Exploratory model for SynNotch CAR-Tregs system

CPU CHINA iGEM 2017

October 31, 2017

1 Introduction

The purpose of designing synthetic bio-systems is to obtain the desired product or a new biological function. Therefore, the biological module, as a functional unit, is inevitably implanted in a heterologous or homologous organism and is bound to be affected by many signals in the organism. Synthetic biology emphasizes the "plug and play" of a biological module. To ensure that the design of a good biological module can be easily integrated into other functional modules, the biological module must have a very good stability.

Given that it is difficult to isolate and culture the Treg cells from human body, that animal experiments and B cell co-culture experiments are difficult to carry out, and that the Jurkat cell line used in the experiment is very special, our modeling of solving the above problems is very limited. Therefore, the modeling must be carried out on the basis of almost no experimental data. In synthetic biology, the general approach to modeling of synthetic bio-systems is engineering. And there are several main indicators to analyze the performance of synthetic bio-systems:

1. Stability
2. Robustness
3. Fastness response

Among them, the stability of the system is very important. Because our system is an attempt for treatment, the first consideration should be the safety of the problem. We determined the working parameters of the black box system by experiments to set the treatment plan or optimize the system design, which has important reference value of the development of our project in the future. After considering the safety of the system, the next important indicator must also be taken into account: effectiveness. Effectiveness and safety for a system is like the mathematical proof of the adequacy and necessity. In the guarantee of safety conditions, the chief consideration for our system is the effectiveness. There are various methods to evaluate the effectiveness. Generally it can be obtained from the experiments between the experiment group and the control group. In this case, the control group is usually validated to evaluate the effectiveness. But it is difficult for our system. Thus, we began simply from the verification system in the operation of the various components of the state. We created the simulation modeling, corrected our models with experimental data, and calculate the state in which the system is running under parameters.

