2018 International Genetically Engineered Machine Competition (iGEM)

Models Based on Stochastic Process and Differential Equations to Optimize System Properties

Team Fudan-CHINA

Summary

The core of our project is the signal transducer system on the cell membrane, which consists of a transcription factor chain (TC chain) and a protease chain (PC chain). The two chains can be combined with the help of an exogenous ligand, thus producing an transcription factor, which is regarded as the output. This is the main operating mechanism of our system[1].

There are several properties measuring the faculty of the system: background noise, system (saturated) output, dynamic range and system efficiency. The system output without exogenous ligand, regarded as the background noise, comes from the random collision between the two chains[1]. And the system saturated output, on the contrary, characterizes the maximum response strength of our system. Without question, the system will be superior if we maintain the saturated output as well as reduce background noise, namely the increasing of dynamic range.

Moreover, the TC chain without transcription factor may still be combined with a PC chain, which makes the PC chain occupied but does not contribute to system output. This phenomenon reduces the utilization of system resources, which is measured by the indicator called system efficiency. The practical significance of promoting system efficiency lies in achieving the same level of system output with fewer resources (i.e., lower chain generation rates).

Based on the above, this is the main content of our modeling:

- In Section 1.1, **as is partly verified by wet lab**, we quantitatively prove that the maximum of dynamic range can be reached by changing the production rates of PC and TC chains[1] and demonstrate how to do it.
- In Section 1.2, **for the high background noise from wet lab results**, our model explains why the excessively high chain production rates enhance background noise. Furthermore, we provide a method, that can be used in further experiment, to simultaneously achieve greater saturated output and lower background noise.
- In Section 2.1, we quantitatively prove that the dynamic range is negatively correlated with protease cleavage ability under normal condition, **which gives a explicit guidance on further experiment**.
- In Section 2.2, we demonstrate that it is feasible to increase system efficiency by changing binding abilities, as long as we design a dedicated degradation pathway of free TC chains without transcription factor. What's more, we obtain a method to improve system efficiency while preventing system saturation output from decreasing. **This will be one of the ideas for our further experiment**.

Contents

1 Differential Equation Model **1**

1.1 Effect of production rates of the two chains on dynamic range 1

1.1.1 Introduction 1

1.1.2 Notations 1

1.1.3 Assumption and establishment 1

1.1.4 Solution 3

1.1.5 Conclusion 4

1.2 Explanation of the too large background noise and ways to reduce it 4

1.2.1 Problem Introduction 4

1.2.2 Explanation 5

2 Stochastic Process Model **6**

2.1 Effect of protease cleavage ability on dynamic range 6

2.1.1 Introduction 6

2.1.2 Notations 6

2.1.3 Assumption and establishment 6

2.1.4 Solution 7

2.1.5 Conclusion and simulation 8

2.2 Effect of binding abilities between chains and the exogenous ligand on system efficiency 10

2.2.1 Introduction 10

2.2.2 Notations 10

2.2.3 Assumption and establishment 10

2.2.4 Solution 11

2.2.5 Conclusion and Simulation 13

References **15**

1 Differential Equation Model

1.1 Effect of production rates of the two chains on dynamic range

1.1.1 Introduction

In the wet experiment, we can control the expression rate of TC chain and PC chain by regulating the content of foreign genes transferred into the cells, and then indirectly control the quantitative ratio of the two chains on the membrane. Although the expression rate and the number of receptor chains are not strictly proportional, there is no doubt that there is a positive correlation between them.

In this section, we will establish a system of differential equations that can reflect the reaction mechanism of the entire system in detail and use it to **quantitatively prove that the local maximum of dynamic range can be reached by changing the production rates of PC and TC chains, under normal condition**. In addition, via the model, **we provide a specific method of changing the two values mentioned above**.

1.1.2 Notations

The primary notations used in this section are listed in **Table 1**.

Table 1: Notations of Section 2.1

Symbol	Definition
A	PC chains
B	TC chains
P	The coalition of A and B
B_l	The TC chain whose transcription factor is lost
P_l	The coalition of A and B_l
R	Transcription factor
υ, μ, σ	The rate constants in mass-action kinetics
α	The production rate of A
β	The production rate of B
ζ	Dilution constant
$y(t)$	System output
DR	Dynamic range of the system

1.1.3 Assumption and establishment

For convenience, we use A to represent for the PC chain, B to represent for the TC chain carrying the transcription factor, P to represent for their coalition, and R to represent for the transcription factor. Meanwhile, we use the footnote l to indicate that the substance has lost the transcription factor. That is, a TC chain of which the transcription factor is lost is represented by B_l, and a coalition of B_l and A is represented by P_l.

Firstly, the reaction mechanism of our system can be expressed by

$$A + B \underset{\mu}{\overset{\upsilon}{\rightleftharpoons}} P \xrightarrow{\sigma} A + B_l + R \tag{1}$$

and

$$A + B_l \underset{\mu}{\overset{\upsilon}{\rightleftharpoons}} P_l \tag{2}$$

where υ is positively correlated with the concentration of the exogenous ligand and it has a small base value υ_0 when the exogenous ligand concentration is zero. Secondly, A and B is produced at two constant rates α and β respectively. Thirdly, all of the above substances are diluted with a dilution constant ζ. So there are five more chemical reaction equations written as

$$\phi \underset{\zeta}{\overset{\alpha}{\rightleftharpoons}} A, \qquad \phi \underset{\zeta}{\overset{\beta}{\rightleftharpoons}} B, \qquad B_l \xrightarrow{\zeta} \phi, \qquad P \xrightarrow{\zeta} \phi, \qquad P_l \xrightarrow{\zeta} \phi. \tag{3}$$

