Mathematical models for gene expression and regulation based on integer delay difference equations

基于整数型延时差分方程组的基因表达调控的数学模型研究

Gao Zhu

Instructor Gao Junhui

Xishan High School of Jiangsu Province

Abstract

Differential equations usually used to establish the mathematical model for the gene expression and regulation. However, the molecular discreteness and the time delay effect are not negligible for the biochemical reactions and gene netwoks. In order to describe a more realistic cell world, we attempt to use integer delay difference equations to study the processes of gene expression and regulation, with the number of discrete molecules instead of continuous molecule concentration, taking into consideration the delay effect in transcription, translation, and regulation.

The models constructed include single gene expression model, regulatory network model of multiple genes, and signal pathway-gene regulation model. Some processes, such as process of genes activated by transcription factors, process of membrane receptor activated by extracellular ligands, process of transcription factor activated by membrane receptor, are similar but somewhat different. Therefore, this paper constructs a framework for the competition model between the ligands to bind one receptor. The above models are carried out with numerical simulation in this paper reproducing biological phenomena such as ‘transcriptional bursting’, ‘translational bursting’, and ‘gene switch’. In addition, the paper has discovered and proved average concentration formula of mRNA associated with the delay parameters under the assumption that the gene is activated and inactivated alternately.

Key words : delay difference equations, integer, gene expression and regulation, signal pathway, numerical simulation

CONTENTS

1 Topics background

It is very interesting that the cellular molecular quantitative is changing, which is related to the gene expression and regulation. There are usually two mathematical models for gene expression, which is deterministic model based on the ordinary differential equation and stochastic model based on the stochastic differential equation. However, these two models either ignore the discreteness of molecular number or ignore the delay effect. We try to use integer delay difference equations to study the process of gene expression and gene regulation, with the number of discrete molecules instead of continuous molecule concentration, taking into consideration the delay effect in transcription, translation, and regulation.

In fact, gene expression and regulation in cells is more complex, which involves many factors such as the multiple ligands competing to bind one same receptor. We constructed a framework for the competition model between the ligands to bind one receptor.

By combing the two models above we simulated regulatory network of multiple genes and a very important process in cells, signal pathway. We think more complicated biology process can be simulated by using our method, which can help us to understand the biology mechanism more and more intuitively.

2 Single gene expression model

2.1 Molecular biology background

Gene expression refers to the process during which the genetic information storing in DNA sequence is changed to protein by transcription and translation.

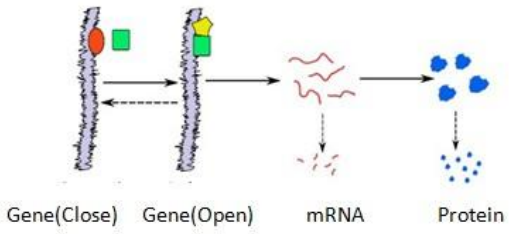


Figure 1 Schematic diagram for gene expression

Figure 1 describes the process of single gene expression which involves a series of biochemical reactions such as the state transition of promoter between inactivation and activation, DNA transcribing mRNA, mRNA translating protein, and the degradation of mRNA and protein. Here we omit the modification and processing for mRNA, and the impaction of the transcription factors on promoter, etc.

2.2 Mathematical model

The mathematical model for single gene expression by using integer delay difference equations is as follows.

Suppose gene is activated and inactivated periodically and the period is τ_1. The activation and inactivation time is τ_1^+ and τ_1^- respectively. That is:

$$\tau_1 = \tau_1^+ + \tau_1^- \tag{2-1}$$

Further, we assume that one mRNA molecule is transcribed per time unit, will mature after time τ_2 and be degraded after time τ_3 (include τ_2). Similarly, the mature mRNA molecules are translated to proteins by one to one per time unit, the protein will mature by time τ_4 and to be degraded by time τ_5 (include τ_4). Suppose at time t, the number of activated gene, mature mRNA molecules and proteins is $x_1(t)$, $x_2(t)$, $x_3(t)$, respectively. Then we have:

$$\begin{cases} x_1(t) = x_1(t-\tau_1) \\ x_2(t) = x_2(t-1) + x_1(t-\tau_2) - x_1(t-\tau_3) \\ x_3(t) = x_3(t-1) + x_2(t-\tau_4) - x_2(t-\tau_5) \end{cases} \tag{2-2}$$

The initial conditions for single gene model are as follow:

$$\begin{cases} x_1(t) = 0, t < 0, i = 1,2,3 \\ x_1(t) = 1, t \in \left[0, \tau_1^+ - 1\right] \\ x_1(t) = 0, t \in \left[\tau_1^+, \tau_1 - 1\right] \\ x_2(0) = 0 \\ x_3(0) = 0 \end{cases}$$

So $x_1(t) \in \{0,1\}$. Notice that, $x_1(t) = 1 - x_1\left(t - \tau_1^+\right)$ when the activation time is same as the inactivation time.

2.3 Analysis

2.3.1 Properties of the activated gene number

Apparently, $x_1(t)$ is periodic function with the period τ_1 and the average value $\frac{\tau_1^+}{\tau_1^+ + \tau_1^-}$. We suppose that the variance of $x_1(t)$ is $D(x_1)$, then :

$$\begin{aligned} & D(x_1) \\ &= \frac{\tau_1^+\left(1-\frac{\tau_1^+}{\tau_1^+ + \tau_1^-}\right)^2 + \tau_1^-\left(0-\frac{\tau_1^+}{\tau_1^+ + \tau_1^-}\right)^2}{\tau_1} \\ &= \frac{\tau_1^+\left(\frac{\tau_1^-}{\tau_1^+ + \tau_1^-}\right)^2 + \tau_1^-\left(\frac{\tau_1^+}{\tau_1^+ + \tau_1^-}\right)^2}{\tau_1} \\ &= \frac{\tau_1^+\tau_1^-\left(\tau_1^+ + \tau_1^-\right)}{\left(\tau_1^+ + \tau_1^-\right)^2 \tau_1} = \frac{\tau_1^+\tau_1^-}{\tau_1^2} \end{aligned}$$

2.3.2 Properties of the mature mRNA

Property 1: $x_2(t)$ **is a periodic function with period** τ_1 **from time** τ_3**.**

Proof 1:

$$x_2(t) = x_2(t-1) + x_1(t-\tau_2) - x_1(t-\tau_3)$$

so only to prove $f(t) = x_1(t-\tau_2) - x_1(t-\tau_3)$ is a periodic function($t \geq \tau_3$).

When $t = t + \tau_1$,

$$f(t+\tau_1) = x_1(t+\tau_1-\tau_2) - x_1(t+\tau_1-\tau_3) = x_1(t-\tau_2) - x_1(t-\tau_3) = f(t)$$

Proved.

Proof 2:

$x_1(t)$ is a periodic function with period τ_1 from time τ_3. When $t=\tau_1$,

$$x_2(t+\tau_1) = x_2(t) + \sum_{i=t+1}^{t+\tau_1} x_1(i-\tau_2) - \sum_{i=t+1}^{t+\tau_1} x_1(i-\tau_3) = x_2(t) + 0 = x_2(t)$$

Proved.