2 procedure analysis

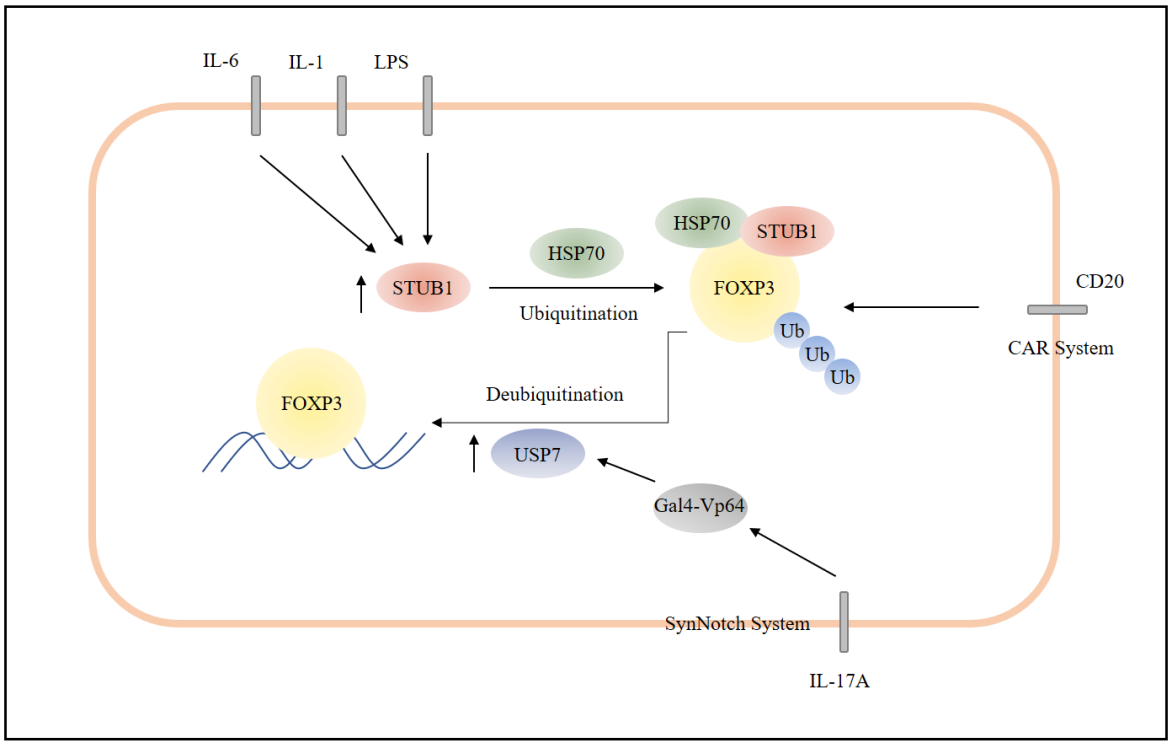

Figure 1: Our system

Our system consists of two parts: one is SynNotch system, and the other is CAR system (Figure 1). The system contains the following biological processes:

1. The interaction between the receptor and its ligand
    - The binding of IL-17A to its receptors
2. Signal transduction kinetics after binding
    - the release of VP64
3. Gene expression modeling
    - recognition of the promoter
    - the dynamics of transcription and translation
4. Enzyme kinetic process
    - the ubiquitination and de-ubiquitination of FoxP3

According to the above biological process, we mainly selected the gene expression modeling and enzyme kinetic process to have a quantitative analysis, and try to establish the parameters with experimental data.

2.1 The binding of IL-17A to its receptors

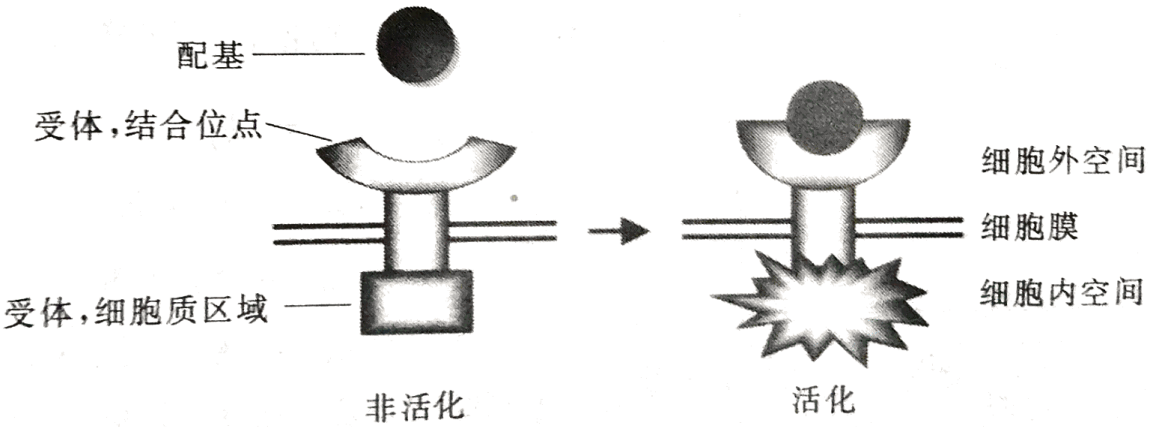

Figure 2: Receptor and ligand binding mode

The combination of the syn-Notch protein with the Gal-VP64 is disrupted after the stimulation of the signal. This process can be described mathematically as a set of differential equations (ODE).

$$\frac{dR}{dt} = f(S, r) \tag{1}$$

Where S is the signal strength (Eg. IL-17A concentration), R is the output signal strength (VP64 concentration), r is the receptor concentration.

For such an ODE system modeling, what we are most concerned about is the choice of parameters. To solve this problem, we try to query the relevant data from the literature or database and model the parameters based on parametric and nonparametric tests.

2.2 Gene expression modeling

Modeling of gene expression is a prime example of the success of obtaining great results using different techniques in the scientific field. The results and dynamics of gene expression are mathematically described by Boolean network, Bayesian network, directed graph, ordinary and partial differential equation, stochastic equation and rule-based formal method.

2.2.1 The recognition of the promoter

For our designed promoter, we searched for desired parameters associated with the promoter intensity, the transcription factor and the promoter sequence on the relevant bioinformatics database.

