ICROS-SICE International Joint Conference 2009
August 18-21, 2009, Fukuoka International Congress Center, Japan

Validation of a model of regulation in the tryptophan operon against multiple experiment data using global optimisation

Jongrae Kim[1], Jung-Su Kim[2], Ian Postlethwaite[3], Pat Heslop-Harrison[3], and Declan G. Bates[3]

[1] Dept. of Aerospace Engineering, University of Glasgow, Glasgow, UK, jkim@aero.gla.ac.uk
[2] Dept. of Control & Instrumentation Eng., Seoul National University of Technology, Seoul, Korea, jungsu@snut.ac.kr
[3] Systems Biology Lab, Depts. of Engineering and Biology, University of Leicester, LE1 7RH, UK, {ixp,phh4,dgb3}@le.ac.uk

Abstract: This paper is concerned with validating a mathematical model of regulation in the tryptophan operon using global optimization. Although a number of models for this biochemical network are proposed, in many cases only qualitative agreement between the model output and experimental data was demonstrated, since very little information is currently available to guide the selection of parameter values for the models. This paper presents a model validating method using both multiple experimental data and global optimization.

Keywords: Model validation, global optimization, tryptophan regulation, parameter identification

1. INTRODUCTION

Many cellular regulation systems employ multiple feedback loops to allow fast and efficient adaptation to uncertain environments. The various feedback mechanisms used by prokaryotes such as *E. coli* to regulate the expression of proteins involved in the production of the amino acid tryptophan combine to form an extremely complex, but highly effective, feedback control system. This system has been the subject of numerous modelling studies in recent Systems Biology research, with the result that a plethora of different mathematical models of tryptophan regulation may now be found in the literature, see for example [1, 3-5] and references therein. In each of these modelling studies, the dynamics of the proposed model were compared with an extremely limited set of experimental data. In addition, in many cases only qualitative agreement between the model output and experimental data was demonstrated, [3], since very little information is currently available to guide the selection of parameter values for the models. Since many of these models have been derived using diverse assumptions about the exact workings of the underlying feedback mechanisms involved, the lack of any strong validation (or invalidation) of a particular model has severely limited the degree of insight achieved into the underlying design principles of this system.

In this paper, it is claimed that if there is a properly structured model, there must be a parameter set which make the model produce responses close to the experiment data. We address this issue by showing how global optimisation methods may be used to validate Systems Biology models against multiple experiment data. We focus on one particular model of the tryptophan control system, [3], which includes regulation of the trp operon by feedback loops representing repression, feedback inhibition, and transcriptional attenuation [3]. The model also incorporates the effect of tryptophan transport from the growth medium as well as the various time delays involved in the transcription and translation processes. We use global optimisation methods to investigate whether, for the proposed model structure, realistic (i.e. biologically plausible) parameter values can be found so that the model reproduces the dynamic response of the *in vitro* system to a number of different sets of experimental data. The experimental data is extracted from [6].

2. MATHEMATICAL MODEL FOR REGULATION OF TRYPTOPHAN OPERON

2.1 Mathematical model

The mathematical model of the tryptophan control system considered in this study is taken from [3], and consists of the set of nonlinear differential equations (6). In Equation 6, R is total repressor concentration, O is total operon concentration, P is mRNA polymerase (mRNAP) concentration, $O_F(t)$ is free operon concentration, $M_F(t)$ is free mRNA concentration, $E(t)$ is total enzyme concentration, $T(t)$ is tryptophan concentration, K_r is the repression equilibrium constant, K_t is the the rate equilibrium constant between the total repressor and the active repressor, μ is the growth rate, k_p is the DNA-mRNAP isomerisation rate, b and c are the constants for the transcriptional attenuation, k_ρ is the mRNA-ribosome isomerisation rate, ρ is the ribosomal concentration, k_d is the mRNA destroying rate, D is the mRNA destroying enzyme, γ is the enzymatic degradation rate constant, K is the tryptophan production rate, which is proportional to the active enzyme concentration, K_i is the equilibrium constant for the *Trp* feedback inhibition of anthranilate synthase reaction, which is modelled by a Hill equation with the coefficient, n_H, g is the maximum tryptophan consumption rate, the internal tryptophan consumption is modelled by a Michaelis-Menten type with the constant K_g, T_{ext} is the external tryptophan uptake, d, e, and f are the parameters for modelling the external tryptophan uptake rate, τ_p is the time taking that mRNAP binding the DNA moves away and frees the operon, τ_m is the time taking that actually mRNA produces after mRNAP