Property 2: $x_2(t)$ **is a nondecreasing function in** $[0,\tau_3-1]$ **and gets max value when** $t=\tau_3-1$**.**

Proof:

(1) According to the initial condition $\begin{cases} x_i(t)=0, t<0 \\ x_3(0)=0 \end{cases}$, we can get that $x_1(t-\tau_3)$ identically equals 0 in $[0,\tau_3-1]$ while $x_1(t-\tau_2)=0$ or 1 in $[0,\tau_3-1]$, therefore,

$x_2(t)=x_2(t-1)+x_1(t-\tau_2)-x_1(t-\tau_3)$ is a nondecreasing function.

(2) On the other hand, we have $\begin{cases} x_1(t)=1, t\in[0,\tau_1^+-1] \\ x_1(t)=0, t\in[\tau_1^+,\tau_1-1] \end{cases}$, so we get that $x_2(\tau_3-1)\geq x_2(\tau_3)$.

(3) We have known $\begin{cases} x_1(t)=1, t\in[0,\tau_1^+-1] \\ x_1(t)=0, t\in[\tau_1^+,\tau_1-1] \end{cases}$, with the fact that in one period which starts at the moment τ_3, the value sequence of $x_1(t-\tau_3)$ must be like: 111..100..0 or 00...011...1, and the amount of 1 is τ_1^+, the amount of 0 is τ_1^-, but $x_1(t-\tau_2)$ is not similar with $x_1(t-\tau_3)$, its value sequence may like this: 00...011...100...0 or 11...100...011...1, the amount of 1 is τ_1^+, the amount of 0 is τ_1^-, so according to $x_2(t)=x_2(t-1)+x_1(t-\tau_2)-x_1(t-\tau_3)$, we get $x_2(t)\leq x_2(\tau_3-1)$ while t is in $[\tau_3,\tau_3+\tau_1-1]$.

At last, from (1),(2),(3) we can know that $x_2(t)$ gets max value at the moment τ_3-1.

Property 3: The average of $x_2(t)$ **is** $\frac{\tau_1^+(\tau_3-\tau_2)}{\tau_1^++\tau_1^-}$ **from time** τ_3**.**

Proof:

Suppose that $\tau_3-\tau_2=n\tau_1+k$, (n,k are nonnegative integer, $k<\tau_1$). Let $B=x_2(\tau_3-1)$, we mark S as total number of every moment in one period, according to the definition, $S=\sum_{t=\tau_3}^{\tau_3+\tau_1}x_2(t)$. So S is no more than $B\tau_1$, suppose $L=B\tau_1-S$, let's do some analysis of the relationship between k and B,S,L;

(1) $k=0$, $\tau_3-\tau_2=n\tau_1$, $B=n\tau_1^+$, $L=0$;

so we have $S=B\tau_1-L=n\tau_1^+\tau_1-0=n\tau_1^+\left(\dfrac{\tau_3-\tau_2}{n}\right)=\tau_1^+\left(\tau_3-\tau_2\right)$.

(2) $0<k<\tau_1^+$, $B=n\tau_1^++k$;

If $\tau_1^+\geq\tau_1^-$, consider the period starts at time τ_3, the value sequence of $x_1\left(t-\tau_2\right)$ must be like this format: first τ_1^+-k positions are '1', next τ_1^- positions are '0', last k positions are '1', for $x_1\left(t-\tau_3\right)$, its value sequence must be: first τ_1^+ positions are '1' and next τ_1^- positions are '0', according to the definition, $x_2\left(t\right)=x_2\left(t-1\right)+x_1\left(t-\tau_2\right)-x_1\left(t-\tau_3\right)$, so we can get table 1.

Table 1 The change of $x_2\left(t\right)$ on $\left[\tau_3-1,\tau_3+\tau_1\right]$

t	$x_2\left(t-1\right)$	$x_1\left(t-\tau_2\right)$	$x_1\left(t-\tau_3\right)$	$x_2\left(t\right)$	Row number	
τ_3-1	B-1	1	0	B	1	
τ_3	B	1	1	B	τ_1^+-k	
	B	…	…	B		
	B	1	1	B		
	B	0	1	B-1	k	τ_1^-
	B-1	0	1	B-2		
	B-2	…	…	…		
	…	0	1	B-k		
	B-k	0	0	B-k	τ_1^--k	
	B-k	…	…	B-k		
	B-k	0	0	B-k		
	B-k	1	0	B-(k-1)	k	
	B-(k-1)	…	…	…		
	…	1	0	B-2		
	B-2	1	0	B-1		
$\tau_3+\tau_1$	B-1	1	0	B		

Here we have $L=(1+2+\ldots+k)+k\left(\tau_1^- -k\right)+(1+2+\ldots+k-1)]=k^2+k\left(\tau_1^- -k\right)$, then

$$\begin{aligned}S&=B\tau_1-L=(n\tau_1^+ +k)\tau_1-\left[k^2+k(\tau_1^- -k)\right]\\&=n\tau_1^+\tau_1+k\tau_1-k\tau_1^- =n\tau_1^+\tau_1+k\tau_1^+ =\tau_1^+(n\tau_1+k)\\&=\tau_1^+\left(\tau_3-\tau_2\right)\end{aligned}$$

If $\tau_1^+ <\tau_1^-$, similarly, we can get $S=\tau_1^+\left(\tau_3-\tau_2\right)$.

(3) $k=\tau_1^+$, $B=n\tau_1^+ +\tau_1^+ =(n+1)\tau_1^+$;

similarly, $L=(\tau_1^-)^2+\tau_1^-(\tau_1^+ -\tau_1^-)$, then

$$\begin{aligned}&S=B\tau_1-L\\&=(n+1)\tau_1^+\tau_1-\left[(\tau_1^-)^2+\tau_1^-(\tau_1^+ -\tau_1^-)\right]\\&=(n+1)\tau_1^+\tau_1-\tau_1^-\tau_1^+ =\tau_1^+(n\tau_1+\tau_1-\tau_1^-)\\&=\tau_1^+(n\tau_1+\tau_1^+)=\tau_1^+(n\tau_1+k)=\tau_1^+(\tau_3-\tau_2)\end{aligned}$$

(4) $k>\tau_1^+$, $B=n\tau_1^+ +\tau_1^+ =(n+1)\tau_1^+$;

similarly, $L=(\tau_1-k)^2+(\tau_1-k)\left[\tau_1^+ -(\tau_1-k)\right]$,so

$$\begin{aligned}&S=B\tau_1-L=(n+1)\tau_1^+\tau_1-\left\{(\tau_1-k)^2+(\tau_1-k)\left[\tau_1^+ -(\tau_1-k)\right]\right\}\\&=(n+1)\tau_1^+\tau_1-\left[(\tau_1)^2-2k\tau_1+k^2+\tau_1\tau_1^+ -(\tau_1)^2+k\tau_1-k\tau_1^+ +k\tau_1-k^2\right]\\&=(n+1)\tau_1^+\tau_1-\left(\tau_1\tau_1^+ -k\tau_1^+\right)=\tau_1^+\left[(n+1)\tau_1-(\tau_1-k)\right]\\&=\tau_1^+\left(n\tau_1+\tau_1-\tau_1+k\right)=\tau_1^+\left(n\tau_1+k\right)=\tau_1^+\left(\tau_3-\tau_2\right)\end{aligned}$$

In summary, average of $x_2(t)$ is $\dfrac{S}{\tau_1}=\dfrac{\tau_1^+\left(\tau_3-\tau_2\right)}{\tau_1^+ +\tau_1^-}$, proof is completed.