We assume all of these reactions follow mass-action kinetics, with rate constants written near the arrow for each reaction. Although our reactions occur on the cell membrane as a three-dimensional surface, rather than in solution, we can still calculate the density of each reactant on the cell membrane as the concentration of the reactants in the law of mass action. The result should be similar. Hence the system of differential equations that the reaction process satisfies is written as

$$\begin{cases} \dfrac{\mathrm{d}A}{\mathrm{d}t} = \alpha + \mu P + \sigma P + \mu P_l - \zeta A - \upsilon AB - \upsilon AB_l \\ \dfrac{\mathrm{d}B}{\mathrm{d}t} = \beta - \zeta B - \upsilon AB + \mu P \\ \dfrac{\mathrm{d}B_l}{\mathrm{d}t} = \sigma P - \zeta B_l - \upsilon AB_l + \mu P_l \\ \dfrac{\mathrm{d}P}{\mathrm{d}t} = \upsilon AB - \mu P - \sigma P - \zeta P \\ \dfrac{\mathrm{d}P_l}{\mathrm{d}t} = \upsilon AB_l - \mu P_l - \zeta P_l \end{cases} \tag{4}$$

where the capital letters indicate the density of the corresponding reactants on the membrane. Using the Taylor expansion of binary function

$$f(x,y) \approx f(x_0,y_0) + (x - x_0)\frac{\partial f(x_0,y_0)}{\partial x} + (y - y_0)\frac{\partial f(x_0,y_0)}{\partial y} \tag{5}$$

we derive

$$AB \approx b_0 A + a_0 B - a_0 b_0 \tag{6}$$

and

$$AB_l \approx b_{l0} A + a_0 B_l - a_0 b_{l0} \tag{7}$$

where the constants a_0, b_0, b_{l0} are approximations of the densities of A, B, and C, respectively.

Substituting Eq. (6) and (7) into Eq. (4), the linear approximation system of Equation Set (4) is written as

$$\begin{cases} \dfrac{\mathrm{d}A}{\mathrm{d}t} = \alpha + \mu P + \sigma P + \mu P_l - \zeta A - \upsilon(b_0 A + a_0 B - a_0 b_0) - \upsilon(b_{l0} A + a_0 B_l - a_0 b_{l0}) \\ \dfrac{\mathrm{d}B}{\mathrm{d}t} = \beta - \zeta B - \upsilon(b_0 A + a_0 B - a_0 b_0) + \mu P \\ \dfrac{\mathrm{d}B_l}{\mathrm{d}t} = \sigma P - \zeta B_l - \upsilon(b_{l0} A + a_0 B_l - a_0 b_{l0}) + \mu P_l \\ \dfrac{\mathrm{d}P}{\mathrm{d}t} = \upsilon(b_0 A + a_0 B - a_0 b_0) - \mu P - \sigma P - \zeta P \\ \dfrac{\mathrm{d}P_l}{\mathrm{d}t} = \upsilon(b_{l0} A + a_0 B_l - a_0 b_{l0}) - \mu P_l - \zeta P_l \end{cases} \tag{8}$$

Although Lyapunov Stability Theory states that the stability of the primary system can be represented by the stability of its linear approximation system only if some conditions are met[6][7], there is no doubt that the whole system will eventually reach a steady state[4] because the production rates of A and B can be balanced with the dilution rate ζ, as shown in Figure 1.

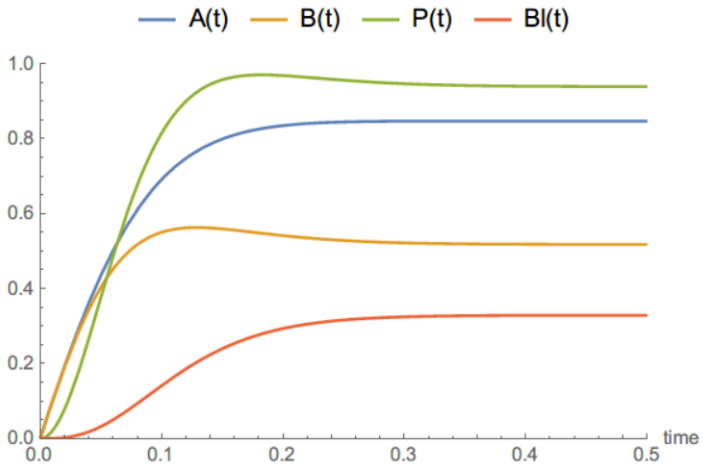

Figure 1: A simulation of Eq. (4) with appropriate parameters

1.1.4 Solution

In order to study the steady state of the system, let the equations in Equation Set (6) be equal to 0[6] so we can get a nonhomogeneous linear equation set, whose matrix form is written as

$$\boldsymbol{L}\boldsymbol{x} = \boldsymbol{F} \tag{9}$$

where

$$\boldsymbol{L} = \begin{pmatrix} -\zeta - \upsilon\left(b_0 + b_{l0}\right) & -\upsilon a_0 & -\upsilon a_0 & \mu + \sigma & \mu \\ -\upsilon b_0 & -\zeta - \upsilon a_0 & 0 & \mu & 0 \\ -\upsilon b_{l0} & 0 & -\zeta - \upsilon a_0 & \sigma & \mu \\ \upsilon b_0 & \upsilon a_0 & 0 & -\mu - \sigma - \zeta & 0 \\ \upsilon b_{l0} & 0 & \upsilon a_0 & 0 & -\mu - \zeta \end{pmatrix} \tag{10}$$

and

$$\boldsymbol{x} = \begin{pmatrix} A \\ B \\ B_l \\ P \\ P_l \end{pmatrix} \qquad \boldsymbol{F} = \begin{pmatrix} -\alpha - \upsilon a_0\left(b_0 + b_{l0}\right) \\ -\beta - \upsilon a_0 b_0 \\ -\upsilon a_0 b_{l0} \\ \upsilon a_0 b_0 \\ \upsilon a_0 b_{l0} \end{pmatrix}. \tag{11}$$

