Analyses of metabolic system dynamics for time series data of small samples

DAISUKE TOMINAGA^{1,a)} HIDEO KAWAGUCHI² YOSHIMI HORI² TOMOHISA HASUNUMA³ CHIAKI OGINO⁴ SACHIYO ABURATANI¹

Abstract: Measuring concentrations of metabolites and estimating reaction rates of each reaction step consisting of metabolic pathways are significant for the improvement of microorganisms such as production maximization of materials. Although the reaction pathway must be identified for improvement, that is not easy. Numerous reaction steps have been reported, but the actually activated reaction steps vary or change according to conditions. Furthermore, reaction mechanisms and parameter values must be known to build mathematical models for dynamical analysis, but sufficient information has not been published for many cases to date. In addition, experimental observations are expensive. A new mathematical approach that is applicable to small sample data and which requires no detailed information of reactions is strongly needed. The S-system is one such model that can use smaller samples than other ODE models can. We propose a simplified S-system to apply minimal quantities of samples for dynamic analysis of metabolic pathways. We applied the model to the phenyl-lactate production pathway of *Escherichia coli*. The obtained model suggests that actually activated reaction steps and feedback inhibitions in the pathway.

Keywords: Metabolic pathway, Pathway dynamics, Small sample, S-system, Michaelis-Menten

1. Introduction

1.1 Mathematical model for metabolic pathways

Artificial and industrial uses of microorganisms for material production have a long history of over a thousand years. Recently, genetic operations are applied widely to improve production. Two generally considered approaches introduce enzymes that have higher activities from other organisms or species and introduce enzymes to realize metabolic pathways that the microorganisms do not have naturally. The former is a popular method because the operation is simpler and improvements are more predictable than the latter one. Conventional gene modifications using ultraviolet or other radiation are easy. They have been widely applied in many industries. Nevertheless, the efficiency of the improvements is quite low because gene modifications occur accidentally and uncontrollably. Progress is made serendipitously. Therefore, currently, gene introduction is used in addition to conventional means.

Target genes for modification are chosen based on information including the reaction rates of respective reaction steps in the metabolic pathway, and include production materials and substrates of microorganisms, as well as changes in reaction rates by changes in concentrations of metabolites that consist of the pathway. Bottleneck reaction steps and feedback loop inhibitions are suggested by this information. Genes of enzymes of such reactions are candidates for modification.

Rates of enzymic reactions are generally defined as the limit of changes of compounds over time[5], [6]. Several formulae are established for types of enzymic reactions, such as inhibition schemes. The most popular is the Michaelis-Menten law[1], [2] for a simple one-to-one enzymic reaction without inhibition and catalysis by an enzyme. The reaction rate is modeled based on the following reaction scheme:

$$S + E \rightleftharpoons ES \to P + E \tag{1}$$

in which S stands for the substrate of the reaction, E signifies the enzyme, and P denotes the product. The bi-directional arrow represents a reversible reaction. The one-directional arrow signifies a one-way reaction. The reaction rate is modeled as the following ordinary differential equation (ODE).

$$\frac{d[P]}{dt} = \frac{V_{max}[S]}{K_m + [S]} \tag{2}$$

Therein, a pair of square brackets denotes concentration of the compound, *t* is time, V_{max} and K_m are parameters that define the kinetic character of the enzyme. All reaction steps in a metabolic pathway can be represented by the ODE above if all reactions are simple enzymic reactions and if parameter values are defined. Then the pathway is modeled as simultaneous ODEs, or the ODE system. Consequently, simulations can be done of concentration changes of metabolites, stability analysis, steady state estimation,

¹ Computational Bio Big-Data Open Innovation Laboratory (CBBD-OIL), National Institute of Advanced Industrial Science and Technology (AIST), 3-4-1 Okubo, Shinjuku, Tokyo 169-8555, Japan

² Graduate School of Science, Technology and Innovation, Kobe University, 1-1 Rokkodai, Nada, Kobe, Hyogo 657-8501, Japan

³ Organization of Advanced Science and Technology, Kobe University, 1-1 Rokkodai, Nada, Kobe, Hyogo 657-8501, Japan

⁴ Graduate School of Engineering, Kobe University, 1-1 Rokkodai, Nada, Kobe, Hyogo 657-8501, Japan

a) tominaga-daisuke@aist.go.jp

and bottleneck finding[12], [15].

