

RESEARCH

Open Access

Stochastic modelling of biochemical systems of multi-step reactions using a simplified two-variable model

Qianqian Wu¹, Kate Smith-Miles¹, Tianshou Zhou², Tianhai Tian^{1*}

From IEEE International Conference on Bioinformatics and Biomedicine 2012
Philadelphia, PA, USA. 4-7 October 2012

Abstract

Background: A fundamental issue in systems biology is how to design simplified mathematical models for describing the dynamics of complex biochemical reaction systems. Among them, a key question is how to use simplified reactions to describe the chemical events of multi-step reactions that are ubiquitous in biochemistry and biophysics. To address this issue, a widely used approach in literature is to use one-step reaction to represent the multi-step chemical events. In recent years, a number of modelling methods have been designed to improve the accuracy of the one-step reaction method, including the use of reactions with time delay. However, our recent research results suggested that there are still deviations between the dynamics of delayed reactions and that of the multi-step reactions. Therefore, more sophisticated modelling methods are needed to accurately describe the complex biological systems in an efficient way.

Results: This work designs a two-variable model to simplify chemical events of multi-step reactions. In addition to the total molecule number of a species, we first introduce a new concept regarding the location of molecules in the multi-step reactions, which is the second variable to represent the system dynamics. Then we propose a simulation algorithm to compute the probability for the firing of the last step reaction in the multi-step events. This probability function is evaluated using a deterministic model of ordinary differential equations and a stochastic model in the framework of the stochastic simulation algorithm. The efficiency of the proposed two-variable model is demonstrated by the realization of mRNA degradation process based on the experimentally measured data.

Conclusions: Numerical results suggest that the proposed new two-variable model produces predictions that match the multi-step chemical reactions very well. The successful realization of the mRNA degradation dynamics indicates that the proposed method is a promising approach to reduce the complexity of biological systems.

Background

The advances in systems biology have raised the importance of quantitative methods for studying various systems in molecular biology. In recent years, various research methods, including mathematical modeling, statistical analysis, computer simulation and visualization, have been employed to investigate the dynamic or statistical properties of regulatory networks. In particular, mathematical

models have been widely used to describe the dynamics of complex systems inside the cell, including genetic regulatory networks, cell signalling transduction pathways and metabolic pathways [1,2]. However, these substantial progresses have further raised a number of fundamental and challenging issues that require to be addressed imperatively.

One of the major challenges in systems biology is how to use simple mathematical models to describe complex biological systems. To address this issue, a number of modelling techniques have been designed. Among them, a widely used approach is to use one-step reaction to

* Correspondence: tianhai.tian@monash.edu

¹School of Mathematical Sciences, Monash University, VIC 3800 Melbourne, Australia

Full list of author information is available at the end of the article

represent multi-step reactions, which is also called slow reaction. This technique is very important because recent theoretical and experimental studies have shown that a wide variety of biochemical events involve multi-step reactions [3]. Perhaps the most important example of multi-step reactions is transcriptional and translational processes that produce mRNA transcripts and proteins, respectively. Other examples include molecules (e.g. mRNA and protein) degradation and telomere length shortening processes. In fact, the process of multi-step reactions also exists in other areas such as organic chemistry and biophysical chemistry [4,5]. Therefore the major aim of this research work is to design simplified models to accurately characterize biological systems with multi-step reactions.

A widely used approach to simplify multi-step chemical reactions in the literature is to use one-step reaction. For example, the degradation process of mRNA or protein has been modelled by a first order reaction. However, since the one-step reaction cannot provide consistent description of the multi-step reactions, chemical reactions with time delay have been designed recently to model the multi-step chemical events more accurately [6-9]. Another important factor is noise in biological networks that may influence the system dynamics substantially. The deterministic modelling methods, which approximate molecular numbers using continuous concentrations [10,11], may not be appropriate to describe systems that contain species with small population numbers. To model stochastic systems more accurately, there are a few other ways. For example, we can use discrete Markov processes where the density of states of a well-stirred chemical reaction system at each time point can be represented by the chemical master equation (CME) [12,13]. One of the most well-known methods is called the stochastic simulation algorithm (SSA), which is a statistically exact method for simulating trajectories of the CME as the system evolves in time [14].

Furthermore, to deal with the intrinsic noise in reactions with time delay, the delay stochastic simulation algorithm (DSSA) was designed by introducing time delay into the SSA [15,16]. Unlike the SSA, which assumes that biochemical reactions are instantaneous and independent, the DSSA characterizes chemical systems that contain both fast and slow reactions. This delayed modelling approach has been applied to many physical and biological systems [16]. The DSSA was also extended to describe chemical events that have multiple delays or stochastic delay that follows a given probabilistic distribution [17,18]. In recent years, the DSSA has been widely used to simulate the dynamics of genetic regulatory networks and cell signalling pathways [7][19-22]. In addition, a number of effective simulation methods have been proposed to reduce the huge computing load

of the DSSA [23-26]. Recently the work done by Luis Mier-y-Terán-Romero et al. opened some new aspects for the application of time delays in biological systems. Time delay may not be a constant that was assumed before [27]. Other modelling techniques proposed recently include the slow-scale linear noise approximation and stochastic quasi-steady-state assumption [28,29]. Most recently a new modelling approach has been proposed to simulate chemical reaction systems with memory reactions [30].

The degradation process of mRNA molecules is an important step in the regulation of gene expression, which also represents a typical system with multi-step reactions [31]. Although the mechanisms of mRNA degradation have been studied extensively during the last ten years, there are still a number of open problems with respect to the function of enzymes, structure of pathways and role of P-bodies, etc. in the regulation of mRNA degradation [32-34]. A major step in the quantitative study of mRNA degradation was the development of mathematical models based on the detailed chemical processes. A linear multi-component model was designed to investigate the nonsense-mediated decay of mRNA molecules in yeast [35,36]. This deterministic model for mRNA degradation process consists of 23 first-order reactions that describe transcription, translocation, poly (A) shortening, decapping and digestion process. Computer simulations suggested that the widely used concept of half-life underestimated the averaged life-span of mRNA molecules; however, it is still a major factor that determines the life-span of different steps in the degradation pathway. In addition, robustness analysis showed that the change of degradation rate constant led to large variations of mRNA copy numbers. To interpret the complexity of mRNA degradation in a simpler manner, we proposed a multi-step reaction model using a chain of 11 chemical reactions, which gave very good approximation to the detailed one [37].

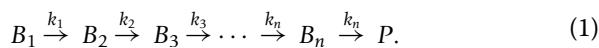
Chemical reactions with time delay has been used to further simplify mathematical models of mRNA degradation. Here time delay represents the time required in the multi-step reactions except the first reaction [37]. This simplified model was also extended to using stochastic time delay. However, numerical results showed that these first-order reaction models with delay did not give good approximation to the detailed degradation process [37]. Instead of using time delay to represent the missing intermediate reactions in the multi-step reaction, we recently proposed a new modelling approach by introducing a novel concept, namely the length of a molecule indicating its location in the multi-step reactions. Deterministic models using ordinary differential equations have been used to find the optimal value in a non-linear probability function [38]. However, it is still a challenge

to apply this concept to stochastic models that are much more important than deterministic models for chemical reaction systems. Thus this work further validates the proposed model using stochastic simulations. We first introduce a new stochastic modelling method with two variables for describing chemical events with multi-step reactions, and then propose a stochastic simulation algorithm to numerically calculate the probability of the firing of the last reaction in the multi-step events. The efficiency and accuracy of the proposed method are examined by studying the mRNA degradation process of gene PRL30 based on experimental data.

