

Research on the modeling and control methods of nonlinear hybrid system of polymerase chain reaction

ZHANG NAN², MENG FANYUN², WANG JINHE^{2,3}

Abstract. Polymerase chain reaction (PCR) is a new technology of DNA amplification. The interaction between biological macromolecules in the DNA amplification has a variety of possible dynamic process, it is difficult to characterize fully through the existing experiment method. In this thesis, the principles of DNA amplification have been studied and put forward the mathematical model in the process of DNA amplification, nonlinear hybrid system for the PCR process has been established, optimization algorithm has been presented based on bundle optimization for nonlinear hybrid system, and system model has carried on the PCR platform. Experimental results show that the proposed algorithm is robust.

Key words. Polymerase chain reaction, pid control, nonlinear hybrid system.

1. Introduction

Polymerase chain reaction (PCR) is a major specific technique for gene testing, and has been widely used in universities, laboratories, medical institutions, and clinics. PCR consists of 3 stages: denaturation, annealing and extension. In the first stage, the mixture of DNA molecules is heat to a high temperature T_1 , and the double stranded DNA can be decomposed into single one and primers will be bound with single DNA under the action of polymerase; then, the temperature dropped to a low temperature of T_2 , The binding has the stable state with matching primers on the stage; T_3 is the stage of Polymerase elongation, a large number of nucleotides added to the primer template^[1–4]. For example, through the above cycle, it will produce two times of amplification, after two cycles will produce 4

¹Acknowledgement - This paper is supported by National Natural Science Foundation of China with project number 31271077 and Shandong Province Higher Educational Science and Technology Program with project number J17KA061.

²Workshop 1 - School of Computer Engineering, Qingdao university of Technology, Qingdao 266033, China

³Corresponding author: Wang Jinhe; e-mail: jinhewang@163.com

times of amplification, in this way, after 30 or so cycles, billion DNA fragments will be produced. The process of DNA amplification is a dynamic process with multiple possibilities, it is difficult to depict completely in the present experimental method, therefore, a nonlinear hybrid system is used to describe the PCR process, and then the bundle optimization method is used to solve the control problem. Finally, the experimental results are verified based on a PCR platform.

2. Nonlinear system modeling of PCR enzymatic reaction system

2.1. Some Definitions

Each reaction stage of DNA amplification can be divided into two periods, called temperature control period. DNA amplification process is the cycle of temperature control process, the temperature control period is a time interval. The temperature control period TCP C is usually expressed as: $C := [t'_1, t'_2]$. The cost of each temperature control period of time on DNA, let $t_{C_{i_1}}^N, t_{C_{i_2}}^N$ denote the two endpoints of $C_i (i \in \{1, 2, \dots, 6\})$ N denote the circle number, $\forall t_i^N$ on C_i $t_{C_i}^N \in [t_{C_{i_1}}^N, t_{C_{i_2}}^N]$, the $t_{C_i^N}^L, t_N^{CIRCLE}, t^{TOTAL}$ and t^{END} can be denoted respectively as:

$$t_{C_i^N}^L = t_{C_{i_2}}^N - t_{C_{i_1}}^N \tag{1}$$

t_N^{CIRCLE} can be denoted as:

$$t_N^{CIRCLE} = \sum_{i=1}^6 t_{C_i^N}^L \tag{2}$$

t^{TOTAL} can be denoted as:

$$t^{TOTAL} = \sum_{i=1}^N t_i^{CIRCLE} \tag{3}$$

t^{END} can be denoted as:

$$t^{END} = t^{START} + t^{TOTAL} \tag{4}$$

So the time interval of the reaction stage is $[t^{START}, t^{END}]$, shorthand as $[t^S, t^E]$.