This system of equations has a unique solution when the ranks of its coefficient matrix and augmented matrix are both 5. The analytical solution of P is written as

$$P = \frac{h_{12}\upsilon^2 + h_{11}\upsilon}{h_{22}\upsilon^2 + h_{21}\upsilon + h_{20}} \tag{12}$$

where

$$\begin{cases} h_{12} = a_0\left(a_0\beta + \alpha b_0 + \beta b_{l0} - a_0 b_0 \zeta\right) \\ h_{11} = \left(a_0\beta + \alpha b_0 - a_0 b_0 \zeta\right)\left(\mu + \zeta\right) \\ h_{22} = a_0\left(a_0 + b_0 + b_{l0}\right)\left(\sigma + \zeta\right) \\ h_{21} = \left(\left(a_0 + b_{l0}\right)\zeta + a_0\left(\zeta + \mu\right)\right)\sigma + \left(2a_0 + b_0 + b_{l0}\right)\left(\zeta + \mu\right)\zeta \\ h_{20} = \zeta\left(\sigma + \zeta + \mu\right)\left(\zeta + \mu\right) \end{cases}. \tag{13}$$

We consider the production rate of the transcription factor R as the system output $y(t)$, that is,

$$y\left(t\right) = \frac{\mathrm{d}R}{\mathrm{d}t} = \sigma P. \tag{14}$$

When the system output is saturated (i.e. the concentration of exogenous ligand is large enough so that υ is a large positive real number), the value of P depends mainly on the square of υ on the numerator and denominator, which means

$$y\left(t\right)|_{\upsilon\ \max} \doteq \sigma\frac{h_{12}}{h_{22}}. \tag{15}$$

And when the concentration of exogenous ligand is nearly zero, note that υ has a small base value υ_0 in this condition. So the system output is written as

$$y\left(t\right)|_{\upsilon=\upsilon_0} = \sigma\frac{h_{12}\upsilon_0^2 + h_{11}\upsilon_0}{h_{22}\upsilon_0^2 + h_{21}\upsilon_0 + h_{20}}. \tag{16}$$

If we only consider the effect of α and β on the dynamic range (i.e. dynamic range is the function of α and β), and note that h_{22}, h_{21} and h_{20} have nothing to do with α and β. Hence, the dynamic range can be written as

$$\begin{aligned} DR\left(\alpha,\beta\right) &= \frac{y\left(t\right)|_{\upsilon\ \max}}{y\left(t\right)|_{\upsilon=\upsilon_0}} \doteq \frac{h_{12}\left(h_{22}\upsilon_0^2 + h_{21}\upsilon_0 + h_{20}\right)}{h_{22}\left(h_{12}\upsilon_0^2 + h_{11}\upsilon_0\right)} \\ &== \frac{k_1 h_{12}}{h_{12}\upsilon_0^2 + h_{11}\upsilon_0} = \frac{k_1 k_2}{\upsilon_0 + \frac{a_0\beta+\alpha b_0 - a_0 b_0\zeta}{a_0\beta+\alpha b_0+\beta b_{l0}-a_0 b_0\zeta}\left(\mu+\zeta\right)} \\ &= k_1 k_2\left(\upsilon_0 + \left(1+\frac{b_{l0}}{a_0 + b_0\frac{\alpha-a_0\zeta}{\beta}}\right)^{-1}\left(\mu+\zeta\right)\right)^{-1} \end{aligned} \tag{17}$$

where k_1 and k_2 are scale factors.

1.1.5 Conclusion

Note that a_0 is approximation of density of free A and α is A's production rate. Since the production rate of A should be equal to the total dilution rate of A whatever its state is when the system is steady, we have $\alpha - a_0\zeta > 0$. Thus, it is obvious from Eq. (17) that, **positive real number α is always negatively correlated with dynamic range and β is always positively correlated with it**.

In order to be more realistic, we further consider the effects of random errors, downstream leakage, measurement errors and other factors on DR. Let all these factors, which are not significantly related to the experimental parameters, can be attributed to a total interference term D. We can regard D as a random variable and simply consider its influence by adding the same constant, the mean value of D, to both the numerator and the denominator of the definition of DR. Figure 2 is the graph of DR function considering the effects of D, where the horizontal and vertical coordinates indicate the production rates of A and B, respectively (the transfection concentrations of the two chains can be used to replace them in wet experiment), and the color indicates the size of DR.

The figure shows the existence of the maximum value of dynamic range, which is verified by the wet experiment. For instance, we can get the curve of DR when the ratio of the horizontal and vertical coordinates is constant, which is equivalent to the vertical cutting through the origin of Figure 2, as shown in Figure 3.

In conclusion, our model reveals the intrinsic link between production rates and dynamic range. Thanks to it, we now have a clearer way to achieve a higher dynamic range.

1.2 Explanation of the too large background noise and ways to reduce it

1.2.1 Problem Introduction

Our wet experiment shows a relatively large output of our system in spite of low concentration of the exogenous ligand, which is not quite as expected. This phenomenon reduces the dynamic range of the system, resulting in a decrease in the sensitivity of the system to the concentration of exogenous ligands.

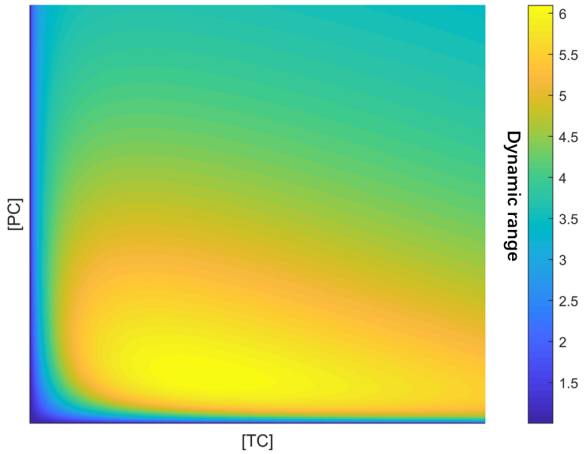

Figure 2: Effect of production rates of the two chains on dynamic range

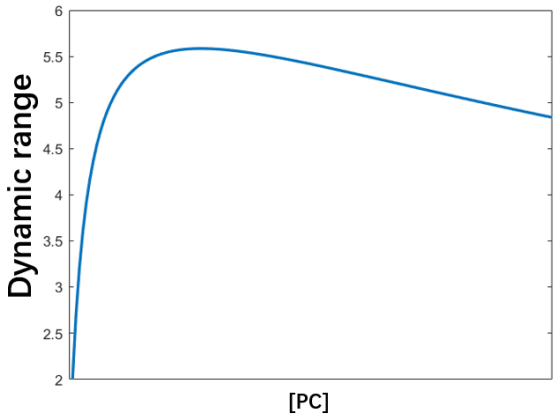

Figure 3: A way to maximize dynamic range

In this section, we will use the solution to the model in the previous section to **explain why the excessively high chain generation rate enhances the background noise and provide a method to simultaneously achieving greater dynamic range and smaller background noise**.