Generally, finding parameter values is difficult and expensive because it requires enzyme isolation and measurements of reaction rates in test tubes (in vitro measurement). Although enzyme information in literatures and public databases are growing[11], these kinetic parameter values are not published or accumulated sufficiently. An enzyme generally has different values of parameters for conditions and species of organisms. Moreover, parameter values differ generally between in vitro and in vivo (in living cells) conditions[3], [4]. For most industrial applications, dynamical analysis of pathways must be done without reaction scheme information or kinetic parameter values.

1.2 Canonical ODE model

ODE systems in canonical forms are applicable because they are independent of the molecular mechanism the reaction scheme. The S-system[15] is one such canonical ODE model. For a reaction scheme with two reactions as

$$X_1 + X_2 + \ldots + X_l \to P \to X_{l+1} + X_{l+2} + \ldots + X_m$$
 (3)

where $X_1, X_2 \cdots$, and X_l and P respectively denote the substrates and product of the first reaction, and P and $X_{l+1}, \cdots X_m$ respectively denote the substrate and products of the second reaction, the form of the S-system is represented as

$$\frac{d[P]}{dt} = \alpha \prod_{j=1}^{n} [X_j]^{g_j} - \beta \prod_{j=1}^{n} [X_j]^{h_j}$$
(4)

where X_j stands for the concentration of the metabolite j, g_j signifies the kinetic parameter that represents the influence of X_j to the increasing processes of P, h_j denotes the influence of X_j to the decreasing processes of P, and α and β represent rate constants. The first term of the left hand of the equation represents the total rate of increasing or composing processes of P. The second term is the total rate of decreasing or decomposing processes. l and m in the reaction (3) are numbers of composing and decomposing processes of P respectively. All l and m compounds are suffixed sequentially in eq. (4).

An S-system above is a simplified form of the general mass action law[15], which can be presented as follows.

$$\frac{d[P]}{dt} = \sum_{i=1}^{m} \alpha_i \prod_{j=1}^{n} [X_{ij}]^{g_{ij}}$$
(5)

Therein the *i* suffix denotes each composing and decomposing reaction of *P*. Simplification for the S-system is summarizing composing reactions of *P* into a term with α and decomposing reactions into β term in eq. (4). Parameters g_j and h_j respectively correspond to reaction orders in the mass action law, g_{ij} , and represent dependencies between the metabolites *P* and X_j . Consequently, they represent the network scheme of the reaction pathway. No direct dependence exists between *P* and X_j when g_j and h_j in eq. (4) are equal to zero. X_j suppresses production of *P* when g_j is negative.

Parameter values of g_j , h_j , α , and β can be estimated using numerical optimization methods that find parameter values by which the calculated time series of *P* by numerical integration of eq. (4) matches the observed time series of the concentration of P. Found values of g_j and h_j might be considered that represent orders of each reaction between X_j and P. However, parameter optimization is the inverse problem[14] because several sets of different parameter values are found generally for the given observed time series data. Restrictions and limitations are effective for difficulties such as fixing some g_j to zero based on biological knowledge.

1.3 Method for small sample

Numerical optimizations require a sufficient number of observed samples. Smaller needs are better because observations entail costs. A mathematical model with fewer parameters requires fewer samples. We propose a canonical ODE model for small samples by simplification of the second term of eq. (4), as shown below.

$$\frac{d[P]}{dt} = \alpha \prod_{j=1}^{n} [X_j]^{g_j} - \beta [P]^h.$$
 (6)

The decomposition rate of a compound depends only on its concentration in many biological processes like Michaelis-Menten type reactions shown in eq. (2). We introduce this idea as an assumption in eq. (6). Although the decomposition reactions are often modeled as linear ODE like

$$\frac{d[P]}{dt} = A \exp(-[P]t).$$

our model includes a nonlinear decomposition term because we suspect that the linear term might be too simple for the metabolite in the complex biological network system that contains many unkown reactions.