2.2. The nonlinear system model of PCR enzymatic reaction system

In this thesis, a cluster of nonlinear hybrid systems is established in order to reveal the dynamic process of biological macromolecules interaction (the formula (5)). The model is the nonlinear equation take the concentration of template DNA, the

concentration of primers, the concentration of DNA polymerase, the concentration of polymerase activator Mg2 + as status variable:

$$(P) \begin{cases} \dot{x}(t) = f(x, v, t), t \in [t^S, t^E] \\ x|_{t=t_0} = x^0 \end{cases} \tag{5}$$

In the formula $v \in R^k$ is parameter vector, $f = \{f_1, f_2, \dots, f_8\}^T, x \in R^8$ is the state vector, Its components represent the biomass concentration x_1 , concentration of the template DNA; x_2 , concentration of DNA polymerase; x_3 , concentration of the primer; x_4 , concentration of polymerase activator Mg2+; x_5 , concentration of dNTP; x_6 , concentration of PCR buffer regulate PH x_7 , concentration of dipolymer which is nonspecific product in the reaction process; x_8 , temperature of the biomass reaction liquid during the reaction process.

The initial values that are involved in formula (5) and the following formula are all calculated by experiment or got by system characteristic.

Set temperature control section $[t^S, t^E] \in R^1, x_j^i(t) (j = 1, 2, 3, 4, 5)$ respectively represent the concentrations of template DNA, the dynamics model of individual equation of the formula (5) was established by using the M-M equation (M - M), v is parametric vectors, its components are implicit in the function. For template DNA concentration, equation is:

$$f_1(x, v, t) = \varphi_{11}(i, t) \times E_1 \times (E_2(1 + P_{K_1})^{k_1}) + \varphi_{12} \times x_2(t) \tag{6}$$

In the equation, φ_{11} a subfunction of the template concentration, Its value set is bounded closed set, the initial value of the experiment is 1.7×10^{-2} , E_1 is the amplification factor for DNA, the initial value of the experiment is 1.3×10^{-3} , P_{k_1} is the average expansion efficiency of the first k_1 times, the initial value of the experiment is 0, if the average molecular weight of the base is 230, E_2 is $2 \times 230 \times L$, among it, L is DNA fragment length, φ_{12} is the polymerase association factor initial value is 1. The initial values of φ_{11} and other parameters were substituted into the template DNA concentration and DNA polymerase concentration formula, the expressions of template DNA concentration and DNA polymerase concentration at the beginning is

$$f_1(x, v, t) = 1.0166 \times 10^{-2} \varphi_{11}(i, t) + \varphi_{12} \times x_2, f_2(x, v, t)$$

In the equation, φ_{21} is the enzyme factor (subfunction of $x_2(t)$), its value set is bounded closed set, the initial value of the experiment is 3.7, φ_{22} is the measure of the enzyme, its initial value of the experiment is 0.57×10^2 , φ_{23} is the primer correlation factor, its initial value of the experiment is 0.34×10^2 , g is the enzyme association function: $g(x, v, t) = (k - k_1) \times (1 - kt' + P_{k_1} t') \text{div } 2e^{\omega T}$. The initial-time expression of the concentration of the DNA polymerase and the primer were obtained.

$$\begin{aligned} f_2(x, v, t) &= 28.5 \varphi_{21}(i, t) \times (k - k_1) \times (1 - kt' + P_{k_1} t') \text{div } 2e^{\omega T} + 34 \times x_3(t), \\ f_2(x, v, t) &= \varphi_{31}(i, t) \times x_4(t) + \varphi_{32}(i, t) \times x_5(t) + \varphi_{33}(i, t) \times x_6(t). \end{aligned} \tag{7}$$

In the equation, φ_{31} is metric factor for the primer subfunction of $x_3(t)$, its initial value of the experiment is 0.227×10^2 , φ_{32} is triphosphate deoxyribonucleotide factor, its initial value of the experiment is 0.9576×10^{-2} , φ_{33} is measure of the buffer, its initial value of the experiment is 0.0227×10^3 , then expression of initial moments of primer concentration were obtained.

Assumed that $f_j(x, v, t)$ ($j = 1, 2, \dots, 8$) is continuous on R^{k+1} . Used symbol $\|\cdot\|$ to express a measure of n -dimension space, thereupon $f_j(x, v, t)$ meet the linear growth rule: There are constants $\delta_1 > 0, \delta_2 > 0$, which make $f_j(x, v, t)$ satisfied $\|f_j(x, v, t)\| \leq \delta_1 x + \delta_2$.

Then, we can obtain the following theorem:

Theorem 1 linear growth theorem of PCR system $f(x, v, t)$ satisfied linear growth conditions: For constants $\delta_1 > 0, \delta_2 > 0$, there are

$$\|f_j(x, v, t)\| \leq \delta_1 x(t) + \delta_2 \quad (8)$$

This theorem is crucial for the control system, which reflects the PCR system is bounded, controllable and existing solution set. If system was controlled by this constraint, the system control is robust.