2.3.3 Properties of the mature protein

After a sufficient time, $x_3(t)$ is a periodic function with period τ_1;

Proof:

Because from moment τ_3, $x_2(t)$ is a periodic function with period τ_1, while $t=\tau_1$, we have

$$\begin{aligned}&x_3\left(t+\tau_1\right)\\&=x_3(t)+\sum_{i=t+1}^{t+\tau_1}x_2\left(i-\tau_4\right)-\sum_{i=t+1}^{t+\tau_1}x_2\left(i-\tau_5\right)\\&=x_3(t)+0=x_3(t)\end{aligned}$$

Proved.

2.4 Numerical simulation and application

In recent ten years, researches about 'transcriptional bursting' and 'translational bursting' phenomenon continuously sprung up [1][2] and we reconstruct these phenomenon by numerical simulations.

Transcriptional bursting

τ_1 is called survival time for activated gene which describes the period for gene switch between the activation and inactivation. τ_3 and τ_5 represents the average survival time for mRNA molecules and proteins respectively.

If the time of gene activation or inactivation process is much more longer than the average survival time for mRNA and protein, and the activation efficiency is low, the static fluctuation of protein number is large. Figure 2 simulates the model (2-2) in which the phenomenon is called transcriptional bursting, caused by long time activation and inactivation of promoter.

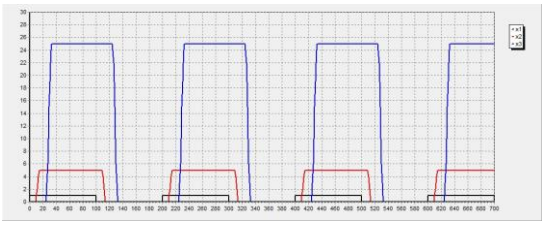
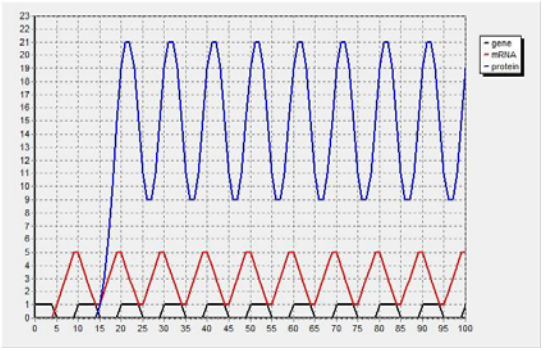


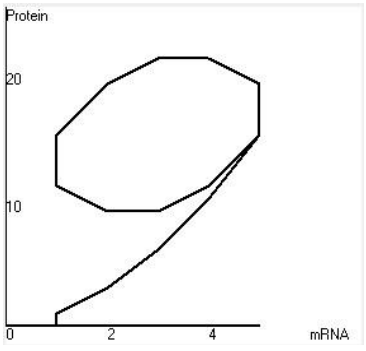
Figure 2 Transcriptional bursting， $\tau_1 : \tau_3 : \tau_5 = 200 : 15 : 20$

Translational bursting

The fluctuation of protein number is mainly associated with the fluctuation of mRNA number and this phenomenon is called translational bursting.



(a) $\tau_5 = 15, \tau_4 = 10$



(b) Number variation of mRNA and protein

Figure 3 Numerical simulation for the phenomena of ' translational bursting '

Simulation with yeast cell parameters

We extracted the experimental parameters of the E.coli and yeast cells from reference [3], which are shown in table 2. Then we re-simulate the equations (2-2) according to the parameters of yeast in table 2 and obtained the results which are given in figure 4.

Table 2 Relevant parameters of Bacteria and yeast (extracted from reference [3])

Property	E. coli	Yeast
Time to transcribe a gene	~ 1min	~ 1min
Time to translate a protein	~ 2min	~ 2min
Typical mRNA lifetime	2 − 5min	10 min to over 1 h
Timescale of transcription factor binding to DNA site	~ 1sec	~ 1sec

In figure 4, the blue curve represents number of mRNA molecules, the red curve represents number of protein molecules, the simulation parameters include gene switch cycle of 800 seconds, the life of mRNA molecules of 600 seconds, cell cycle of 2 hours, transcription time of 30 seconds, and translation time of 60 seconds.

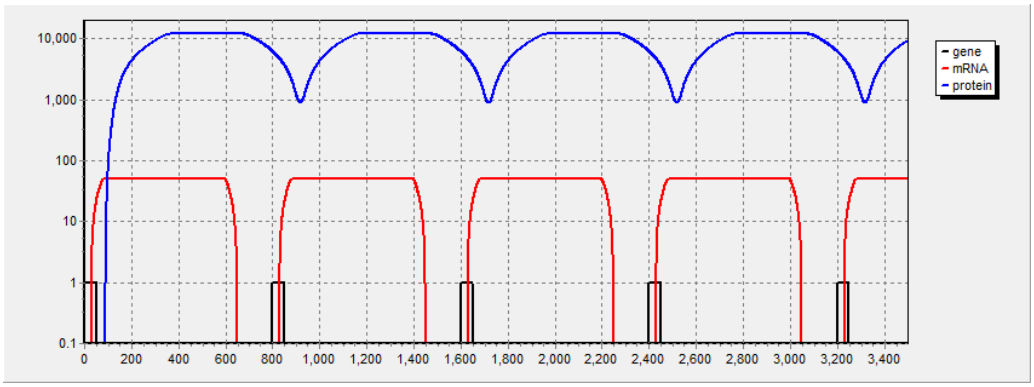


Figure 4 The simulation results using yeast parameters

Random switch of gene status

If the time of gene activation and inactivation is random, we assume that the ratio of activation and inactivation time is 8:2 and the simulation results are shown in figure 5.

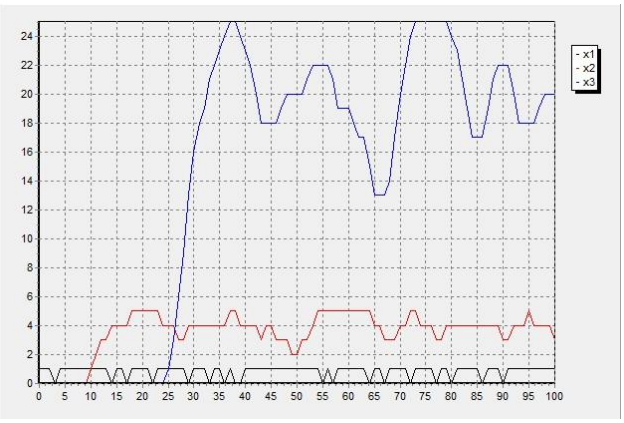


Figure 5 The simulation results of random switching of gene status

$(\tau_1^+ : \tau_1^- = 8:2\ ,\ \ \tau_2 = 10, \tau_3 = 15, \tau_4 = 15, \tau_5 = 20)$

3 Competition model between the ligands to bind one receptor

3.1 Molecular biology background

In the previous model of single gene expression, we assume that gene is activated and inactivated periodically. In fact, quite a part of the genes are activated because a kind of specific protein (known as the gene transcription factor) combines with the gene's promoter region to activate the gene. And there are usually more than one transcription factor around the promoter at a moment, but only one can combine with the promoter region.