Resulting time series of the model by numerical integration varies greatly by change of the initial value. Finding the best initial value is difficult because the observed initial value often has errors, especially for small sample datasets. Therefore, we compare the model and data in differential spaces. One can calculate the differential of observed data by numerical differentiation and evaluate parameter values by comparing the differential values and values of eq. (6).

We evaluated the proposed method according to its application to phenyl lactate production pathway from glucose by *Escherichia coli*. For the pathway includes branches and feedback loops, we estimated the actually activated reaction steps and activities of feedback inhibitions that suggest strategies for production improvement.

2. Method

First we built a pathway map based on information from the literature and databases. Then we chose some metabolites in the map as observation targets. The pathway map is reconstructed with target metabolites only. Observations are measurements of the target metabolite concentration at sampling time points with equal intervals. The simplified S-system model is defined based on the reconstructed pathway map by fixing some g_j to zero if the link to *P* from X_j does not exist in the reconstructed pathway map.

Observed time series data of the concentration of metabolites



Fig. 1 Main pathway of phenyl lactate metabolism of *Escherichia coli*. Dashed line means that the path consists of plural reactions. Solid one-directional line indicate a single enzymic reaction. Bi-directional line indicate two reactions; forward and backword processes catalyzed by same or different enzymes. Double lined circle indicate a selected target metabolite for observation.



Fig. 2 Reconstructed pathway consists of observation target metabolites.



Fig. 3 Observed time series of concentrations of target metabolites on six time points. SA, PP, PA and PL are for Shikimic acid, Phenyl pyruvate, Phenylalanine and Phenyl lactate respectively.

are numerically differentiated. Optimal values of g_j , h, α , and β are sought using a nonlinear numerical optimization method such as the genetic algorithm[14] or the differential evolution[13] to minimize the difference between differentials of the observed data by numerical differentiation and those of by eq. (6). Optimal g_j are considered. They represent activity of the reaction from X_j to *P*. Building the formula of eq. (6) and optimization of parameters are done for each target metabolite. We chose the phenyl lactate production metabolic pathway[7], [8], [9], [10] (**Fig. 1**) as the application target and chose six metabolites for the observation target. Then, we reconstructed the pathway only with the observation target metabolites (**Fig. 2**). The number of sampling time points is set to 6. Changes in concentration of metabolites are observed in the log phase of cell growth (**Fig. 3**).

2.1 Pathway map construction and observation

The pathway from glucose to phenyl lactate consists roughly of the glycolysis and the shikimic acid pathway. Phenyl lactate is from phenyl pyruvate. Phenyl pyruvate is from phenylalanine. Phenylalanine is from prephenic acid. Also, prephenic acid is from chorismic acid in the shikimic acid pathway. We chose phosphoenolpyruvate (PEP), erythrose 4-phosphate (E4P), shikimic acid (SA), phenylalanine (PA), phenyl pyruvate (PP), and phenyl lactate (PL) as the observation targets. Then we reconstructed the pathway for these six metabolites (Fig. 2). The pathway includes a branching point and two feedback loops.

2.2 ODE models for respective metabolite

Changes of concentration of metabolites in the reconstructed pathway are modeled mathematically using the following ODE models.

$$\frac{d[SA]}{dt} = \alpha_{SA}[PEP]^{g_{SA,PEP}}[E4P]^{g_{SA,E4P}} - \beta_{SA}[SA]^{h_{SA}}$$

$$\frac{d[PP]}{dt} = \alpha_{PP}[SA]^{g_{PP,SA}}[PA]^{g_{PP,PA}}[PL]^{g_{PP,PL}} - \beta_{PP}[PP]^{h_{PP}}$$

$$\frac{d[PA]}{dt} = \alpha_{PA}[SA]^{g_{PL,SA}}[PP]^{g_{PA,PP}} - \beta_{PA}[PA]^{h_{PA}}$$

$$\frac{d[PL]}{dt} = \alpha_{PL}[PP]^{g_{PL,PP}} - \beta_{PL}[PL]^{h_{PL}}$$
(7)