Results and discussion

A new two-variable model

The starting-point of this research work is the chemical events with multi-step reactions. Using the notation proposed in [3], we consider the following chemical reactions



where B_i are molecular species and k_i rate constants. It is assumed that each molecule in the system will eventually turn to the product P or degrade if $P = ()$. During this process, each molecule will pass through a number of states B_1, B_2, \dots, B_n via the multi-step reactions.

When the number of reaction step n is large, we need to design a smaller scale model to simplify the multi-step reactions. We first consider the total number of molecules in the system, defined by

$$X = \sum_{i=1}^n [B_i]. \quad (2)$$

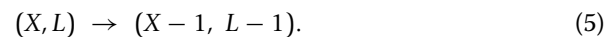
Here we introduce a new concept to describe the system state. The number of reactions for a molecule to reach the product P is termed as the length of that molecule. Thus the length of molecule B_i is $(n - i + 1)$ and the total length of the molecules in the system is

$$L = \sum_{i=1}^n (n - i + 1)[B_i]. \quad (3)$$

According to the total molecule number, chemical reactions in the system can be classified into two groups. If one of the first $(n - 1)$ step reactions occurs, namely $B_i \xrightarrow{k_i} B_{i+1}$, the total number of molecules X is unchanged but the total length L is decreased by one,



However, if the last reaction $B_n \xrightarrow{k_n} P$ fires, both the total number and total length will decrease by one,



In this work we use reactions (4) and (5) to design the two-variable reaction model.

The key question now is how to determine whether reaction (4) or (5) will fire if one of the reactions in the multi-step process (1) happens. We denote the probability for the degradation of one molecule, namely the firing of reaction (5), as $f(X, L, n)$, and then the corresponding probability for reaction (4) as $1 - f(X, L, n)$. It is clear that, when all molecules are of full length ($X = nL$), the probability of f is zero; while when $X = L$, the probability is one. For the molecules with other lengths, we developed an algorithm, namely Algorithm I in the Method section, to calculate the probability of molecule degradation. With the help of this algorithm, we numerically calculated the exact probability $f(X, L, n)$ using $n = 8$ and $X = 15$ as an example. The probability is represented in Figure 1 as the solid line.

Next we find an appropriate probability function to approximate the calculated curve in Figure 1. Note that the total length L of X molecules satisfies $X \leq L \leq nX$. When $L = X$, all molecules have length 1, the probability of firing of the last step reaction is 1, i.e. $f(X, L, n) = 1$; when $L = nX$, all molecules have length n , there is no chance for the final reaction to occur in the next step, i.e. $f(X, L, n) = 0$. Therefore we suggested a probability function to approximate the curve in Figure 1 in the following format:

$$f(L, X, n) = 1 - \frac{L - X}{X(n - 1)}. \quad (6)$$

The approximated probability through the proposed function (6) is plotted as the straight dashed line in Figure 1. It shows that the approximated values are not close to the exact probability values, and the exact probability curve is in a quadratic-like form. Hence, instead of using a linear probability f in terms of X , L and n (6), we introduced another parameter q into this approximation, and proposed the following two expressions for the probability function f in terms of L , X , n , and q . One candidate is

$$\text{Type I: } f(X, L, n, q) = 1 - \left(\frac{L - X}{X(n - 1)} \right)^q, \quad (7)$$

and the alternative expression is

$$\text{Type II: } f(X, L, n, q) = \left(1 - \frac{L - X}{X(n - 1)} \right)^q. \quad (8)$$

Determination of probability function

The major work of this research is to select a probability function from (7) and (8) and also search the optimal

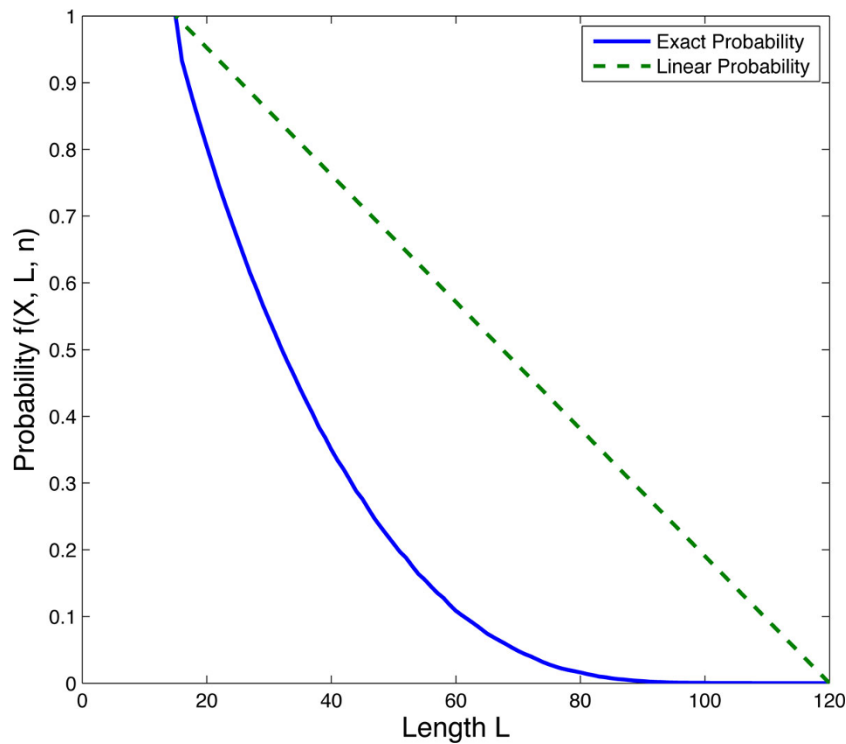


Figure 1 The probability for the firing of the last reaction $B_n \xrightarrow{k_n} P$.

value of parameter q in the probability function. Using Algorithm I in the Method section, we first calculated the probability $f(X, L, n, q)$ with different values of the total molecule number X ($X = 3 \sim 20$), different numbers of reaction step n ($n = 3 \sim 20$) and various values of the total length L ($L = X \sim nX$). The calculated probability was used as the exact value to search the optimal q in the proposed probability functions. To select a better probability function, we used both type I and II functions to calculate the probability $f(X, L, n, q)$ using the same initial condition ($n = 8$ and $X = 15$) but different values of q ($q = 0.01 \sim 15$) in a step size of 0.01. By searching for the smallest difference between the exact probability values and those obtained from approximated functions with different q , the optimal values of q for two approximations were achieved which are 0.27 and 3.91 respectively in this particular case. The exact and approximated probability values are shown in Figure 2. We found that the type II approximation is closer to the exact probabilities than the type I approximation. Then we only used the probability function (8) for the following studies.

To establish a more general formula for defining q under different conditions of X, L, n , we extended the simulations to various initial conditions of n, X both varying from 3 to 20 together with different values of q . The optimal q values acquired under these conditions

are illustrated by Figure 3 (A) and (B). Figure 3 (A) shows that when n increases, the optimal value of q increases for a fixed X value; while Figure 3 (B) indicates that there is no significant variation for the optimal q when X increases for a fixed n value. Therefore, we calculated the averaged optimal q values under various value of n for each given X . A plot of this averaged optimal \bar{q} against n is shown in Figure 3 (C). We suggested a linear relationship between n and q . A linear regression analysis suggested that this relationship is

$$\bar{q} = 0.3146n + 1.3615. \quad (9)$$

We have developed deterministic models of ordinary differential equations (ODEs) based on the multi-step reactions (1) and the two-variable model (4, 5) [38]. Simulation results of the deterministic models gave some similar patterns such that the optimal value of q increases when the number of chemical reactions n increases. The established relationship between the optimal value of q and related model parameters, which is also shown in Figure 4, is given by

$$\bar{q} = 0.4570n + 0.8567. \quad (10)$$

The above equation is slightly different from expression (9). For example, when $n = 5$, the averaged value of optimal q is found to be 2.8494 using probability