PR0002/09/0000-0209 ¥400

$$\frac{dO_F(t)}{dt} = \frac{K_r}{K_r + \dfrac{T(t)}{T(t)+K_t}R}\left\{\mu O - k_p P\left[O_F(t) - O_F(t-\tau_p)e^{-\mu\tau_p}\right]\right\} - \mu O_F(t) \tag{6a}$$

$$\begin{aligned}\frac{dM_F(t)}{dt} &= k_p P O_F(t-\tau_m)e^{-\mu\tau_m}\left[1 - b\left(1 - e^{T(t)/c}\right)\right] \\ &\quad - k_\rho \rho\left[M_F(t) - M_F(t-\tau_\rho)e^{-\mu\tau_\rho}\right] - \left(k_d D + \mu\right) M_F(t)\end{aligned} \tag{6b}$$

$$\frac{dE(t)}{dt} = \frac{1}{2}k_\rho \rho M_F(t-\tau_e)e^{-\mu\tau_e} - (\gamma+\mu)\,E(t) \tag{6c}$$

$$\frac{dT(t)}{dt} = K\frac{K_i^{n_H}}{K_i^{n_H} + T^{n_H}(t)}E(t) - g\frac{T(t)}{T(t)+K_g} + d\frac{T_{\text{ext}}}{e + T_{\text{ext}}\left[1 + T(t)/f\right]} - \mu T(t). \tag{6d}$$

bound to the DNA, τ_ρ is the time taking that ribosome binds to mRNA and initiates translation, and τ_e is another ribosome binding rate delay for the enzyme. More detail about the model can be found in [3].

All 25 independent parameters are given in Table 1 and the dependent parameters are calculated as follows: $\bar{T} = K_i$, $k_\rho = 1/(\rho\tau_\rho)$, $kdD = \rho k_\rho/30$, $K_g = \bar{T}/20$, $g = T_{\text{cr}}(\bar{T} + K_g)/\bar{T}$, $\bar{E}_A = \bar{E}K_i^{n_H}/(K_i^{n_H} + \bar{T}^{n_H})$, $\bar{G} = g\bar{T}/(\bar{T} + K_g)$, and $K = (\bar{G} + \mu\bar{T})/\bar{E}_A$, where $\bar{T}$ and $\bar{E}$ are the steady state of tryptophan and enzyme, respectively, and T_{cr} is the tryptophan consumption rate.

2.2 Model structure: feedback loops

This subsection describes the important regulatory mechanism in regulatory network in the tryptophan operon and shows the mathematical model under consideration includes these mechanism properly. As previously mentioned, there are three feedback mechanisms in the regulatory network: repression, feedback inhibition, and transcriptional attenuation [3]. Repression means that when the produced tryptophan bounds to the repressor, the repressor becomes active which prevents the mRNA polymerase from bounding to Operon. The term (active repressor:=)$\dfrac{T(t)}{T(t)+K_t}R$ belonging to Equation 6a indicates the repression. Feedback inhibition implies that the anthranilate synthase is feedback inhibited by the produced tryptophan. The term (active anthranilate:=)$\dfrac{K_i^{n_H}}{K_i^{n_H} + \bar{T}^{n_H}}\bar{E}$ in Equation 6d denotes the feedback inhibition. Transcriptional attenuation means that the produced tryptophan makes transcription abort. Equation 6b represents the transcriptional attenuation. We can see that the mathematical model in Equation 6 has a proper structure in the sense that the model includes most important regulatory network. In order to build a useful model for the network, the next step is to determine biochemically plausible parameters with which the model response is close to the experiment data. In [3], they identified the nominal parameters based on various biological reasonings and approximations. In Figure 1, the dash-dot lines denotes the model responses according to three different experimental setups. As shown in the figure, the steady state values are close to the experimental data but there are large discrepancies in the transient period, which means that the model with the nominal parameters taken from [3] is not good enough. In the next section, we present a parameter identification method using global optimization and multiple experiment data.

3. MODEL VALIDATION USING MULTIPLE EXPERIMENT DATA SET

3.1 Multiple experiment data

The experimental data is extracted from [6], which reports the results of a number of experiments with wild and mutant strains of the *E. coli* CY15000 strain. These experiments consisted of growing bacteria in various media which included tryptophan until the culture reached a steady state. Then the bacteria were washed and put into the same media without tryptophan. The response of the activity of the enzyme anthranilate synthase (the key enzyme in tryptophan biosynthesis) to these nutritional shifts was then measured as a function of time. The activity is proportional to the production rate of tryptophan.

3.2 Parameter identification

Three sets of experimental data are used in this paper and the optimal parameter set is found by minimising the square sum of errors between the dynamics of the active enzyme concentration produced by the model and the experimental data as follows

$$\min_p J = \sum_{j=1}^{N} \left|\bar{x}(t_j) - \tilde{\bar{x}}_i(t_j)\right|^2 \tag{1}$$

where p denotes the parameters in the model, the subscript i the index of the experiment, $\tilde{\bar{x}}_i(t_j)$ measurement at time t_j, and $\bar{x}_i(t_j)$ model response at time t_j. This nonlinear and nonconvex optimisation problem is solved using a hybrid Genetic Algorithm based on the one developed by the authors in [2] for each of the three different sets of experimental data. Results of our model validation are shown in Figure 1. As can be seen from the figure, while the original model shows quite a poor agreement with the data, the optimised model is able to almost exactly reproduce the responses of the *in vitro* system for each different experiment. Importantly, the optimal model parameters are all within biologically plausi-

