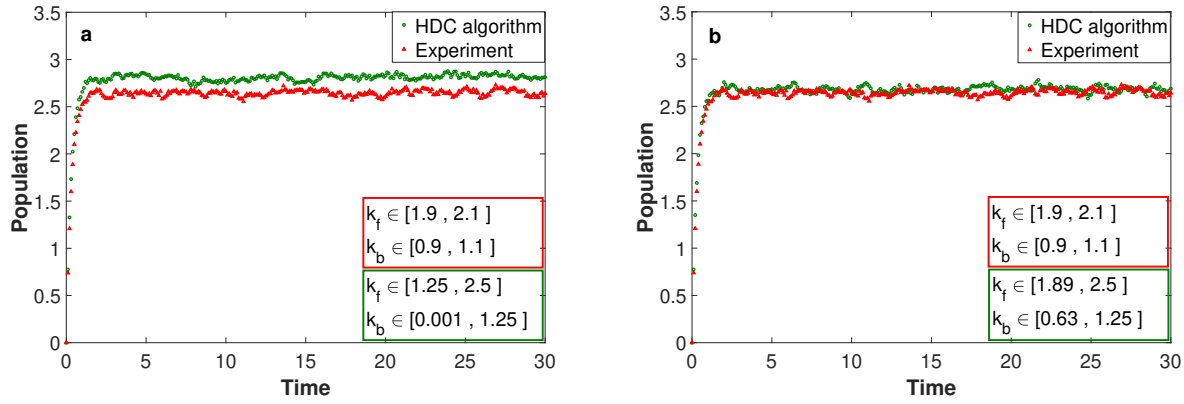

Figure 6.6: **Average population of state X using the estimated and actual parameter ranges.** The average dynamics of state variable X is plotted from both experiment and simulation. For experiment we draw the parameters from the actual ranges (in green box) and for the simulation from the estimated parameter ranges (in red box). We compared the estimated ranges by the HDC algorithm from iteration 3 (on left hand side) with iteration 4 (on right hand side). The average trajectories are computed using the results from 1,000 sets of parameters.

Extending the simulation results for $c > 1$. Previously in this section, we have considered two states (i.e., $c = 1$) in the mono-molecular system described in (6.1). As elaborated earlier, the reason for this assumption was to derive the system parameters analytically and to show that using the transition matrix, we can estimate the parameter vector. Next, we relax this assumption by extending the number of states to 5 (i.e., $c = 4$). Figure 6.6 compares the average population of molecule X from experiment with simulations. To construct the experimental trajectories, we sample 1,000 parameter vectors from the actual parameter range (displayed inside a red rectangular in Fig. 6.6) and use the SSA to generate the temporal trajectories of X. Then we average over those trajectories to demonstrate the overall behavior of the system at steady-state. A similar procedure is followed to construct the trajectories from the simulation, except only for simulations, we sample from the parameter ranges estimated by the HDC algorithm. In Fig 6.6a we use the estimated parameter range when the HDC has stopped at iteration 3 while Fig 6.6b shows the result when the HDC algorithm has refined the parameter ranges further in iteration 4. We can observe that the proposed HDC algorithm can be applied to the extended system. In addition, the results show that

extending the simulation to one more iteration leads to a more accurate result.

6.5 Discussion

We have presented a novel theory-guided ML approach to solve the inverse problem for stochastic models in systems biology. In particular, due to the lack of sufficient experimental data, we have leveraged the forward stochastic model that generates labeled data to train the inverse ML-based model. We have proved the proposed HDC algorithm useful for estimating the parameter vector in a simple stochastic model of the monomolecular reaction system. We have built the inverse model based on an ANN, a widely used universal function approximator, to establish the relationship between the complex and highly stochastic trajectories of the state variables and the corresponding parameter vector. In fact, the proposed framework confers the advantages of model-free learning together with the domain knowledge from the systems biology.

Furthermore, experimental measurements are often conducted using polymerase chain reaction (PCR), immunoblotting assays, florescent markers; all of which involve a substantial level of measurement uncertainty. Therefore, models that are constructed based on a single set of estimated parameter values will not be robust enough to capture the inherent measurement noise observed in the experimental data. To this end, our proposed scheme returns a range for each model parameter whereas most of the existing optimization methods serve as point estimators.

Nonetheless, we must note that the computational cost of training ML-based models can be high depending on the complexity of the ANN and the quality of the labeled data. In this regard, one can use the transfer learning technique [136] to leverage the knowledge obtained from the training of simpler models to address the parameter estimation problem for more complex systems. For example, as we have mentioned in Sec. 6.2, the weights, and biases of the ANN model are assigned randomly in the first iteration. We can use the learned model from the first iteration

rather than selecting random initial weights and biases in the second iteration. According to our observations in the simulation of the monomolecular system, the number of iterations for training the ANN reduces significantly when we use such transfer learning. Moreover, a subset of the labeled training data generated in one iteration can be relabeled and reused for training the ANN model in the subsequent iterations.

The results presented in this chapter have demonstrated the high accuracy of the proposed ANN model when applied to the considered example system. However, for more complicated systems, various strategies could be adopted to maintain the prediction accuracy. In such classification problems, the inevitable biases and variances in modeling and data generation can cause prediction errors [137]. We notice that since the ANN is a model-free approach, the bias is inherently low. Therefore, it is imperative to adopt methods such as the ensemble method [138] or weight regularization [139] to reduce the variance. In Chapter 7, we elaborate more on these strategies and also present some possible directions to extend this research.