2.3. Analysis of the existence of nonlinear system solutions for DNA polymerase chain reaction DNA

We discussed the existence and uniqueness of the solution of DNA polymerase chain reaction model (6) as follow. It can be obtained from theorem 1 that the function of the model (6) has a very good property, and the following conclusions obviously established.

Theorem 2 existence theorems of solutions if $f(x, v, t)$ was continuous on $R^8 \times R^{k+1}$, and satisfied local Lipschitz condition with respect to x , and for constants $\delta_1 > 0, \delta_2 > 0$, there was $\|f(x, v, t)\| \leq \delta_1 x + \delta_2$,

for $\forall (x_0, v_0, t_0) \in R^8 \times R^{k+1}$, the solution of initial value problem (6) existed on the interval $(0, +\infty)$.

By theorem 2, we can solve the initial value problem (5). Because of the equivalence, we can solve for the problem (7). Supposed

$$\Omega := \{t \mid \dot{x}(t) = f(v, t)x(t), t \in [t^S, t^E]\} \quad (9)$$

Now, we need to set the performance index of the problem (7) to solve the unknown parameters.

2.4. Numerical calculation of nonlinear system of DNA polymerase chain reaction

1) Determined concrete expression of $\varphi_{j1}(i, t)$

In numerical calculation, a linear system of equations for solving unknown parameters set up first, for each $x_j^i(t), \varphi_{j1}(i, t)$ that is subfunction of

$x_j^i(t)$ ($j = 1, 2, \dots, 8; i = 1, 2, \dots, 6$) is the temperature control period). It is diffi-

cult to determine their expression, but in the experiment, interpolation simulation can be performed according to the experimental data, and the interpolation function is used instead of the original function to calculate the data. The specific calculation steps are:

Step 1: figure out the concentration of each component in temperature control segment i .

- (1) formulate process, determine time point
- (2) measure component concentration

According to above six processes the absorbance $A_j^i(t)$ at t moment of component template DNA, polymerase, primers, Mg^{2+} and dNTP were measured by spectrophotometer, following Moore's law, the concentration $x_j^i(t)$ was solved by linear relation between absorbance and concentration $x_j^i(t)$ ($i, j = 1, 2, \dots, 6$):

$$x_j^i(t) = A_j^i(t) / \epsilon_j \quad (10)$$

In the equation, ϵ_j is molar absorption coefficient of component respectively.

The molar absorption coefficient of the component is relatively fixed and does not change with the concentration of the measured material. The molar absorption coefficient of each component is analyzed in the following.

For DNA product, Freifelder^[6] provide the relation of absorbance A and molar absorption coefficient OD by experiments measure.

With regard to double-strand DNA, the relation is $A^{260} = OD^{260} = 50 \mu g/mL$;

With regard to single-strand DNA, the relation is $A^{260} = OD^{260} = 33 \mu g/mL$.

Plug the last two expressions in equation (10), we obtained:

(1) With regard to double-strand DNA, the molar absorption coefficient of component was measured in temperature control section 4, T_1' is $\epsilon_1' = \frac{A_1^4(t)}{x_1^4(t)} = 20 \cdot OD/g$.

(2) With regard to single-strand DNA, the molar absorption coefficient of component was measured in temperature control section 2, T_1 is $\epsilon_1 = \frac{A_1^2(t)}{x_1^2(t)} = 30.3030 \cdot OD/g$.

Plug measured absorbance OD and the expression above in formula 10, the concentration of DNA products can be figured out, constitute number pairs $(t_1^k, x_1^k(t_1^k))$ ($k = 1, 2, \dots, 16$) with t moment.

For the coefficient of primer molar absorption ϵ_2 , Because the primer is different, the base composition is different, therefore, we suppose $U_1 U_2 U_3 U_4$ as numbers of four base pairs A C G T, there are^[6]

$$\epsilon_2 = 16000U_1 + 12000U_2 + 7000U_3 + 9600U_4$$

Plug measured absorbance OD and the expression above in formula 10, the concentration of primer and the relevant time t can be figured out, constitute number pairs $(t_2^k, x_2^k(t_2^k))$ ($k = 1, 2, \dots, m$) with t moment, Where m is the number of calculations.