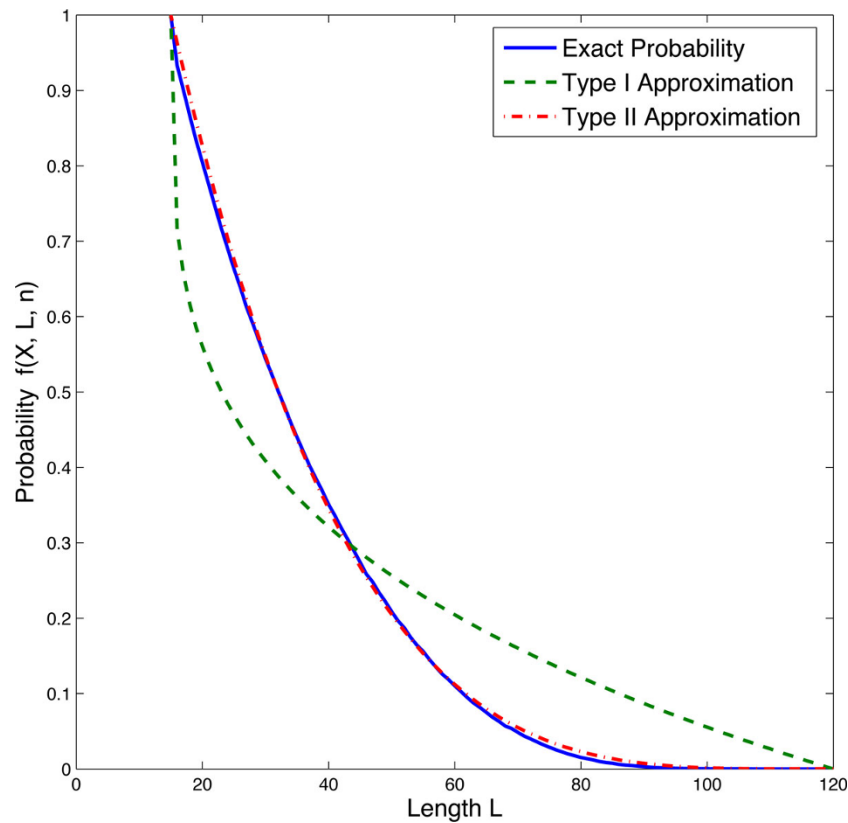


Figure 2 Simulated exact probabilities and two approximated probabilities for the firing of the last reaction.

simulation while it is 3.06 from the ODE simulation. The possible reason of the difference is that the ODE model is not the best approach for describing chemical reaction systems with molecules of small copy numbers and some model errors may arise from the ODE simulations. A combination of regression analyses is shown in Figure 4.

Even with the different formulations for the optimal q values, we still find that the ODE method confirms the conclusion derived from stochastic simulations. Based on both stochastic and deterministic simulations, we found that the value of q is associated with the number of reaction step n , but not connected to the total molecule number X . Our results also suggested that, when the value of q approaches the optimal one, simulation error of the two-variable model using q is very close to that using the optimal q value. Using the function derived from stochastic simulations, the probability function of molecule degradation is given by

$$f(L, X, n) = \left(1 - \frac{L - X}{X(n - 1)}\right)^{0.3146n + 1.3615} \quad (11)$$

Using this probability function, we designed an algorithm, namely Algorithm II in the Method section, to

simulate the two-variable reaction model based on the SSA.

mRNA decay dynamics: case study for gene RPL30

In this section, we apply the established theory in the previous section to study the dynamics of mRNA degradation. Here we use gene ribosomal protein L30 (*RPL30*) as the test system with a dataset generated from experiments. In these experiments, two constructs of *RPL30* were used to demonstrate the decay kinetics of the mRNA transcripts [39]. The first construct ("construct A") contains the *ACT1* UAS (upstream activating sequence), and the other ("construct B") contains the *RPL30* UAS. The mRNA molecule decay dynamics was monitored after blocking transcription by using drug 1,10-phenanthroline [39]. Thus we assumed that there was no further transcription during the monitoring process. The decay dynamics was normalized by the *RPL30* transcript level at time zero (namely before adding the drug), which was set to 100%. Using the endogenous *RPL30* mRNA levels obtained from the two construct [39], we first used the one-step differential equation model