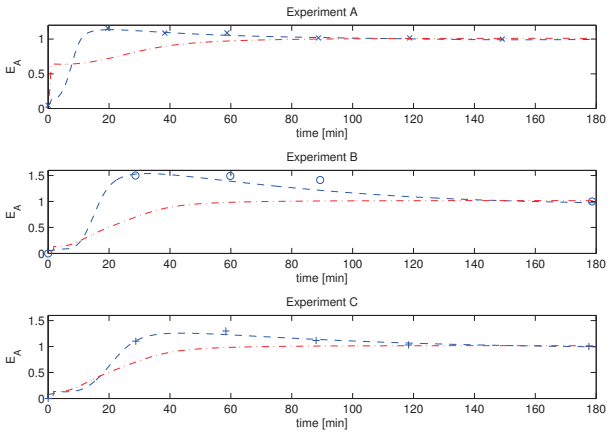

Fig. 1 Optimised (dashed line) versus original (dash-dot line) model responses for data from experiment A (x), experiment B (o) and experiment C (+), experimental data taken from [6]

.

ble ranges. See the rightmost column in Table 1 (we presented the resulting optimal parameter only for the case where the experiment data A was used). Thus the proposed method represents a strong validation of the proposed model of the tryptophan control system. Consequently, this supports the claim that the global optimization based method can search at least one parameter set which makes the model produce responses very close to the experiment data as long as the assumed model structure is proper.

4. SUMMARY AND OUTLOOK

This paper presents a global optimization based parameter identification method using multiple experiment data. Thanks to the global optimization, it was shown that at least one plausible parameter set can be found with a properly structured model.

As mentioned, the structure of the model is assumed (i.e. fixed) but the model shows robust performance against different experimental setups, i.e. environmental changes. This observation implies that some parameters need to change for the purpose of providing the model with robust functioning. This issue is one of our future research topic along this line.

REFERENCES

[1] Sharad Bhartiya, Subodh Rawool, and K. V. Venkatesh. Dynamic model of *escherichia coli* tryptophan operon shows an optimal structural design. *Journal of European Biochemistry*, 270(12):2644–2651, June 2003.

[2] Prathyush P. Menon, Jongrae Kim, Declan G. Bates, and Ian Postlethwaite. Clearance of nonlinear flight control laws using hybrid evolutionary optimisation. *IEEE Transactions on Evolutionary Computation*, 10(6):689–699, December 2006.

[3] Moisés Santillán and Michael C. Mackey. Dynamic reguiation of the tryptophan operon: A modeling study and comparison with experimental data. *Proceedings of the National Academy of Sciences*, 98(4):1364–1369, February 2001.

[4] Moisés Santillán and Eduardo S. Zeron. Analytical study of the multiplicity of regulatory mechanisms in the tryptophan operon. *Bulletin of Mathematical Biology*, 68(2):343–359, February 2006.

[5] Zhi-Long Xiu, Zeng-Yi Chang, and An-Ping Zeng. Nonlinear dynamics of regulation of bacterial *trp* operon: Model analysis of integrated effects of repression, feedback inhibition, and attenuation. *Biotechnology Progress*, 18(4):686–693, August 2002.

[6] Charles Yanofsky and Virginia Horn. Role of regulatory features of the *trp* operon of *Escherichia coli* in mediating a response to a nutritional shift. *Journal of Bacteriology*, 176(20):6245–6254, October 1994.

5. APPENDIX

Table 1 Nominal parameter values

	Unit	Original in [3]	Optimal parameters
d	[·]	23.5	23.5
e	[·]	0.9	0.9
f	[·]	380	380
R	[μM]	0.8	1.357
O	[μM]	0.0033	0.0059
P	[μM]	2.6	3.22
$\bar{E}$	[μM]	0.378	0.338
T_{cr}	[μM/min]	22.7	14.07
K_r	[μM]	0.0026	0.0015
K_t	[μM]	60.34	64.55
K_i	[μM]	4.09	6.93
n_H	[·]	1.2	1.00
b	[·]	0.85	0.53
c	[·]	0.04	0.0083
ρ	[μM]	2.9	3.33
γ	[1/min]	0.0	0.0113
μ	[1/min]	0.01	0.0264
τ_p	[min]	0.1	0.0267
τ_m	[min]	0.1	0.0277
τ_ρ	[min]	0.05	0.0874
τ_e	[min]	0.66	1.1284
$O_F(0)$	[μM]	4.8765×10^{-5}	7.6444×10^{-5}
$M_F(0)$	[μM]	1.2037×10^{-4}	0.4304×10^{-4}
$E(0)$	[μM]	0.0119	0.0238
$T(0)$	[μM]	16.571	13.962