Chapter 7

Conclusions

7.1 Concluding Remarks and Future Research Directions

In this dissertation, we have developed a novel framework for efficient modeling, simulation, and analyses of the stochastic effects in biochemical reaction networks. In particular, we have focused on addressing the following key challenges in constructing and validating a realistic stochastic model in biochemical systems:

- First, the computational cost of simulating a stochastic model is very high. That is because realistic regulatory networks are multiscale and complex. Not only a large number of biochemical species are involved, but also both species and reactions present different scales of abundances and firing frequencies. To address this challenge, in Chapter 3, we have proposed a new hybrid ODE/SSA model to address the computational complexity of modeling molecular mechanisms that control the budding yeast cell cycle. The proposed hybrid method maintains the trade-off between accuracy and efficiency by integrating deterministic and stochastic simulations for a multiscale system of the budding yeast. The results have

shown that the proposed model can reduce the simulation time from 34 hours to 15 minutes while maintaining the same accuracy.

- Second, compared to wild-type cells, mutant cells are more vulnerable to environmental alternations such as the growth rate of their media. Hence, it is imperative to study the non-robust dynamics and stochastic characteristics of mutant cells. In Chapter 4, we have extended our model to study the stochastic phenotypes of more than 100 mutant cases. We have proposed a new quantitative rule to evaluate the viability of a mutant cell that accounts for stochastic phenotypes such as partial viability. The results have shown that using the proposed rule, our model can successfully reproduce 85% of mutant cases.

- Third, the size control mechanism in the cell cycle imposes substantial variabilities to different properties of the cell cycle. To understand the control mechanism that maintains the cell size, there is a need for efficient quantitative models that can capture the effect of such variabilities in progression through the cell cycle. In this regard, Chapter 5 has focused on quantifying the variability of the G1 phase properties in both wild-type cells as well as several mutant cases. To this end, we have extended the proposed hybrid stochastic model by incorporating messenger RNAs and proteins of the G1 phase. The results have shown that the model can successfully identify the size and time regulation modules in the size control mechanism of both wild-type and several early G1 mutants.

- Finally, estimating the parameters of stochastic models in systems biology is a challenging problem due to the complexity of biochemical networks and sparsity of measurement data. In Chapter 6, we have proposed a novel theory-guided machine learning approach that leverages model-free learning as well as domain knowledge from the systems biology to solve the parameter estimation problem. The results have proved the proposed hierarchical deep classification algorithm useful for estimating the model parameters in a simple monomolecular reaction system. Simulation results have also shown that the proposed approach can

successfully identify a range for the parameters of a monomolecular system.

While the proposed algorithm in Chapter 6 has been applied to a simple monomolecular system, the provided results shed light on the potentials of our theory-guided ML scheme for addressing the parameter estimation of stochastic models in systems biology. Following, we propose several research directions to extend our proposed framework and achieve higher generalizability when dealing with more complex systems. *First*, while we have adopted a simple feed-forward NN model, more advanced types of NNs such as recurrent, Long Short Term Memory (LSTM), or convolutional NNs can be considered for more complex biochemical systems with various types of input (such as time-series data). *Second*, ensemble methods, or combining multiple classifiers could be used to reduce the variance of the inverse model. There are various strategies to form an ensemble network, including but not limited to: using multiple network architectures, using the same architecture trained with different algorithms, making decisions based on the outcomes of neural networks along with traditional statistical classifiers. *Third*, the proposed hierarchical classification method could be used in a hybrid setting along with conventional optimization methods such as the QNSTOP method [128].

7.2 List of Publications

As a byproduct of the above contributions, this dissertation has led to the following key publications:

1. **M. Ahmadian**, S. Wang, J. J. Tyson, Y. Cao "A Hybrid Stochastic Model of the Budding Yeast Cell Cycle," *npj Systems Biology and Applications*, npj Systems Biology and Applications, 6 (7), 2020.

2. M. Chen, **M. Ahmadian**, L. Watson, J. J. Tyson, Y. Cao "Finding Acceptable Parameter Re-

gions of Stochastic Hill Equations for Cooperative Binding Mechanisms," *Journal of Chemical Physics*, vol. 14, no. 11, pp. 5927-5940, 2020.

3. **M. Ahmadian**, J. J. Tyson, Y. Cao "Stochastic Model of Size Control in the Budding Yeast Cell Cycle," *BMC Bioinformatics*, vol. 20, no. 12, pp. 322, 2019.

4. **M. Ahmadian**, J. J. Tyson, Y. Cao, "A Stochastic Model of Size Control in the Budding Yeast Cell Cycle", extended abstract, *In proceeding of ACM-BCB 2018*, Washington DC, August 2018.

5. **M. Ahmadian**, S. Wang, J. J. Tyson, Y. Cao, "Hybrid ODE/SSA Model of the Budding Yeast Cell Cycle Control Mechanism with Mutant Case Study", *In proceeding of ACM-BCB 2017*, Boston, MA, August 2017.

6. S. Wang, **M. Ahmadian**, M. Chen, J. J. Tyson, Y. Cao, "A Hybrid Stochastic Model of the Budding Yeast Cell Cycle Control Mechanism", *In Proceeding of ACM-BCB 2016*, Seattle, WA, October 2016.