The similar situation occurs when the receptor proteins in the cell membrane combine with the small ligands molecule outside the membrane. The receptor is a kind of protein which can combine with the extracellular specific signal molecule (ligand) and cause cell response. And it can be divided into cell surface receptor and intracellular receptor. The binding of receptor and ligand will change the molecular conformation to cause cell response, such as mediating signal transduction between cells, cell,adhesion, cell endocytosis and so on.

Here, gene and transcription factor is regarded as the receptor and the ligand. We call this model competition model between the ligands to bind one receptor, for only one ligand can bind with the receptor at a moment.

3.2 Mathematical model

The famous function, Hill function used to described the ligands competing for same receptor form as $\frac{g_a}{1+k[B]^n}$, which means the effect of n molecules B (ligands) to molecule A (receptor). Therein, g_a is the maximum productivity of A, k is the intensity of the repression and n is called Hill coefficient. Considering the time delay, integer number of molecules and only one ligand can bind with the receptor, we can not use this function. Therefore, we have to construct an associated model.

First, we defined a random function with value 0 or 1;

$$r\left(k^+,k^-\right)=\begin{cases}1, random\left(k^++k^-\right)<k^+\\0, random\left(k^++k^-\right)\geq k^+\end{cases} \quad (3\text{-}1)$$

Both k^+,k^- are positive integers, k^+,k^- denote average binding capacity and separation capacity of ligand and receptor (Eg, protein and operon) respectively and $random(N)$ produce a random integer in $[0,N]$. Denote $r\left(k^+\right)=r\left(k^+,k^-\right),r\left(k^-\right)=r\left(k^-,k^+\right)$ for convenience.

Simply suppose that there is only one receptor in free or the binding status with ligand and the number of ligands is from 0 to N . Then suppose at time $t-1$, the amount of free ligands is $x_f(t-1)$ and the amount of ligands binding with receptor is $x_b(t-1)$. Then at the time t, we have:

$$\begin{cases}x_f(t)=x_f(t-1)-C(x,k^+,t)[1-x_b(t-1)]+r(k^-)x_b(t-1)\\x_b(t)=x_b(t-1)+C(x,k^+,t)[1-x_b(t-1)]-r(k^-)x_b(t-1)\end{cases} \quad (3\text{-}2)$$

The second item of the right side of the equation above is $C\left(x,k^+,t\right)[1-x_b\left(t-1\right)]$, representing the deceasing number of free ligands which bind with the receptor at time t. k^+,k^- denote the average binding capacity and average separation capacity of protein and operon respectively. $C\left(x,k^+,t\right)$ represents the binding capacity of protein x with corresponding operon. Two implementations of $C\left(x,k^+,t\right)$ are as follow:

Implementation 1:

$$C\left(x,k^+,t\right)=r(k^+)Q_x(t) \quad (3\text{-}3)$$

Function $Q_x(t)$ with value 0 or 1 is used to judge whether there are ligands x arriving at the surface of corresponding receptor at time t.

Implementation 2:

$$C(x,k^+,t)=1-\prod_{i=1}^{\min(N,Q_x(t))}\left[1-r_i(k^+)\right] \quad (3\text{-}4)$$

Function $Q_x(t)$ represents the amount of all ligands which have arrived at the surface of receptor, N is the amount of molecules at saturation. If the amount of molecules is larger than N, the binding capacity of ligand and receptor won't increase because the surface of receptor (active site) is limited. As long as there is a ligand molecule binding with receptor successfully, the second item of the right side of the equation will be 0, then $C(x,k^+,t)$ will be 1.

4 The double gene regulation model

4.1 Molecular biology background

Auto-regulation is the simplest regulation form for single gene and there are many complicated regulation forms such as dual genes regulation or multiple genes regulation. In 2000, Timothy S. Gardner and other scientists at Boston University successfully synthesized a gene switch in E. coli, known as the toggle switch system [4]. This gene switch consists of two kinds of inhibition proteins and two promoters. As shown in figure 6, the protein X binds to the operon of gene y, inhibiting the expression of the gene y, while the protein Y binds to the operon of gene x, inhibiting the expression of the gene x.

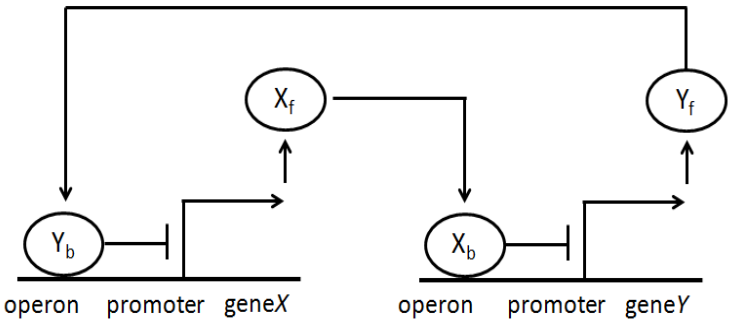


Figure 6 Schematic diagram for dual gene regulatory networks

4.2 Mathematical model

We try to establish a mathematical model for dual gene regulation system using the integer delay difference equations. For simplicity, we ignore the stage of mRNA, unified the process of transcription and translation as the protein synthesis process. Let $x_b(t)$, $y_b(t)$ denote the number of proteins binding to gene y and gene x at time t, $x_f(t)$, $y_f(t)$ denote the number of proteins in free state at time t, assume each gene is a single copy. Then we have:

$$0 \le x_b(t), y_b(t) \le 1 \qquad (4\text{-}1)$$

Let τ_1 denote the time from transcription and translation to mature protein generated (include mRNA splicing, modification, transport, the extension of the peptide chain on the ribosome, protein folding, etc). Let τ_2 (include τ_1) denote the survival time of the protein and τ_3 denote the time that protein reaches the corresponding operon position (include τ_1); Assume that one mRNA molecule is transcribed per time unit, then the number of protein X in free state at time t ($x_f(t)$) can be expressed as:

$$x_{\mathrm{f}}(t) = x_{\mathrm{f}}(t-1) + [1 - y_{\mathrm{b}}(t-\tau_1)] - [1 - y_{\mathrm{b}}(t-\tau_2)] - C\left(x, k_x^+, t\right)[1 - x_{\mathrm{b}}(t-1)] + r\left(k_x^-\right)x_{\mathrm{b}}(t-1) \qquad (4\text{-}2)$$

The second item of the right side of the equation above **(**$[1 - y_{\mathrm{b}}(t-\tau_1)]$**)** denotes the number of protein generated by gene from time τ_1 to t. Because at time $t-\tau_1$, we think that the number of protein Y in binding state is 0 and the gene x can generate one protein per time unit; while if the number of protein Y in binding state is 1, the gene x will be suppressed and can not express to generate protein X and the number of protein X generated by gene x is 0. Accordingly, the third item of the right side of the equation above denotes the number of degraded protein which is 0 or 1 at current time (time t). The fourth item of the right side of the equation above **(**$C\left(x, k_x^+, t\right)[1 - x_{\mathrm{b}}(t-1)]$**)** denotes the number of reduced protein X combing with the operon of gene y at time t with number 0 or 1. k^+, k^- denote average binding capacity and average separation capacity of protien and operon respectively. $C\left(x, k^+, t\right)$ denotes the binding capacity of protein X and corresponding operon at time t. There are two solutions for $C\left(x, k^+, t\right)$**.** One is

$C\left(x,k^{+},t\right)=r(k^{+})Q_{x}(t)$, where $Q_{x}\left(t\right)=1-\prod_{s=t-\tau_2}^{t-\tau 3}y_{b}\left(s\right),\prod_{t-\tau_2}^{t-\tau 3}y_{b}\left(t\right)$ denotes whether there is new protein X synthesized from $t-\tau_2$ to $t-\tau_3$ in which 1 indicates no and 0 indicates yes and $\tau_1<\tau_3<\tau_2$.