1.2.2 Explanation

Considering the statement related to Equation (16) in the previous section, the background noise, which represents the positive system output without exogenous ligand, can be represented by $y(t)\mid_{v=v_0}$, where $v_0 \neq 0$ because of random collision between chains. Reiterate the content of Equation (16)

$$y(t)\mid_{v=v_0} = \sigma \frac{h_{12}v_0^2 + h_{11}v_0}{h_{22}v_0^2 + h_{21}v_0 + h_{20}} \tag{18}$$

and

$$\begin{cases} h_{12} = a_0\left(a_0\beta + \alpha b_0 + \beta b_{l0} - a_0 b_0 \zeta\right) \\ h_{11} = \left(a_0\beta + \alpha b_0 - a_0 b_0 \zeta\right)(\mu + \zeta) \\ h_{22} = a_0\left(a_0 + b_0 + b_{l0}\right)(\sigma + \zeta) \\ h_{21} = \left(\left(a_0 + b_{l0}\right)\zeta + a_0(\zeta + \mu)\right)\sigma + \left(2a_0 + b_0 + b_{l0}\right)(\zeta + \mu)\zeta \\ h_{20} = \zeta(\sigma + \zeta + \mu)(\zeta + \mu) \end{cases} . \tag{19}$$

Intuitively, the variables α and β appear only on the numerator, with positive constants multiply them. Therefore, these two production rates of chains are both positively correlated with $y(t)\mid_{v=v_0}$, the background noise. So **we can simultaneously achieve greater dynamic range and**

smaller background noise by reducing α and β at the same time, as long as the factor $\frac{\alpha-a_0\zeta}{\beta}$ in Eq. (17) gets local minimum.

2 Stochastic Process Model

2.1 Effect of protease cleavage ability on dynamic range

2.1.1 Introduction

Due to random collision or exogenous ligand, a PC chain and a TC chain will be combined. There are two situations after that. In the first case, after a period of time T the protease cleaves the transcription factor, and after a while, the two chains are separated. In the second case, the protease has no time to cleave the transcription factor before the two chains are separated. This whole process happens many times on the cell membrane, and T is different each time. Obviously T is a random variable, and the more likely the cleavage occurs, the smaller its mathematical expectation $E(T)$ is.

Whether the cleavage is prone to occur is mainly determined by the catalytic efficiency of the protease, and is also affected by other factors such as the length of the intracellular segment of the PC chain and the TC chain. For the sake of convenience, we say that protease cleavage ability is stronger if the cutting is more likely to occur.

In this section, we will prove that **the dynamic range of the system is negatively correlated with protease cleavage ability**.

2.1.2 Notations

The primary notations used in this section are listed in Table 2.

Table 2: Notations of Section 1.1

Symbol	Definition
$T_{i,j}$	The time a TC chain spends transitioning directly from State i to j
T_i	The time that a random variable stays in State i for a single time
$P_{i,j}$	The probability of a random variable transitioning from State i to j
M_i	The total time a random variable spend transitioning from State i to j
p	The exponential distribution parameter related to the concentration of PC chain
d	The exponential distribution parameter related to the binding ability of two kinds of chains
c	The exponential distribution parameter related to protease cleavage ability

2.1.3 Assumption and establishment

To research dynamic range we need to consider two typical cases. In the first case the concentration of exogenous ligand is nearly 0 and the combination events of TC chains and PC chains are

almost caused by random collision; in the second case the concentration of exogenous ligand is large enough so that the system output is saturated and the combination events of the two chains are almost caused by exogenous ligands.

Let us focus on how a TC chain behaves in the two situations. At first, it moves freely on the membrane (State 1, the free sate). After a period of time it is combined with the PC chain (State 2, the bound state) due to random collision (in the first situation) or an exogenous ligand (in the second situation), and then it is either cleaved by protease (State 3, the inactive state) or separated from the PC chain without being cut(returns to State 1). Once it reaches State 3, it loses the transcription factor, so whether it is combined with the PC chain, it will no longer increase the output of the system. In other words, it cannot be moved from state 3 to state 1 or 2.

We use T_{ij} to indicate the time taken for the TC chain to transition from state i to state j ($i, j \in \{1, 2, 3\}$ and naturally assume T_{ij} is memoryless[5], that is, T_{ij} obeys the exponential distribution. Let the stochastic process $\{D(t), t \geqslant 0\}$ indicate the state of the TC chain and the state space is $\{1, 2, 3\}$. Since T_{ij} obeys the exponential distribution, this is a continuous-time Markov chain. The state transition diagram is shown in Figure 4 below.

Figure 4: State transition diagram of $D(t)$

Let $T_{12} \sim \text{Exp}(p)$, $T_{21} \sim \text{Exp}(d)$ and $T_{23} \sim \text{Exp}(c)$. A TC chain cannot transition from State 3 to 2, which means $E(T_{32}) \to \infty$, hence we can say $T_{21} \sim \text{Exp}(0)$ although this is not very rigorous. Since the mean of a exponentially distributed variable is the reciprocal of its exponential distribution parameter, the longer the mean value is, the smaller the corresponding probability parameter is. Therefore, it is easy to understand that in the case of stable external environment (temperature, pH, etc.), mainly, p is positively correlated with the concentration of PC chain, d is negatively correlated with the binding ability of the two chains, and c is positively correlated with protease cleavage ability.

2.1.4 Solution

Since $D(t)$ in a certain state can only be transferred to an adjacent state at most, this is a birth-death process. As for an ordinary birth-death process $\{X(t), t \geqslant 0\}$, the state transition diagram is shown in Figure 5.