Bibliography

[1] Stefan Bornholdt, "Boolean network models of cellular regulation: prospects and limitations," *Journal of the Royal Society Interface*, vol. 5, no. suppl_1, pp. 85–94, 2008.

[2] Maria I Davidich and Stefan Bornholdt, "Boolean network model predicts cell cycle sequence of fission yeast," *PloS One*, vol. 3, no. 2, pp. e1672, 2008.

[3] Rajat Singhania, R Michael Sramkoski, James W Jacobberger, and John J Tyson, "A hybrid model of mammalian cell cycle regulation," *PLoS Computational Biology*, vol. 7, no. 2, pp. e1001077, 2011.

[4] Yurie Okabe and Masaki Sasai, "Stable stochastic dynamics in yeast cell cycle," *Biophysical Journal*, vol. 93, no. 10, pp. 3451–3459, 2007.

[5] Stefan Braunewell and Stefan Bornholdt, "Superstability of the yeast cell-cycle dynamics: ensuring causality in the presence of biochemical stochasticity," *Journal of Theoretical Biology*, vol. 245, no. 4, pp. 638–643, 2007.

[6] Hao Ge, Hong Qian, and Min Qian, "Synchronized dynamics and non-equilibrium steady states in a stochastic yeast cell-cycle network," *Mathematical Biosciences*, vol. 211, no. 1, pp. 132–152, 2008.

[7] Adrien Fauré, Aurélien Naldi, Fabrice Lopez, Claudine Chaouiya, Andrea Ciliberto, and Denis Thieffry, "Modular logical modelling of the budding yeast cell cycle," *Molecular BioSystems*, vol. 5, no. 12, pp. 1787–1796, 2009.

[8] John J Tyson, "Modeling the cell division cycle: cdc2 and cyclin interactions.," *Proceedings of the National Academy of Sciences*, vol. 88, no. 16, pp. 7328–7332, 1991.

[9] Katherine C Chen, Laurence Calzone, Attila Csikasz-Nagy, Frederick R Cross, Bela Novak, and John J Tyson, "Integrative analysis of cell cycle control in budding yeast," *Molecular Biology of the Cell*, vol. 15, no. 8, pp. 3841–3862, 2004.

[10] Pavel Kraikivski, Katherine C Chen, Teeraphan Laomettachit, TM Murali, and John J Tyson, "From START to FINISH: computational analysis of cell cycle control in budding yeast," *NPJ Systems Biology and Applications*, vol. 1, pp. 15016, 2015.

[11] David Ball, Neil Adames, Nadine Reischmann, Debashis Barik, Christopher Franck, John J Tyson, and Jean Peccoud, "Measurement and modeling of transcriptional noise in the cell cycle regulatory network," *Cell Cycle*, vol. 12, no. 19, pp. 3392–3407, 2013.

[12] Stefano Di Talia, Jan M Skotheim, James M Bean, Eric D Siggia, and Frederick R Cross, "The effects of molecular noise and size control on variability in the budding yeast cell cycle," *Nature*, vol. 448, no. 7156, pp. 947–951, 2007.

[13] Daniel T Gillespie, "A rigorous derivation of the chemical master equation," *Physica A: Statistical Mechanics and its Applications*, vol. 188, no. 1-3, pp. 404–425, 1992.

[14] Daniel T Gillespie, "Stochastic simulation of chemical kinetics," *Annu. Rev. Phys. Chem.*, vol. 58, pp. 35–55, 2007.

[15] Daniel T Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of computational physics*, vol. 22, no. 4, pp. 403–434, 1976.

[16] Daniel T Gillespie, "Exact stochastic simulation of coupled chemical reactions," *The journal of physical chemistry*, vol. 81, no. 25, pp. 2340–2361, 1977.

[17] Michael A Gibson and Jehoshua Bruck, "Efficient exact stochastic simulation of chemical systems with many species and many channels," *Journal of Physical Chemistry A*, vol. 104, no. 9, pp. 1876–1889, 2000.

[18] Yang Cao, Hong Li, and Linda Petzold, "Efficient formulation of the stochastic simulation algorithm for chemically reacting systems," *Journal of Chemical Physics*, vol. 121, no. 9, pp. 4059–4067, 2004.

[19] James M McCollum, Gregory D Peterson, Chris D Cox, Michael L Simpson, and Nagiza F Samatova, "The sorting direct method for stochastic simulation of biochemical systems with varying reaction execution behavior," *Computational Biology and Chemistry*, vol. 30, no. 1, pp. 39–49, 2006.

[20] Hong Li and Linda Petzold, "Logarithmic direct method for discrete stochastic simulation of chemically reacting systems," *Journal of Chemical Physics*, vol. 16, 2006.

[21] Alexander Slepoy, Aidan P Thompson, and Steven J Plimpton, "A constant-time kinetic Monte Carlo algorithm for simulation of large biochemical reaction networks," *Journal of Chemical Physics*, vol. 128, no. 20, pp. 05B618, 2008.

[22] Daniel T Gillespie, "Approximate accelerated stochastic simulation of chemically reacting systems," *Journal of Chemical Physics*, vol. 115, no. 4, pp. 1716–1733, 2001.

[23] Daniel T Gillespie and Linda R Petzold, "Improved leap-size selection for accelerated stochastic simulation," *Journal of Chemical Physics*, vol. 119, no. 16, pp. 8229–8234, 2003.