$Q_{x}\left(t\right)$ denotes whether there is a protein X reaches the corresponding operon (gene y) at time t. Let τ_3 denote the time that protein reaches the corresponding operon position (include τ_1), Obviously, only the proteins generated before $t-\tau_3$ have the chance. Another is $C\left(x,k^{+},t\right)=1-\prod_{i=1}^{\min\left(N,Q_{x}(t)\right)}\left[1-r_{i}\left(k^{+}\right)\right]$, where $Q_{x}\left(t\right)=\sum_{s=t-\tau_2}^{t-\tau 3}\left[1-y_{b}\left(s\right)\right]$. Accordingly, the fifth item of the right side of the equation (4-2) $\left(r\left(k^{-}\right)x_{\mathrm{b}}\left(t-1\right)\right)$ denotes the number of protein which is separated from the operon of gene y with number of 0 or 1.

Obviously, all of the above five items are integers. Therefore, using integer delay difference equations, the entire system is described as :

$$\begin{cases}x_{b}\left(t\right)=x_{b}\left(t-1\right)+C\left(x,k^{+},t\right)[1-x_{\mathrm{b}}\left(t-1\right)]-r\left(k^{-}\right)x_{\mathrm{b}}\left(t-1\right)\\ y_{b}\left(t\right)=y_{b}\left(t-1\right)+C\left(y,k^{+},t\right)[1-y_{\mathrm{b}}\left(t-1\right)]-r\left(k^{-}\right)y_{\mathrm{b}}\left(t-1\right)\\ x_{\mathrm{f}}\left(t\right)=x_{\mathrm{f}}\left(t-1\right)+[1-y_{\mathrm{b}}\left(t-\tau_1\right)]-[1-y_{\mathrm{b}}\left(t-\tau_2\right)]-C\left(x,k^{+},t\right)[1-x_{\mathrm{b}}\left(t-1\right)]+r\left(k^{-}\right)x_{\mathrm{b}}\left(t-1\right)\\ y_{\mathrm{f}}\left(t\right)=y_{\mathrm{f}}\left(t-1\right)+[1-x_{\mathrm{b}}\left(t-\tau_1\right)]-[1-x_{\mathrm{b}}\left(t-\tau_2\right)]-C\left(y,k^{+},t\right)[1-y_{\mathrm{b}}\left(t-1\right)]+r\left(k^{-}\right)y_{\mathrm{b}}\left(t-1\right)\end{cases}$$

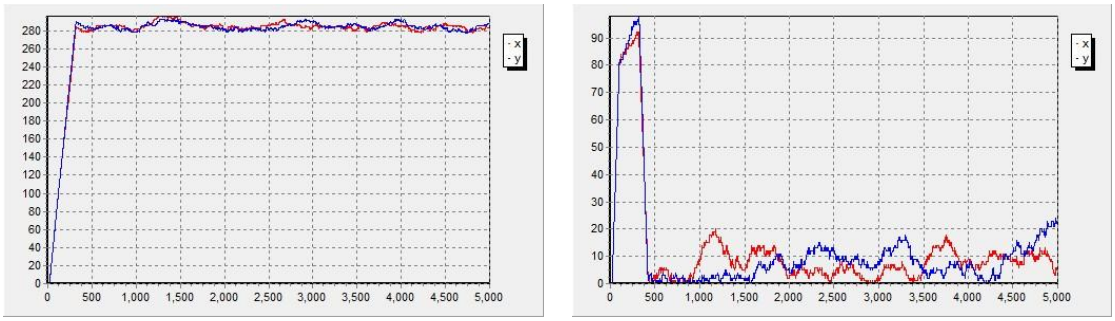
（4-3）

Further, when the promoters of the two genes are overlap, there is only one gene that can express and the system gives a mutually exclusive relationship with $[1-x_{\mathrm{b}}\left(t-1\right)]$ and $[1-y_{\mathrm{b}}\left(t-1\right)]$ changing to $[1-y_{\mathrm{b}}\left(t-1\right)-x_{\mathrm{b}}\left(t-1\right)]$ and $[1-x_{\mathrm{b}}\left(t-1\right)-y_{\mathrm{b}}\left(t-1\right)]$ in formula (4-3).

4.3 Numerical simulation

When comparing conditons of weak inhibition ($k_{+}/k_{-}=0.05$) with strong inhibition ($k_{+}/k_{-}=20$), we get results which are shown in figure 7 (a), (b). The figure shows that the system has only one stable equilibrium state for weak inhibition and can not be transferred to other states. But for strong inhibition condition, the system has three possible states: 1) the X is dominant, 2) the Y is dominant and 3) X and Y

inhibit each other.

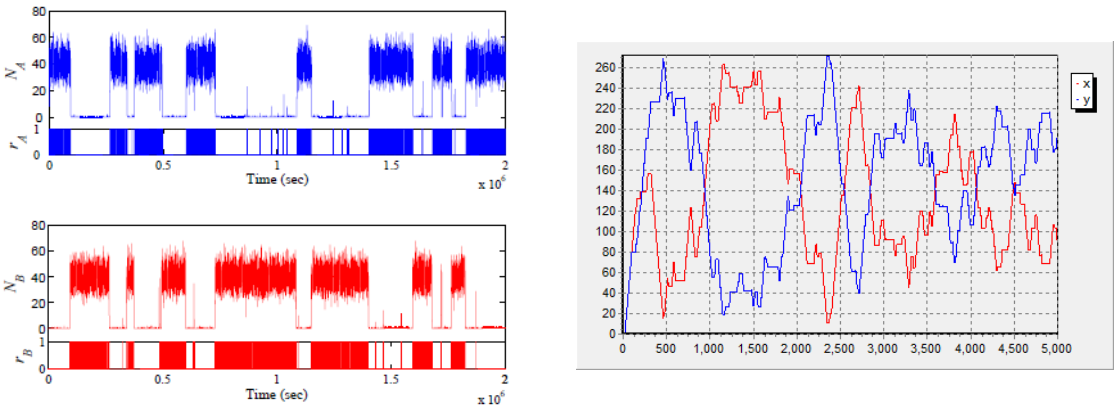


(a) $k_+ / k_- = 0.05$ (b) $k_+ / k_- = 20$

Figure 7 The dual gene regulation system ($\tau_1 : \tau_2 : \tau_3$=20:320:100)

The red and blue curves represent the change of the number of protein X and Y, showing the phenomenon of 'gene switch'. The numerical simulation shows that as k_+ / k_- increases, the average number decreases.