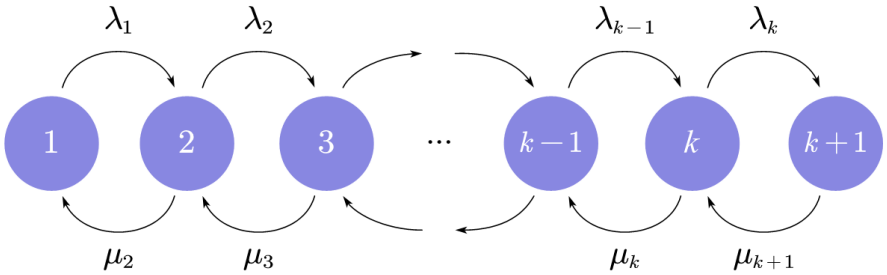

Figure 5: State transition diagram of an ordinary birth-death process

If $X(t)$ is in State i, it can only transition to State $i-1$ or State $i+1$ next time. We still use $T_{i,j}$ to indicate the time taken for $X(t)$ in State i to transition to state j ($i \in \{1, 2, 3, ...\}$, $j = i+1$ or $j = i-1$) next time. Let random variable T_i to indicate the time that $X(t)$ stays in State i for a single time. Obviously $T_i = \min(T_{i,i+1}, T_{i,i-1})$, so $T_i \sim \text{Exp}(\lambda_i + \mu_i)$. Hence the state transition probabilities can be written as

$$P_{i,i+1} = \frac{\lambda_i}{\lambda_i + \mu_i}, \qquad P_{i,i-1} = \frac{\mu_i}{\lambda_i + \mu_i} \tag{20}$$

where

- we specify $\mu_1 = 0$,
- $P_{i,i+1}$ is the probability of $X(t)$ transitioning from State i to State $i+1$,
- $P_{i,i-1}$ is the probability of $X(t)$ transitioning from State i to State $i-1$.

According to the foregoing, $X(t)$ in State i may directly transition to State $i+1$ (the probability is $P_{i,i+1}$). It is also possible to transition to State $i-1$ first (the probability is $P_{i,i-1}$), and then it may transition to State i again (the probability is $P_{i-1,i}$) or transition to State $i-2$ (the probability is $P_{i-1,i-2}$)... To examine the mean value of M_i, which is the total time at which $X(t)$ transitions from State i to State $i+1$, the recursive formula is written as

$$E\left(M_i\right) = P_{i,i+1}E\left(T_{i,i+1}\right) + P_{i,i-1}\left(E\left(M_{i-1}\right) + E\left(M_i\right)\right) \tag{21}$$

where $E\left(T_{i,i+1}\right) = \frac{1}{\lambda_i}$ and $E(M_1) = \frac{1}{\lambda_1}$.

Going back to the process shown in Figure 4, as a birth-death process, considering

$$\lambda_1 = p, \qquad \lambda_2 = c, \qquad \mu_2 = d, \qquad \mu_3 = 0, \tag{22}$$

we have

$$E\left(M_1\right) = \frac{1}{p} \tag{23}$$

and

$$E\left(M_2\right) = \frac{1}{c+d} + \frac{d}{c+d}(E(M_1) + E(M_2)). \tag{24}$$

Hence, the mathematical expectation of M_e, which is the average time at which $X(t)$ transitions from State 1 to State 3, can be derived as

$$E\left(M_e\right) = E\left(M_1\right) + E\left(M_2\right) = \frac{c+p+d}{pc}. \tag{25}$$

In the two cases mentioned above, the form of the above formula is constant, but the values of the parameters may be different. Let the values of these three parameters to be p_1, d_1, c_1 in the first case and p_2, d_2, c_2 in the second case.

The binding of the two chains is mainly caused by exogenous ligands in the second case, which makes their binding ability stronger and the system more active. However, the expression rates of new TC and PC chains, which are mainly controlled by the activity of the corresponding genes, do not change much, so there are fewer unbound PC chains in this case.

Note that d is negatively correlated with the binding ability of the two chains and p is positively correlated with the concentration of PC chain, hence $d_1 > d_2 > 0$ and $p_1 > p_2 > 0$. In addition, there is no difference in proteases cleavage ability in both cases, i.e. $c_1 = c_2 = c$.

Since $E(M_e)$ represents the mean of the total time that a TC chain has undergone from being placed on the cell membrane to being cleaved by the protease, intuitively it reflects the activity of the system (i.e., the greater the output of the system is, the bigger the reciprocal of $E(M_e)$ is). Let the M_es in both cases be M_{e1} and M_{e2} respectively. Then we can derive that dynamic range is positively correlated with

$$\frac{E\left(M_{e1}\right)}{E\left(M_{e2}\right)} = \frac{c+p_1+d_1}{p_1 c} \Big/ \frac{c+p_2+d_2}{p_2 c} = \frac{p_2}{p_1} \cdot \frac{c+p_1+d_1}{c+p_2+d_2}. \tag{26}$$

2.1.5 Conclusion and simulation

Since

$$\frac{\mathrm{d}}{\mathrm{d}c}\left(\frac{E\left(M_{e1}\right)}{E\left(M_{e2}\right)}\right) = \frac{p_2}{p_1} \cdot \frac{(p_2-p_1)+(d_2-d_1)}{\left(c+p_2+d_2\right)^2} < 0, \tag{27}$$

it follows that the dynamic range of the system is negatively correlated with protease cleavage ability under normal condition.