In the same way, the coefficient of molar absorption ϵ_j ($j=3,4,5,6$) of other components can be calculated or examined, plug measured absorbance OD in formula 10, the concentration of component can be figured out, constitute number pairs $(t_j^k, x_j^k(t_j^k))$ ($k = 1, 2, \dots, 16; j = 3, 4, 5, 6$) with t moment.

For the concentration when j equal to 7,8, we can obtain through the indirect calculation method, the calculation procedure is elliptical.

In conclusion, We've calculated the number pairs $(t_j^k, x_j^k(t_j^k)) (k = 1, 2, \dots, 16)$; We will express data pair that we calculate as $\Omega_j (j=1, 2, \dots, 8)$

$$\Omega_j = \left\{ (t_j^{k_{pq}}, x_j^{k_{pq}}(t_j^{k_{pq}})) \mid p = 1, 2, 3, 4; q = 1, 2, 3, 4 \right\} \tag{11}$$

Now we're going to use data pair of Ω_j to approximate $\varphi_{j1}(i, t)$.

Step 2: Expressed $\varphi_{j1}(i, t)$ function with Ω_j

First, determine the relationship between i and t : in the function $\varphi_{j1}(i, t)$ are cycle index i and time t , they relationship is derived from formula (3): $t = \sum_{k=1}^{i-1} t_k^{CIRCLE} + t'$. In the equation, t' is execution time for the current cycle, $t_k^{CIRCLE} (k=1, 2, \dots, i-1)$ is each time the cycle takes for the previous $i-1$ times cycles, based on the time of the current cycle t' , circulation section $C_{i^*} (i^* \in \{1, 2, \dots, 6\})$ of the current cycle can be calculated, and we can calculate the current temperature control section and vice versa.

For each j , we solve the expression of function $f_j(x, v, t)$. Because each loop in the detection process has only four temperature control segments, the target function is set up as:

$$\bar{\varphi}_{11}(i, t) = b_{10}^i \times t + b_{11}^i \times x_1^i(t) e_1(i, t) = \left(\sum_{q=1}^4 x_1^{k_{iq}}(t_1^{k_{iq}}) - b_{10}^i \times t - b_{11}^i \times x_1^i(t) \right)^2, i \in I$$

If we figure out the minimum of the above formula, we can figure out the coefficients of it. We take partial derivative of b_{10}^i, b_{11}^i in above formula at moment $t_1^{k_{iq}}$, that is

$$\frac{\partial e_1(i, t)}{\partial b_{10}^i} = -2 \left(\sum_{q=1}^4 x_1^{k_{iq}}(t_1^{k_{iq}}) - b_{10}^i \times t - b_{11}^i \times x_1^i(t) \right)$$

$$\frac{\partial e_1(i, t)}{\partial b_{11}^i} = -2 x_1^i(t) \left(\sum_{q=1}^4 x_1^{k_{iq}}(t_1^{k_{iq}}) - b_{10}^i \times t - b_{11}^i \times x_1^i(t) \right)$$

Let

$$\frac{\partial e_1(i, t)}{\partial b_{10}^i} = 0 \quad \frac{\partial e_1(i, t)}{\partial b_{11}^i} = 0, \tag{12}$$

According to the collection Ω_1 , we can figure out the coefficients b_{10}^i and b_{11}^i . then, we obtained the expression of $\varphi_{11}(i, t)$: $\varphi_{11}(i, t) = b_{10}^i t + b_{11}^i \times x_1^i(t)$. Plug the equation above in expression of $f_1(x, v, t)$, because of φ_{12} is constant, it can be given initial values in the algorithm (without affecting subsequent optimization calculations), then, we obtain concrete expression of $f_1(x, v, t)$.