[24] Eric L Haseltine and James B Rawlings, "Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics," *Journal of Chemical Physics*, vol. 117, no. 15, pp. 6959–6969, 2002.

[25] Christopher V Rao and Adam P Arkin, "Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the gillespie algorithm," *Journal of Chemical Physics*, vol. 118, no. 11, pp. 4999–5010, 2003.

[26] Howard Salis and Yiannis Kaznessis, "Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions," *Journal of Chemical Physics*, vol. 122, no. 5, pp. 054103, 2005.

[27] Zhen Liu, Yang Pu, Fei Li, Clifford A Shaffer, Stefan Hoops, John J Tyson, and Yang Cao, "Hybrid modeling and simulation of stochastic effects on progression through the eukaryotic cell cycle," *Journal of Chemical Physics*, vol. 136, no. 3, pp. 034105, 2012.

[28] Ashyraliyev M, Fomekong-Nanfack Y, Kaandorp JA, Blom JG., "Systems biology: parameter estimation for biochemical models.," *FEBS J*, vol. 276, no. 11, pp. 886–902, 2009.

[29] P Mendes and D Kell., "Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation.," *Bioinformatics*, vol. 14, no. 10, pp. 869–883, 1998.

[30] Pedro Mendes and Douglas B. Kell, "Non-linear optimization of biochemical pathways:applications to metabolic engineering and parameter estimation," *Bioinformatics*, vol. 14, no. 10, pp. 869–883, 1998.

[31] Claude Gérard, John J Tyson, and Béla Novák, "Minimal models for cell-cycle control based on competitive inhibition and multisite phosphorylations of cdk substrates," *Biophysical journal*, vol. 104, no. 6, pp. 1367–1379, 2013.

[32] David A Ball, Tae-Hyuk Ahn, Pengyuan Wang, Katherine C Chen, Yang Cao, John J Tyson, Jean Peccoud, and William T Baumann, "Stochastic exit from mitosis in budding yeast: model predictions and experimental observations," *Cell Cycle*, vol. 10, no. 6, pp. 999–1009, 2011.

[33] Linda R Petzold, "Description of dassl: a differential/algebraic system solver," Tech. Rep., Sandia National Labs., Livermore, CA (USA), 1982.

[34] Yung-Yao Chen, Yu-Hsiu Lin, Chia-Ching Kung, Ming-Han Chung, I Yen, et al., "Design and implementation of cloud analytics-assisted smart power meters considering advanced artificial intelligence as edge analytics in demand-side management for smart homes," *Sensors*, vol. 19, no. 9, pp. 2047, 2019.

[35] Kevin P Murphy, *Machine learning: a probabilistic perspective*, MIT press, 2012.

[36] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *2004 IEEE international joint conference on neural networks (IEEE Cat. No. 04CH37541)*. IEEE, 2004, vol. 2, pp. 985–990.

[37] Xavier Glorot and Yoshua Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249–256.

[38] José Orozco and Carlos A Reyes García, "Detecting pathologies from infant cry applying scaled conjugate gradient neural networks," in *European Symposium on Artificial Neural Networks, Bruges (Belgium)*, 2003.

[39] Mark T Borisuk and John J Tyson, "Bifurcation analysis of a model of mitotic control in frog eggs," *Journal of Theoretical Biology*, vol. 195, no. 1, pp. 69–85, 1998.

[40] Bela Novak and John J Tyson, "Modeling the cell division cycle: M-phase trigger, oscillations, and size control," *Journal of Theoretical Biology*, vol. 165, no. 1, pp. 101–134, 1993.

[41] Bela Novak and John J Tyson, "Quantitative analysis of a molecular model of mitotic control in fission yeast," *Journal of Theoretical Biology*, vol. 173, no. 3, pp. 283–305, 1995.

[42] Akos Sveiczer, John J Tyson, and Bela Novak, "A stochastic, molecular model of the fission yeast cell cycle: role of the nucleocytoplasmic ratio in cycle time regulation," *Biophysical Chemistry*, vol. 92, no. 1-2, pp. 1–15, 2001.

[43] John J Tyson and Bela Novak, "Regulation of the eukaryotic cell cycle: molecular antagonism, hysteresis, and irreversible transitions," *Journal of Theoretical Biology*, vol. 210, no. 2, pp. 249–263, 2001.

[44] Fangting Li, Tao Long, Ying Lu, Qi Ouyang, and Chao Tang, "The yeast cell-cycle network is robustly designed," *Proceedings of the National Academy of Sciences*, vol. 101, no. 14, pp. 4781–4786, 2004.

[45] Andrew W Murray, "Recycling the cell cycle: cyclins revisited," *Cell*, vol. 116, no. 2, pp. 221–234, 2004.

[46] Paul Jorgensen and Mike Tyers, "How cells coordinate growth and division," *Current Biology*, vol. 14, no. 23, pp. 1014–1027, 2004.

[47] John J Tyson and Bela Novak, "Temporal organization of the cell cycle," *Current Biology*, vol. 18, no. 17, pp. 759–768, 2008.

[48] Jenna E Gallegos, Neil R Adames, Mark F Rogers, Pavel Kraikivski, Aubrey Ibele, Kevin Nurzynski-Loth, Eric Kudlow, TM Murali, John J Tyson, and Jean Peccoud, "Genetic interactions derived from high-throughput phenotyping of 6589 yeast cell cycle mutants," *NPJ systems biology and applications*, vol. 6, no. 1, pp. 1–15, 2020.