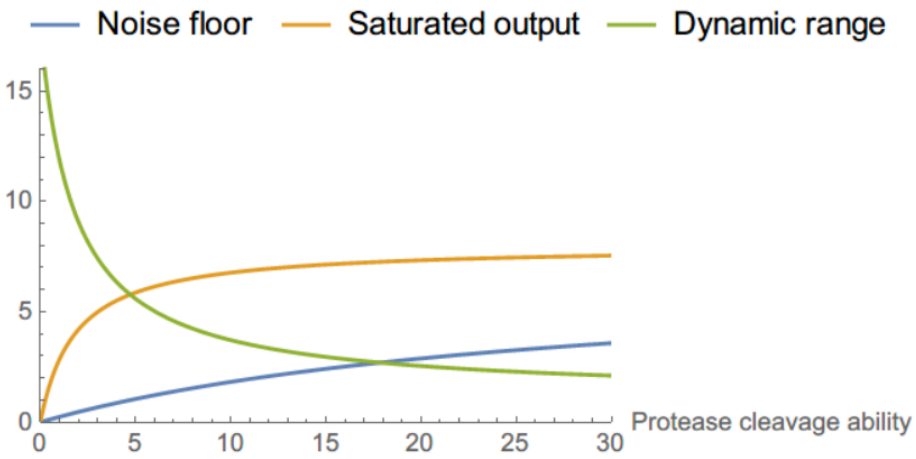

Figure 6: Effect of protease cleavage ability on dynamic range

In order to show the relationship between protease cleavage ability and dynamic range better, we assume $DR = k\frac{E(M_{e1})}{E(M_{e2})}$ without loss of generality, where k is a scale factor. Then the graph of $1/E(M_{e1})$ (representing background noise), $k/E(M_{e2})$ (representing saturated output), and DR are shown in Figure 6, where $p_1 = 6.9, p_2 = 1.6, d_1 = 21, d_2 = 0.2$. Note that in order for the three curves to be compared to each other on the ordinate, we also multiply $1/E(M_{e2})$ by a scale factor $k = 5$.

It is clear from the figure that **the negative correlation between dynamic range and protease cleavage ability is caused by different growth rates of background noise and saturated output when protease cleavage ability rises**.

2.2 Effect of binding abilities between chains and the exogenous ligand on system efficiency

2.2.1 Introduction

After losing its transcription factor, a TC chain may still be combined with a PC chain, which makes the latter occupied but does not contribute to system output. We use the indicator called system efficiency (SE) to measure the extent, to which TC chains out of transcription factor occupy system resources. By promoting SE, we can achieve the same level of system output with fewer resources (i.e., lower chain generation rates).

It should be noted that since the TC chain is no longer likely to contribute to system output after losing its transcription factor, using the TC chain to measure SE is a bit of a hassle. However, the structure of the PC chain is always the same, which means that we can easily measure SE with parameters related to PC chain.

In this section, we will demonstrate that **it is feasible to increase system efficiency by changing the binding abilities, as long as we design a dedicated degradation pathway for free TC chains without transcription factor,** .

2.2.2 Notations

The primary notations used in this section are listed in Table 3.

Table 3: Notations of Section 1.2

Symbol	Definition
$A(t)$	The serial numbers of those states a PC chain can be in
I	The state space of $A(t)$
$p_{i,j}(t)$	The transition probability of $A(t)$ transitioning from State i to j after a period of time t
Q	The transfer rate matrix of $A(t)$
q_{ij}	The elements of the j-th column in the i-th row of Q
λ_{ij}	The parameter of the exponential distribution obeyed by the time that A transfers from State i to State j by one step
A_n	The embedded chain of $A(t)$
K	The one-step transition probability matrix of A_n
p_j	The probability that $A(t)$ is in State j when $t \to \infty$
$\boldsymbol{\pi}$	The stationary distribution of discrete-time Markov chain $\{A_n\}$

2.2.3 Assumption and establishment

Similar to the modeling approach adopted in Section 1.1, now let us research the state transition process of PC chain when system output is large enough (i.e., The concentration of the exogenous ligand is large enough so that our system is running effectively). Under this condition, the binding of the two chains is mostly caused by the exogenous ligand[2]. For convenience, we use A for the PC chain, B for the TC chain carrying the transcription factor and L for the exogenous ligand. Meanwhile, we use the footnote l to indicate that the substance has lost the transcription factor. That is, a TC chain of which the transcription factor is lost is represented by B_l.

Let the stochastic process $\{A(t), t \geqslant 0\}$ indicate the serial number of those states a PC chain can be in. We still assume that the time required for $A(t)$ to move directly from State i to State j obeys the exponential distribution with parameter λ_{ij}. Then the state space of $A(t)$ is $I = \{1, 2, 3, 4\}$ and the meaning of each state is as follows:

- State 1: (free state) A moves freely on the cell membrane,
- State 2: (semi-bonded state) A binds to an exogenous ligand,
- State 3: (effective binding state) A binds to B so that they may produce transcription factors,
- State 4: (invalid binding state) A binds to B_l, which means A is occupied but cannot help produce transcription factors.

In addition, the state transiton diagram is shown in Figure 7. It follows that $\{A(t)\}$ is a time-

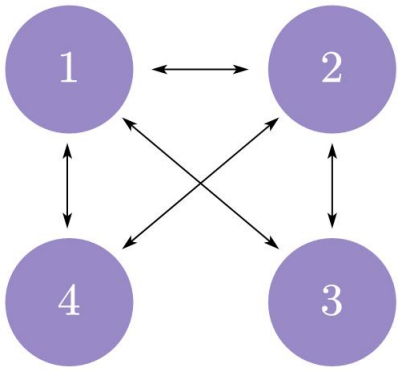

Figure 7: State transition diagram of $A(t)$

homogeneous continuous-time Markov Chain. Let A be free at $t = 0$, that is, the initial distribution of $\{A(t)\}$ is

$$P(A(0) = i) = \begin{cases} 1, i = 1 \\ 0, else \end{cases} . \tag{28}$$

And its transition probability matrix is

$$P(t) = (p_{ij}(t))_{i,j \in I} \tag{29}$$

where

$$p_{ij}(t) = P(A(t+s) = j \mid A(s) = i) \tag{30}$$

which has nothing to do with start time s.

2.2.4 Solution

The transfer rate matrix (Q-matrix) of $A(t)$ can be written as

$$Q = (q_{ij})_{i,j \in I} \tag{31}$$

where

$$q_{ij} = p_{ij}^{'}(0) . \tag{32}$$

Hence, for those States i, j where there is a positive possibility of one-step transfer between them,

$$q_{ij}(t) = \lambda_{ij}; \tag{33}$$

for those States i, j where it is impossible for $A(t)$ to transfer between them directly (such as $i = 3$ and $j = 4$),

$$q_{ij}(t) = 0; \tag{34}$$

specially,

$$q_{ii}(t) = -\sum_{j \in I, j \neq i} p_{ij}(t). \tag{35}$$

For convenience, if A transitions from a relatively free state to a state that the substance combined with A has become more, we called this an ascended order transfer. Conversely, we call it a reduced-order shift.

