

SCIENTIFIC REPORTS



OPEN

Stochastic modeling of biochemical systems with multistep reactions using state-dependent time delay

Qianqian Wu^{1,2} & Tianhai Tian¹

Received: 11 April 2016

Accepted: 29 July 2016

Published: 24 August 2016

To deal with the growing scale of molecular systems, sophisticated modelling techniques have been designed in recent years to reduce the complexity of mathematical models. Among them, a widely used approach is delayed reaction for simplifying multistep reactions. However, recent research results suggest that a delayed reaction with constant time delay is unable to describe multistep reactions accurately. To address this issue, we propose a novel approach using state-dependent time delay to approximate multistep reactions. We first use stochastic simulations to calculate time delay arising from multistep reactions exactly. Then we design algorithms to calculate time delay based on system dynamics precisely. To demonstrate the power of proposed method, two processes of mRNA degradation are used to investigate the function of time delay in determining system dynamics. In addition, a multistep pathway of metabolic synthesis is used to explore the potential of the proposed method to simplify multistep reactions with nonlinear reaction rates. Simulation results suggest that the state-dependent time delay is a promising and accurate approach to reduce model complexity and decrease the number of unknown parameters in the models.

The advances in systems biology have raised a number of key challenges for modeling large-scale biochemical networks. Although a trend in mathematical modeling is to construct more and more mechanistically detailed models, the complexity of biological network, lack of experimental data and requirement of computing power have put a limitation on the complexity of mathematical models^{1–3}. Recently various methods have been developed to reduce model complexity^{4,5}. Simultaneously research works have also been conducted to explore the conditions and assumptions of these simplified models in order to obtain accurate simulations^{6,7}. Among them, one of the important biological processes is multistep reactions that has implication in a wide range of biochemical processes, including synthesis of mRNA for a series of strands of DNA, protein synthesis when ribosome reading a series of codons of mRNA⁸, signaling transduction for the activation of a sequence of kinases from growth factor receptor to transcriptional factors⁹, degradation of polymeric carbohydrates and synthesis of metabolic^{10,11}, cancer initiation that can be regarded as a series of gene mutations¹², as well as telomere shortening process¹³. Although a number of mathematical models in recent years have been designed to describe these multistep pathways accurately, due to the complex nature of molecular networks, more sophisticated models are needed to simplify the multistep reaction processes.

It has been widely accepted that biochemical processes are stochastic. Recent advances in experimental technology have provided the ability to measure cellular heterogeneity in single cells^{14,15}. Experimental studies have shown that gene expression is subject to stochastic fluctuations that lead to considerable differences in the level of expression between genetically identical cells. In addition, variation in protein levels arises from fluctuations in mRNA levels¹⁶. Stimulated by the pioneer work of stochastic modelling for gene expression¹⁷, the last ten years have seen an explosion in stochastic modelling to predict protein fluctuations in terms of the frequencies of probabilistic event^{18,19}. Although the expression process can be modelled by a series of detailed chemical events, the model structure may be too complex to get predictive insights. To address this issue, a number of modelling techniques have been proposed to simplify the complexity of mathematical model^{20–22}. Among them, differential equations with time delay have been used to simplify processes of multistep reactions²³. To explore the combined effects of time delay and intrinsic noise on genetic regulation, delay stochastic simulation algorithm (delay-SSA)^{24,25} has been proposed to simulate discrete chemical kinetic systems. The advances in

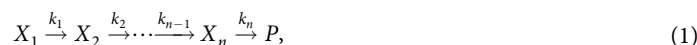
¹School of Mathematical Sciences, Monash University, Melbourne, VIC 3800, Australia. ²School of Mathematics Hefei University of Technology, Hefei, Anhui 230009 China. Correspondence and requests for materials should be addressed to T.T. (email: Tianhai.tian@monash.edu)

delayed modelling approaches include mathematical models for spatial effects in gene expression²⁶, and stochastic reaction systems with distributed delays²⁷. Other modelling techniques proposed recently include the slow-scale linear noise approximation and stochastic quasi-steady-state assumption^{28–30}. Recently, we have proposed the memory stochastic simulation algorithm for memory reactions and a two-variable model with a new concept of length^{5,31}.

It has been widely assumed in literature that time delay is either a constant or distributed delay with constant mean. We have used a one-step reaction model with constant, exponentially distributed or Erlang distributed delay to realize the mRNA turnover dynamics³². Simulation results suggest that the value of time delay may depend on the system state, rather than be a constant value. In fact, the state-dependent time delay has already been used in various research areas such as optimal control and population dynamics^{33,34}. Although these ideas were proposed about 40 years ago, the relationship between time delay and system state remains uncertain for discrete chemical reaction settings. Recently, a model with non-constant time delay has been derived to simplify the translational process of multistep reactions³⁵. In addition, research work has been conducted for exact model reduction with time delay; and closed form distribution and extension have been derived for fully bi-directional monomolecular reactions^{36,37}. In spite of these advances, more work is still needed to address the issue of accuracy for modelling multistep reaction systems. In this work we will develop a new method using chemical reaction with state-dependent time delay to simplify multistep reaction systems accurately. The proposed method will be validated by two models for the degradation process of mRNA molecules and one model for multistep metabolic synthesis pathway.

Results

Mathematical model. This study considers the following system with a series of chemical reactions:



where X_i represents the i -th state of a molecule with copy number x_i and k_i is rate constant. Here “ P ” is the product, and it may also be “ $()$ ” if it is a degradation process. Denote s as the total copy number of molecules in all states, namely $s = \sum_{i=1}^n x_i$. The dynamics of system (1) can be described by an ordinary differential equation (ODE) model in Supplementary Information. For simplicity, it is assumed that rate constant in each step is the same (i.e. $k_i = k$). Then the exact solution of this ODE model can be derived analytically. In particular, the total molecule number is represented by

$$s = s_0 e^{-kt} + (s_0 - x_{n0}) k t e^{-kt} + (s_0 - x_{(n-1)0} - x_{n0}) \frac{(kt)^2}{2!} e^{-kt} + \dots + \frac{x_{10}}{(n-1)!} (kt)^{n-1} e^{-kt} \quad (2)$$

where $s_0 = \sum_{i=1}^n x_{i0}$. We assume that the initial condition satisfying $x_{20} = x_{30} = \dots = x_{n0} = y_0 / (n-1)$ with $y_0 = \sum_{i=2}^n x_{i0}$. Then the total molecule number is represented by

$$s = e^{-kt} \left\{ (x_{10} + y_0) \left[1 + kt + \dots + \frac{(kt)^{n-1}}{(n-1)!} \right] - \frac{y_0 kt}{n-1} \left[1 + kt + \dots + \frac{(kt)^{n-2}}{(n-2)!} \right] \right\}. \quad (3)$$

Using the Taylor polynomial and remainder term of exponential function e^{kt} , the above solution is approximated by

$$s = e^{-kt} \left\{ (x_{10} + y_0) \left[e^{kt} - \frac{(kt)^n}{n!} e^{k\xi_1} \right] - \frac{y_0 kt}{n-1} \left[e^{kt} - \frac{(kt)^{n-1}}{(n-1)!} e^{k\xi_2} \right] \right\}, \quad (4)$$

where $\xi_1, \xi_2 \in (0, t)$ are functions of time t in the remainder terms. If the number of reactions n is large, such as the model of mRNA degradation in which the decay dynamics is described by a chain of eight-step, poly (A)-shortening reactions and one-step terminal deadenylation reaction (namely $n=9$)³², we further assume that $\xi_1 = \xi_2 = \xi$.

We use a reaction with time delay to simplify the multistep process (1), which is described as follows:



Here reaction (5) represents the first reaction of system (1), while delayed reaction (6) is a simplification of the process from state X_2 to product P . Time delay in reaction (6) is the sum of waiting time experiencing $n-1$ consecutive reactions from state X_2 to product P . Thus the imaginary state G represents anyone of the intermediate states X_2, \dots, X_n and its molecular number is $y = \sum_{i=2}^n x_i$. Here reactions (5) and (6) are termed as consuming and nonconsuming reactions³⁸. In this work we consider systems in which the rate constants are relatively close to each other. If the multistep reaction chain involves different time-scales, a two-step reaction system will be a better approach to approximate the multistep reaction system.

