Abstract—Complex assemblies of interacting proteins carry out most of the interesting jobs in a cell, such as metabolism, DNA synthesis, mitosis and cell division. These physiological properties play out as a subtle molecular dance, choreographed by underlying regulatory networks that control the activities of cyclin-dependent kinases (CDK). The network can be modeled by a set of nonlinear differential equations and its behavior predicted by numerical simulation. In this paper, an innovative approach has been proposed that uses genetic algorithms to mine a set of behavior data output by a biological system in order to determine the kinetic parameters of the system. In our approach, the machine learning method is integrated with the framework of existing biological information in a wiring diagram so that its findings are expressed in a form of system dynamic behavior. By numerical simulations it has been illustrated that the model is consistent with experiments and successfully shown that such application of genetic algorithms will highly improve the performance of mathematical model of the cell division cycle to simulate such a complicated bio-system.

Keywords—Cell cycle, Cyclin-dependent kinase, Fission yeast, Genetic algorithms, Mathematical modeling, Wiring diagram

I. INTRODUCTION

The cell cycle is the sequence of events during which a growing cell replicates all its components and divides them more or less evenly between two daughter cells, so that each daughter contains the information and machinery necessary to repeat the process [1]–[2]. Cell proliferation underlies all biological growth, reproduction, and development, and its misregulation results in serious human diseases. As might be expected of a process so central to cell viability, the molecular machinery regulating crucial events of the cell cycle (DNA synthesis and mitosis) is highly conserved across eukaryotic organisms (Nurse, 1990). Hence, thorough genetic studies of cell cycle regulation in budding yeast [3]–[4] and fission yeast [5]–[6] have paid handsome dividends in genetic studies of cell cycle regulation in budding yeast [3]–[4].

The fast progress of biology development has accumulated a tremendous amount of experimental data, which becomes a big challenge to efficiently extract valuable knowledge hidden behind. In a series of researches in two past decades, scientists have been trying to provide simulations of complex biochemical networks and show the general principles of eukaryotic cell cycle regulation.

Mathematical models of cell cycle controls have kept pace with the advances of molecular genetics (Goldbeter 1991, Tyson 1991, Novak and Tyson 1993, Novak and Tyson 1995). In these models, biochemical mechanisms are translated, by the law of mass action, into systems of non-linear differential equations, and dynamical systems theory is used to uncover the qualitative behavior of these equations and to bridge the gap between mechanisms and physiology. A number of mathematical models [7]–[20] have been developed to illuminate the workings of the cell cycle, based on various dynamical mechanisms including limit cycle oscillation, bistability, and transient processes. In the past decade, Tyson and colleagues developed models for the yeast cell cycle and the Xenopus egg cell cycle which have greatly improved our understanding of cell cycle dynamics.

One of the difficulties of solving a mathematical model in a bio-system is parameter identification of the model so that the results provide a suitable fit to the experimental data. The parameter values proposed in previously mentioned models have been selected by a painstaking process of trial-and-error in order to achieve the best fit to real results.

The mathematical model used in this research is described by a dozen differential equations, involving ~40 kinetic parameters. Concerning plurality of parameters this method of trial-and-error is very frustrating. A possible way to deal with the problem is to use robust mathematical methods like optimization control algorithms or genetic algorithms that can contribute substantially in this area by generating optimum solutions to save the time and effort of a scientist.

In this research genetic algorithms will be applied to discover the kinetic parameters of the model. The proposed method is applied to experiment on the mathematical model of the fission yeast cell cycle with checkpoint controls at the G1/S, G2/M and metaphase/anaphase transitions [21]. The method is applied to find a parameter set that accounts for the properties of wild-type cells. After assigning values to these constants, the differential equations can be solved to produce a...
simulation of progression through the cell cycle in fission yeast. The results are then compared with the original data to evaluate the effectiveness of the approach.

II. MATERIALS AND METHODS

A. A Quantitative Mathematical Model

The molecular mechanism in Fig. 1 is a hypothetical account of the chemical reactions among the genes and proteins known to play principal roles in controlling the cell cycle of fission yeast. The mechanism summarizes information from many publications on the individual genes, their patterns of expression, and the interactions among their encoded proteins.