For $f_2(x, v, t) (i \in \{1, 2, \dots, 6\})$, Based on the characteristics of polymerase, the linear objective function of polymerase is established as follows:

$$\bar{\varphi}_{21}(i, t) = b_{20}^i t + b_{21}^i \times x_2^i(t)$$

Let

$$e_2(i, t) = \left(\sum_{q=1}^4 x_2^{k_{iq}} (t^{k_{iq}}) - b_{20}^i \times t - b_{21}^i \times x_2^i(t) \right)^2$$

If we figure out the minimum of the above formula, we can solve the coefficients of it. We take partial derivative of b_{20}^i and b_{21}^i in above formula at moment $t_2^{k_{iq}}$, let the derivative be equal to zero, according to the collection Ω_2 we can figure out the coefficients b_{20}^i and b_{21}^i . then, we obtained the expression of $\varphi_{21}(i, t)$: $\varphi_{21}(i, t) = b_{20}^i t + b_{21}^i \times x_2^i(t)$. Plug the equation above in expression of $f_2(x, v, t)$, because of φ_{22} and φ_{23} are constants, then, we obtain concrete expression of $f_2(x, v, t)$.

Step 3: Set performance indicators for solving unknown parameters

The most important indicators in the experiment are $x_1^i(t)$. For the six-segment temperature control segment, the product indicator extraction time is $i=6$; For the four-segment temperature control segment, the product indicator extraction time is $i=4$, the data collections of other temperature control section $i=1, 2, 3$ are taken the same magnitude, then, There are the following indicator functions:

$$\max J(x, v) = \left(c_1 \sum_{i=1}^6 \int_{t_{i-1}}^{t_i} x_1^i(t) dt - \int_{t_{i-1}}^{t_i} (c_2 x_7^i(t) + c_3 x_8^i(t)) dt \right)^2.$$

The nonlinear equation determined by formula (5) is multi-solvable, its solutions is the unknown parameters that we need, the parameter identification can be taken by the maximum function with respect to variable x to find out the better parameters.

2.5. The nonlinear system properties of PCR enzymatic reaction system

We assumed that the measurement of the difference between template DNA and PCR products in the PCR reaction system is bounded; the difference between the other reactants and the current consumption concentration is bounded, that is, $\exists M > 0$ so that $|x_j(t_0) - x_j(t)| < M, j = 1, 2, \dots, 6$

Therefore, for equation (7), there are the following theorems:

Theorem 3 continuity theorem $f(x, v, t)$ ($i=1, 2, \dots, 8$) meet the following properties:

- 1) It is continuous on V with respect to v ;
- 2) It is Lipschitz's continuous on W with respect to t .

Proof: slightly.

2.6. Nonlinear system analysis of PCR enzymatic reaction system

The PCR response system model P (7) reflects the substance of the system, after solving this model, the mathematical expression of the system model is solved.

3. Establish Control model

3.1. Calculate controller parameter

According to the PID control model, the control problem of the reaction process of the PCR nonlinear system in formula (5) can be expressed as the following form of non-smooth optimization problem^[6-7]:

$$(\text{VP}) \begin{cases} \min g(t) = \varepsilon_1 t \sum_{m=0}^t |e(m)| + \varepsilon_2 e(t)^2 \\ \text{s.t. } t \in \Omega \end{cases} \quad (13)$$

In the formula, $e(j) = u(j) - r(j)$ $u(j)$ is output for the controller, $r(j)$ is input for the controller.

The optimization problem of the above formula can be solved by using the bundle set, and its solution is given by algorithm 1:

Algorithm 1A control problem solving algorithm based on constrained optimization method

Step 1, Initialize take $\varepsilon > 0, \rho \in (0, 1), M_k = 1, k = 0, x_1, \xi_1,$

Step 2, **loop begin**;

Step 3, $k = k + 1$;

Step 4, Calculated $d_k, g(x_k), \xi_k \in \partial g(x_k)$

Step 5, Update the slice linear function: $\widehat{g}_k(x) := \max_{j \in J_k} \{g(x_k) + \langle \xi_j, d_k \rangle - \alpha_j^k\}$

Step 6, $y_{k+1} = d_k + x_k$??calculate $\widehat{g}_k(y_{k+1})$;

Step 7, Judge $\widehat{g}_k(y_{k+1}) - g(x_k) \leq \rho\varepsilon$;

Step 8, If $x_{k+1} = y_{k+1}, t_k^* = x_{k+1}$ output bundle set, go to step 11;

Step 9, Else $x_{k+1} = x_k$, output bundle set;

Step 10, **loop end**

Step 11, End.