Note that there is only ONE PC chain studied in this process. Therefore, the exponential distribution parameter corresponding to an ascended-order transfer (such as λ_{12}, λ_{14} and λ_{23}) is related to the concentration of the newly combined substance. For example, λ_{24} is positively correlated with the concentration of B_l and λ_{13} is positively correlated with the concentration of the combination of L and B on the membrane. Besides, the exponential distribution parameter corresponding to a reduced-order transfer (such as λ_{21}, λ_{32} and λ_{41}) is related to the binding ability between the corresponding substances. For example, λ_{21} and λ_{41} are both negatively correlated with the binding ability between A and L, so we assume they are equal. The factors that affect the value of each parameter λ are shown in the Table 4.

Table 4: The factors that affect the values of parameter λs

Parameter λ	Factor
λ_{12}	The concentration of L
λ_{13}	The concentration of the combination of L and B
λ_{14}	The concentration of the combination of L and B_l
λ_{23}	The concentration of B
λ_{24}	The concentration of B_l
$\lambda_{41}=\lambda_{31}=\lambda_{21}$	The binding ability between A and L
$\lambda_{42}=\lambda_{32}$	The binding ability between L and B(or B_l)

Therefore, the Q-matrix is written as

$$Q=\begin{pmatrix} -\lambda_{12}-\lambda_{13}-\lambda_{14} & \lambda_{12} & \lambda_{13} & \lambda_{14} \\ \lambda_1 & -\lambda_1-\lambda_{23}-\lambda_{24} & \lambda_{23} & \lambda_{24} \\ \lambda_1 & \lambda_2 & -\lambda_1-\lambda_2 & 0 \\ \lambda_1 & \lambda_2 & 0 & -\lambda_1-\lambda_2 \end{pmatrix} \tag{36}$$

where $\lambda_1=\lambda_{41}=\lambda_{31}=\lambda_{21}$ and $\lambda_2=\lambda_{42}=\lambda_{32}$.

Using the theory of Markov chain[5], the embedded chain $\{A_n\}$ of stochastic process $\{A(t)\}$ has one-step transition probability matrix, written as $K=(k_{ij})_{i,j\in I}$. Let $q_i=|q_{ii}|$ and $\delta_{ij}=\begin{cases}1,\ i=j\\0,\ i\neq j\end{cases}$. Since

$$k_{ij}=\begin{cases} q_{ij}/q_i, & \text{when } q_i>0,\ j\neq i \\ 0, & \text{when } q_i>0,\ j=i \\ \delta_{ij}, & \text{when } q_i=0 \end{cases}, \tag{37}$$

it follows that

$$K=\begin{pmatrix} 0 & \frac{\lambda_{12}}{\lambda_{12}+\lambda_{13}+\lambda_{14}} & \frac{\lambda_{13}}{\lambda_{12}+\lambda_{13}+\lambda_{14}} & \frac{\lambda_{14}}{\lambda_{12}+\lambda_{13}+\lambda_{14}} \\ \frac{\lambda_1}{\lambda_1+\lambda_{23}+\lambda_{24}} & 0 & \frac{\lambda_{23}}{\lambda_1+\lambda_{23}+\lambda_{24}} & \frac{\lambda_{24}}{\lambda_1+\lambda_{23}+\lambda_{24}} \\ \frac{\lambda_1}{\lambda_1+\lambda_2} & \frac{\lambda_2}{\lambda_1+\lambda_2} & 0 & 0 \\ \frac{\lambda_1}{\lambda_1+\lambda_2} & \frac{\lambda_2}{\lambda_1+\lambda_2} & 0 & 0 \end{pmatrix}. \tag{38}$$

Since the number of the states of $\{A_n\}$ is finite, there is its stationary distribution $\boldsymbol{\pi}=[\pi_1,\pi_2,\pi_3,\pi_4]$ where

$$\sum_{j\in I}\pi_j=1 \qquad \text{and} \qquad \boldsymbol{\pi}=\boldsymbol{\pi}K. \tag{39}$$

Through Eq. (38) and (39) we get $\boldsymbol{\pi}$. Then using[5]

$$p_j \equiv \lim_{t\to+\infty} p_{ij}(t) = \frac{\pi_j/q_j}{\sum_{i\in I}\pi_i/q_i}\ , j \in I \tag{40}$$

we obtain p_j, which is the the probability that $\{A(t)\}$ is in State j when $t \to \infty$. Specially, p_3 and p_4 are the stable probabilities of $A(t)$ in State 3 and State 4, respectively, when the elapsed time is long enough. Note that $A(t)$'s being in State 3 means that A binds to B with an exogenous ligand, so that they contribute to system output. And $A(t)$'s being in State 4 means that A binds to B_l with an exogenous ligand, which means A is occupied but cannot help produce transcription factors. Hence,

$$S = \frac{p_3}{p_4} = \frac{\pi_3 q_4}{\pi_4 q_3} \tag{41}$$

can be used as a representation of system efficiency. Through the calculation method mentioned above, we obtain

$$S = \frac{\lambda_{13}\lambda_1^2 + (\lambda_{12}\lambda_{23} + \lambda_{13}\lambda_{23} + \lambda_{13}\lambda_{24} + \lambda_{13}\lambda_2)\lambda_1 + (\lambda_{12}+\lambda_{13}+\lambda_{14})\lambda_{23}\lambda_2}{\lambda_{14}\lambda_1^2 + (\lambda_{12}\lambda_{24} + \lambda_{14}\lambda_{23} + \lambda_{14}\lambda_{24} + \lambda_{14}\lambda_2)\lambda_1 + (\lambda_{12}+\lambda_{13}+\lambda_{14})\lambda_{24}\lambda_2}. \tag{42}$$

2.2.5 Conclusion and Simulation

As mentioned earlier, the key reason for the decrease in system efficiency is that B_l, which has lost the transcription factor, occupies part of A's resources. Therefore, a natural idea is whether we can improve system efficiency by reducing the combined ability of the TC chain (B and B_l) and the exogenous ligand L (i.e., increasing λ_2).