State dependent time delay. To demonstrate the dependence of time-delay on system state, we use Supplementary Algorithm 1 to numerically calculate the value of time delay under various initial conditions. We first test the case with different values of initial molecular number x_{10} but fixed $y_0 (=0)$. Figure 1(A) provides three

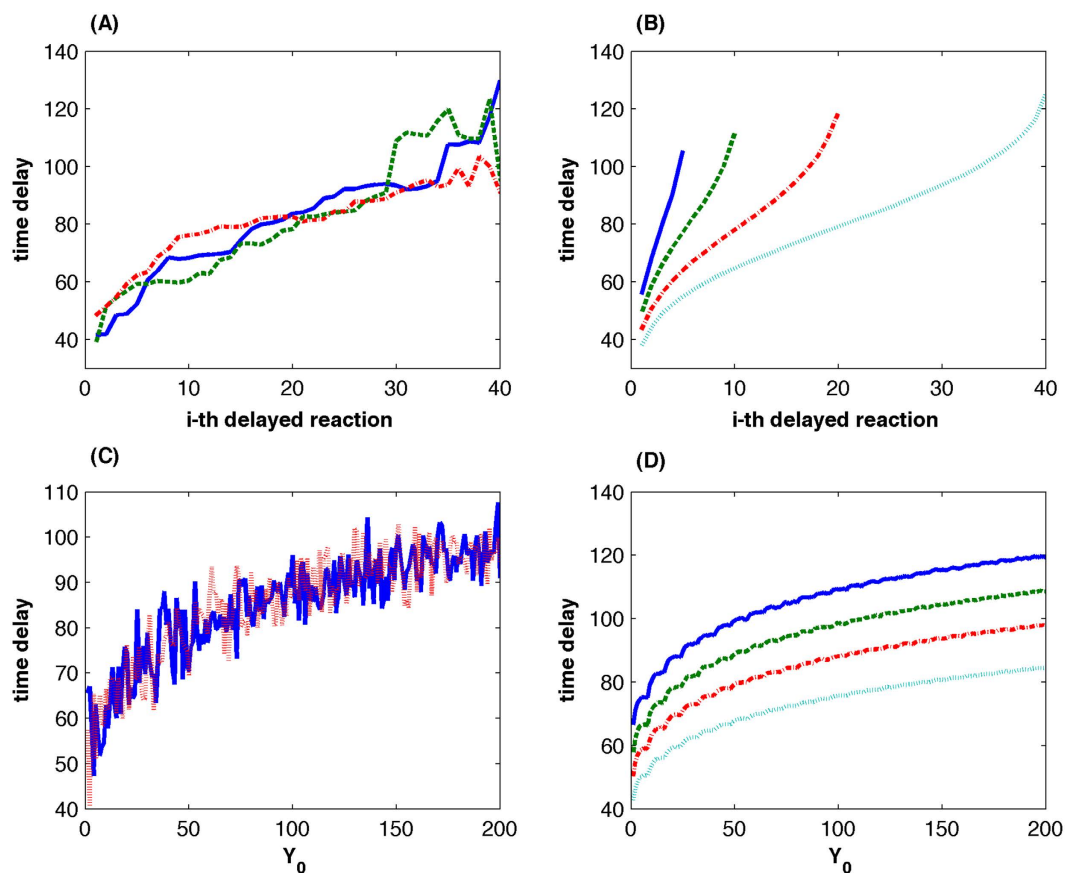


Figure 1. Calculated time delay using stochastic simulations of the multistep reactions process (1).

(A) Three simulations of time delay for the decay of each molecule using initial molecular numbers $x_{10} = 20$ and $y_0 = 0$. (B) Averaged time delay using 1000 simulations for the decay of each molecule based on different initial number x_{10} but null initial imaginary species y_0 . Index i means the delay of the i -th molecule. (C) Two simulations of time delay for the decay of the first molecule using initial molecular number $x_{10} = 20$ and different values of y_0 . (D) Averaged time delay using 1000 simulations for the decay of the first molecule based on different values of x_{10} and y_0 . ((B,D): Solid-line: $x_{10} = 5$, dash-line: $x_{10} = 10$, dash-dot-line: $x_{10} = 40$, dot-line: $x_{10} = 20$).

stochastic simulations of time delay using $x_{10} = 20$. We also obtain 1000 stochastic simulations of time delay and present the averaged values in Fig. 1(B) based on $x_{10} = 5, 10, 20, 40$. For each initial condition, the value of time delay increases when the total molecular number decreases. The reason is that a smaller value of propensity function may lead to a larger waiting time of chemical reactions. For example, when $x_{10} = 40$, the time delay for the decay of the first molecule is $t = 38.2$, while that for the last molecule is $t = 125.3$. Similarly, if the initial molecular number x_{10} is larger, the delay time for the molecule of the same order is smaller. These results clearly suggest that the value of time delay depends on the values of propensity functions that are determined by the system state.

To further demonstrate the dependence of time delay on the imaginary molecule, we use Supplementary Algorithm 2 to calculate time delay for the decay of the first molecule based on different x_{10} and $y_0 (> 0)$. In this algorithm the value y_0 is transferred to the initial molecule numbers ($x_{20}, x_{30}, \dots, x_{n0}$). It is assumed that the initial number x_{i0} satisfies $x_{20} \geq x_{30} \geq \dots \geq x_{n0}$ and the difference between these numbers is at most 1. This requirement is consistent with the assumption in the deterministic setting ($x_{20} = x_{30} = \dots = x_{n0}$). For example, if $n = 5$ and $y_0 = 6$ the initial system state is $(x_{10}, \dots, x_{50}) = (x_{10}, 2, 2, 1, 1)$ for a given initial molecule number x_{10} . Figure 1(C) gives two stochastic simulations of time delay for the decay of the first molecule based on $x_{10} = 20$ but different values of y_0 . The averaged value of time delay in Fig. 1(D) using 1000 stochastic simulations suggests that time delay also depends on the number of imaginary species in the system.

Formula for calculating time delay. Methods section derives a formula for calculating time delay based on a given system state. However, an unsolved question is the value of C_2 in (25) explicitly includes time t . Here we find an approximation of C_2 through numerical computation. We first search for the optimal value of C_2 for different values of x_{10} and y_0 using the derived expression (25) to match the calculated time delay in Fig. 1(D). Note that this computation is based on different values of $x_{10} (= 5, 10, \dots, 100)$ but Fig. 2(A) only shows the results for 4 values of x_{10} . Each line in Fig. 2(A) represents the optimal value of C_2 for a particular value of x_{10} but different values of y_0 . For a fixed value of y_0 , the smaller the value of x_{10} is, the smaller the value of C_2 becomes. The optimal

