Modeling and Simulation of Production of Metallothionein and Red Fluorescent Fusion Protein by Recombinant *Escherichia coli* Using LabVIEW

Liangyu Su

School of Electrical and information Engineering
Xu Chang University
Xu Chang, China
sulyu@qq.com

Abstract—Genetic algorithm(GA) is a powerful tool for the analysis of the mechanism and dynamic responses of the fermentation process,and for process optimization and automatic control.In this research,the structured and unstructured mathematical models and computer simulation of MT production and RFP fluorescence are made,and it is implemented,verified and analyzed in LabVIEW.The result shows that the advantages and the usefulness of the data driven and graphical programming software of LabVIEW in modeling and simulation of biological phenomena.

Keywords-Genetic algorithm; LabVIEW; MT production; RFP fluorescence; graphical programming

I. INTRODUCTION

Metallothioneins(MTs) are a family of low molecular weight,cysteine rich,metal binding proteins,and exist in a variety of organisms ranging from bacteria to mammals.MTs play significant roles in clearance of free radical against oxidative damage,trace element metabolism and heavy metal detoxification,cell metabolism,disease processes,rare metal accumulation,heavy metal removal and environment protection.The important applications stimulate the interest of large scale production of MT.Compared with the conventional method by extraction from animal organs,MT production by using recombinant E.coli has obvious advantages in reducing cost,increasing efficiency and product quality[1].However,recombinant MT is easily hydrolyzed and unstable for the thiol group of cysteine when produced by recombinant microorganism.By integrating the MT gene with fusion tags,MT was successfully expressed in E.coli.On the other hand,rapid measurement of MT production is important for optimization of fermentation process.MT was reported to have light absorption at the wave length of 250nm after binding with Cd,which could be used for MT measurement.But,according to our experience,this method has low accuracy and poor reproducibility[2].In order to detect MT production in an easy and rapid way during fermentation process,in this study,MT gene is fused with red fluorescent protein(RFP) gene and coexpressed in recombinant E.coli.Modeling and simulation is a powerful tool for the analysis of the mechanism and dynamic responses of the fermentation process,and for process optimization and automatic control[3].The production of recombinant protein using recombinant E. Coli involves the regulations of aimed gene transcription and translation,which process can be well described by using the white box structured model in voling operon regulation and is done in this research. But, the structured model has the disadvantage of extreme complex not suitable for the purpose of process optimization.Therefore,a gray box unstructured model is also constructed, which model is simple enough for process variable prediction in real applications but still sufficient for modeling the major regulation effects on the aimed genes of the recombinant strain.LabVIEW is a data driven and graphical programming software tool,widely used in process simulation,data acquisition and automatic control.The MT production mathematical model developed using LabVIEW is easily incorporated into the model based on-line optimization and automatic control system of MT fermentation process developed using LabVIEW in the future[4].In this research,the structured and unstructured mathematical models and computer simulation of MT production and RFP fluorescence are made.The final developed mathematical model includes substrate consumption,the growth of the recombinant E.coli cells,the repression and induction of the aimed genes,the production of the aimed protein,and the maturation of the expressed RFP.The mathematical model consists of two levels:the intracellular level for the expression and induction of the aimed genes and the reactor level for the dynamic changes of substrate and cell concentrations as well as the fluorescence. This research shows the advantages and the usefulness of the data driven and graphical programming software of

LabVIEW in modeling and simulation of biological phenomena[5].

II. EXPERIMENTAL PRODUCTION OF MT FUSION PROTEIN

The human MT(hMT-1A) gene sequence was obtained from Genebank.Some rare codons for E.coli were modified,and the sequence was synthesized chemically.The plasmid pGHM-RFP was used for MT production.DsRed2 gene was cloned and ligated with hMT-1A gene to construct the fusion protein expression vector.In order to minimize the conformational change and final clearence of RFP,a Gly-Gly-Gly tri-peptide and thrombin restriction site were designed and inserted between DsRed2 and hMT-1A genes in pGHM-RFP construction.E.coli BL21(DE3) was used as the host cell to construct E.coli(pGHM-RFP).The plasmid pGHM-RFP has a tac promotor (Ptac),which is a combination of -10 reagion of lac promotor and -35 reagion of trp promotor[6].