Actually, PEP and E4P are not controlled by any other metabolite. These are independent variables in the ODE system above. Their respective dynamics are not modeled. Parameter $g_{PP,SA}$ in the ODE system represents the summarized actual activity of the reaction chain to phenyl pyruvate from shikimic acid, consisting of several reaction steps. $g_{a,b}$ represents the rate of concentration change of *a* is raised by *b* when the sign of $g_{a,b}$ is positive. A negative value of $g_{a,b}$ means that *b* suppresses the composition or consumption of the metabolite. Negative *h* means that the metabolite suppresses decomposition itself. Therein, the α and β are fixed rate coefficients. For each metabolite, activities of the respective reaction step of composition of the metabolite can be compared.

Time differential values of metabolite concentration are cal-



Fig. 4 Estimated actual reaction activities on the reconstructed pathway. Numerical values shown in the figure are g_j in eq. (6) that represent incoming links for each metabolite.



Fig. 5 Values of differentials calculated from observed data by quadratic interpolation and calculated from the optimized ODE model. Dashed line shows differential values that calculated from the ODE model with optimized parameters. Solid line is values of numerical differentiation from observed data. SA, PP, PA and PL are for Shikimic acid, Phenyl pyruvate, Phenylalanine and Phenyl lactate respectively.

culable with the ODE system by determining all parameter values. We introduce the differential evolution algorithm[13] to find the parameter values that minimize differences between differential of concentration values calculated using the ODE system, as shown below.

$$E = \sum_{t=1}^{T} (D_{ct} - D_{ot})$$
(8)

Therein, *E* stands for the summarized differences, D_{ct} signifies the calculated differentials by the ODE system with parameter values at time point *t*, and D_{ot} denotes the differential at time point t of observed data. D_{ot} is calculated using quadratic interpolation of the observed data. *E* is minimized by searching parameter values of *g*, *h*, α , and β for each metabolite. The optimization algorithm introduces uniformly distributed random numbers for initial values of searching parameter values.

3. Result

3.1 Model interpretation

The ODE system with found values of parameters is shown below (**Fig. 4**).

$$\frac{d[SA]}{dt} = 21.1[PEP]^{65.4}[E4P]^{33.9} - 99.2[SA]^{-0.0674}$$

$$\frac{d[PP]}{dt} = 9.38[SA]^{-9.44}[PA]^{8.77}[PL]^{-0.390} - 57.2[PP]^{-1.95}$$

$$\frac{d[PA]}{dt} = 0.000223[SA]^{2.12}[PP]^{-0.307} - 90.7[PA]^{0.207}$$

$$\frac{d[PL]}{dt} = 5.60[PP]^{1.13} - 1.51 \times 10^{-6}[PL]^{2.82}$$

The ODE system suggests that the reaction chain to shikimic acid from phosphoenolpyruvate is more active than that from erythrose 4-phosphate. Also, phenyl pyruvate is inhibited by feedback loops from both phenylalanine and phenyl lactate.

4. Discussion

4.1 Numerical Advantages of our Model

Metabolite concentration changes in time can be modeled by the Michaelis-Menten law as shown below.