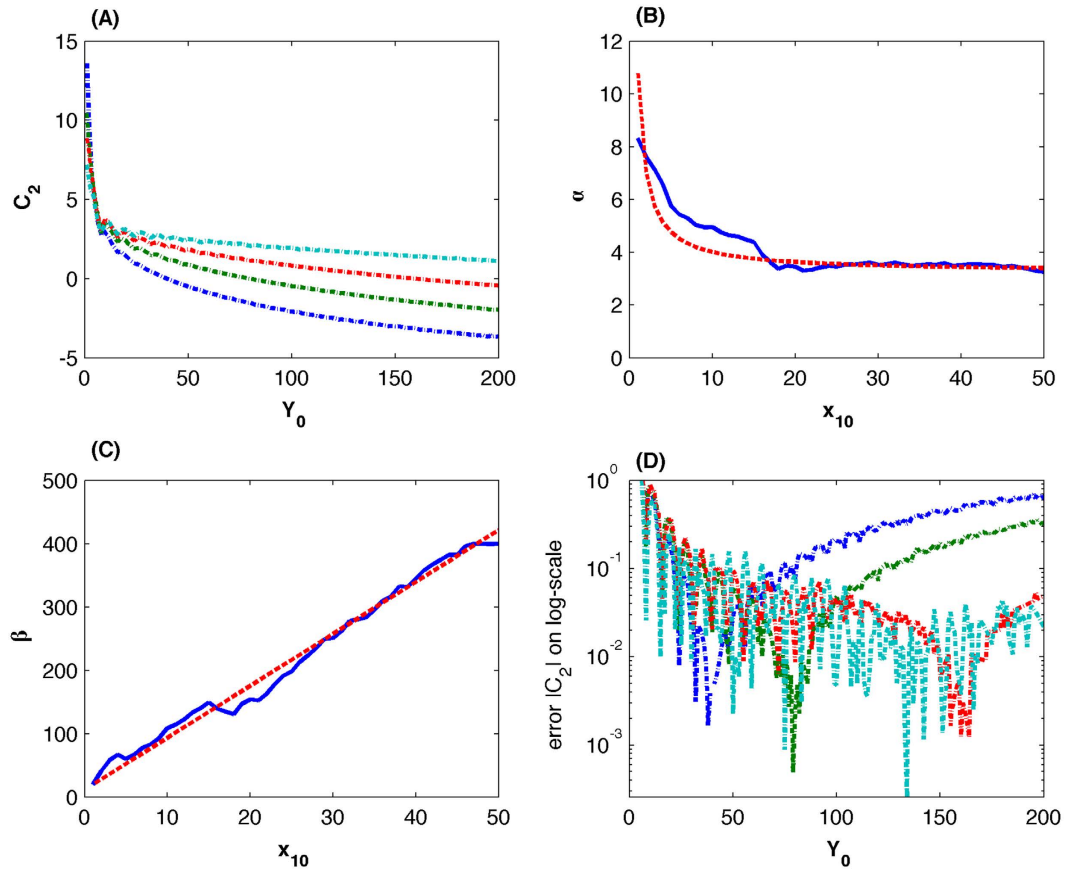


Figure 2. Algorithm for calculating time delay that is dependent on system state. (A) Estimated optimal values of C_2 based on different system states (x_{10}, y_0) that match time delay showing in Fig. 1(D). (Solid-line: $x_{10} = 5$, dash-line: $x_{10} = 10$, dash-dot line: $x_{10} = 20$, dot-line: $x_{10} = 50$). (B) Estimated values of α based on different values of x_{10} . (Solid line: estimated values based on simulated time delay in Fig. 1(B); dash line: $\alpha = 3.25 + 7.5/x_1$). (C) Estimated values of β based on different values of x_{10} . (Solid line: estimated values based on simulated time delay in Fig. 1(B); dash line: $\beta = 11.8 + 8.2x_1$). (D) Difference between the predicted values of C_2 and optimal values of C_2 in Fig. 2(A) on a logarithmic scale. (Solid-line: $x_{10} = 5$, dash-line: $x_{10} = 10$, dash-dot line: $x_{10} = 20$, dot-line: $x_{10} = 50$).

value of C_2 in Fig. 2(A) suggests that it is a monotonically decreasing function of y_0 . In addition, the value of C_2 is $-1/y$ when $y = x_1 = (n - 1)$. Thus it is assumed that

$$C_2 = \frac{\alpha(x_1(n - 1) - y)}{\beta + y} - \frac{1}{y}. \tag{7}$$

To determine the values of α and β , we further estimate these values by matching the determined time delay using equation (25) with those shown in Fig. 1(D). The estimated values in Fig. 2(B,C) suggest that the values of α and β may also be functions of x_1 . Based on the values in Fig. 2(B,C), we use the following two functions to approximate α and β , namely $\alpha = 3.25 + 7.5/x_1$ and $\beta = 11.8 + 8.2x_1$. Thus the final expression of the approximated C_2 is

$$C_2 = \frac{\left(3.25 + \frac{7.5}{x_1}\right)(x_1(n - 1) - y)}{11.8 + 8.2x_1 + y} - \frac{1}{y} \tag{8}$$

To validate the proposed approach (8), we compare the optimal value of C_2 in Fig. 2(A) with that determined by (8). Figure 2(D) shows the difference between these two values under different values of x_1 and y on a logarithmic scale. The optimal value of C_2 for small value of y_0 in Fig. 2(A) is larger than that for large value of y_0 . The estimated value of α from $\alpha = 3.25 + 7.5/x_1$ in Fig. 2(B) matches the optimal value of α very well only for large value of y_0 . Thus, when y_0 is small, the error of C_2 is relatively large, but it is still quite small, which suggests that approach (8) provides accurate approximation to the optimal value of C_2 .

Time delay model for mRNA degradation. The degradation process of mRNA molecules is a typical multistep reactions system. In experimental studies, a large sample of cells are genetically modified or treated

with inhibitors to stop transcription and thus kinetic information of a decaying mRNA species can be obtained³⁹. Recently single-cell and single-molecule techniques have advanced our understanding of mRNA turnover⁴⁰. The accuracy of decay measurement varies with the technique used^{41–43}. A detailed mechanistic model has been designed to describe the degradation process exactly^{44,45}. In addition, a simplified model of multistep reactions was proposed by combining a number of terminal deadenylation reactions into a single reaction³². However, it is difficult to derive accurate information of half-life from detailed mechanistic models.

Next we apply our state-dependent delay model to study the mRNA degradation process of gene ribosomal protein L30 (*RPL30*). Experimental studies have demonstrated the transcript decay dynamics of two constructs for this gene, namely construct A-ACT1 UAS (upstream activating sequence) and construct B-RPL30 UAS⁴⁶. In experiments, mRNA molecule decaying dynamics was monitored after blocking transcription by using drug 1,10-phenanthroline⁴⁶. Therefore, it is assumed that no further transcription occurs after drug application. Since there is no explicit information regarding the mRNA copy number, we test the case with initial total mRNA number $s_0 (=100)$.

Here we use the delayed reactions (5, 6) to represent the degradation dynamics, where X_1 is mRNA molecule with full length of poly(A)-tail and imaginary species G represents transcripts in the poly(A)-shortening process. The initial number of imaginary species y_0 and degradation rate k are unknown parameters that will be estimated to match experimental data. In addition, the manifesting time of these initial imaginary species is uniformly distributed in time interval $[0, MT]$ and

$$MT = \text{delay}(x_{10}, y_0, k, n)/D \quad (9)$$

where $\text{delay}(x_{10}, y_0, k, n)$ is the time delay determined by the initial system state (x_{10}, y_0) , degradation rate k , and number of steps n in system (1). For initial imaginary species, a few reactions of the multistep reactions may have already taken place, and the time to reach the product should be smaller than the calculated time delay $\text{delay}(x_{10}, y_0, k, n)$. Thus we use a factor ($D > 1$) to adjust the time delay of initial imaginary species. We use an Approximate Bayesian Computation (ABC) rejection sampling algorithm⁴⁷ to search for optimal parameters of y_0 , k and D . The time delay of each delayed reaction is calculated based on the current system state (x_1, y) . We select 150 sets of the inferred model parameters with smaller error and use the set with minimal error as our final estimation.

Based on 1000 simulations, Fig. 3 shows that the state-dependent delay model is able to provide accurate description of mRNA degradation dynamics for the two constructs of gene *RPL30*. We also present simulation results of the one-step model and two-variable model that were shown in ref. 5. For each model, we use absolute error to measure the difference between simulation and experimental data. In addition, we provide error bars of experimental data and count the number of observation time points for each simulation that is beyond the error bar. Compared with the one-step and two-variable models, STable 1 suggests that this new model with state-dependent delay provides more accurate simulations to the experimental data. Distributions of inferred parameters in SFigure 1(A and D) suggest that ~25% of initial mRNA molecules are imaginary species, namely the transcripts in the poly(A)-shortening process. In addition, distributions of value D in SFigure 1(B and F) suggest that the degradation time points of these shortened transcripts are distributed in an interval that is ~60% of the normal time delay interval. Thus these imaginary species may already exist in the middle of the shortening process.