Cultivations of E.coli(pGHM-RFP) were made at 37℃,shaken at 180r/min.The volume of 250ul of seed culture was inoculated into 40ml of flask containing 300ml of LB medium(10g/l tryptone,5g/l yeast extract,10g/l NaCl,pH 7.0) and 100ug/ml of ampicillin.Isopropyl-D-thiogalactoside(IPTG) was added to the final concentration of 0.2mM when OD600 reached 0.9~1.1.Samples were made every hour and the cells were harvested by centrifugation at 7000×g for 10min at 4oC.Proteins were analyzed using 12% SDS_page,and stained by 0.25% Coomassie brilliant blue R250.The amount of MT produced was analyzed by Image Quant TL using BSA(bovine serum albumin) as the standard.Fluorescence intensity was measured using RF-5301PC.The results were shown in Fig.2.The results showed that the fusion protein of GST-MT-RFP expressed immediately after IPTG induction,but the fluorescence was detected about 6 hours after the induction.The fluorescence was detected about one hour after MT production reached the maximum,and then increased.It indicated that RFP fluorescence could be used as the suboptimal indicator for termination of MT fermentation process.In addition,the fluorescence intensity of RFP was quite strong,which ensured the accuracy of the detection.The possible reason for the delay of the appearance of RFP fluorescence should be that the maturation of red chromophore undergoes a period of oxidation,which costs some time.The time delay of the RFP fluorescence made it unsuitable for online monitoring and optimization of MT fermentation process, but could be used as the suboptimal indicator for termination of the MT fermentation process[7].

Figure 1. The structure of plasmid pGHM-RFP

III. STRUCTURED MODEL FOR PLASMID BASED MT PRODUCTION

The detailed structure and regulation mechanism of tac promotor-operon was shown in Fig.3.From Fig.3 for the control mechanism,it shows that the operon is coordinately controlled by two regulatory sites positioned upstream of the aimed genes,which are the regulatory sites of cAMP-CAP complex and the lacI repressor protein.For the lacI repressor mechanism,when the active lacI repressor protein is binding to the operator gene,mRNA polymerase can not pass through the operator gene to transcribe the aimed gene.When the inducer molecules (lactose or its structure analogue IPTG) binds to lacI to inactivate the lacI repressor and make it get off from the operator,the mRNA polymerase can pass through the operator gene to transcribe the aimed gene.On the other hand, the operon is also controlled by the mechanism of carbon catabolite repression.The binding of mRNA polymerase to the promotor gene (at the -10 and -35 reagions) is loose,needs to be strengthened by the cAMP-CAP complex.The catabolite activator(CAP) protein is activated when it binds to cAMP to form the complex.When glucose concentration is high in the medium,the intracellular cAMP concentration is low,as a result,the concentration of cAMP-CAP complex decreases,and the binding of mRNA polymerase to the promotor becomes loose and the aimed gene is untranscribed[8].
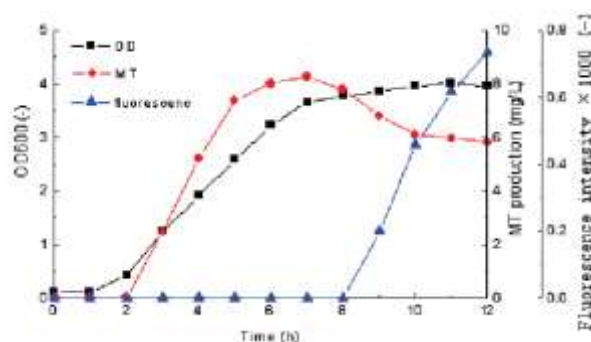

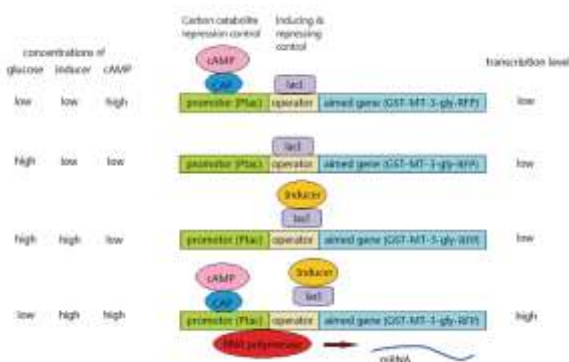Figure 2. The time courses of cell growth,MT fusion protein production and RFP fluorescence.