It can be seen that S is a function of multiple parameters, meaning that even if the binding abilities between A, B (or B_l), and L (i.e., λ_1 and λ_2) is constant, S may change at different system output levels (usually caused by different exogenous ligand concentrations). Hence, in order to make it meaningful to compare the value of S after changing the binding abilities with the value before the change, we stipulate that all parameters except λ_1 and λ_2 are constant. That is, S is a function of λ_1 and λ_2.

According to Eq. 42, we have

$$\frac{\partial S}{\partial \lambda_2} = \frac{\lambda_1(\lambda_{14}\lambda_{23} - \lambda_{13}\lambda_{24})(\lambda_1(\lambda_{13}+\lambda_{14}) + (\lambda_{12}+\lambda_{13}+\lambda_{14})(\lambda_{23}+\lambda_{24}))}{(\lambda_{14}\lambda_1^2 + (\lambda_{12}\lambda_{24} + \lambda_{14}\lambda_{23} + \lambda_{14}\lambda_{24} + \lambda_{14}\lambda_2)\lambda_1 + (\lambda_{12}+\lambda_{13}+\lambda_{14})\lambda_{24}\lambda_2)^2} \tag{43}$$

and

$$\frac{\partial S}{\partial \lambda_1} = -\frac{(\lambda_{14}\lambda_{23} - \lambda_{13}\lambda_{24})\left(\lambda_1^2\lambda_{12} + (2\lambda_1 + \lambda_2 + \lambda_{23} + \lambda_{24})(\lambda_{12}+\lambda_{13}+\lambda_{14})\lambda_2\right)}{(\lambda_{14}\lambda_1^2 + (\lambda_{12}\lambda_{24} + \lambda_{14}\lambda_{23} + \lambda_{14}\lambda_{24} + \lambda_{14}\lambda_2)\lambda_1 + (\lambda_{12}+\lambda_{13}+\lambda_{14})\lambda_{24}\lambda_2)^2}. \tag{44}$$

It follows that whether the partial derivatives is greater than 0 depends on $\lambda_{14}\lambda_{23} - \lambda_{13}\lambda_{24}$. Under the condition that our basic system does not contain a specific degradation pathway to free B_l, the formula $\lambda_{14}\lambda_{23} - \lambda_{13}\lambda_{24}$ is symmetric so that the value of it is (nearly) zero, which means that we cannot increase system efficiency by lowering binding abilities. However, if we design a dedicated degradation pathway for free B_l, λ_{24} will decrease, which allows $\lambda_{14}\lambda_{23} - \lambda_{13}\lambda_{24} > 0$. In this situation, we can easily increase system efficiency by lowering the binding ability of the TC chain and the exogenous ligand or increasing the binding ability of the PC chain and the exogenous ligand.

Moreover, the latter can help promote the combination of TC and PC chains, which leads to a higher output. It follows that **we obtain a method to improve the system efficiency while preventing system saturation output from decreasing: improving the binding ability of the PC chain and the exogenous ligand while reducing the binding ability of the TC chain and the exogenous ligand at the same time. This will be one of the ideas for our further experiment**.

Figure 8 shows the relationships between system efficiency and λ_1, λ_2, respectively, where $\lambda_{12} = 20, \lambda_{13} = 6, \lambda_{14} = 3, \lambda_{23} = 10$ and $\lambda_{24} = 1$. For comparability, one of λ_1 and λ_2 is equal to 5 when the other is the independent variable.

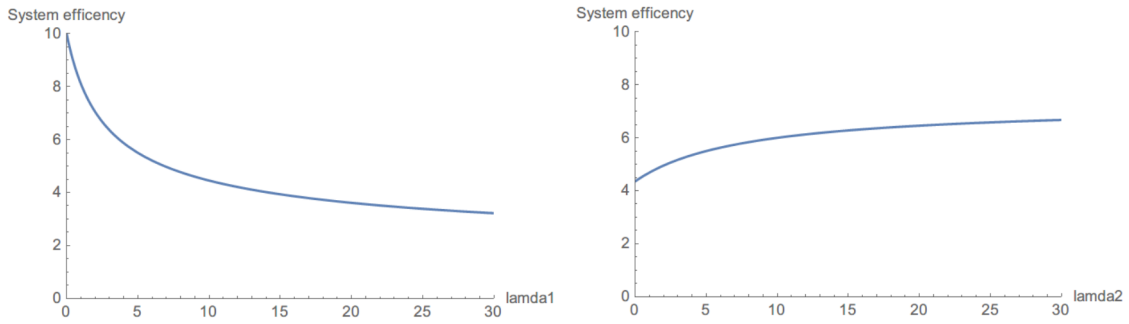

Figure 8: Relationship between system efficiency and λ_1, λ_2

Note that λ_1 and λ_2 is negatively correlated with the binding ability between A and L and the binding ability between B and L, respectively. So what the figure shows is consistent with the modeling result.

References

[1] Schwarz, K. A., et al. *Rewiring human cellular input-output using modular extracellular sensors*. Nature Chemical Biology 13.2(2016):202.

[2] Daringer, N. M., et al. *Modular extracellular sensor architecture for engineering mammalian cell-based devices*. Acs Synthetic Biology 3.12(2014):892-902.

[3] Schwarz, Kelly A., and J. N. Leonard. *Engineering cell-based therapies to interface robustly with host physiology*. Advanced Drug Delivery Reviews 105.Pt A(2016):55-65.

[4] Oppenheim, Alan V., et al. *Signals and Systems, 2nd Edition*. Englewood: Prentice Hall, 1996.

[5] He, Shuyuan, et al. *Stochastic Process*. Beijing: Peking University Press, 2008.

[6] Cai, Suilin, et al. *Ordinary Differential Equation, 3rd Edition*. Hangzhou: Zhejiang University Press, 2012.

[7] Liu, Wending, et al. *The Principle of Automatic Control, 3rd Edition*. Beijing: Publishing House of Electronics Industry, 2013.