For the case of two gene promoters overlapping, a numerical simulation was carried out in literature [5] by using stochastic differential equations, and obtained the time sequence which is shown in figure 8 (a).The vertical coordinate denotes the number of molecules of gene-binding protein and free protein. And figure 8 (a) shows the phenomenon of the 'gene switch'.

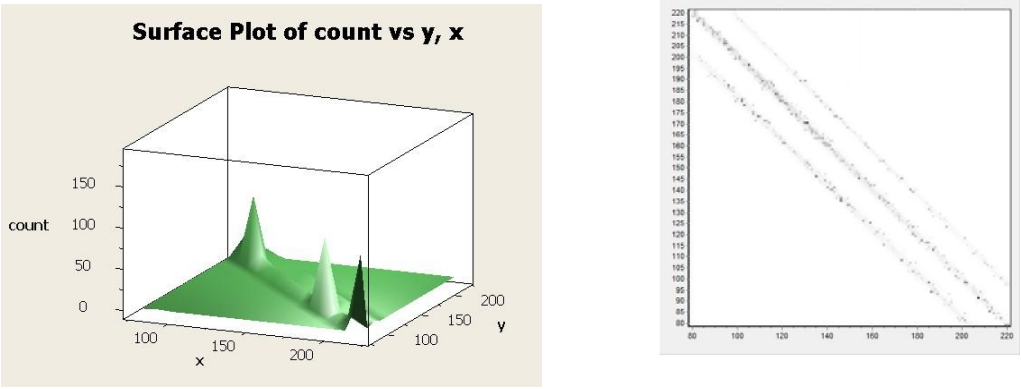


(a) Gene switch (Taken from [5]) (b) Result of integer delay differencr equations

Figure 8 Numerical simulations for the dual gene promoter overlapping

In this paper, we simulated the dual-gene promoter overlapping system by using integer delay difference equations and the result is shown in figure 8 (b). The changes of the number of two kinds of proteins display symmetry around the mean value considerably. Because there is no consideration for time delay, the number of average molecule soon reaches to form a peak in figure 8 (a). But the molecules increase and decrease gradually in figure 8 (b), which accords with the actual situation.

Through statistic analysis of the data (x,y), we get the results showing in figure 9. Figure 9 (a) is three-dimensional with a few peaks on the diagonal ((200,100) to (100,200)). Figure 9 (b) is displayed in a two-dimensional manner with the gray scale representing the number in coordinate, the darker the color is, the greater the number is. Interestingly, the points with deep color are basically distributed in three straight lines which are almost parallel.



(a) three-dimensional manner (b) two-dimensional manner (grayscale)

Figure 9 The value distribution of data pairs ($x(t), y(t)$), ($k_+ / k_- = 995:5$)

5 Three-gene regulatory network model

5.1 Mathematical model

There are several cases of three-gene regulatory network and we consider only one case, that is three genes circulating repression (x repress y, y repress z and z repress x). According to the two-gene model, three-gene regulatory network model is:

$$\begin{cases} x_b(t) = x_b(t-1) + C(x,k^+,t)[1-x_{\mathrm{b}}(t-1)] - r(k^-)x_{\mathrm{b}}(t-1) \\ y_b(t) = y_b(t-1) + C(y,k^+,t)[1-y_{\mathrm{b}}(t-1)] - r(k^-)y_{\mathrm{b}}(t-1) \\ z_b(t) = z_b(t-1) + C(z,k^+,t)[1-y_{\mathrm{b}}(t-1)] - r(k^-)y_{\mathrm{b}}(t-1) \\ x_{\mathrm{f}}(t) = x_{\mathrm{f}}(t-1) + [1-z_{\mathrm{b}}(t-\tau_1)] - [1-z_{\mathrm{b}}(t-\tau_2)] - C(x,k^+,t)[1-x_{\mathrm{b}}(t-1)] + r(k^-)x_{\mathrm{b}}(t-1) \\ y_{\mathrm{f}}(t) = y_{\mathrm{f}}(t-1) + [1-x_{\mathrm{b}}(t-\tau_1)] - [1-x_{\mathrm{b}}(t-\tau_2)] - C(y,k^+,t)[1-y_{\mathrm{b}}(t-1)] + r(k^-)y_{\mathrm{b}}(t-1) \\ z_{\mathrm{f}}(t) = z_{\mathrm{f}}(t-1) + [1-y_{\mathrm{b}}(t-\tau_1)] - [1-y_{\mathrm{b}}(t-\tau_2)] - C(z,k^+,t)[1-y_{\mathrm{b}}(t-1)] + r(k^-)y_{\mathrm{b}}(t-1) \end{cases}$$

If $C(i,k^+,t), i = x,y,z$ uses implementation 1, we have:

$$\begin{cases} Q_x(t) = 1 - \prod_{t-\tau_2}^{t-\tau 3} z_b(t) \\ Q_y(t) = 1 - \prod_{t-\tau_2}^{t-\tau 3} x_b(t) \\ Q_z(t) = 1 - \prod_{t-\tau_2}^{t-\tau 3} y_b(t) \end{cases}$$

5.2 Numerical simulation

Under the condition of strong inhibition, the result of numerical simulation is shown in figure 10, where the red, blue, yellow curves denote the number of molecules of the protein x,y,z respectively.

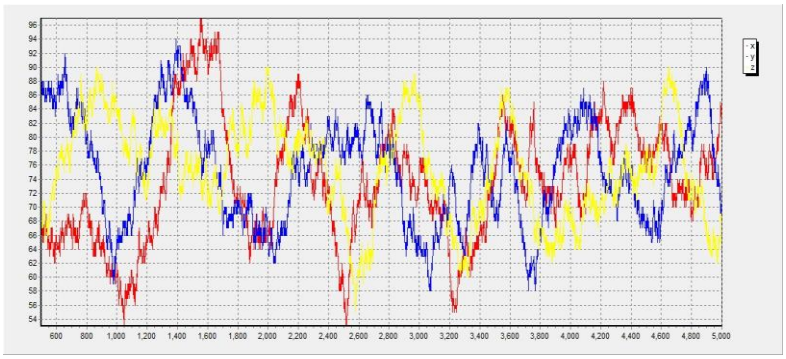


Figure 10 The simulation results of the three gene cycle repression

6 Models for cell signal pathways

We only consider the intracellular regulation mechanism in previous study, omitting the effect of environment to cell. In fact, extracellular stimulation can transfer to cell through multiple signal pathways which will regulate the related genes to guide the cell to adapt to environment. Taking the JAK-STAT pathway [6] as an example, we construct the corresponding delay difference equations model.

6.1 Molecular biology background

JAK-STAT is a signal pathway which closely related to cell growth, proliferation and differentiation. In recent years, the JAK-STAT pathway is deeply studied especially for the change in the hematopoietic process. The JAT-STAT signal pathway is shown in fugure 11.