2.2.2 The kinetic process of transcription and translation

The kinetic process of transcription and translation is similar to that of the previously mentioned receptor ligand interaction model, and is also described by the ODE system, except that the equation is slightly different. After reviewing the relevant books and

literature, it can be seen that the modeling of specific processes of eukaryotic gene expression can be used in a variety of ways, but for the quantitative model ODE system,it has an irreplaceable advantage.

Using ODE system to determine the parameters (mainly in the Milestone dynamics) is also difficult. Usually we access the literature and database to find the parameters, but after analysis, we find using data mining method can help us determine the parameters more reasonable and accurate.

2.3 Ubiquitination and de-ubiquitination of FoxP3

This part mainly establishes the kinetic model of enzymatic reaction under different mechanism. We try to use the data to explore the mechanism or to compute the kinetic parameters from the experimental data. Similar to the above model, the de-ubiquitination process of proteins can also be expressed as a series of ODE systems based on the Hill equation. However, depending on the mechanism of the reaction, the equation is different. And the kinetic parameters of the two enzymes (USP7 and STUB1) involved in the above reaction can be obtained from the BRENDA database.

3 The establishment of model

3.1 Experimental Data Mining and Parameter Calculation based on machine learning

3.1.1 Machine Learning Summary

Machine Learning is to study how computers simulate or implement human learning behavior to acquire new knowledge or skills, and it reorganizes existing knowledge structures to continually improve its performance. There are several different types of learning algorithms. And we introduce two machine learning methods.

(1) Unsupervised learning

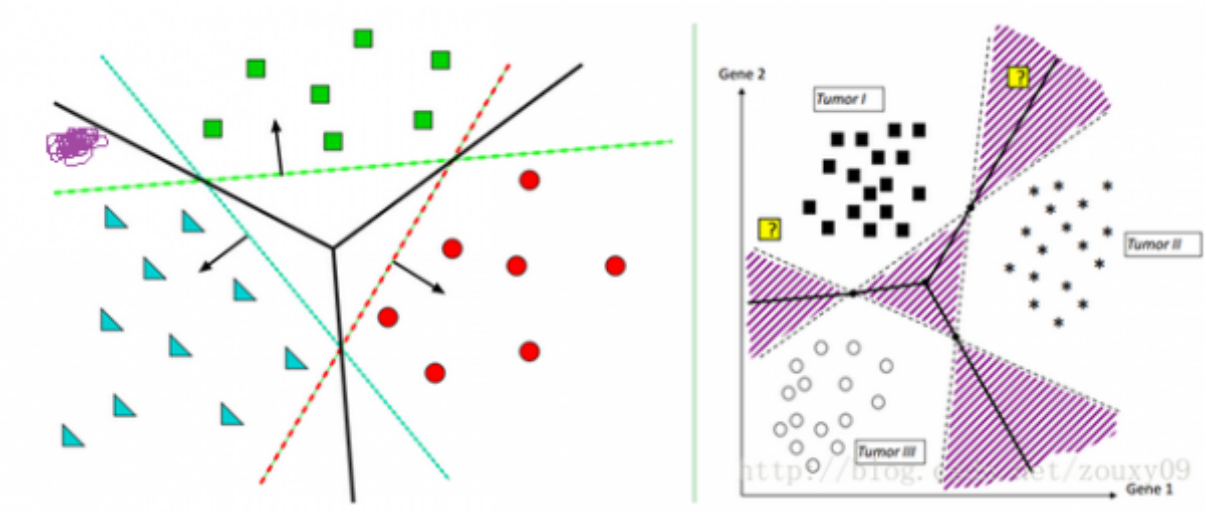

Figure 3: Unsupervised learning two-dimensional spatial classification diagram

The most important use of unsupervised learning is to study the classification problem. We use cluster and principal components analysis to identify the key factors and find the potential categories in the data, which is very difficult in classical statistics. However, with the high-speed computing power of modern computers, we can let the computer learn data, and then classify the data.