$$\frac{dX}{dt} = -kX \quad (12)$$

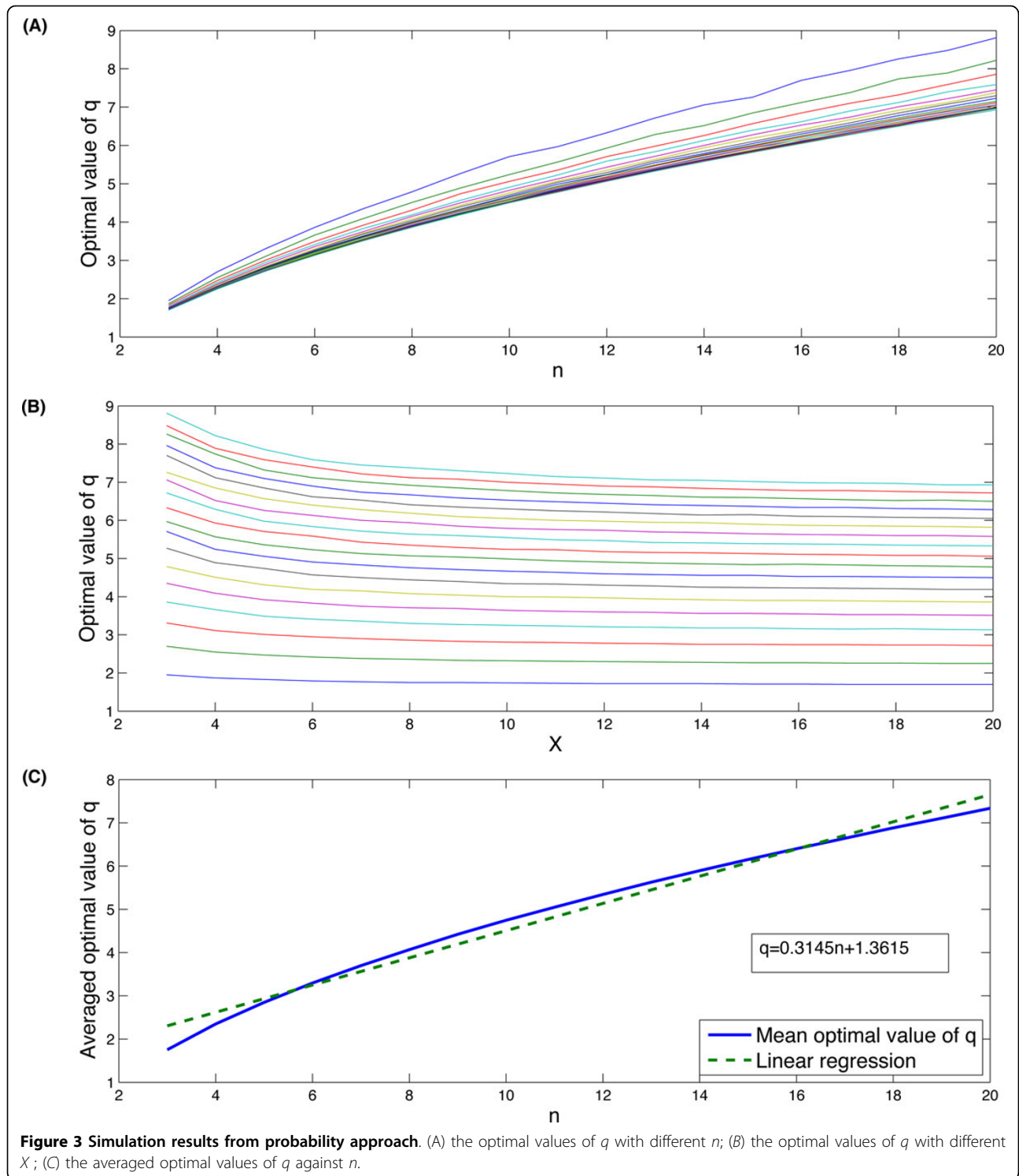
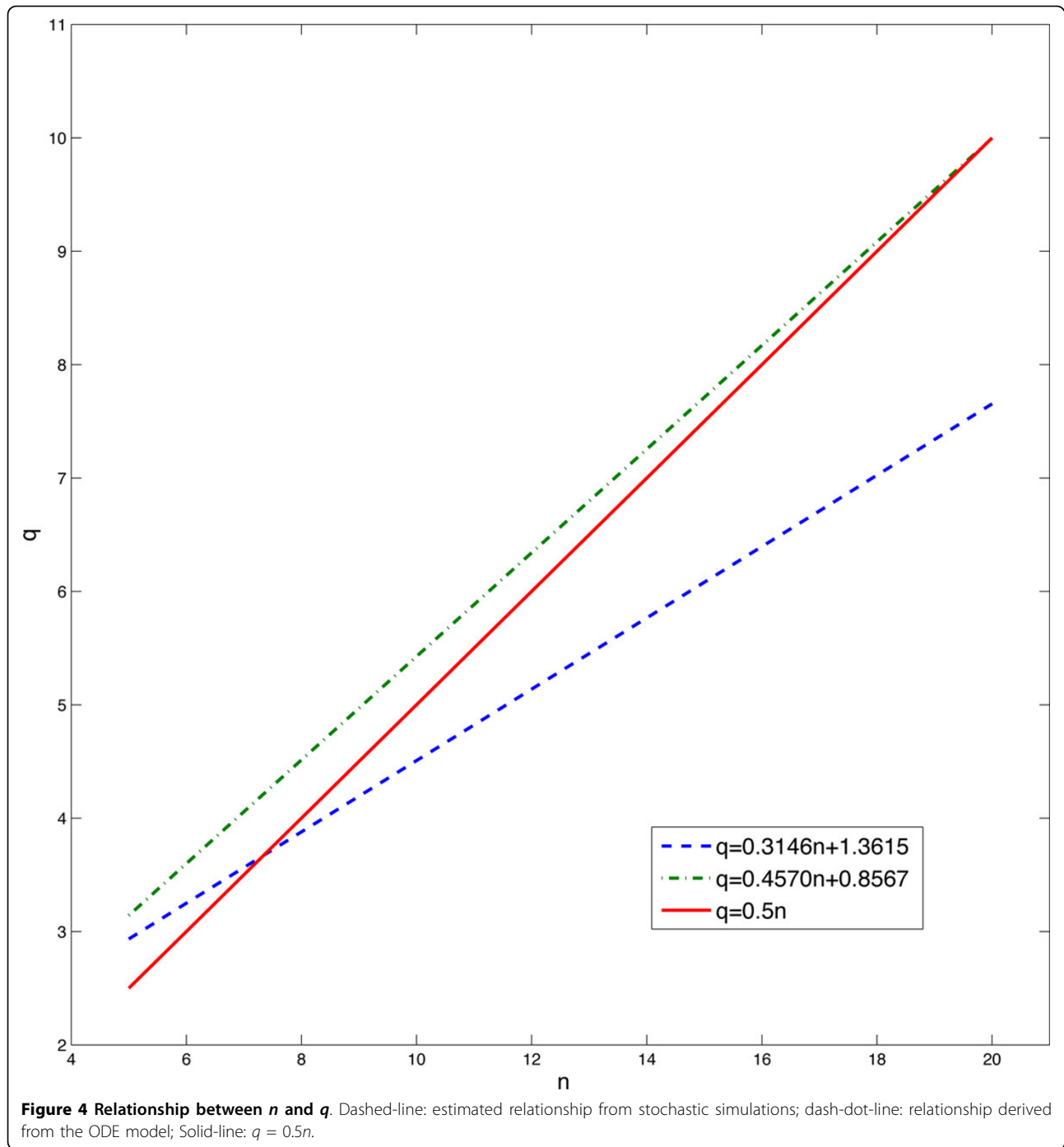


Figure 3 Simulation results from probability approach. (A) the optimal values of q with different n ; (B) the optimal values of q with different X ; (C) the averaged optimal values of q against n .

To simulate the decay dynamics [40].

Figure 5 (A) and (B) show that the one-step model failed to describe the dynamics of the first 25 minutes accurately. The simulated mRNA levels are always smaller than the experimental observation.

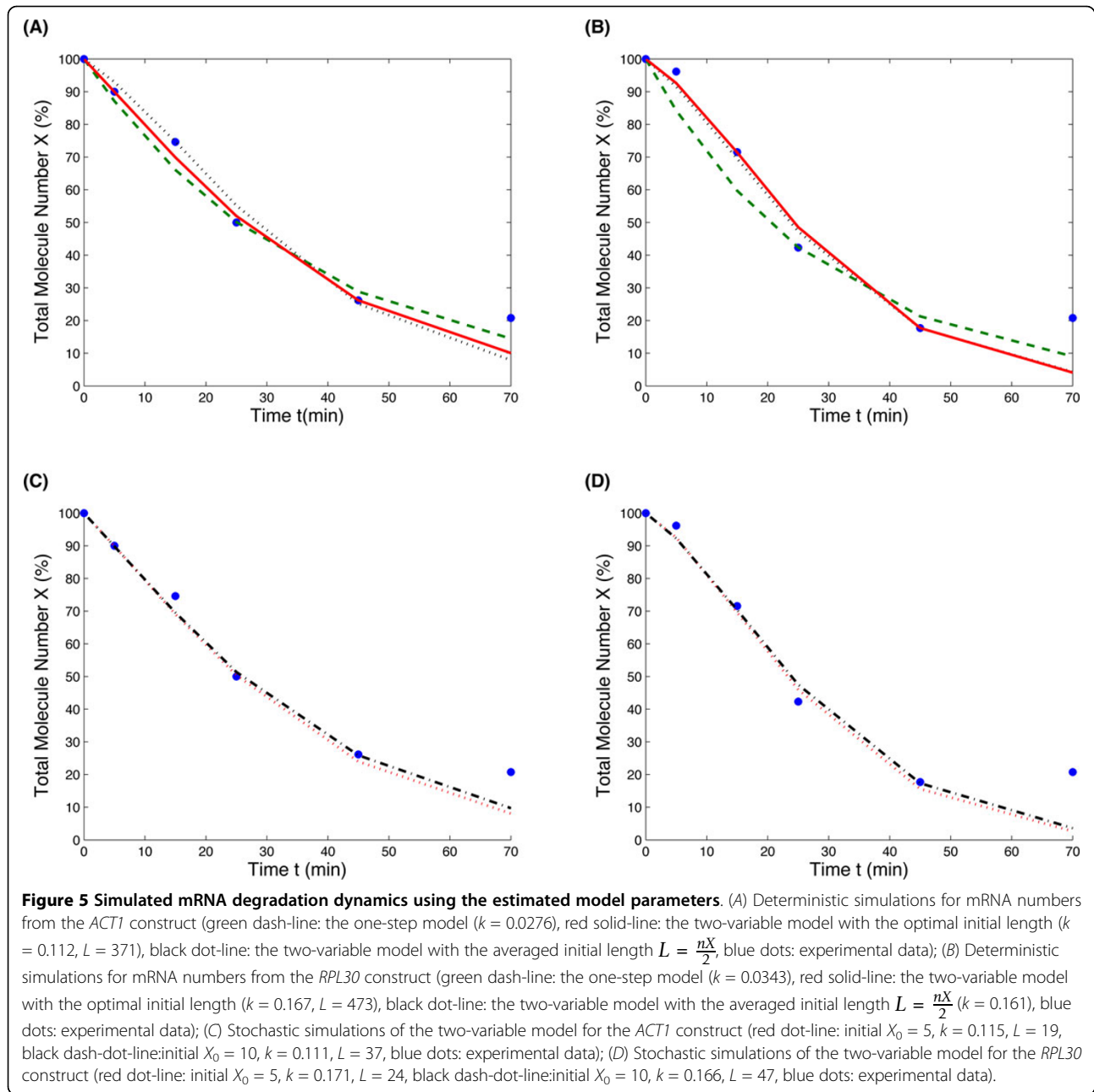
To model mRNA degradation, Cao and Parker proposed a multi-component model that includes mRNA transcript synthesis, mRNA translocation, poly(A)-shortening process, and terminal deadenylation [35]. We have proposed a simplified model by putting a number