Simulation results in Fig. 3 are based on the assumption that the initial total transcripts number is $s_0 = 100$. The next question is whether this assumed initial total mRNA number influences the estimated model parameters. To answer this question, we simulate the delay model using the estimated parameters and different initial total mRNA $s_0 [=10, 50, 150, 200]$. The rate constant k and parameter D remain unchanged, but the value of y_0 is rescaled according to s_0 to maintain a fixed ratio y_0/s_0 . Simulation results in SFigure 2 suggest that our estimated parameters can also produce accurate simulations for various initial mRNA numbers. The difference between simulation and experimental data is relatively large when the initial total mRNA number is small.

Time delay model in gene expression. The success of our proposed method for simplifying a multistep reaction system leads to the next study to model the expression of a cell cycle-regulated gene (e.g. *SWI5*) based on the measured change in the mRNA turnover during a cell cycle⁴⁸. *SWI5p* is a transcription regulator of late mitosis genes and its expression is tightly regulated during the cell cycle. It was measured to degrade with 8 min half-lives⁴³. In addition, *NDD1* (nuclear division defective) is an essential gene for the expression of gene *SWI5*. It has been shown that overexpression of *NDD1* enhances the expression of *SWI5*⁴⁹. The expression of gene *NDD1* peaks during the S phase and is essential for the expression of its target gene *SWI5* during the G2/M phase^{49,50}.

A simple mathematical model has been proposed to describe the expression of gene *SWI5* based on experimental data measured in single cells. In this model the degradation of mRNA molecules is described by a one-step reaction and simulation is used to measure the half-life of mRNA molecules⁴⁸. To accurately measure the half-life of mRNA transcripts, we propose a delayed model to describe the expression of gene *SWI5*. It is assumed that the transcription of this gene is activated by TF *NDD1* with transcriptional rate

$$k_1 = \frac{a[NDD1]}{b + [NDD1]} \quad (10)$$

where a and b are parameters for genetic regulation. In addition, the elongation process needs time for RNAP II polymerase travelling along the template DNA. Due to the fixed number of DNA in a single cell, it is assumed that the time of elongation process is relatively fixed and a delay reaction with constant time delay is used to represent the synthesis of mRNA transcripts. Then mRNA transcripts translocate from nucleus to cytosol, which is also

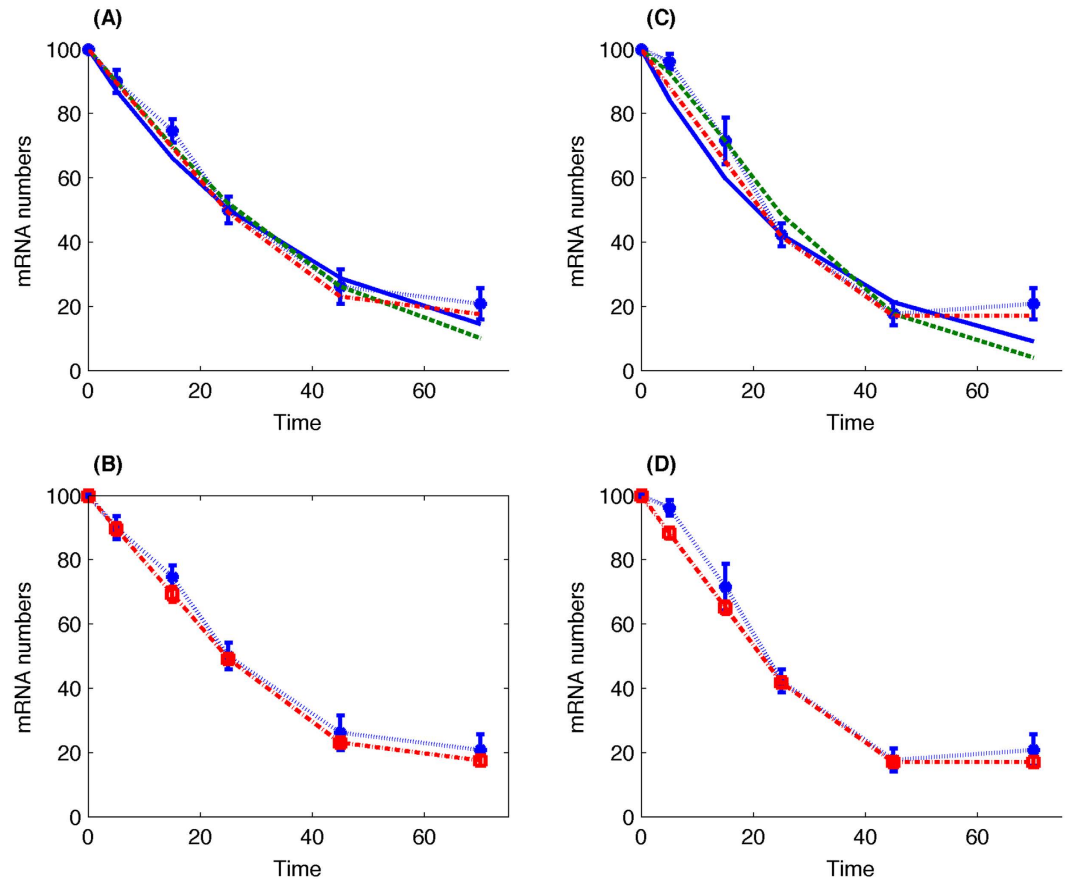
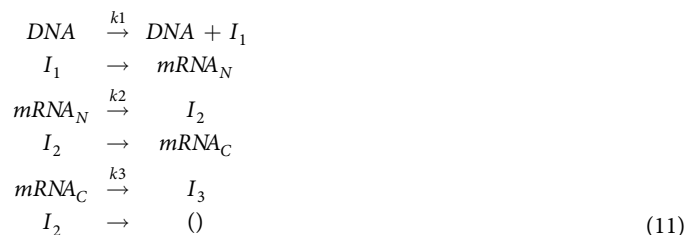


Figure 3. Simulation of mRNA degradation for gene *RPL30* using the state-dependent delay model.

Numerical results are the averaged molecular numbers using 1000 stochastic simulations. (A,B) Construct ACT1 using estimated parameters $k=0.1260$, $y_0=23$, $D=1.7184$. (C,D) Construct RPL30 using estimated parameters $k=0.1260$, $y_0=17$, $D=1.7525$. (A,C) Star-dash line with error bar: experimental data. Solid-line: one-step model; dash-line: two-variable model. dash-dot line: time-dependent delay model). (B,D: star-dash line with error bar: experimental data, square-dash line with error bar: time-dependent delay model).

modelled by a delay reaction with constant time delay for simplicity. Finally mRNA molecules in cytosol decay via a multistep process that is simplified as a state-dependent delay reaction (5, 6). Thus the proposed model for the expression of gene *SWI5* is given below:



where $mRNA_N$ and $mRNA_C$ are mRNA molecules in nucleus and cytosol; I_1 , I_2 , and I_3 are imaginary species for $mRNA_N$, $mRNA_C$, and shortening mRNA, respectively. We use the inferred concentration of [NDD1] in ref. 51 as the activity of this TF, which is consistent with the drafted TF activity in ref. 48. In addition, experimental studies show that gene expression is regulated by mechanisms of cell cycle. In yeast, the mitosis process at ~ 49 min of each cell cycle terminates the process of transcription⁵². This regulatory mechanism is realized by the assumption that the activity of [NDD1] is zero after 49 min of each cell cycle.