[49] Mostafa Herajy, Fei Liu, and Monika Heiner, "Efficient modelling of yeast cell cycles based on multisite phosphorylation using coloured hybrid petri nets with marking-dependent arc weights," *Nonlinear Analysis: Hybrid Systems*, vol. 27, pp. 191–212, 2018.

[50] Sandip Kar, William T Baumann, Mark R Paul, and John J Tyson, "Exploring the roles of noise in the eukaryotic cell cycle," *Proceedings of the National Academy of Sciences*, vol. 106, no. 16, pp. 6471–6476, 2009.

[51] Teeraphan Laomettachit, Katherine C Chen, William T Baumann, and John J Tyson, "A model of yeast cell-cycle regulation based on a standard component modeling strategy for protein regulatory networks," *PloS One*, vol. 11, no. 5, pp. e0153738, 2016.

[52] Yuping Zhang, Minping Qian, Qi Ouyang, Minghua Deng, Fangting Li, and Chao Tang, "Stochastic model of yeast cell-cycle network," *Physica D: Nonlinear Phenomena*, vol. 219, no. 1, pp. 35–39, 2006.

[53] Debashis Barik, William T Baumann, Mark R Paul, Bela Novak, and John J Tyson, "A model of yeast cell-cycle regulation based on multisite phosphorylation," *Molecular Systems Biology*, vol. 6, no. 1, 2010.

[54] Debashis Barik, David A Ball, Jean Peccoud, and John J Tyson, "A stochastic model of the yeast cell cycle reveals roles for feedback regulation in limiting cellular variability," *PLoS Computational Biology*, vol. 12, no. 12, pp. e1005230, 2016.

[55] Akos Sveiczer, John J Tyson, and Bela Novak, "A stochastic, molecular model of the fission yeast cell cycle: role of the nucleocytoplasmic ratio in cycle time regulation," *Biophysical Chemistry*, vol. 92, no. 1-2, pp. 1–15, 2001.

[56] Ralf Steuer, "Effects of stochasticity in models of the cell cycle: from quantized cycle times to noise-induced oscillations," *Journal of Theoretical Biology*, vol. 228, no. 3, pp. 293–301, 2004.

[57] Mukund Thattai and Alexander Van Oudenaarden, "Intrinsic noise in gene regulatory networks," *Proceedings of the National Academy of Sciences*, vol. 98, no. 15, pp. 8614–8619, 2001.

[58] Peter S Swain, Michael B Elowitz, and Eric D Siggia, "Intrinsic and extrinsic contributions to stochasticity in gene expression," *Proceedings of the National Academy of Sciences*, vol. 99, no. 20, pp. 12795–12800, 2002.

[59] Juan M Pedraza and Johan Paulsson, "Effects of molecular memory and bursting on fluctuations in gene expression," *Science*, vol. 319, no. 5861, pp. 339–343, 2008.

[60] Katherine C Chen, Attila Csikasz-Nagy, Bela Gyorffy, John Val, Bela Novak, and John J Tyson, "Kinetic analysis of a molecular model of the budding yeast cell cycle," *Molecular Biology of the Cell*, vol. 11, no. 1, pp. 369–391, 2000.

[61] Mohsen Sabouri-Ghomi, Andrea Ciliberto, Sandip Kar, Bela Novak, and John J Tyson, "Antagonism and bistability in protein interaction networks," *Journal of Theoretical Biology*, vol. 250, no. 1, pp. 209–218, 2008.

[62] Ivan Mura and Attila Csikász-Nagy, "Stochastic petri net extension of a yeast cell cycle model," *Journal of Theoretical Biology*, vol. 254, no. 4, pp. 850–860, 2008.

[63] R Bundschuh, F Hayot, and C Jayaprakash, "Fluctuations and slow variables in genetic networks," *Biophysical Journal*, vol. 84, no. 3, pp. 1606–1615, 2003.

[64] Juan M Pedraza and Johan Paulsson, "Effects of molecular memory and bursting on fluctuations in gene expression," *Science*, vol. 319, no. 5861, pp. 339–343, 2008.

[65] Christian Miller, Björn Schwalb, Kerstin Maier, Daniel Schulz, Sebastian Dümcke, Benedikt Zacher, Andreas Mayer, Jasmin Sydow, Lisa Marcinowski, Lars Dölken, et al., "Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast," *Molecular systems biology*, vol. 7, no. 1, 2011.

[66] Sina Ghaemmaghami, Won-Ki Huh, Kiowa Bower, Russell W Howson, Archana Belle, Noah Dephoure, Erin K O'shea, and Jonathan S Weissman, "Global analysis of protein expression in yeast," *Nature*, vol. 425, no. 6959, pp. 737–741, 2003.

[67] Stefano Di Talia, Hongyin Wang, Jan M Skotheim, Adam P Rosebrock, Bruce Futcher, and Frederick R Cross, "Daughter-specific transcription factors regulate cell size control in budding yeast," *PLoS Biology*, vol. 7, no. 10, pp. e1000221, 2009.