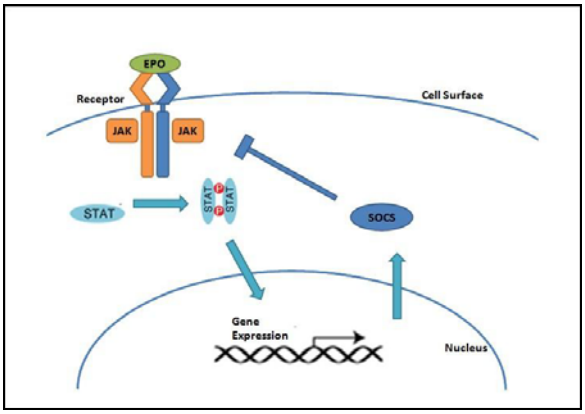


Figure 11 Signal pathways of JAK-STAT

When ligand EPO combines with receptor EPOR, the structure domain conformation in the membrane receptor will change, which can activate the JAK kinase family members related to the receptor. Activated JAK, mediates the tyrosine residue phosphorylation of the specific receptor sites, and becomes the anchor point for STAT molecule and other signal molecules. Once the tyrosine residue site of protein STAT is activated by JAK, the STAT will dissociate from the receptor. Then it changes into polymerization and translocate into the nucleus, regulates gene transcription in nucleus. This induces the expression of gene SOCS, which turns down the expression of JAK-STAT, inhibiting its biological response.

6.2 Mathematical model

To simplify the model, we see JAK and STAT as one molecule called JAK-STAT and combine translation and transcription to one process. Suppose the amount of EPO, activated EPOR, activated JAK-STAT, activated gene SOCS, and SOCS protein are $x_1(t), x_2(t), x_3(t), x_4(t), x_5(t)$ respectively, the model will be:

$$\begin{cases} x_1(t) = x_1(t-1) - C\left(x_1, k_{x_1}^+, t\right)\left[1 - x_2(t-1)\right] + r\left(k_{x_1}^-\right)x_2(t-1) - \min\left[x_1(t-\tau_1), x_1(t-1)\right] \\ x_2(t) = x_2(t-1) + C\left(x_1, k_{x_1}^+, t\right)\left[1 - x_2(t-1)\right] - r\left(k_{x_1}^-\right)x_2(t-1) \\ x_3(t) = x_3(t-1) + C\left(n_3 - x_3(t-1), k_{x_3}^+, t\right)x_2(t-1) - Q_{x_3}(t)Q_{x_5}(t) \\ x_4(t) = x_4(t-1) + C\left(x_3, k_{x_3}^+, t\right)\left[1 - x_4(t-1)\right] - r\left(k_{x_3}^-\right)x_4(t-1) \\ x_5(t) = x_5(t-1) + x_4(t-\tau_3) - x_4(t-\tau_4) \end{cases}$$

（6-1）

And the initial conditions are:

$$\begin{cases} x_i(t)=0, t<0 \\ x_1(0)=10 \\ x_2(0)=0 \\ x_3(0)=0 \\ x_4(0)=0 \\ x_5(0)=0 \end{cases}$$

In the equation of $x_1(t)$, the second,third, and fourth item of the right side represents the decreasing of free EPOR for combining to complex $x_2(t)$, the increasing of it for isolating from complex $x_2(t)$ and the decreasing for the degradation respectively.

In the equation of $x_2(t)$, the second and third item of the right side represents the increasing of combination by EPO and free EPOR, and the decreasing caused by separation of EPO.

In the equation of $x_3(t)$, the second item of the right side represents the increasing of activated JAK-STAT molecules, which are activated by the activated EPOR molecules at time $t-1$; the third item of the right side represents increasing of inactivated JAK-STAT molecules caused by SOCS protein molecules, which can also be marked as $\min\{1, \max[0, x_3(t-1)*x_5(t-1)]\}$. n_3 represents the sum of inactivated JAK-STAT and activated JAK-STAT.

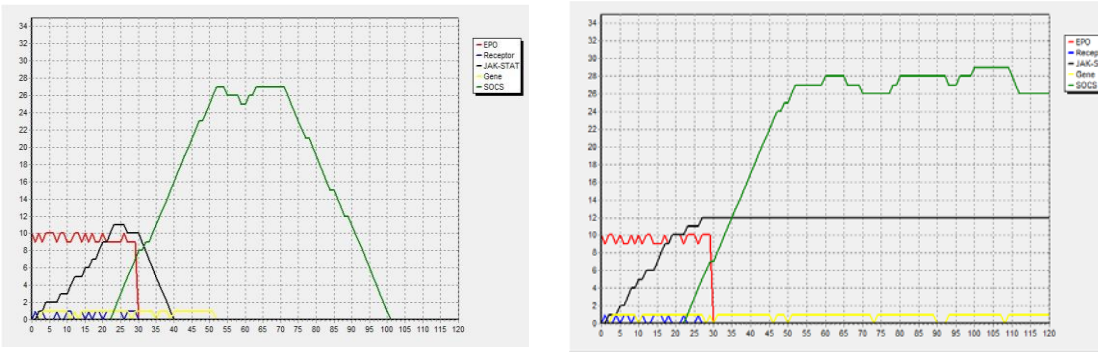
In the equation of $x_4(t)$, the second item of the right side represents the increasing of activated gene SOCS caused by JAK-STAT molecules at time $t-1$. The third item of the right side represents the decreasing of inactivated SOCS gene molecules caused by separation of JAK-STAT.

In the equation of $x_5(t)$, the second item of the right side represents the increasing caused by the expression of gene SOCS, the third item of the right side represents the decreasing caused by the degradation after survival time.

There are 3 delay parameters in the model (6-1), corresponding to τ_1 as time of EPO molecular degradation, τ_3 as time of transcription and translation of gene SOCS, τ_4 as survival time of SOCS protein respectively.

6.3 Numerical simulation

We simulate the model (6-1), taking $n_3 = 10$, $k_{x_1}^+ = 50, k_{x_1}^- = 50, k_{x_3}^+ = 98, k_{x_1}^- = 2$, $\tau_1 = 30, \tau_3 = 20, \tau_4 = 50$, the simulation results are shown in figure 12.



(a) Normal cells (b) Leukemia cells

Figure 12 Numerical simulation for signal pathways of JAK-STAT

In figure 12, curves of red, blue, black, yellow, and green denote the number of molecules of EPO, activated EPOR, activated JAK-STAT, activated gene SOCS, and SOCS protein changed with time respectively. From figure 12 (a), the number of JAK-STAT molecules is 0 at the initial time and gradually increases as the receptor is activated. Then the JAK-STAT molecules activate the gene SOCS to transcribe and translate SOCS protein. At time of 30 time units, all EPO molecules are degraded. As the number of JAK-STAT molecules in activated state gradually reduces, gene SOCS eventually closes, the number of SOCS protein molecules is no longer increase and maintains at a certain level for some time, then turns to 0 because molecules degrade one by one. The entire response process lasts about 100 time units.

Leukemia is a malignant disease of the hematopoietic tissue, also known as "blood cancer". There is sustained activation phenomenon of signal pathway of JAK-STAT in leukemia cells, because the point mutations of JAK gene make the SOCS molecules unable to close the activated JAK-STAT molecules and return to the inactivated state. The numerical simulation results are shown in figure 12(b).