The measured mRNA copy numbers in single cells⁴⁸ are used to infer regulation parameters a , b , rate constant k_3 , and transcription and translocation delays. We use the ABC rejection sampling algorithm to search for optimal model parameters. Using simulation error to both cytosol and nucleus data as the criterion, we select 150 sets of model parameters with smaller simulation error. The parameter set with the minimal error is the final inference result. Figure 4 shows that the numerical simulation matches experimental data very well. In addition, the distribution of transcriptional time delay in SFigure 3(A) is consistent with the experimental estimations showing

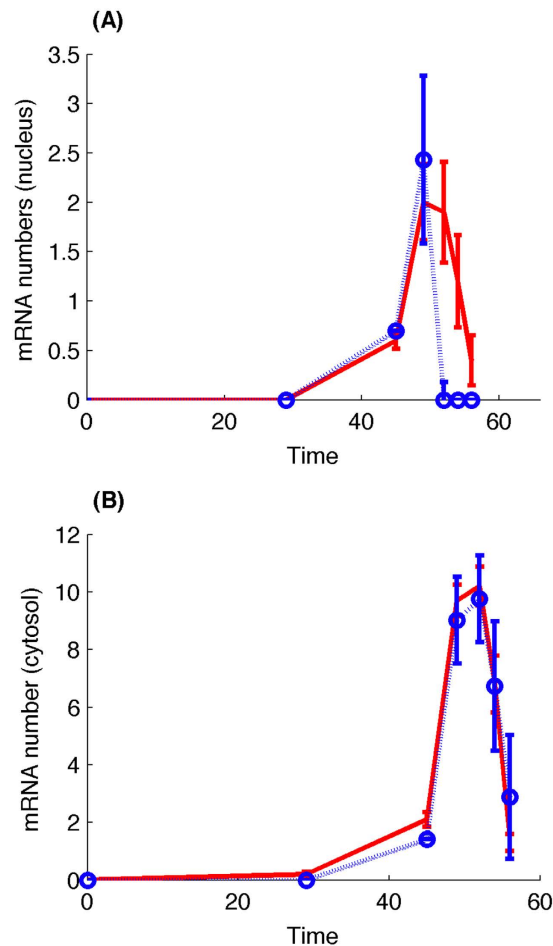
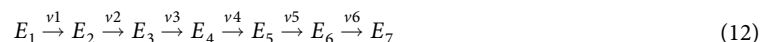


Figure 4. Simulation of gene transcription for gene SWI5 using the state-dependent delay model.

Numerical results are the averaged molecular numbers using 1000 stochastic simulations. (A) mRNA copy number in nucleus. (B) mRNA copy number in cytosol. (dot line: experimental data; solid line: simulations). Estimated parameters are $a = 9.148$, $b = 3.390$, $\tau_1 = 46.665$, $\tau_2 = 0.733$, $k_2 = 2906.04$, $k_3 = 1.297$.

that the time delay in transcription is ~ 35 min⁵³. An interesting observation is the degradation rate of mRNA is $k_3 = \sim 1.29/\text{min}$.

Multistep pathway of metabolic synthesis. The previous two systems are used to study the accuracy of state-dependent time delay for simplifying multistep reaction processes with a fixed rate constant. The next question is whether the state-dependent time delay is able to approximate multistep reactions with varied reaction rates with good accuracy. Here we discuss a system that is a simplified representation of the pathway for aliphatic glucosinolate biosynthesis¹¹. This system considers the chain elongation process as a series of sequential events. Each chain elongation cycle is simplified into a one step reaction $E_i \rightarrow E_{i+1}$. Here we only consider multistep reactions in wild-type cells and thus the process for reducing the conversion of chain elongated 2-oxo acids into final glucosinolates is not included in this model¹¹. This pathway contains six reactions



where E_1 is the input and E_7 the product. In each step the reaction rate is a Michaelis-Menten function

$$v_i = \frac{V_i}{K_i + E_i} E_i \quad (13)$$

where V_i and K_i are the maximal synthesis rate and equilibrium constant, respectively. The detailed values of these parameters in terms of concentrations are given in ref. 11. Since a plant cell stretches from 10 to 100 micrometers, we assume that the size of a plant cell is 4×10^{-14} liter. Thus a concentration of $1 \mu\text{M}$ is about 25600 molecules⁵⁴. The values of V_i and K_i are converted into those with unit of molecular numbers as $(V_1, \dots, V_6) = (37.07, 38.27, 73.44, 35.84, 9.31, 2.08)$ and $(K_1, \dots, K_6) = (23859200, 12185600, 11852800, 9164800, 6476800, 2073600)$. For this process of six-step reactions, we use a delayed reaction to simplify the model. Here the first reaction $E_1 \xrightarrow{v_1} U$ remains unchanged. For the delayed reaction $U \rightarrow E_7$, we use (25) to calculate time delay using reaction rate

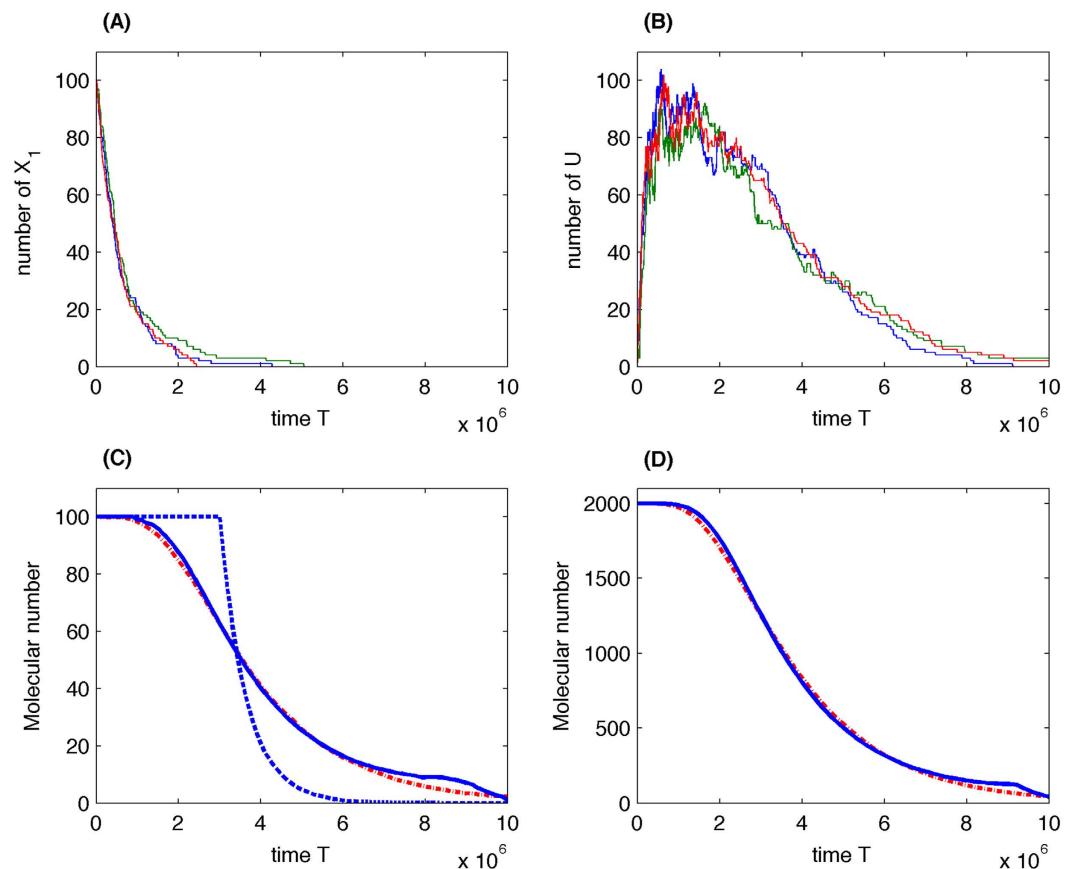


Figure 5. Simulation of substrate competition of metabolic pathway. (A) Three stochastic simulations of E_1 . (B) Three stochastic simulations of $\sum_{i=2}^6 E_i$. (C) Simulations of metabolic pathway with initial condition $E_1 = 100$ (Solid-line: multistep reaction pathway; dash-dot line: delayed reaction with state-dependent time delay; dish-line: delayed reaction with constant time delay). (D) Simulations of metabolic pathway with initial condition $E_1 = 2000$ (Solid-line: multistep reaction pathway; dash-dot line: delayed reaction with state-dependent time delay).