According to the constrained optimization method (algorithm 1), we calculate with MATLAB, obtain collection W :

$$W := \{(t, \mu) | t_k \in [t^S, t^E], \mu_i^k \in R, i = 1, 2, 3, 4\} \quad (14)$$

3.2. Example

In this example, the DNA amplification reaction of 1 cyclic processes was tested on the PCR platform, and 15 temperature test points were taken for the reaction process temperature data sampling. Control node is 93.2, the time interval of constant temperature stage is 360s, and initial temperature is 36.2. There are three constant temperature stages in this example. The aim is to control the noise of the environment by feedforward control with feedback.

Calculated example:

In this example PID parameters are calculated in two calculations, first the parameters of p-mode is calculated, then the control parameters of PID mode (constant temperature) is calculated. Steps as follows:

(1) Find the solution set of the VP model; For the temperature control section of 1, 3 and 5, the model VP and algorithm 1 were used to calculate the parameters of the proportional mode, and the P mode control parameters μ_1^k and t_k^* $k=1,2,\dots,15$, were obtained.

(2) For the temperature control section of 2, 4 and 6, the model VP and algorithm 1 were used to calculate the parameters of the PID controller, the P mode control parameters $\mu_1^k, \mu_2^k, \mu_3^k$ and the corresponding moment t_k^* $k=16,17,\dots,30$, were obtained.

(3) In the 1th, 3th, and 5th temperature control segments, First, the BANG-BANG mode control is adopted, then the p-proportional mode control is adopted, In the 2th temperature control segments, PID controller is adopted to control. The output of the system is shown in figure 1, the left of the figure is the output of the temperature data of sampling value, the right of the figure is voltage of the control curve. The control data curve reaches the design index, curve is superimposed, the temperature uniformity is less than 0.2°C .

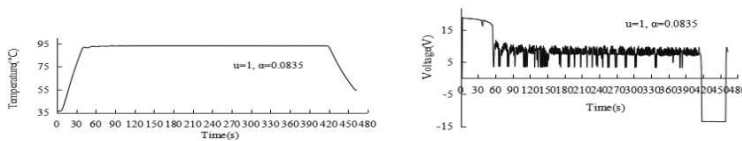


Fig. 1. PCR Control Curve Chart with Algorithm 1 from Sampling Points

4. Conclusion

This thesis proposed basic concepts of DNA amplification, analyzed the principle of PCR enzymatic reaction, established PCR enzymatic reaction system model and analyzed system model, decomposed the model of PCR, obtained the existence of solutions to the system model, numerically calculated the DNA polymerase chain reaction model. According to the characteristics of PCR this thesis proposed PID control mathematical model of PCR enzymatic reaction system, proposed a constrained optimization method for solving this model and given the algorithm steps, The experimental results show that the control effect of the proposed control algorithm is obvious, loop control time is short, control precision is reached 0.1°C .

References

- [1] Y. AN, K. REIMERS, C. ALLMELING: *Validation of differential gene expression in muscle engineered from rat groin adipose tissue by quantitative real-time PCR*. Biochemical and Biophysical Research Communications 1 (2012), No. 421, 736–742.
- [2] A. IWABI, D. SEBAH: *A multiplex real-time PCR method for the quantitative determination of equine (horse) fractions in meat products*. Food Control 1 (201), No. 74, 89–97.
- [3] S. CUEVAS-ROMEROA, A. L. BLOMSTROM, A. ALVARADOB: *Development of a real-time RT-PCR method for detection of porcine rubulavirus (PoRV-LPMV)*. Journal of Virological Methods 189 (2013), No. 1, 1–6.

- [4] C. D. VILJOEN, G. G. THOMPSON, S. SREENIVASAN: *Stability of ultramer as copy number standards in real-time PCR*. *Gene* 516 (2013), No. 1, 143–145.
- [5] H. G. OUATTARA, S. REVERCHON, S. L. NIAMKE: *Regulation of the synthesis of pulp degrading enzymes in Bacillus isolated from cocoa fermentation*. *Food Microbiology* 63 (2017), No. 1, 255–262.
- [6] D. M. FREIFELDE: *Physical Biochemistry: Applications to Biochemistry & Molecular Biology* W.H. Freeman and Company. CA 52 (1976), No. 2, 177–183.

Received November 16, 2017