(2) Abnormal value detection

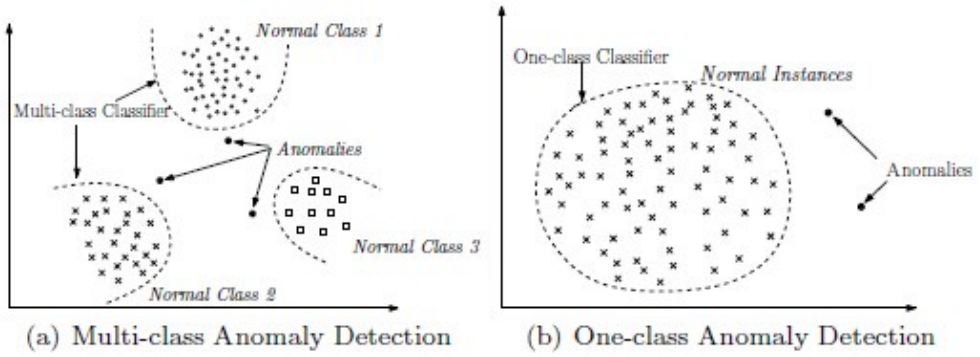

Figure 4: Schematic diagram of abnormal value detection

Abnormal value detection is a machine learning algorithm based on the multivariate Gaussian distribution. It allows us to use the training set to train the model, and through the cross-test set to improve the model, which leads to the improvement of prediction accuracy and the reduction of false alarm rate. Finally, the test set can be used to evaluate the model, in order to assess the safety and stability of the system.

3.1.2 Data mining and parameter calculation

(1) Use unsupervised learning to find the potential steady state of the system

For such a complex system, there will inevitably be a lot of steady state, but found that these steady state is difficult. So we use the unsupervised learning in machine learning to excavate the steady state of the system. Through clustering analysis, data points can be clustered into several classes in multidimensional space. These classes are likely to be some potential systemmetical steady-state, because if the data can be clustered, the corresponding kinetics parameters of the center are mostly stable (can be explained by the vector field theory of differential equations). So we use unsupervised learning to classify the data to guide the later dynamics modeling.

(2) Use the anomaly detection method to evaluate the stability of the system

Abnormal value detection methods are widely used in medicine, for example, we use some disease patients and normal blood routine training to practice the model. When a patient's new data is entered, the model can indicate whether the patient's blood is normal and what disease it is. In this model, we practice a machine to do those things. We use the anomaly detection model to predict whether our system is working properly, if not, what the problem is.

3.2 Stability analysis

At present, the stability analysis of the system is mainly based on the system model, and it depends on the accuracy of the model. There are two general ways to analyze its stability: analytic and numerical methods. Analytical analysis usually requires cumbersome operations, and it does not work for the model without analytical solutions. So we usually use numerical solutions to analyze the stability of the system. The phase plane analysis is a common method of numerical analysis.