Figure 3. The structure and regulation mechanism of the operon of pGHM-RFP

IV. UNSTRUCTURED MODEL AND SIMULATION OF MT PRODUCTION

As discussed in last section,the structured model is too sophisticated for bioprocess modelling and simulation in practical applications.In contrast,unstructured models without detailed intracellular mechanisms are convenient

and widely used in this field.But,one major drawback of the unstructured model is that the unstructured black box model sometimes can not reflect the real dynamics of the bioprocess.For example,the Monod model and the Luedeking-Piret model are most often used unstructured models for the cell growth and product production,which are not sufficient to model the gene induction and repression effects.In this section,an unstructured model will be constructed for recombinant MT fusion protein production,and improvements are made in this model so that it can describes the induction and repression effects on the aimed genes exerted by the bioreactor level process variables of the concentrations of related substrates in the culture medium.

The inducer and repressor control mechanism of the operon is summarized into Q_1,and the carbon catabolite repression control mechanism is summarized into Q_2 by using several intracellular reactions and signal transductions.Hear,unstructured model will be constructed for Q_1 and Q_2 .IPTG is used in replace of lactose as the inducer,which is unable to be degraded and has strong inducing effect than lactose.Therefore,the induction effect exerted by IPTG is in the extremes,which can be regarded as "1" or "yes" with IPTG induction, and "0" or "no" without IPTG induction.In addition,the signal transduction of the induction and repression effects have a time scale of several seconds from on-state to off-state or vise versa.Therefore,Q1 can be set to "1" or "0" for the cases of IPTG addition or not described by Eq.(1).

$$Q_1 = \begin{cases} 0(noIPTG) \\ 1(IPTG) \end{cases} \quad (1)$$

## V. MODEL PARAMETERS AND THE OPTIMIZATION USING GENETIC ALGORITHM (GA)

Some of the model parameters have clear physiological meanings and can be obtained by theoretical calculation or by experimental measurements. X_{max} and t_L can be obtained directly from Fig.2;μ_{max} can be calculated from the growth curve of Fig. 2;$\gamma_{X/S}$ can be calculated from the measured data; $\gamma_{MT/S}$ can be estimated from the biochemical pathways of protein synthesis; $N_{P \cdot max}$ is obtained from general knowledge about the plasmid copy number of E.coli; $k_{P max}$ is chosen arbitaryly as 2.5 times of μ_{max} .The other model parameter values are optimized using GA.The differential equations are solved using 4th order Runge-Kutta method[9].

Figure 4. The structure of the chromosome and the $m \times 10n$ array for m chromosomes of n genes

Genetic algorithm is a search heuristic that mimics the process of natural evolution,which is proved useful in solving sophisticated optimization problems as well as in optimization of biological model parameters.A standard GA is used in the optimization,which has the operations of coding,initialization,mutation,hybridization,decoding,fitness,calculation,selection,and reproduction procedures exerted on individuals(or chromosomes) in a population[10].The potential solution(individual) is coded as a binary vector,called a chromosome,the elements of which were called genes and situated in predefined positions,indicated as alleles[11].One gene codes for one model parameter.The gene number n of one chromosome equals to the number of model parameters,so that each chromosome codes for all parameters of the model and enable to calculate the model prediction error with the set of model parameters coded by the chromosome.In GA,one population contains m individuals (or chromosomes) so that m model prediction errors can be calculated for the m chromosomes,and the fitness of each chromosome can then be calculated and used for the selection procedure.In this study,one gene is coded by a binary string of 10 bits.Then,n parameters are represented by a binary string of $n \times 10$ bits.The population of m chromosomes can be represented by a $m \times 10n$ array.The initialization procedure is performed to randomly assign initial values of "0" or "1" to the $m \times 10n$ array.The hybridization procedure is applied to two randomly selected chromosomes called parents,and two new chromosomes created by the exchange of one or more parts randomly selected from the parent chromosomes,which occurred based on a probability called the hybridization rate.The mutation procedure is applied by turning over one or more randomly selected bits in the $m \times 10n$ array from "0" to "1" or from "1" to "0" ,which occurred based on a probability called the mutation rate,it is shown in Fig.4.Decoding was performed to transform the binary number of each gene in the chromosome to a decimal system using the following equation:

$$D_{mn} = D_{min} + \frac{B_{mn}}{2^Z - 1} \cdot (D_{max} - D_{min}) \quad (2)$$

Where, m ,the ordinal rank of the chromosome; n ,the ordinal rank of the gene in the chromosome; Z ,the digits of the binary vector for one gene; B ,the binary value coded by the corresponding gene; D ,the decimal value transformed from B ;and D_{max} and D_{min} , the decimal values of the maximum and minimum limits of the corresponding parameter.Selection and reproduction are performed to select individuals (chromosomes) to create a mating pool for reproducing offspring. The selection procedure is stochastic with the fitted chromosomes, judged by fitness (or objective function J), indicating a better chance of being selected. A low value of J is defined as a high fitness, with J defined as follows:

789
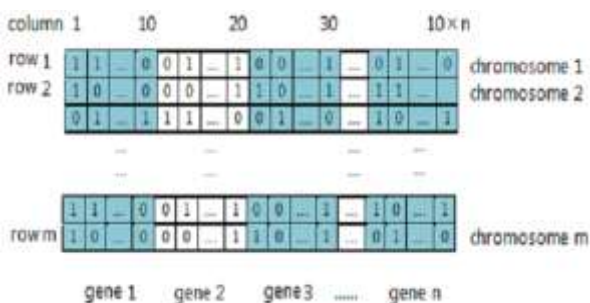
$$J = \frac{1}{R \cdot K \cdot L} \cdot \sum_{r}^{R} \sum_{k}^{K} \sum_{l}^{L} \left[\frac{\gamma_{rkl}^{*} - \gamma_{rkl}}{\gamma_{rkl}} \right]^{2} \qquad (3)$$

Where, R ,the number of experiments; K ,the number of measured variables; L ,the number of measured points of one variable; γ_{rkl}^{*} ,the model output; and γ_{rkl} ,the measured value in the experiments.In order to give each variable the same importance in the optimization, the relative error $(\gamma_{rkl}^{*} - \gamma_{rkl})/\gamma_{rkl}$ is used so that J had no unit. Another advantage using the relative error instead of the absolute error $(\gamma_{rkl}^{*} - \gamma_{rkl})$ in calculating J is to enable the data points of one variable at different sampling time have the same importance in the optimization.A flow diagram of the GA used for the model parameter optimization is shown in Fig.5.In this research,the 4th order Runge-Kutta method is used to solve the differential equations and GA is used to minimize J in Eq.(3).A population size of 50,hybridization rate of 0.2,and mutation rate of 0.05 are used in GA optimization.GA programming using LabView for model parameter optimization was reported.In LabView,there are sub VIs of GA and also the 4th order Runge-Kutta method,it is shown in Fig.6 and Fig.7.In addition,the softwares written in MatLab,VisualBasic et al. can also be linked with LabVIEW and used for the same purpose[12].
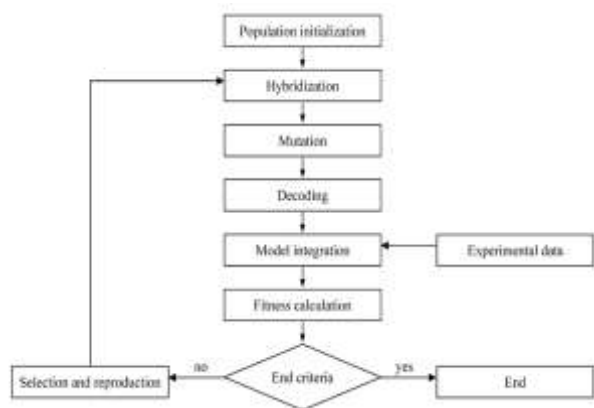

Figure 5. The flow diagram of GA calculation in model parameter optimization.