The cell cycle is controlled by a single CDK, namely Cdc2 (the protein encoded by the cdc2' gene), in combination with three B-type cyclins (Cdc13, Cig1, and Cig2 [22]). The most important cyclin partner of Cdc2 protein is Cdc13. The complex of Cdc2 and Cdc13 (known as M-phase promoting factor (MPF)) is absolutely essential to initiate mitosis, and in the absence of other cyclins, this complex can trigger S-phase as well [23]. The Cdc13 level fluctuates dramatically during the cell cycle, reaching a maximum as cells enter mitosis, dropping precipitously as cells exit mitosis, and reappearing after S-phase is initiated [24]. To simplify this molecular machinery Cig1 and Cig2 are ignored as demonstrated by [21].

The wiring diagram represented in Fig. 1 consists of a set of boxes (components) interconnected by arrows (reactions). An instantaneous state of the system is a specification of the current concentrations of all its components. Given a state of the system, the chemical reactions (synthesis, degradation, activation, inhibition, binding, and release) indicate how the state will change in the next moment of time.

Each reaction proceeds at a rate determined by the state of the system and by kinetic parameters (e.g., rate constants and binding constants). By applying the general principles of biochemical kinetics, the mechanism in Fig. 1 can be converted into a set of differential and algebraic equations that determine how the state of the control system evolves in time.

A mathematical description of this control system could be start with rate equations for Cdc2/Cdc13 (MPF) concentration as follows:

\[ \frac{d}{dt}[\text{MPF}] = k_{i, mass} - k_{wee}[\text{MPF}] + k_{\text{Cig1}}[\text{MPF}] - k_{i, Rum}[\text{MPF}] \]

Where, \( k_1 \) illustrates the influence of concentration of cyclin in nucleus through growing cell. \( k_{wee} \) is the specific rate of degradation of Cdc2/Cdc13 (dependent on the activity of wee1 enzyme) and \( k_{wee5} \) shows the formation rate of this complex due to Cdc25 activity. The influence of Rum1 is given by the \( k_r \) rate and its reverse effect, namely arising Cdc2/Cdc13 through disengaging Rum1, is modeled by \( (k_{wee}+k_1) \). Finally the effect of APC on degradation of cyclin and consumption of active complex is presented by \( k_2 \) parameter.

Similar equations must be written for each temporally varying protein in the reaction mechanism. The following equation is applied for preMPF (tyrosine-phosphorylated dimmers) concentration;

\[ \frac{d}{dt}[\text{preMPF}] = k_{i, mass}[\text{MPF}] - k_{wee}[\text{preMPF}] - k_i[\text{preMPF}] \]

Eq. 1 in combination with Eq.s (8), (9) and (10);

\[ \frac{d}{dt}[\text{weeMPF}] = k_{\text{Cig1}}[\text{weeMPF}] - k_{i, Rum}[\text{weeMPF}] \]

And finally the variable parameters are given by following equations;

\[ k_1 = v_1(1-\text{APC})+v_1\text{APC} \]
The concentrations variables are scaled so that all rate constants (k’s and V’s) have units of min$^{-1}$ and all Michaelis constants (J’s) are dimensionless. The aforementioned differential equations simply capture, in mathematical terms, the intuitive ideas about protein synthesis and degradation, phosphorylation and dephosphorylation. They allow us to test a hypothesis (for example, the network in Fig. 1) by computing how the concentration of each protein will rise and fall, and then comparing the simulated behavior of the model with observed behavior of the cell.

There are few direct kinetic measurements of the individual steps of the mechanism in fission yeast, so in the next section robust mathematical methods like genetic algorithm will be applied to identify the parameters of the model so that the results provide a suitable fit to experimental data.

**B. Application of Genetic Algorithm in Mathematical Modeling**

Genetic algorithms, introduced by John Holland in 1975 to mimic the mechanisms of natural adaptation are evolutionary methods widely used in different aspects of optimization and system biology modeling because of their unique capabilities of finding global optimum in highly multi-modal and/or non-differentiable search space [25]–[26]. Such stochastic methods can be used in the modeling of cell division cycles in terms of associated weights or coefficients which successfully perform better than traditional gradient-based techniques [27]. In particular, the search space of possible solutions is infinitely large, complex and not necessarily differentiable which makes the evolutionary methods more attractive for efficiently searching such complex search spaces [25].