$$k = \frac{\bar{V}}{\bar{K} + U/5} \quad (14)$$

where $U/5$ represents the averaged molecular number of E_i ($i = 2 \sim 6$). We use $\bar{V} = 15.22$ and $\bar{K} = 5481279.50$ are the harmonic mean of parameters V_i and K_i ($i = 2 \sim 6$), respectively. For example, the value of \bar{V} is given by

$$\bar{V} = \frac{5}{\frac{1}{V_2} + \frac{1}{V_3} + \frac{1}{V_4} + \frac{1}{V_5} + \frac{1}{V_6}}. \quad (15)$$

Figure 5 gives three stochastic simulations of the multistep pathway using the SSA and initial condition $E_1 = 100$ and $E_i = (i = 2 \sim 6)$. Simulated values of E_1 in Fig. 5(A) show that the molecules are converted into molecules E_2 soon but the molecular number of $\sum_{i=2}^6 E_i$ in Fig. 5(B) stays in the high level for quite a long time. Figure 5(C) gives the averaged total molecule number $\sum_{i=1}^6 E_i$ over 1000 simulation using the multistep pathway and state-dependent time delay. Numerical results suggested that the state-dependent time delay reaction approximates the multistep reactions accurately. Simulations with an initial molecular number $E_1 = 2000$ in Fig. 5(D) confirm this result. We have also tested the accuracy of delayed reaction with constant time delay. Simulation in Fig. 5(C) using constant delay ($=3000000$) suggests that the total molecular numbers stays at 100 before $t = 3000000$ and then decreases much quicker than that obtained by using state-dependent time delay. Simulations suggest that it is not appropriate to use constant time delay reaction to approximate this test system.

Conclusions

In this work, we propose a new algorithm to calculate time delay in chemical reaction systems according to the system state. Using the process of multistep reaction systems as the test problem, we utilize both analytical solution of ODE model and stochastic simulation of chemical reaction systems to determine the relationship between the system state and value of time delay. The proposed method is applied to model the degradation process of mRNA molecules based on experimental data measured in single cells and a multistep pathway for metabolic

synthesis. For the first test system of mRNA degradation, our model gives simulations with better accuracy than those of the existing models. For the second test system of gene expression, our model provides simulated dynamics with very good accuracy for both synthesis and degradation of mRNA transcripts. Simulation of the third system suggests that the state-dependent time-delay can be applied to approximate multistep reactions with non-linear reaction rates with very good accuracy. Simulation results in this work suggest that the proposed method is an effective approach to approximate multistep reaction systems more accurately. Compared with the full multistep reaction model, it is also an efficient approach to reduce computing time of stochastic simulation, save computer storage, and decrease the number of unknown parameters that should be estimated from experimental data.

Half-life is an important concept to measure the degradation process in biological studies. It is the time required for the amount of a species to fall to a half of the initial value. Based on the widely used assumption that the quantity follows an exponential decay, the decay rate constant can be determined by the half-life value or vice versa. However, for many biological molecules such as mRNA transcripts, the decaying dynamics may not follow an exponential process; rather it may be an event of multistep reactions. Thus, molecules in the intermediate states may also be important to determine the value of half-life. This may be a reason to explain the difference between the determined half-life values under different experimental conditions. Using the inferred degradation rate in the state-dependent delay model, our results suggest that more work is needed to establish the relationship between half-life and degradation rate constant of biological species.

This work is based on the assumption that rate constants in the multistep reactions are the same. However, rate constants in biological systems are usually different from each other. We have conducted further computation for testing the influence of parameter variation on the simulation accuracy. Numerical results suggested that the difference between simulation using the same rate constant and that using different rate constants is proportional to the variance of rate constants. Thus, the proposed method with state-dependent time delay is applicable to model pathways in which the rate constants are relatively close to each other, such as the three systems discussed in this work. However, if the multistep reaction chain involves different time-scales, namely the difference between the rate constants is large, we may need to use multiscale approaches by dividing the pathway into two or more subsystems. Then a two-step delayed reaction system may be a better approach to approximate the multistep reaction system.

Using the multistep reaction system as the test system, this work represents a further step forward in developing accurate delayed models for chemical reaction systems. However, more research work is still strongly needed to study other types of multistep reaction systems as well as the complex systems that include multistep reactions processes as subsystems. For genetic regulation, for example, it would be important to study TF regulation by including DNA/mRNA/protein interactions and also explore the mechanisms of transcriptional elongation. In addition, the proposed approach is based on the mass action law kinetics. Delay models based on other approaches, such as the Hill function for catalyzing enzyme kinetics or Shea-Ackers model for genetic regulation, would also be interesting research problems. Another significant challenge is the possible large variations of estimated parameter values that all can faithfully realize experimental data, in particular for inferring unknown parameters in stochastic models. Currently a number of approaches use important system properties (e.g. robustness property) as additional criteria to select estimation candidature^{2,55}. More system properties and research work are needed to address this issue. All these interesting problems will be potential topics of future research.

Methods

State-dependent time delay. For delayed model (5,6), we need to determine the value of time delay based on current system state (X_1, G) with molecule numbers (x_1, y) . When the first reaction fires, a molecule of X_1 moves into the queue structure of time delay L in which there are already y imaginary molecules. When the newly added molecule turns to product P , it is assumed that all y molecules queued before the newly added molecule already turn into the product. Thus, when the first molecule from X_1 state turns to product, the total molecule number should be reduced from $x_1 + y$ to $x_1 - 1$. The time required for this process, namely time delay, is defined as

$$\tau = \tau_2 - \tau_1 \quad (16)$$

where τ_1 is the firing time of the first reaction $X_1 \rightarrow X_2$, and τ_2 is the firing time of the last reaction $X_n \rightarrow P$ to let the system state be $s = x_1 - 1$.

We use computational simulations to determine the value of time delay. Here τ_1 is determined by the stochastic simulation algorithm (SSA)⁵⁶. The key issue is how to determine the value of τ_2 . Given a system state (x_1, y) at time t , the time τ_2 for the first X_1 molecule turns into product P is

$$\begin{aligned} x_1 - 1 &= e^{-k\tau_2} \left\{ (x_1 + y) \left[e^{k\tau_2} - \frac{(k\tau_2)^n}{n!} e^{k\xi} \right] - \frac{yk\tau_2}{n-1} \left[e^{k\tau_2} - \frac{(k\tau_2)^{n-1}}{(n-1)!} e^{k\xi} \right] \right\} \\ &= (x_1 + y) - \frac{yk\tau_2}{n-1} + \frac{(k\tau_2)^n}{n!} e^{k(\xi-\tau_2)} \left[\frac{ny}{n-1} - (x_1 + y) \right] \end{aligned} \quad (17)$$

which can be simplified as

$$e^{k(\xi-\tau_2)} (k\tau_2)^n \left((x_1 + y) - \frac{ny}{n-1} \right) = \left[1 + y \left(1 - \frac{k\tau_2}{n-1} \right) \right] n!. \quad (18)$$

Denote $C_1 = x_1 + y - ny/(n-1)$, $C_2 = 1 - k\tau_2/(n-1)$, $C = (1 + C_2y)n!/C_1$, the above equation is simplified as

$$e^{k(\xi-\tau_2)}(k\tau_2)^n = C \quad (19)$$

There are a number of undetermined coefficients in the above equation. Thus, we first use a special case to determine the value of ξ by letting $y=0$. In this case, $C_1 = x_1$, $C = n!/C_1$. We rewrite the above equation as

$$e^{k(\xi-\tau_2)}\tau_2^n = \frac{n!}{x_1 k^n}, \quad (20)$$

Here $\xi \in (0, \tau_2)$ is a function of τ_2 . To determine the optimal value of ξ , we compared the time delays obtained by using a number of values [$\xi = (0, 0.1, \dots, 1)\tau_2$] in (20) with those obtained from stochastic simulations. We found that, when $\xi = \tau_2/2$, formula (20) provides more accurate estimate for time delay than other values. In this case, it becomes

$$e^{-k\tau_2/2}\tau_2^n = \frac{n!}{x_1 k^n} \quad (21)$$

Using the Lambert W function, the solution of τ_2 is given by

$$\tau_2 = -\frac{2n}{k} W\left[-\frac{1}{2n} \left(\frac{n!}{x_1}\right)^{\frac{1}{n}}\right]. \quad (22)$$