Initialization of the GA

Decodification, obtention of the fitness and selection   Crossover and Mutation
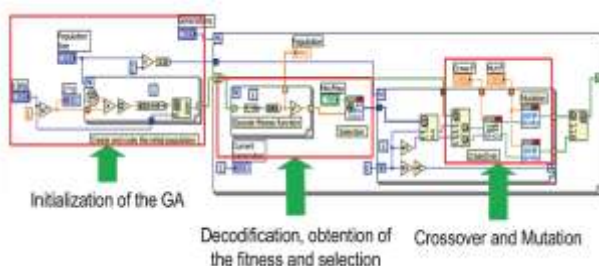
Figure 6. The VI of GA from LabVIEW

Figure 7. The VI image of 4th order Runge-Kutta method from the manuals of LabVIEW.

In optimization of the unknown model parameter values using GA,the known value model parameters are also refined in the mean time to eliminate the possible errores made in the measurements.Each of the measured value is set as the middle value of a small range and the value is searched within the small range by GA in the refining process to minimize the aimed function.

VI. CONCLUSIONS

The production of MT fusion protein is modelled using structured and unstructured models.The unstructured model developed in this research is suitable for bioprocess prediction and optimization.The data driven and graphical programming software of LabView has the advantages of high efficiency in programming,having many handy sub VIs for various applications,and especially easy in constructing data acquitation and on-line monitoring and control system. The mathematical models developed in LabView is easily incorporated into the future developed on-line model based predictive control system.

REFERENCES

[1] Cherian,M.G.&Apostolova,M.D,"Nuclear localization of metallothionein during cell proliferation and differentiation",Cell Mol Biol,46(5),pp.347-356,2012.

[2] Lin J,Lee S.M,Koo Y.M,"Model development for lactic acid fermentation and parameter optimization using genetic algorithm",J.Microbiol.Biotechnol, 14(6),pp.1163-1369,2004.

[3] Moore,"Artificial intelligence programming with LabVIEW:genetic algorithms for instrumentation control and optimization",Computer Methods and Programs in Biomedicine,47(3), pp.76-79,2005.

[4] Su Y,Lin J,Lin J,Hao D,"Bioaccumulation of arsenic in recombinant Escherichia coli expressing human metallothionein.Biotechnol Bioprocess Eng",14(1),pp.565-570,2009.

[5] Verkhusha,Chudakov,Gurskaya,"Common pathway for the red chromophore formation in fluorescent proteins and chromoproteins".Chem Biol,11(2),pp.845-854,2004.

[6] Yang F,Zhou M,He Z,"High-yield expression in Escherichia coli of soluble human MT2A with native functions".Protein Expr Purif,53(4),pp.186-194,2007.

[7] Hong S H,Toyama M,"High yield expression and single step purification of human thionein/metallothionein".Protein Expr Purif,21(1),pp.243-250,2005.

[8] Lee S.B, Bailey J.E,"Genetically structured models for lac promotor-operotor function in the Escherichia coli chromosome and in multicopy plasmids:lac promotor function",Biotechnol Bioeng,26(2),pp.1381-1389,1994.

[9] Moffatt M,"Metallothionein in physiological and physiopathological processes",Drug Metabolism Reviews,29(3),pp.261-307,1997.

[10] Nordberg M,"Metallothioneins:historical review and state of knowledge",Talanta,46(1),pp.243-254,2008.

[11] HUA,LIN Shuwen,"New Knowledge-based Genetic Algorithm for Excavator Boom Structural Optimization.Chinese Journal of Mechanical Engineering",1(4),pp.181-185,2014.

[12] CHU,WANG FuLi,WANG XiaoGang,"A model for parameter estimation of multistage centrifugal compressor and compressor performance analysis using genetic algorithm",Science China,12(11),pp.247-251,2012.