[68] Tracy L Laabs, David D Markwardt, Matthew G Slattery, Laura L Newcomb, David J Stillman, and Warren Heideman, "ACE2 is required for daughter cell-specific G1 delay in saccharomyces cerevisiae," *Proceedings of the National Academy of Sciences*, vol. 100, no. 18, pp. 10275–10280, 2003.

[69] Mary E Miller and Frederick R Cross, "Mechanisms controlling subcellular localization of the G1 cyclins Cln2p and Cln3p in budding yeast," *Molecular and Cellular Biology*, vol. 21, no. 18, pp. 6292–6311, 2001.

[70] Frederick R Cross, Vincent Archambault, Mary Miller, and Martha Klovstad, "Testing a mathematical model of the yeast cell cycle," *Molecular Biology of the Cell*, vol. 13, no. 1, pp. 52–70, 2002.

[71] Yang Cao, Daniel T Gillespie, and Linda R Petzold, "The slow-scale stochastic simulation algorithm," *Journal of Chemical Physics*, vol. 122, no. 1, pp. 014116, 2005.

[72] Stefan Hoops, Sven Sahle, Ralph Gauges, Christine Lee, Jürgen Pahle, Natalia Simus, Mudita Singhal, Liang Xu, Pedro Mendes, and Ursula Kummer, "COPASI—a complex pathway simulator," *Bioinformatics*, vol. 22, no. 24, pp. 3067–3074, 2006.

[73] Shuo Wang, Minghan Chen, Layne T Watson, and Yang Cao, "Efficient implementation of the hybrid method for stochastic simulation of biochemical systems," *Journal of Micromechanics and Molecular Physics*, vol. 2, no. 02, pp. 1750006, 2017.

[74] Minghan Chen, Shuo Wang, and Yang Cao, "Accuracy analysis of hybrid stochastic simulation algorithm on linear chain reaction systems," *Bulletin of Mathematical Biology*, vol. 81, pp. 3024–3052, 2018.

[75] Robert A Weinberg and Robert A Weinberg, *The biology of cancer*, Garland science, 2013.

[76] Frederick R Cross, "Two redundant oscillatory mechanisms in the yeast cell cycle," *Developmental Cell*, vol. 4, no. 5, pp. 741–752, 2003.

[77] CB Epstein and FR Cross, "Genes that can bypass the CLN requirement for saccharomyces cerevisiae cell cycle START.," *Molecular and Cellular Biology*, vol. 14, no. 3, pp. 2041–2047, 1994.

[78] Herman Wijnen and Bruce Futcher, "Genetic analysis of the shared role of CLN3 and BCK2 at the G1-S transition in saccharomyces cerevisiae," *Genetics*, vol. 153, no. 3, pp. 1131–1143, 1999.

[79] Vincent Archambault, Caihong X Li, Alan J Tackett, Ralph Wasch, Brian T Chait, Michael P Rout, and Frederick R Cross, "Genetic and biochemical evaluation of the importance of cdc6 in regulating mitotic exit," *Molecular biology of the cell*, vol. 14, no. 11, pp. 4592–4604, 2003.

[80] Frederick R Cross, "Two redundant oscillatory mechanisms in the yeast cell cycle," *Developmental Cell*, vol. 4, no. 5, pp. 741–752, 2003.

[81] Leland H Hartwell and Michael W Unger, "Unequal division in saccharomyces cerevisiae and its implications for the control of cell division.," *The Journal of cell biology*, vol. 75, no. 2, pp. 422–435, 1977.

[82] BLA Carter, "The control of cell division in saccharomyces cerevisiae," *The cell cycle*, pp. 99–117, 1981.

[83] Kim Nasmyth, "Evolution of the cell cycle," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 349, no. 1329, pp. 271–281, 1995.

[84] Kim Nasmyth, "At the heart of the budding yeast cell cycle," *Trends in Genetics*, vol. 12, no. 10, pp. 405–412, 1996.

[85] Michael Botchan, "Coordinating dna replication with cell division: current status of the licensing concept.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 19, pp. 9997, 1996.

[86] J.J Tyson, B Novak, GM Odell, K Chen, and CD Thron, "Chemical kinetic theory as a tool for understanding the regulation of m-phase promoting factor in the cell cycle," *Trends Biochem. Sci*, vol. 21, pp. 89–96, 1996.

[87] Angelika Amon, "Regulation of b-type cyclin proteolysis by cdc28–associated kinases in budding yeast," *The EMBO Journal*, vol. 16, no. 10, pp. 2693–2702, 1997.

[88] Sue Biggins and Andrew W Murray, "Sister chromatid cohesion in mitosis," *Current opinion in cell biology*, vol. 10, no. 6, pp. 769–775, 1998.

[89] Janet Leatherwood, "Emerging mechanisms of eukaryotic dna replication initiation," *Current opinion in cell biology*, vol. 10, no. 6, pp. 742–748, 1998.

[90] Wolfgang Zachariae, Michael Schwab, Kim Nasmyth, and Wolfgang Seufert, "Control of cyclin ubiquitination by cdk-regulated binding of hct1 to the anaphase promoting complex," *Science*, vol. 282, no. 5394, pp. 1721–1724, 1998.

[91] Peter G Lord and Alan E Wheals, "Asymmetrical division of saccharomyces cerevisiae.," *Journal of Bacteriology*, vol. 142, no. 3, pp. 808–818, 1980.