Next we return to the general case when $y > 0$. When $C_1 = 0$, the left hand side of equation (18) is zero. By letting the right-hand side be zero, we have that

$$\tau_2 = \frac{1 + x_1 n - x_1}{k x_1}. \quad (23)$$

When $C_1 \neq 0$, using the optimal value $\xi = \frac{t}{2}$, the time to reach the system state with $x_1 - 1$ molecules is

$$\tau_2 = -\frac{2n}{k} W\left[-\frac{1}{2n} C^{\frac{1}{n}}\right]. \quad (24)$$

In summary, we have the following expression for calculating time delay $\tau = \tau_2 - \tau_1$, where the value of τ_1 is determined by the SSA, and

$$\tau_2 = \begin{cases} \frac{1 + x_1 n - x_1}{k x_1} & \text{if } C_1 = 0 \\ -\frac{2n}{k} W\left[-\frac{1}{2n} C^{\frac{1}{n}}\right] & \text{if } C_1 \neq 0 \end{cases}, \quad (25)$$

Note that C_2 is dependent on the values of x_i , y , and time t , which is determined in Results section by numerical simulations.

SSA with state-dependent time delay. This work proposes the following modelling framework with state-dependent time delay to simplify multistep reaction events. Here we need to simulate a well-stirred mixture of $N(\geq 1)$ molecular species $\{X_1, \dots, X_N\}$ that chemically interact inside some fixed volume Ω at a constant temperature and through M reaction channels $\{R_1, \dots, R_M\}$, which includes M_1 elementary reactions and M_2 delayed reactions ($M = M_1 + M_2$). Here a delayed reaction may be a reaction with constant time delay, distributed delay that follows a distribution, or state-dependent time delay that is simplified from the lumped multistep chemical reactions (1). The system state is denoted as $X \equiv \{x_1, \dots, x_N\}^T$ where $x_i(t)$ is the copy number of species X_i . For each delayed reaction, we define an imaginary species G_i to represent the corresponding intermediate species. We also define a stoichiometric vector ν_j for non-delayed reactions, as well as consuming and manifesting stoichiometric vectors ν_j and u_j for delayed reactions (5) and (6), respectively. For each reaction channel, a propensity function $a_j(X)$ is defined and $a_j(X)dt$ represents the probability that this reaction will fire inside Ω in the next infinitesimal time interval $[t, t + dt]$. Detailed algorithm is given below.

Algorithm: State-dependent Delay SSA (SD-SSA)

Set initial molecular numbers at $t=0$, and an empty queue structure L for storing the information of delayed reactions.

Step 1. Calculate propensity functions $a_i(X)$, $i = 1, \dots, M$ and $a_0 = \sum_{i=1}^M a_i(X)$.

Step 2. Generate a uniform random number $r_1 \in U(0, 1)$ and determine the waiting time μ of next reaction by

$$\mu = \frac{1}{a_0} \ln \frac{1}{r_1} \quad (26)$$

where $r_1 \sim U(0,1)$.

Step 3. Compare μ with the least update time δ_{min} in the queue structure L to check whether there is any delayed reaction that is scheduled to finish within $[t, t + \mu)$.

Step 4. IF $\delta_{min} < \mu$ (update the delayed reaction with index k at δ_{min})

$$X(t + \delta_{min}) = X(t) + u_k \quad (27)$$

ELSE: Generate a sample $r_2 \sim U(0,1)$ to determine the index j of next reaction

$$\sum_{k=1}^{j-1} a_k(X) < r_2 a_0 \leq \sum_{k=1}^j a_k(X) \quad (28)$$

update the system state by

$$X(t + \mu) = X(t) + v_j. \quad (29)$$

If reaction with index j is a delayed reaction, use the constant delay; or generate a sample for the distributed delay reaction; or use (25) to calculate the delay value τ if R_j is a reaction with state-dependent time delay. Then add index j and update time $\delta = t + \mu + \tau$ to the queue structure L .

Step 5. Go to Step 2.

Note that this algorithm is based on the so-called rejection delay-SSA⁵⁷. A more precise algorithm can be developed if we consider the change of propensity functions due to the update of a delayed reaction in step 2⁵⁸. In addition, the calculated value of τ_2 is deterministic but the value of time delay in Fig. 1(A,C) is stochastic. Similar to other approaches using distributed delay⁵⁹, we can generate random samples of τ_2 from a random variable whose mean is the calculated deterministic value.

References

1. Le Novère, N. Quantitative and logic modelling of molecular and gene networks. *Nat Rev Genet* **16**, 146–158 (2015).
2. Villaverde, A. F. & Banga, J. R. Reverse engineering and identification in systems biology: strategies, perspectives and challenges. *J R Soc Interface* **11** (2014).
3. Ghosh, S., Matsuoka, Y., Asai, Y., Hsin, K. Y. & Kitano, H. Software for systems biology: from tools to integrated platforms. *Nat Rev Genet* **12**, 821–832 (2011).
4. Mackey, M. C., Santillan, M., Tyrann-Kaminska, M. & Zeron, E. S. The utility of simple mathematical models in understanding gene regulatory dynamics. *In silico biology* **12**, 23–53 (2015).
5. Wu, Q. Q., Smith-Miles, K., Zhou, T. S. & Tian, T. H. Stochastic modelling of biochemical systems of multi-step reactions using a simplified two-variable model. *BMC Syst Biol* **7** (2013).
6. Thomas, P., Straube, A. V. & Grima, R. The slow-scale linear noise approximation: an accurate, reduced stochastic description of biochemical networks under timescale separation conditions. *BMC Syst Biol* **6**, 39 (2012).
7. Schnell, S. Validity of the Michaelis-Menten equation--steady-state or reactant stationary assumption: that is the question. *Febs J* **281**, 464–472 (2014).
8. Gilman, A. & Arkin, A. P. Genetic “code”: representations and dynamical models of genetic components and networks. *Annual review of genomics and human genetics* **3**, 341–369 (2002).
9. Posas, F. *et al.* Yeast HOG1 MAP kinase cascade is regulated by a multistep phosphorelay mechanism in the SLN1-YPD1-SSK1 “two-component” osmosensor. *Cell* **86**, 865–875 (1996).
10. Kartal, O., Mahlow, S., Skupin, A. & Ebenhoh, O. Carbohydrate-active enzymes exemplify entropic principles in metabolism. *Mol Syst Biol* **7**, 542 (2011).
11. Schauble, S., Stavrum, A. K., Puntervoll, P., Schuster, S. & Heiland, I. Effect of substrate competition in kinetic models of metabolic networks. *FEBS Lett* **587**, 2818–2824 (2013).
12. Sun, S. H., Klebaner, F. & Tian, T. H. A new model of time scheme for progression of colorectal cancer. *BMC Syst Biol* **8** (2014).
13. op den Buijs, J., van den Bosch, P. P., Musters, M. W. & van Riel, N. A. Mathematical modeling confirms the length-dependency of telomere shortening. *Mechanisms of ageing and development* **125**, 437–444 (2004).
14. Raj, A. & van Oudenaarden, A. Single-Molecule Approaches to Stochastic Gene Expression. *Ann Rev Biophys* **38**, 255–270 (2009).
15. Spiller, D. G., Wood, C. D., Rand, D. A. & White, M. R. H. Measurement of single-cell dynamics. *Nature* **465**, 736–745 (2010).
16. Kaufmann, B. B., Yang, Q., Mettetal, J. T. & van Oudenaarden, A. Heritable stochastic switching revealed by single-cell genealogy. *PLoS Biology* **5**, 1973–1980 (2007).
17. Arkin, A., Ross, J. & McAdams, H. H. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics* **149**, 1633–1648 (1998).
18. Burgess, D. J. Synthetic biology: Cut up to bring together. *Nat Rev Genet* **15**, 365 (2014).
19. Weng, L. *et al.* Rosetta error model for gene expression analysis. *Bioinformatics* **22**, 1111–1121 (2006).
20. Burrage, K., Tian, T. & Burrage, P. A multi-scaled approach for simulating chemical reaction systems. *Prog Biophys Mol Biol* **85**, 217–234 (2004).
21. Bokes, P., King, J. R., Wood, A. T. & Loose, M. Multiscale stochastic modelling of gene expression. *J Math Biol* **65**, 493–520 (2012).
22. Tian, T. H. & Burrage, K. Stochastic models for regulatory networks of the genetic toggle switch. *P Natl Acad Sci USA* **103**, 8372–8377 (2006).
23. Monk, N. A. Oscillatory expression of Hes1, p53, and NF-kappaB driven by transcriptional time delays. *Current Biology* **13**, 1409–1413 (2003).
24. Barrio, M., Burrage, K., Leier, A. & Tian, T. Oscillatory regulation of Hes1: Discrete stochastic delay modelling and simulation. *PLoS Comput Biol* **2**, e117 (2006).
25. Bratsun, D., Volfson, D., Tsimring, L. S. & Hasty, J. Delay-induced stochastic oscillations in gene regulation. *Proc Natl Acad Sci USA* **102**, 14593–14598 (2005).
26. Marquez-Lago, T. T., Leier, A. & Burrage, K. Probability distributed time delays: integrating spatial effects into temporal models. *BMC Syst Biol* **4**, 19 (2010).
27. Brett, T. & Galla, T. Stochastic processes with distributed delays: chemical Langevin equation and linear-noise approximation. *Phys Rev Lett* **110**, 250601 (2013).
28. Thomas, P., Straube, A. V. & Grima, R. The slow-scale linear noise approximation: an accurate, reduced stochastic description of biochemical networks under timescale separation conditions. *BMC Syst Biol* **6** (2012).
29. Srivastava, R., Haseltine, E. L., Mastny, E. & Rawlings, J. B. The stochastic quasi-steady-state assumption: Reducing the model but not the noise. *J Chem Phys* **134** (2011).
30. Ribeiro, A. S. Stochastic and delayed stochastic models of gene expression and regulation. *Math Biosci* **223**, 1–11 (2010).