of terminal deadenylation reactions into a single one [37]. This simple model is a typical multi-step reaction process. In this model, mRNA transcript is synthesized by a zero-order reaction S_1 , then mRNA molecules translocate from the nucleus to cytosol via reaction S_2 . The mRNA molecules in the cytosol produce proteins by the translational process, and in the meantime, the length of mRNA begins to decrease via a number of

poly(A)-shortening reactions S_3, \dots, S_9 . The final reaction in this process is the further exonucleolytic degradation S_{10} , which is regarded as the degradation reaction in this work, since the fragment product (FG) has no function to produce protein molecules.

Based on the reactions listed in Table 1 and rate constants, the propensity functions of these reactions are listed below.



Reaction	Propensity function
$DNA \xrightarrow{s_1} A$	$a_1 = s_1,$
$DNA \xrightarrow{s_1} A$	$a_1 = s_1,$
$B \xrightarrow{s_3} BC1$	$a_3 = s_3 \cdot [B],$
$BC1 \xrightarrow{s_4} BC2$	$a_4 = s_4 \cdot [BC1],$
$BC2 \xrightarrow{s_5} BC3$	$a_5 = s_5 \cdot [BC2],$
$BC3 \xrightarrow{s_6} BC4$	$a_6 = s_6 \cdot [BC3],$
$BC4 \xrightarrow{s_7} BC5$	$a_7 = s_7 \cdot [BC4],$
$BC5 \xrightarrow{s_8} BC6$	$a_8 = s_8 \cdot [BC5],$
$BC6 \xrightarrow{s_9} BC7$	$a_9 = s_9 \cdot [BC6],$
$BC7 \xrightarrow{s_{10}} FG$	$a_{10} = s_{10} \cdot [BC7],$

(13)

Following the experimental conditions, it is assumed that $s_1 = 0$. For simplicity, it is assumed that $s_2 = \dots = s_{10}$. When using the two-variable model to study the mRNA degradation process, we write the total copy number X and total length of mRNA molecules L as

$$X = [A] + [B] + [BC1] + \dots + [BC7],$$

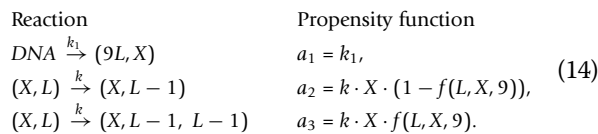
$$L = 9[A] + 8[B] + 7B[BC1] + \dots + [BC7].$$

Here we put the mRNA synthesis as a separate reaction. Then the remaining nine reactions ($n = 9$) form a chemical event of multi-step reactions. The dynamics of variables $x = (L, X)$ is described by the following reactions together with the corresponding propensity functions

Table 1 Reactions and kinetic rates of the simplified stochastic model.

	Reaction	Rate constant s_i	Comment
S_1	$DNA \rightarrow A$	s_1	transcription
S_2	$A \rightarrow B$	s_2	transport
S_3	$B \rightarrow BC1$	s_3	full-length 70A-60A
S_4	$BC1 \rightarrow BC2$	s_4	full-length 60A-50A
S_5	$BC2 \rightarrow BC3$	s_5	full-length 50A-40A
S_6	$BC3 \rightarrow BC4$	s_6	full-length 40A-30A
S_7	$BC4 \rightarrow BC5$	s_7	full-length 30A-20A
S_8	$BC5 \rightarrow BC6$	s_8	full-length 20A-10A
S_9	$BC6 \rightarrow BC7$	s_9	full-length 10A-0A
S_{10}	$BC7 \rightarrow FG$	s_{10}	fragment production

The rate constants s_i are in the unit of 1/sec.



Using the assumption $s_2 = \dots = s_{10}$, the rate constant k (14) is the harmonic mean of rate constants s_2, \dots, s_{10} , given by

$$k = \frac{8}{\sum_{i=2}^{10} \frac{1}{s_i}} = s_i. \quad (15)$$

Next we used the proposed two-variable model to give more accurate simulations. We first estimated the degradation rate constant k and optimal initial total length of transcripts. We have also estimated the degradation rate constant k by assuming that the total initial length is a half of the maximal total length ($L = nX / 2$), which is termed as the averaged total length. To reduce the computing time, we first estimated parameters in the ODE model (12) using different initial transcript numbers ($X_0 = 5, 10, 20, \dots, 100$). Table 2 suggests that the variation between the estimate rate constant k was very small for different initial mRNA numbers. Similar observation is applied to the ratio of the optimal initial total length to the maximal total initial length,

Table 2 Estimated parameters for the stochastic model of RPL30 and ACT1 mRNA degradations (Ratio = L_0/nX).

X_0	ACT1 construct			RPL30 construct		
	Rate k	L_0	Ratio	Rate k	L_0	Ratio
$m = 5$	0.1150	19	0.4222	0.1710	24	0.5333
$m = 10$	0.1110	37	0.4111	0.1660	47	0.5222
$m = 20$	0.1130	75	0.4167	0.1680	95	0.5278
$m = 30$	0.1130	112	0.4148	0.1670	142	0.5259
$m = 40$	0.1120	149	0.4139	0.1680	190	0.5278
$m = 50$	0.1120	186	0.4133	0.1670	237	0.5267

namely $L_0/(nX)$, for the tests with different initial mRNA numbers. Thus our results suggested that the estimated model parameters are independent to the initial mRNA copy numbers.