7 Conclusion

This paper attempts to build mathematical models to study the process of gene expression and regulation with integer delay difference equations. We applied deterministic integer delay difference

equations to single gene expression model. The average concentration formula of mRNA associated with the delay parameters has been discovered and proved, under the assumption that the activation and the inactivation state of gene are alternately conducted. For more complicated models which include the competition for multiple ligands binding to one receptor, we constructed the framework of corresponding models. Then we adopted stochastic integer delay difference equations for dual-gene regulation model in which we discussed two cases of promoter separating and promoter overlapping. After that, the model of cycle repression for three-gene was established. Finally, we constructed a more complex model for signal pathway which took JAK-STAT signal pathway as an example demonstrating the process of signal transduction, gene expression and metabolic responses.

Our study demonstrated that gene expression and regulation can be simulated by integer delay difference equations. Table 3 compares the characteristics of the ordinary differential equation model, the stochastic model and the integer delay difference model in modeling and computing.

Table 3 The comparison of three models

	The ordinary differential equation model	The stochastic model	The integer delay difference model
Discreteness of molecular concentration	NO	YES	YES
Discreteness of time	NO	NO	YES
Delay	YES	NO	YES
Randomness	NO	YES	YES
Analysis methods	GOOD	Moderate	Poor
Computing speed	General	General	Fast

By comparison, we find that the integer delay difference model is superior to other models in considering molecular discreteness, time discreteness and time delay effect, which was accorded to the facts in describing the gene expression and regulation. In addition, faster calculation speed is another advantage of this model. So with the help of the integer delay difference model, we simulated biology phenomenon of ‘transcriptional bursting’, ‘translational bursting’ and ‘gene switch’. Of course, there is deficiency for this model, which is the lack of analysis methods. But we believe as the further research continues, the model will be improved step by step.

Acknowledgements

We thank Mr. Liu Fei and Ms. Liu Baohong from Shanghai Center for BioInformation Technology, Prof. Wang Lireng from the Mathematics Department of East China Normal University, Dr. Chen Ping from Massachusets Institute of Technology, and Dr. Ding Guohui from Shanghai Institutes for Biological Sciences of Chinese Academy of Sciences .

Thanks for the support and help from my mother Li Qiuping, my aunt Gao Haihong, and my teachers and classmates.

References

[1] Paulsson J, Models of stochastic gene expression. Phy Life Rev. 2005;2(2):157-75.

[2] Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. Cell. 2005;123(6):1025-36.

[3] Alon U, An Introduction to Systems Biology:Design Principles of Biological Circuits. London:Chapman & Hall/CRC. 2007.

[4] Gardner TS, Cantor CR, Collins JJ. Construction of a genetic toggle switch in Escherichia coli. Nature. 2000;403(6767):339-42.

[5] Lipshtat A, Loinger A, Balaban NQ, Biham O. Genetic toggle switch without cooperative binding. Phys Rev Lett. 2006;96(18):188101.

[6] Chen X, Xu F. Progress in the studies of JAK-STAT signal pathway. Chin J Nat. 2003;25(3):149-52.

Appendix

Code for numerical simulation of the expression of a single gene（Delphi）

```
unit Unit1;
interface
uses
```

```
  Windows, Messages, SysUtils, Variants, Classes, Graphics, Controls, Forms,
  Dialogs, TeEngine, Series, ExtCtrls, TeeProcs, Chart, StdCtrls;

const
  int=25;

type
  TForm2 = class(TForm)
    Button1: TButton;
    Chart1: TChart;
    Series1: TLineSeries;
    Series2: TLineSeries;
    Series3: TLineSeries;
    Series4: TLineSeries;
    procedure Button1Click(Sender: TObject);
  private
    { Private declarations }
  public
    { Public declarations }
  end;

var
  Form2: TForm2;
  t0,t1,t2,t3,t4,t5:integer;
  x:array[1..3,0..100000]of integer; // x[1]:Gene; x[2]:Mature mRNA; x[3]: Mature Protein;

implementation
{$R *.dfm}

procedure init;
```

```
begin
   t0:=2;
   t1:=2;
   t2:=1;        // Transcription delay of mRNA
   t3:=8;        // Survival time of mRNA
   t4:=5;        // Translation delay of Protein
   t5:=12;       // Survival time of Protein
end;

procedure workoutx;
var t:integer;
begin
   x[1][0]:=1;
   x[2][0]:=0;
   x[3][0]:=0;
   for t:=1 to int do
   begin
      if t<t0 then x[1][t]:=1
         else if t<t0+t1 then x[1][t]:=0
            else x[1][t]:=x[1][t-t0-t1];
      { Genes showed periodic variation, the ratio of activate and shut down time is t0: t1}

      if t<t2 then x[2][t]:=x[2][t-1]
         else if t<t3 then x[2][t]:=x[2][t-1]+x[1][t-t2]
            else x[2][t]:=x[2][t-1]+x[1][t-t2]-x[1][t-t3];
      { the gene transcript mRNA after time t2, the quantity is x[1][t-t2]
      The mRNA dies after time t3, the quantity is x[1][t-t3]}

      if t<t4 then x[3][t]:=x[3][t-1]
         else if t<t5 then x[3][t]:=x[3][t-1]+x[2][t-t4]
```

```
            else x[3][t]:=x[3][t-1]+x[2][t-t4]-x[2][t-t5];
      { mRNA translate protein after time t4, the quantity is x[2][t-t4]
      The protein dies after time t5, the quantity is x[1][t-t5]}

   end;
end;

procedure TForm2.Button1Click(Sender: TObject);
var i:integer;
begin
   init;                    // Initialization, set the parameter values
   workoutx;                // Analog the dynamic changes of genes, mRNA and protein
   for i:=0 to int do
   begin
      Chart1.Series[0].AddXY(i,x[1][i]);
      Chart1.Series[1].AddXY(i,x[2][i]);
      Chart1.Series[2].AddXY(i,x[3][i]);
   end;
end;
end.
```

Introduction of the team member

Gao Zhu, Grade 12 student at Jiangsu Xishan High School.

Publications:

- "A numerical simulation method to verify the shape of the orbit of planetary motion", *China Education Innovation Herald Vol*. 11, April 2012, First Author.
- "Molecular 3D structure learning software design", *China Education Innovation Herald Vol. 29, Oct*. 2011, First Author.
- "Bashful Grass Told me", *Composition World,* 2009.
- "Cooperation Strategies", *Shoots' Footsteps,* Selected Compositions of Wuxi Wuqiao Primary School Students, 2006.

Awards:

- 2012: First Author of "The development of multi-platform molecule 3D structure learning software", 1st Prize Winner, China Adolescents Science & Technology Innovation Contest, Jiangsu Province.
- 2011: Course Designer & Teacher of Chinese Kongfu, "Course Innovation Award" nominated by the Principal of Jiangsu Xishan High School, reported by Wuxi Media including Wuxi Education TV Station, Wuxi Education Network, Jiangnan Evening News, Wuxi Business Newspaper.
- 2011：Performer of "Xiaocui Talk Show" broadcast on Oct. 23rd at China Central TV Station. " Image Ambassador Award" nominated by the Principal of Jiangsu Xishang High School.
- 2011: 2nd Prize Winner of National Olympiad in Informatics, Jiangsu Province.
- 2010: 3rd Prize Winner of National Olympiad in Informatics, Jiangsu Province.
- Prize winners of painting, chess and oral English competitions at city and district levels.

Personal Interests:

Table Tennis, Music, Literature, Popular science works.