[92] Linlin Chen, Yuming Ma, Jingya Zhao, Xuejing Geng, Wenbo Chen, Shengli Ding, Haiyang Li, and Honglian Li, "The bzip transcription factor fpada1 is essential for fungal growth and conidiation in fusarium pseudograminearum," *Current genetics*, pp. 1–9, 2019.

[93] GC Johnston, JR Pringle, and LH Hartwell, "Coordination of growth with cell division in the yeast saccharomyces cerevisiae," *Experimental cell research*, vol. 105, no. 1, pp. 79–98, 1977.

[94] Alexi I Goranov, Michael Cook, Marketa Ricicova, Giora Ben-Ari, Christian Gonzalez, Carl Hansen, Mike Tyers, and Angelika Amon, "The rate of cell growth is governed by cell cycle stage," *Genes & development*, vol. 23, no. 12, pp. 1408–1422, 2009.

[95] Francisco Ferrezuelo, Neus Colomina, Alida Palmisano, Eloi Garí, Carme Gallego, Attila Csikász-Nagy, and Martí Aldea, "The critical size is set at a single-cell level by growth rate to attain homeostasis and adaptation," *Nature communications*, vol. 3, pp. 1012, 2012.

[96] Kurt M Schmoller, JJ Turner, M Kõivomägi, and Jan M Skotheim, "Dilution of the cell cycle inhibitor whi5 controls budding-yeast cell size," *Nature*, vol. 526, no. 7572, pp. 268, 2015.

[97] Ilya Soifer, Lydia Robert, and Ariel Amir, "Single-cell analysis of growth in budding yeast and bacteria reveals a common size regulation strategy," *Current Biology*, vol. 26, no. 3, pp. 356–361, 2016.

[98] Jonathan J Turner, Jennifer C Ewald, and Jan M Skotheim, "Cell size control in yeast," *Current biology*, vol. 22, no. 9, pp. R350–R359, 2012.

[99] Evgeny Zatulovskiy and Jan M Skotheim, "On the molecular mechanisms regulating animal cell size homeostasis," *Trends in Genetics*, 2020.

[100] Shuo Wang, Mansooreh Ahmadian, Minghan Chen, John Tyson, and Young Cao, "A hybrid stochastic model of the budding yeast cell cycle control mechanism," in *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. ACM, 2016, pp. 261–270.

[101] Mansooreh Ahmadian, Shuo Wang, John Tyson, and Young Cao, "Hybrid ode/ssa model of the budding yeast cell cycle control mechanism with mutant case study," in *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. ACM, 2017, pp. 464–473.

[102] Ahmadian Mansooreh, John J Tyson, Jean Peccoud, and Yang Cao, "A hybrid stochastic model of the budding yeast cell cycle," *NPJ Systems Biology and Applications*, vol. 6, no. 1, 2020.

[103] Michael Polymenis and Emmett V Schmidt, "Coupling of cell division to cell growth by translational control of the g1 cyclin cln3 in yeast," *Genes & development*, vol. 11, no. 19, pp. 2522–2531, 1997.

[104] Laura L Newcomb, Duane D Hall, and Warren Heideman, "Azf1 is a glucose-dependent positive regulator of cln3 transcription in saccharomyces cerevisiae," *Molecular and cellular biology*, vol. 22, no. 5, pp. 1607–1614, 2002.

[105] Harvey F Lodish, "Model for the regulation of mrna translation applied to haemoglobin synthesis," *Nature*, vol. 251, no. 5474, pp. 385, 1974.

[106] Gabriele Lillacci and Mustafa Khammash, "Parameter estimation and model selection in computational biology," *PLoS Comput Biol*, vol. 6, no. 3, pp. e1000696, 2010.

[107] Albert Goldbeter and Daniel E Koshland, "An amplified sensitivity arising from covalent modification in biological systems," *Proceedings of the National Academy of Sciences*, vol. 78, no. 11, pp. 6840–6844, 1981.

[108] Martino Barenco, Daniela Tomescu, Daniel Brewer, Robin Callard, Jaroslav Stark, and Michael Hubank, "Ranked prediction of p53 targets using hidden variable dynamic modeling," *Genome biology*, vol. 7, no. 3, pp. R25, 2006.

[109] Bayu Jayawardhana, Douglas B Kell, and Magnus Rattray, "Bayesian inference of the sites of perturbations in metabolic pathways via markov chain monte carlo," *Bioinformatics*, vol. 24, no. 9, pp. 1191–1197, 2008.

[110] Daniel A Henderson, Richard J Boys, Carole J Proctor, and Darren J Wilkinson, "Linking systems biology models to data: a stochastic kinetic model of p53 oscillations," in *The Oxford Handbook of Applied Bayesian Analysis*, pp. 155–187. Oxford University Press, 2010.

[111] Jacqueline M Dresch, Xiaozhou Liu, David N Arnosti, and Ahmet Ay, "Thermodynamic modeling of transcription: sensitivity analysis differentiates biological mechanism from mathematical model-induced effects," *BMC systems biology*, vol. 4, no. 1, pp. 142, 2010.

[112] Jongrae Kim, Declan G Bates, Ian Postlethwaite, Pat Heslop-Harrison, and Kwang-Hyun Cho, "Least-squares methods for identifying biochemical regulatory networks from noisy measurements," *BMC bioinformatics*, vol. 8, no. 1, pp. 8, 2007.