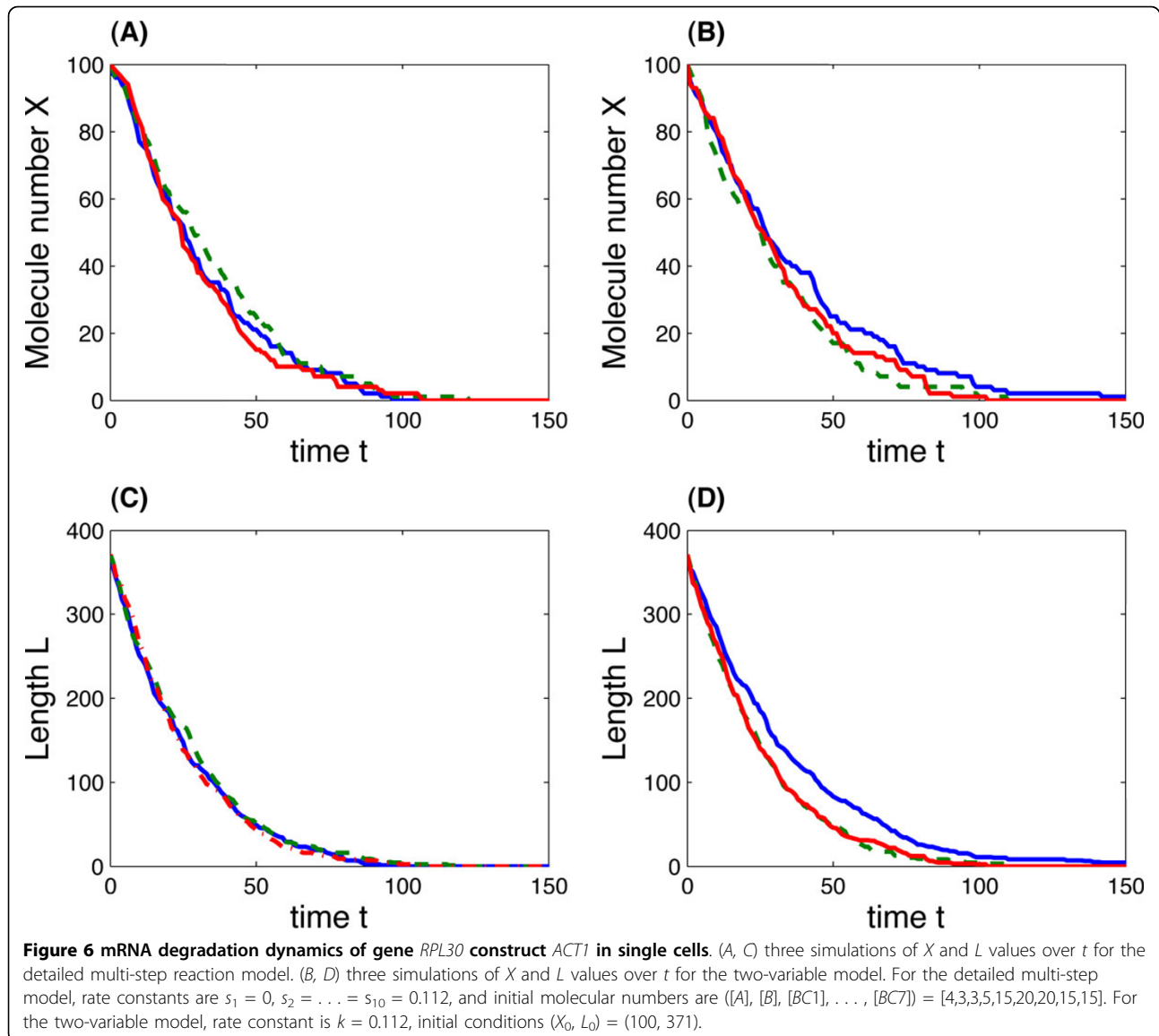
Using the estimated model parameters of the case $X_0 = 100$, simulation results for the two constructs in Figure 5 (A) and (B) show that the two-variable model provides more accurate description of the mRNA degradation dynamics than the one-step model, in particularly for that in the first 25 minutes. For the *ACT1* construct in Figure 5 (A), the optimal length number with ratio 0.412 gave more accurate simulation than the averaged length number. However, in Figure 5 (B) for the *RPL30* transcript, the difference between the simulations using two different length numbers is small. In this case, the optimal ratio is 0.525, which is very close to 0.5.

To further examine the accuracy of the two-variable model, we used the stochastic model to simulate the mRNA dynamics using different initial transcript numbers. For each initial mRNA number, we generated 10,000 simulations and then calculated the averaged mRNA numbers of all stochastic simulations. For both constructs in Figure 5 (C) and (D), our results show that there is small difference between the simulations using $X_0 = 5$ and $X_0 = 10$. However, there is not any significant difference between simulations when the initial mRNA number is larger than 10.

Finally we provided a few stochastic simulations for mRNA degradation dynamics for the construct *ACT1* in single cells. The rate constants of the detailed model were derived from the two-variable model using the relationship (15); and the initial molecular numbers were randomly selected while the length of the initial molecules matches the length in the two-variable model. When the mRNA synthesis rate is $s_1 = 0$, Figure 6 shows that the molecular numbers and lengths approaches to zero at the time point around 100. In addition, compared with the simulations of the detailed model in Figure 6A and 6C, the two-variable model generates simulations with more fluctuations in Figure 6B and 6D. After the time point 100, more simulations of the two-variable model still have non-zero molecular numbers.

Conclusions

This work represents an attempt to use simplified mathematical models to describe complex biological systems. Concentrating on the chemical events of multi-step reactions, we proposed a new concept (e.g. the length of a molecule) as an additional measure to characterize system dynamics. The length of a molecule is defined as the location of a molecule in the multi-step reactions. Using the total molecule number and total length of molecules, we proposed a two-variable model to reduce



the complexity of the multi-step reactions. The major contribution of this work is to design a nonlinear function that represents the probability of the firing of the last reaction in the multi-step reactions. To calibrate this probability function, we proposed a stochastic simulation method to calculate the probabilities of various system states. Numerical results suggested that this probability is dependent on the number of reaction steps but independent of the total molecule number, which suggested that we were able to design a simplified model based on the network structure. Then our proposed two-variable model was applied to simulate the dynamics of mRNA degradation using experimentally observed data. Numerical results suggested that the length of molecules, which is approximately a half of the maximal length initially, played an important role in realizing experimental data.

The potential future work includes the application of the two-variable model to other multi-step reaction systems such as gene expression and telomere length regulation. In addition, the refinement of the two-variable model, such as the accuracy of the probability function, would also be very interesting.

Methods

Simulation algorithm for the probability function

To find the probability for the firing of the last reaction in the multi-step reactions (1), we first designed a Monte-Carlo method to numerically calculate the probability function $f(X, L, n)$ based on the given X and n . By the law of total probability, the formation of probability that the final reaction occurs given by any L and x is defined as following:

$$f(X, L, n) = \sum_{j=0}^X P(R_n | B_n = j) \cdot P(B_n = j | L, X),$$

where R_n represents the occurrence of last reaction $B_n \xrightarrow{k_n} P$. Based on the total molecule number X , we calculate the probability

$$P(R_n | B_n = j) = \frac{j}{X}.$$

The major part of this algorithm is to find frequency of the event $B_n = j$ based on the given length L and total molecule number X , which is explained as the following algorithm I.

Algorithm I

1. Set the total number of molecule X , number of reactions n , and initial full length $L_0 = nX$.
2. Based on the following 10, 000 Monte-Carlo simulations, calculate the frequency $freq(B_n = j | L, X)$ for X molecules with total length L having j molecules, where $j = 0, 1, \dots, X$ and each of the molecule with length 1:
 - (a) Consider X molecules with full length. Denote the length of the i -th molecule as l_i with $i = 0, 1, \dots, X$.
 - (b) Use a random number $r \sim U(0, 1)$ to select one molecule, with index j . If the length of that molecule $l_j > 1$, reduce its length by 1, namely $l_j = l_j - 1$; if $l_j = 1$, then repeat this step until finding a molecule with length greater than 1.
 - (c) Repeat step (b) for $(L_0 - L)$ times to get a set of molecules with total length L .
 - (d) Count the number of molecules in this set with length 1, denote as i , then update

$$freq(B_n = i | L, X) = freq(B_n = i | L, X) + 1.$$

- (e) Repeat steps (a) ~ (d) for 10,000 times.