**C. The Genome for Kinetic Parameters Representation**

The genome or chromosome representation, which shows a set of kinetic parameters simply consists of a binary string, composed of concatenation of 0 or 1 represented of input variables. In this encoding scheme, each coefficient is assigned and defined sub-string and a chromosome is a string of concatenated sub-strings of these parameters as variables. Therefore, for a given input vector K (Kinetic Parameters) a chromosome can be represented as a string of concatenated binary digits in the form of 100010001111000…. It is easily
seen that, for example, in the case of 40 parameters involved in optimization, a strings of concatenated sub-string must be created with different length depended to the desired accuracy and search interval of related parameter.

For instance the domain of variable $k_j$ is $[a_j, b_j]$ and the required precision is five places after the decimal point. The precision requirement implies that the range of domain of each variable should be divided into at least $(b_j - a_j) \times 10^5$ size ranges. The required bits for a sub-string (denoted with $m_j$) for a variable is calculated as follows:

$$2^{m_j-1} < (b_j - a_j) \times 10^5 \leq 2^{m_j}$$  \hspace{1cm} (21)

The mapping from a binary string to a real number for variable $k_j$ is straightforward and completed as follows:

$$k_j = a_j + \text{decimal(substring)} \times \frac{b_j - a_j}{2^m - 1}$$  \hspace{1cm} (22)

where decimal(substring) represents the decimal value of substring $j$ for decision variable $k_j$.

**D. Genetic Operators**

Such genome representation can now be readily used for the most two important genetic operators, namely, crossover and mutation. In this work, the natural roulette wheel selection method is used for choosing two parents producing two offsprings. The crossover operator for two selected individuals is simply accomplished by exchanging the tails of two chromosomes from a randomly chosen point as shown in Fig. 2.

![Fig. 2 Crossover operator](image)

It is very evident from Fig. 2 that the crossover operation can certainly exchange a part of genome of such kinetic parameters and similarly, the mutation operation which is often given little importance in some research papers as reported in [25], can contribute effectively to the diversity of the population. This operation is simply accomplished by charging one or more binary digits as genes in a chromosome to another digit as illustrated in Fig. 3.

![Fig. 3 Mutation operator](image)

The incorporation of genetic algorithm into the mathematical modeling of cell division cycle starts by representing each set of kinetic parameters as a string of concatenated sub-strings of binary digits. A generated solution during the evolution process is ranked by a fitness function that measures the similarity between the target behavior pattern and the actual one. Since the pattern is a time series of real value points, a simplest function of such a measurement is the reciprocal of the Sum of Squared Errors (SSE). The following equation is our fitness function:

$$Fitness = \frac{1}{\sum_{i} ((Y_o - \hat{Y}_o)^2)}$$  \hspace{1cm} (23)

where, $Y_o$ is the desired output; $\hat{Y}_o$ is the actual output of some system variable $y$; $t$ represents the $t^{th}$ time point, and $i$ represents the index of a variable.

The evolutionary process starts by randomly generating an initial population of binary strings each as a candidate solution. Then, using the aforementioned genetic operations of roulette wheel selection, crossover, and mutation, the entire populations of binary strings are caused to improve gradually. In this way, models of cell division cycle with progressively increasing fitness are produced until no further significant improvement is achievable.

**III. DISCOVERING KINETIC PARAMETER VALUES**

Note that, although the aforementioned mathematical models create the initial models of the system, they are just the skeleton with no parametric information and hence cannot be executed. So, the simulated behavior data generated from the original fission yeast cell cycle system created by Tyson in 2001 is used as an input to train the model. The genetic evolutionary process was allowed to run for 1000 generations and each generation has 50 populations. The initial value of each parameter is guessed randomly in an appropriate range.

The best parameter solutions are shown in Table I. In this table, there also show the parameter values introduced by Tyson which have been selected by a painstaking process of trial-and-error in order to achieve the best fit to real results. In Table I, one can see that the parameter values discovered in the process are quite close although some of them have a little larger difference. These reflect the sensitivity of different parameters with respect to the system behavior changes. However, the behavior of the system is not affected by these minor parameter variations.