[113] Mandavilli Srinivas and Lalit M Patnaik, "Genetic algorithms: A survey," *computer*, vol. 27, no. 6, pp. 17–26, 1994.

[114] Maksat Ashyraliyev, Johannes Jaeger, and Joke G Blom, "Parameter estimation and determinability analysis applied to drosophila gap gene circuits," *BMC Systems Biology*, vol. 2, no. 1, pp. 83, 2008.

[115] Scott Kirkpatrick, C Daniel Gelatt, and Mario P Vecchi, "Optimization by simulated annealing," *science*, vol. 220, no. 4598, pp. 671–680, 1983.

[116] Darren J Wilkinson, "Bayesian methods in bioinformatics and computational systems biology," *Briefings in bioinformatics*, vol. 8, no. 2, pp. 109–116, 2007.

[117] Xin Liu, Mahesan Niranjan, "Parameter estimation in computational biology by approximate bayesian computation coupled with sensitivity analysis.," *under submission to Bioinformatics*, 2017.

[118] Stephen Brooks, "Markov chain monte carlo method and its application," *Journal of the royal statistical society: series D (the Statistician)*, vol. 47, no. 1, pp. 69–100, 1998.

[119] Kevin S Brown and James P Sethna, "Statistical mechanical approaches to models with many poorly known parameters," *Physical review E*, vol. 68, no. 2, pp. 021904, 2003.

[120] Dorjsuren Battogtokh, David K Asch, Mary E Case, Jonathan Arnold, and H-B Schüttler, "An ensemble method for identifying regulatory circuits with special reference to the qa gene cluster of neurospora crassa," *Proceedings of the National Academy of Sciences*, vol. 99, no. 26, pp. 16904–16909, 2002.

[121] Scott A Sisson, Yanan Fan, and Mark M Tanaka, "Sequential monte carlo without likelihoods," *Proceedings of the National Academy of Sciences*, vol. 104, no. 6, pp. 1760–1765, 2007.

[122] X Liu and M. Niranjan., "State and parameter estimation of the heat shock response system using kalman and particle filters.," *Bioinformatics*, vol. 28, no. 11, pp. 1501–1507, 2012.

[123] Jane Liu and Mike West, "Combined parameter and state estimation in simulation-based filtering," in *Sequential Monte Carlo methods in practice*, pp. 197–223. Springer, 2001.

[124] Xiaodian Sun, Li Jin, and Momiao Xiong, "Extended kalman filter for estimation of parameters in nonlinear state-space models of biochemical networks," *PloS one*, vol. 3, no. 11, pp. e3758, 2008.

[125] Mansooreh Ahmadian, John J Tyson, and Yang Cao, "A stochastic model of size control in the budding yeast cell cycle," *BMC bioinformatics*, vol. 20, no. 12, pp. 322, 2019.

[126] Stefan Reinker, Rachel M Altman, and Jens Timmer, "Parameter estimation in stochastic biochemical reactions," *IEE Proceedings-Systems Biology*, vol. 153, no. 4, pp. 168–178, 2006.

[127] Brandon D Amos, David R Easterling, Layne T Watson, William I Thacker, Brent S Castle, and Michael W Trosset, "Algorithm 1007: Qnstop—quasi-newton algorithm for stochastic optimization," *ACM Transactions on Mathematical Software (TOMS)*, vol. 46, no. 2, pp. 1–20, 2020.

[128] Minghan Chen, Brandon D. Amos, Layne T. Watson, John J. Tyson, Yang Cao, Clifford A. Shaffer, Michael W. Trosset, Cihan Oguz, Gisella Kakoti., "Quasi-newton stochastic optimization algorithm for parameter estimation of a stochastic model of the budding yeast cell cycle," *IEEE/ACM transactions on computational biology and bioinformatics*, 2017.

[129] M Chen, M Ahmadian, LT Watson, and Y Cao, "Finding acceptable parameter regions of stochastic hill functions for multisite phosphorylation mechanism," *The Journal of Chemical Physics*, vol. 152, no. 12, pp. 124108, 2020.

[130] Anuj Karpatne, Gowtham Atluri, James Faghmous, Michael Steinbach, Arindam Banerjee, Auroop Ganguly, Shashi Shekhar, Nagiza Samatova, Vipin Kumar., "Theory-guided data science: A new paradigm for scientific discovery from data.," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 10, pp. 2318–2331, 2017.

[131] Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP., "Universally sloppy parameter sensitivities in systems biology models.," *PLoS Comput Biol*, vol. 3, no. 10, pp. 1871–78, 2007.

[132] Kurt Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural networks*, vol. 4, no. 2, pp. 251–257, 1991.

[133] Jianhua Lin, "Divergence measures based on the shannon entropy," *IEEE Transactions on Information theory*, vol. 37, no. 1, pp. 145–151, 1991.

[134] Martin F Møller, *A scaled conjugate gradient algorithm for fast supervised learning*, Aarhus University, Computer Science Department, 1990.

[135] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[136] Sinno Jialin Pan and Qiang Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[137] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep learning*, MIT press, 2016.

[138] Guoqiang Peter Zhang, "Neural networks for classification: a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 30, no. 4, pp. 451–462, 2000.

[139] Vivienne Sze, Yu-Hsin Chen, Tien-Ju Yang, and Joel S Emer, "Efficient processing of deep neural networks: A tutorial and survey," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2295–2329, 2017.