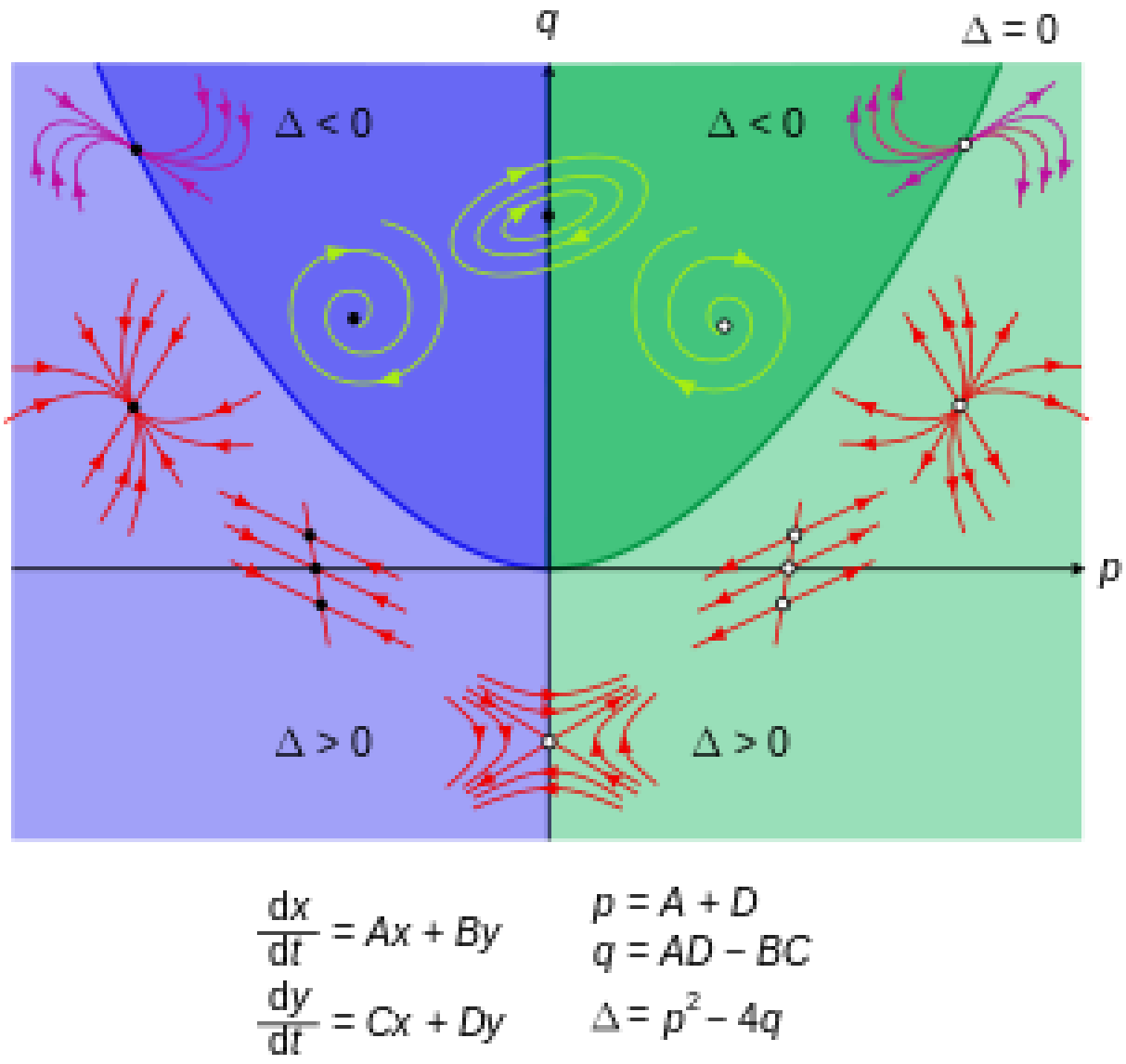

Figure 5: Classification of equilibrium points of a linear autonomous system

4 Discussion

For such a model, our advantage is that the algorithm we used is advanced, but the amount of data is small that is difficult to be measured. We can measure very little data, and the cost of each data measurement is very large. Thus the classical statistical method has failed in that case. We want to find a way that using a small amount

of data can achieve better results by modeling analysis. So we decide to use machine learning. Machine learning is very advantageous in dealing with complex systems because its algorithms are very similar to human thinking.

As can be seen from the previous text, for the whole system, we used a deterministic (continuous) model and statistical (discrete) model, respectively. The deterministic modeling section focuses on the dynamics of the system. In this model, we assume that the biological system is continuous, even if the actual biological system is obviously a random process. Its time-dependent properties (kinetic properties) are only approximate to the dynamical system simulations after long-term measurements. But if we only consider the part of (short-term) changes, the data are often far from the reality. On the contrary, statistical modeling has nothing to do with the process, that is, the prediction of statistic modeling towards experimental data is usually accurate because statistics are descriptive under such conditions. On the whole, it can be said that there must be a loss, we can only try to make the right judgment as much as possible. Of course, it cannot be achieved without data no matter how good the algorithm is. So we can guarantee the completion of the case as much as possible data points, so the model will be more accurate. Therefore, in the guarantee to be able to complete the case, we set up data points as much as possible, so that the model will be accurate.