31. Tian, T. H. Chemical Memory Reactions Induced Bursting Dynamics in Gene Expression. *PLoS One* **8** (2013).
32. Tian, T. H. Simplified stochastic models with time delay for studying the degradation process of mRNA molecules. *Int J Data Min Bioin* **10**, 18–32 (2014).
33. Asher, R. B. & Sebesta, H. R. Optimal Control of Systems with State-Dependent Time Delay. *Int J Control* **14**, 353–& (1971).
34. Cao, Y. L., Fan, J. P. & Gard, T. C. The Effects of State-Dependent Time-Delay on a Stage-Structured Population-Growth Model. *Nonlinear Anal-Theor* **19**, 95–105 (1992).
35. Mier-y-Teran-Romero, L., Silber, M. & Hatzimanikatis, V. The Origins of Time-Delay in Template Biopolymerization Processes. *PLoS Comput Biol* **6** (2010).
36. Leier, A., Barrio, M. & Marquez-Lago, T. T. Exact model reduction with delays: closed-form distributions and extensions to fully bi-directional monomolecular reactions. *J R Soc Interface* **11**, 20140108 (2014).
37. Barrio, M., Leier, A. & Marquez-Lago, T. T. Reduction of chemical reaction networks through delay distributions. *J Chem Phys* **138**, 104114 (2013).
38. Leier, A., Marquez-Lago, T. T. & Burrage, K. Generalized binomial tau-leap method for biochemical kinetics incorporating both delay and intrinsic noise. *J Chem Phys* **128**, 205107 (2008).
39. Passos, D. O. & Parker, R. Analysis of Cytoplasmic Mrna Decay in Saccharomyces Cerevisiae. *Method Enzymol* **448**, 409–427 (2008).
40. Ardehali, M. B. & Lis, J. T. Tracking rates of transcription and splicing *in vivo*. *Nat Struct Mol Biol* **16**, 1123–1124 (2009).
41. Grigull, J., Mnaimneh, S., Pootoolal, J., Robinson, M. D. & Hughes, T. R. Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. *Molecular and Cellular Biology* **24**, 5534–5547 (2004).
42. Holstege, F. C. P. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
43. Wang, Y. L. *et al.* Precision and functional specificity in mRNA decay. *P Natl Acad Sci USA* **99**, 5860–5865 (2002).
44. Cao, D. & Parker, R. Computational modeling of eukaryotic mRNA turnover. *Rna* **7**, 1192–1212 (2001).
45. Cao, D. & Parker, R. Computational modeling and experimental analysis of nonsense-mediated decay in yeast. *Cell* **113**, 533–545 (2003).
46. Bregman, A. *et al.* Promoter Elements Regulate Cytoplasmic mRNA Decay. *Cell* **147**, 1473–1483 (2011).
47. Wu, Q. Q., Smith-Miles, K. & Tian, T. H. Approximate Bayesian computation schemes for parameter inference of discrete stochastic models using simulated likelihood density. *BMC Bioinformatics* **15** (2014).
48. Trcek, T., Larson, D. R., Moldon, A., Query, C. C. & Singer, R. H. Single-Molecule mRNA Decay Measurements Reveal Promoter-Regulated mRNA Stability in Yeast. *Cell* **147**, 1484–1497 (2011).
49. Loy, C. J., Lydall, D. & Surana, U. NDD1, a high-dosage suppressor of cdc28-1N, is essential for expression of a subset of late-S-phase-specific genes in Saccharomyces cerevisiae. *Molecular and Cellular Biology* **19**, 3312–3327 (1999).
50. Veis, J., Klug, H., Koranda, M. & Ammerer, G. Activation of the G(2)/M-specific gene CLB2 requires multiple cell cycle signals. *Molecular and Cellular Biology* **27**, 8364–8373 (2007).
51. Chen, S. F. *et al.* Inferring a transcriptional regulatory network of the cytokinesis-related genes by network component analysis. *BMC Syst Biol* **3**, 110 (2009).
52. Foiani, M., Liberi, G., Lucchini, G. & Plevani, P. Cell cycle-dependent phosphorylation and dephosphorylation of the yeast DNA polymerase alpha-primase B subunit. *Mol Cell Biol* **15**, 883–891 (1995).
53. Ma, L. *et al.* A plausible model for the digital response of p53 to DNA damage. *Proc Natl Acad Sci USA* **102**, 14266–14271 (2005).
54. Tian, T. & Burrage, K. Stochastic models for regulatory networks of the genetic toggle switch. *Proc Natl Acad Sci USA* **103**, 8372–8377 (2006).
55. Kitano, H. Towards a theory of biological robustness. *Mol Syst Biol* **3**, 137 (2007).
56. Gillespie, D. T. Exact Stochastic Simulation of Coupled Chemical-Reactions. *J Phys Chem-Us* **81**, 2340–2361 (1977).
57. Barrio, M., Burrage, K., Leier, A. & Tian, T. H. Oscillatory regulation of hes1: Discrete stochastic delay modelling and simulation. *PLoS Comput Biol* **2**, 1017–1030 (2006).
58. Cai, X. Exact stochastic simulation of coupled chemical reactions with delays. *J Chem Phys* **126**, 124108 (2007).
59. Tian, T. H., Burrage, K., Burrage, P. M. & Carletti, M. Stochastic delay differential equations for genetic regulatory networks. *J Comput Appl Math* **205**, 696–707 (2007).

Acknowledgements

The authors would like to thank the Australian Research Council for the Discovery Project (T.T. DP120104460). T.T. is also an ARC Future Fellow (FT100100748).

Author Contributions

T.T. conceived and designed research; Q.W. and T.T. wrote the programs, conducted research and wrote the paper. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Wu, Q. and Tian, T. Stochastic modeling of biochemical systems with multistep reactions using state-dependent time delay. *Sci. Rep.* **6**, 31909; doi: 10.1038/srep31909 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016