The evolution history of the fitness function is shown in Fig. 4.
IV. SOLUTION OF MATHEMATICAL MODEL

A simulation of ‘wild-type’ fission yeast cell growth and division cycle with trained parameters is shown in Figs. 5 to 8. To solve and simulate the proposed model, the Simulink section of MATLAB 7.4.0 software was applied. When the simulated behaviors with the original ones are compared, one can see that they match quite well in all proteins. It is obvious from the pictures that the proposed model with trained kinetic parameters has identified the bio-system precisely and the phase durations coincide exactly with what extracted from experimental data and the lag appeared in solving mathematical model by parameter values assigned by Tyson has been disappeared (see Fig. 8).

The cell cycle should be viewed as an alternation between two characteristic phases: pre-Start (G1), when APC is active, Rum1 is present, and Cdc13 level is low, and post-Start (S + G2 + M), when APC is inactive, Rum1 is absent, and Cdc13 is accumulating. As model shows the duration of pre-Start is short while the post-Start phase is long, as in wild-type fission yeast cells [28]. In our simulation, the cell is so large in pre-Start that Cdc13/Cdc2, in combination with Puc1/Cdc2, quickly inactivates the APC and phosphorylates (destroys) Rum1. Rising CDK activity in the nucleus initiates DNA synthesis (the cell passes Start). As Cdc13/Cdc2 dimers accumulate after Start, they are at first tyrosine phosphorylated by the unreplicated-DNA surveillance mechanism and by the cell size surveillance mechanism.

Under normal conditions, the DNA-replication requirement is satisfied long before the cell-growth requirement. When the cell grows large enough, the positive feedback loops activate the pool of preMPF (tyrosine-phosphorylated dimers), driving the cell into M-phase. G2-phase is long in wild-type strains because a cell must grow to this minimal size before it can enter mitosis [29].

In this manner, the rising pattern of Cdc13-dependent kinase activity can drive an orderly progression of S- and M-phases, if initiation of DNA synthesis requires a lower CDK activity than initiation of mitosis [23]-[30].

As MPF activity increases dramatically at the end of G2-phase, chromosomes begin to condense, a mitotic spindle forms, and apoAAE begins to accumulate. At the same time, dynamic microtubules search for kinetochores on individual

![Fig. 4 The evolution process of best fitness value for each generation in GA](image-url)
chromosomes, which get pulled back and forth as they attach first to one pole of the spindle and then to the other. High MPF activity tries to activate AAE, but a strong signal coming from unaligned chromosomes inactivates AAE. As the chromosomes line up on the metaphase plate, MPF is able to activate AAE, which in turn activates APC. Cdc13 is degraded, AAE decays, and the cell returns to pre-Start.

Fig. 5 Comparison of the learnt model behavior (Blue line) to the real one (Red line) and the previous model (Green Line) for Mass

Fig. 6 Comparison of the learnt model behavior (Blue line) to the real one (Red line) and the previous model (Green Line) for MPF

Fig. 7 Comparison of the learnt model behavior (Blue line) to the real one (Red line) and the previous model (Green Line) for Pre-MPF

V. DISCUSSION AND CONCLUSION

In this paper, a precise, mathematical connection between the molecular networks is provided that surrounds cyclin-dependent kinase and the classical phases of the cell cycle. The modeling process starts by using the fission yeast’s experimental data and applying genetic algorithm to mine this set of behavior data in order to determine the kinetic parameters of the system. In our approach the machine learning method is integrated with the framework of biological information in a wiring diagram, which has been proposed by Novak and Tyson, so that its findings are expressed in a form of system dynamic behavior. Computer simulations of the model are in accordance with the physiological properties of wild-type cells and it has shown that the simulation results are improved considerably with trained kinetic parameters. The proposed approach is unique in that it provides a means to discover system’s coefficient by genetic algorithm so that the solutions can be shared and transferred in a mathematical model for better integration. Computer simulations of the model are in accordance with the physiological properties of wild-type cells.

Applying genetic algorithm to analysis of bio-information is an important area of study. The fast progress of biology development has accumulated a tremendous amount of experimental data, which becomes a big challenge to efficiently extract valuable knowledge hidden behind. Genetic algorithm can contribute substantially in this area by generating potential solutions to save the time and effort of a biologist. The method proposed in our approach is just an initial step to discover related information from a biological system. The ultimate goal of this line of study can be using data mining to assist model construction and behavior analysis in systems biology.

REFERENCES