3. The probability for the last reaction firing is obtained by

$$f(X, L, n) = \sum_{j=1}^X \frac{freq(B_n = j | L, X)}{10000} \times \frac{j}{X}.$$

Ordinary differential equation model

The most widely used approach to study chemical reaction systems is deterministic model using ordinary differential equations. The approach is valid if the copy numbers of chemical species in the system are large. To confirm the probability function $f(X, L, n)$ derived from stochastic simulations, we designed a deterministic model of ODEs for the multi-step chemical reaction system (1), given by

$$\begin{aligned} \frac{dB_1}{dt} &= -k_1 B_1, \\ \frac{dB_2}{dt} &= k_1 B_1 - k_2 B_2, \\ &\vdots \\ \frac{dB_n}{dt} &= k_{n-1} B_{n-1} - k_n B_n, \\ \frac{dP}{dt} &= k_n B_n. \end{aligned} \tag{16}$$

Using the total molecule number $X (= B_1 + \dots + B_n)$ and the total length of the molecules $L (= B_n + 2B_{n-1} + \dots + nB_1)$, we have a simplified model of the above ODE system

$$\begin{aligned} \frac{dX}{dt} &= -kX, \\ \frac{dL}{dt} &= -kX, \end{aligned} \tag{17}$$

where k is the harmonic mean of the rate constants k_1, \dots, k_n (19), and kB_n represent the probability of molecule degradation which is represented by the probability function $f(X, L, n)$. Using the notations of stochastic simulation, the ODE model with the length of molecules is given by

$$\begin{aligned} \frac{dX}{dt} &= -kX \left(1 - \frac{L - X}{X(n-1)} \right)^q, \\ \frac{dL}{dt} &= -kX. \end{aligned} \tag{18}$$

For a given initial condition $B_i(0)$, we obtained the analytical solution of the detailed system (16) and then solved the two-variable model (18) numerically using a stiff ODE solver *ode23s* in MATLAB. We tested the solution of the two-variable model with different values of q based on different system conditions ranging from $n = 5, 10, 15$ as well as $X = [5 \ 10 \ 50 \ 100 \ 200 \ 500]$. For each system condition, we selected the optimal value of q with which the two-variable model (18) generates simulation that is very close to that of the detailed ODE model (16). Finally we find the relationship between the value of q and system condition (X, L, n) by using a regression method [38].

An algorithm for simulating systems including two-variable model

The SSA is a general framework for simulating biochemical reaction systems. Now we propose an algorithm to incorporate the two-variable model into the SSA. It is assumed that a chemical reaction system is a well-stirred mixture at constant temperature in a fixed volume Ω . This mixture consists of N molecular species $\{S_1, \dots, S_N\}$ that chemically interact through M reaction channels $\{R_1, \dots, R_M\}$. The dynamic state of this

system is denoted as $\mathbf{x} \equiv (x_1(t), \dots, x_N(t))^T$, where $x_i(t)$ is the molecular number of species S_i at time t . For each reaction channel R_j ($j = 1, \dots, M$), a propensity function $a_j(\mathbf{x})$ is defined by a given state $\mathbf{x}(t) = \mathbf{x}$ and the value of $a_j(\mathbf{x})dt$ represents the probability that one reaction will occur somewhere during the infinitesimal time interval $[t, t + dt]$ [14][26][41]. In addition, a state change vector ν_j is defined to characterise the change of molecular numbers due to the reaction R_j . The element ν_{ij} of ν_j represents the change of the copy number of species S_i . The algorithm for simulating chemical reaction systems with two-variable model is given below.

Algorithm II

1. Calculate the values of propensity function $a_j(\mathbf{x})$ based on the system state \mathbf{x} at time t . In particular, for the two-variable reaction with the total molecule number X (2) and total length L (3), the propensity function is $a_j = kX$, where k is the harmonic mean of the rate constants (1), given by

$$k = \frac{n}{\frac{1}{k_1} + \dots + \frac{1}{k_n}}. \quad (19)$$

Then the sum of propensity function values is

$$a_0(\mathbf{x}) = \sum_{j=1}^M a_j(\mathbf{x}).$$

2. Generate a sample r_1 of the uniformly distributed random variable $U(0, 1)$, namely $r_1 \sim U(0, 1)$, and determine the time of next reaction

$$\mu = \frac{1}{a_0(\mathbf{x})} \ln \frac{1}{r_1}.$$

3. Generate another sample r_2 of $U(0, 1)$ to determine the index k of the next reaction occurring in $[t, t + \mu]$,

$$\sum_{j=1}^{k-1} a_j(\mathbf{x}) < r_2 a_0(\mathbf{x}) \leq \sum_{j=1}^k a_j(\mathbf{x})$$

4. If the k -th reaction is not a two-variable model, update the state of the system by

$$\mathbf{x}(t + \mu) = \mathbf{x}(t) + \nu_k$$

Otherwise generate a sample $r_3 \sim U(0, 1)$ to determine which reaction of the followings will occur,

$$(X, L) = \begin{cases} (X, L - 1) & \text{if } r_3 > f(X, L, n), \\ (X - 1, L - 1) & \text{if } r_3 < f(X, L, n), \end{cases}$$

where $f(X, L, n)$ is the probability of the firing of the last reaction. Then the system is updated.

5. Go back to step 1 if $t + \mu \leq T$, where T is the end time point. Otherwise, the system state at T is $x(T) = x(t)$.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

TT conceived and designed the study. QW and TT developed algorithms and carried out research. QW, KS, TZ and TT analyzed the data, interpreted the results and wrote the paper. All authors edited and approved the final version of the manuscript.

Acknowledgements

This work is in part supported by grants from the Australian Research Council (ARC) (DP1094181 and DP120104460). T.T. is also in receipt of an ARC Future Fellowship (FT100100748). T.Z. is in part supported by grants 91230204 and 30973980 from the Natural Scientific Foundation of China.

Declarations

The publication costs for this article were funded by the corresponding author. This article has been published as part of *BMC Systems Biology* Volume 7 Supplement 4, 2013: Selected articles from the IEEE International Conference on Bioinformatics and Biomedicine 2012: Systems Biology. The full contents of the supplement are available online at <http://www.biomedcentral.com/bmcsystbiol/supplements/7/54>.

Authors' details

¹School of Mathematical Sciences, Monash University, VIC 3800 Melbourne, Australia. ²School of Mathematics and Computational Sciences, Sun Yet-Sen University, Guangzhou 510275, P.R. China.

Published: 23 October 2013

References

- Lewis J: **From Signals to Patterns: Space, Time, and Mathematics in Developmental Biology.** *Science* 2008, **322**(5900):399-403 [<http://www.sciencemag.org/content/322/5900/399.abstract>].
- Tomlin CJ, Axelrod JD: **Biology by numbers: mathematical modelling in developmental biology.** *Nat Rev Genet* 2007, **8**(5):331-340 [<http://dx.doi.org/10.1038/nrg2098>].
- Zhou Y, Zhuang X: **Kinetic Analysis of Sequential Multistep Reactions.** *The Journal of Physical Chemistry B* 2007, **111**(48):13600-13610 [<http://pubs.acs.org/doi/abs/10.1021/jp073708%2B>].
- Branz SE: **A Primer to Mechanism in Organic Chemistry (Sykes, Peter).** *Journal of Chemical Education* 1996, **73**(12):A313 [<http://pubs.acs.org/doi/abs/10.1021/ed073pA313.2>].
- Qin F, Li L: **Model-Based Fitting of Single-Channel Dwell-Time Distributions.** *Biophysical Journal* 2004, **87**(3):1657-1671 [<http://www.sciencedirect.com/science/article/pii/S0006349504736474>].
- Monk NA: **Oscillatory expression of Hes1, p53, and NF-kappaB driven by transcriptional time delays.** *Current Biology* 2003, **13**(16):1409-1413 [<http://view.ncbi.nlm.nih.gov/pubmed/12932324>].
- Zhu R, Ribeiro AS, Salahub D, Kauffman SA: **Studying genetic regulatory networks at the molecular level: Delayed reaction stochastic models.** *Journal of Theoretical Biology* 2007, **246**(4):725-745 [<http://www.sciencedirect.com/science/article/pii/S0022519307000513>].
- Ma L, Wagner J, Rice JJ, Hu W, Levine AJ, Stolovitzky GA: **A plausible model for the digital response of p53 to DNA damage.** *Proc Natl Acad Sci USA* 2005, **102**(40):14266-71 [<http://www.biomedsearch.com/nih/plausible-model-digital-response-p53/16186499.html>].
- Burrage K, Hancock J, Leier A, Jr DN: **Modelling and simulation techniques for membrane biology.** *Briefings in Bioinformatics* 2007, **8**(4):234-244 [<http://eprints.qut.edu.au/44378/1>].
- Kaern M, Elston T, Blake W, Collins J: **Stochasticity in gene expression: from theories to phenotypes.** *Nat Rev Genet* 2005, **6**(6):451-64.
- Wilkinson D: **Stochastic modelling for quantitative description of heterogeneous biological systems.** *Nat Rev Genet* 2009, **10**(2):122-33.
- McQuarrie DA: **Stochastic Approach to Chemical Kinetics.** *Journal of Applied Probability* 1967, **4**(3):413-478 [<http://dx.doi.org/10.2307/3212214>].
- Gillespie DT: **A rigorous derivation of the chemical master equation.** *Physica A Statistical Mechanics and its Applications* 1992, **188**:404-425.
- Gillespie DT: **Exact stochastic simulation of coupled chemical reactions.** *J Phys Chem* 1977, **81**(25):2340-2361 [<http://pubs.acs.org/doi/abs/10.1021/j100540a008>].

15. Bratsun D, Volfson D, Tsimring LS, Hasty J: **Delay-induced stochastic oscillations in gene regulation.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**(41):14593-14598 [<http://www.pnas.org/content/102/41/14593.abstract>].
16. Barrio M, Burrage K, Leier A, Tian T: **Oscillatory Regulation of Hes1: Discrete Stochastic Delay Modelling and Simulation.** *PLoS Computational Biology* 2006, **2**(9) [<http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.0020117>].
17. Roussel M, Zhu R: **Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression.** *Phys Biol* 2006, **3**(4):274-84.
18. Tian T, Burrage K, Burrage PM, Carletti M: **Stochastic delay differential equations for genetic regulatory networks.** *Journal of Computational and Applied Mathematics* 2007, **205**(2):696-707 [<http://www.sciencedirect.com/science/article/pii/S0377042706003943>].
19. Schlicht R, Winkler G: **A delay stochastic process with applications in molecular biology.** *Journal of Mathematical Biology* 2008, **57**:613-648 [<http://dx.doi.org/10.1007/s00285-008-0178-y>].
20. Agrawal S, Archer C, Schaffer DV: **Computational Models of the Notch Network Elucidate Mechanisms of Context-dependent Signaling.** *PLoS Comput Biol* 2009, **5**(5):e1000390 [<http://dx.doi.org/10.1371%2Fjournal.pcbi.1000390>].
21. Marquez-Lago T, Leier A, Burrage K: **Probability distributed time delays: integrating spatial effects into temporal models.** *BMC Systems Biology* 2010, **4**:1-16 [<http://eprints.qut.edu.au/42737/>].
22. Marquez-Lago T, Stelling J: **Counter-intuitive stochastic behavior of simple gene circuits with negative feedback.** *Biophys J* 2010, **98**(9):1742-50.
23. Leier A, Marquez-Lago T, Burrage K: **Generalized binomial tau-leap method for biochemical kinetics incorporating both delay and intrinsic noise.** *J Chem Phys* 2008, **128**(20):205107.
24. Pahle J: **Biochemical simulations: stochastic, approximate stochastic and hybrid approaches.** *Brief Bioinform* 2009.
25. Bayati B, Chatelain P, Koumoutsakos P: **D-leaping: Accelerating stochastic simulation algorithms for reactions with delays.** *Journal of Computational Physics* 2009, **228**(16):5908-5916 [<http://www.sciencedirect.com/science/article/pii/S0021999109002435>].
26. Gillespie D: **Stochastic simulation of chemical kinetics.** *Annu Rev Phys Chem* 2007, **58**.
27. Mier-y Terán-Romero L, Silber M, Hatzimanikatis V: **The Origins of Time-Delay in Template Biopolymerization Processes.** *PLoS Computational Biology* 2010, **6**(4) [<http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1000726>].
28. Thomas P, Straube A, Grima R: **The slow-scale linear noise approximation: an accurate, reduced stochastic description of biochemical networks under timescale separation conditions.** *BMC Syst Biol* 2012, **6**:39.
29. Srivastava R, Haseltine EL, Mastny E, Rawlings JB: **The stochastic quasi-steady-state assumption: reducing the model but not the noise.** *J Chem Phys* 2011, **134**(15):154109 [<http://www.biomedsearch.com/nih/stochasticquasi-steady-state-assumption/21513377.html>].
30. Tian T: **Chemical memory reactions induced bursting dynamics in gene expression.** *PLoS One* 2013, **8**:e52029.
31. Mitchell P, Tollervey D: **mRNA turnover.** *Current Opinion in Cell Biology* 2001, **13**(3):320-325 [<http://www.sciencedirect.com/science/article/pii/S0955067400002143>].
32. Gameau NL, Wilusz J, Wilusz CJ: **The highways and byways of mRNA decay.** *Nat Rev Mol Cell Biol* 2007, **8**(2):113-126 [<http://dx.doi.org/10.1038/nrm2104>].
33. Shyu AB, Wilkinson MF, van Hoof A: **Messenger RNA regulation: to translate or to degrade.** *EMBO J* 2008, **2**(3):471-478 [<http://www.nature.com/emboj/journal/v27/n3/full/1601977a.html>].
34. van Hoof A, Parker R: **Messenger RNA Degradation: Beginning at the End.** *Current Biology* 2002, **12**(8):R285-R287 [<http://www.sciencedirect.com/science/article/pii/S0960982202008023>].
35. Cao D, Parker R: **Computational modeling of eukaryotic mRNA turnover.** *RNA* 2001, **7**(9):1192-1212 [<http://majournal.cshlp.org/content/7/9/1192.abstract>].
36. Cao D, Parker R: **Computational Modeling and Experimental Analysis of Nonsense-Mediated Decay in Yeast.** *Cell* 2003, **113**(4):533-545 [<http://www.sciencedirect.com/science/article/pii/S0092867403003532>].
37. Tian T: **Simplified stochastic models with time delay for studying the degradation process of mRNA molecules.** *International Journal of Data Mining and Bioinformatics* 2012.
38. Wu Q, Smith-Miles K, Tian T: **A two-variable model for stochastic modelling of chemical events with multi-step reactions.** *Bioinformatics and Biomedicine (BIBM), 2012 IEEE International Conference on* 2012, 1-6.
39. Bregman A, Avraham-Kelbert M, Barkai O, Duek L, Guterman A, Choder M: **Promoter elements regulate cytoplasmic mRNA decay.** *Cell* 2011, **147**(7):1473-83.
40. Trcek T, Larson DR, Moldón A, Query CC, Singer RH: **Single-Molecule mRNA Decay Measurements Reveal Promoter-Regulated mRNA Stability in Yeast.** *Cell* 2011, **147**(7):1484-1497 [<http://www.sciencedirect.com/science/article/pii/S0092867411014450>].
41. Gillespie DT: **Approximate accelerated stochastic simulation of chemically reacting systems.** *J Chem Phys* 2001, **115**:1716-1733.

doi:10.1186/1752-0509-7-S4-S14

Cite this article as: Wu *et al.*: Stochastic modelling of biochemical systems of multi-step reactions using a simplified two-variable model. *BMC Systems Biology* 2013 **7**(Suppl 4):S14.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