$$\frac{d[SA]}{dt} = \frac{V_{SA,PEP}[PEP]}{K_{SA,PEP} + [PEP]} + \frac{V_{SA,E4P}[E4P]}{K_{SA,E4P} + [E4P]}
- \frac{V_{PP,SA}[SA]}{K_{PP,SA} + [SA]} - \frac{V_{PA,SA}[SA]}{K_{PA,SA} + [SA]}
\frac{d[PP]}{dt} = \frac{V_{PP,SA}[SA]}{K_{PP,SA} + [SA]} + \frac{V_{PP,PA}[PA]}{K_{PP,PA} + [PA]} + \frac{V_{PP,PL}[PL]}{K_{PP,PL} + [PL]}
- \frac{V_{PA,PP}[PP]}{K_{PA,PP} + [PP]} - \frac{V_{PL,PP}[PP]}{K_{PL,PP} + [PP]}
\frac{d[PA]}{dt} = \frac{V_{PA,SA}[SA]}{K_{PA,SA} + [SA]} + \frac{V_{PA,PP}[PP]}{K_{PA,PP} + [PP]} - \frac{V_{PP,PA}[PA]}{K_{PP,PA} + [PA]}
\frac{d[PL]}{dt} = \frac{V_{PL,PP}[PP]}{K_{PL,PP} + [PP]} - \frac{V_{PP,PL}[PL]}{K_{PP,PL} + [PL]}$$

Numbers of parameters for each metabolite are 8 for shikimic acid, 10 for phenyl pyruvate, 6 for phenylalanine, and 4 for phenyl lactate because a model of a reaction step (single enzymic reaction) has two parameters. Phenyl pyruvate cannot be modeled with the six sampling data in this study. The total number of parameters for the reconstructed pathway is 16 because some parameters are common (one of out-going reactions of shikimic acid is an incoming reaction of phenyl pyruvate). Some numerical optimization methods can search 16 parameters simultaneously. However, this simultaneous non-linear numerical optmization is not easy because the optimization difficulty increases explosively with the growing number of parameters ('curse of dimension').

The S-system model (in the original form, not simplified) of the reconstructed pathway is the following.

$$\frac{d[SA]}{dt} = \alpha_{SA}[PEP]^{g_{SA,PEP}}[E4P]^{g_{SA,E4P}} - \beta_{SA}[SA]^{g_{PP,SA}+g_{PA,SA}}$$

$$\frac{d[PP]}{dt} = \alpha_{PP}[SA]^{g_{PP,SA}}[PA]^{g_{PP,PA}}[PL]^{g_{PP,PL}} - \beta_{PP}[PP]^{g_{PA,PP}+g_{PL,PP}}$$

$$\frac{d[PA]}{dt} = \alpha_{PA}[SA]^{g_{PL,SA}}[PP]^{g_{PA,PP}} - \beta_{PA}[PA]^{g_{PP,PA}}$$

$$\frac{d[PL]}{dt} = \alpha_{PL}[PP]^{g_{PL,PP}} - \beta_{PL}[PL]^{g_{PP,PL}}$$

The numbers of S-system parameters for each metabolite are 6 for shikimic acid, 7 for phenyl pyruvate, 5 for phenylalanine and 4 for phenyl lactate: 16 in all because some parameters are common. Phenyl pyruvate cannot be modeled even by the S-system. In our proposed model, the corresponding numbers are 5, 6, 5, and 4: 20 in all. Total number of parameters are not small, however, it is not a problem that the total parameters were more numerous than those of the S-system because no parameters are common to any two metabolites and parameters of a metabolite are optimized independently of other metabolites. Our model has fewer parameters for each metabolite, which means that our model is more robust against error than the S-system and Michaelis-Menten models. Apparently, the optimization precision of parameters (model fitness to the data) shown in **Fig. 5** is sufficiently good, although the model and data do not match perfectly because data generally include errors that exist in probabilistic distributions. According to biochemical engineering sense, the values of some parameters shown in the Fig. 4 are large as the reaction order. perhaps because of fluxes of pathways other than the reconstructed pathway from which we omitted pathways other than the main reaction chain, although glycolysis has many branches to other sub-systems. On the other hand, parameter values for phenylalanine (PA in Fig. 4), phenyl pyruvate (PP), and phenyl lactate (PL) might more reliable because reaction steps occurring naturally around these metabolites may be almost same as those of the reconstructed pathway.

Parameter values of the α term of a metabolite (of the incoming link of a metabolite in the pathway map) are directly comparable. Phenyl pyruvate has three parameters to compare, two of which are negative and one of which is positive. Therefore, phenyl pyruvate, which is inhibited by shikimic acid and phenyl lactate (the final product), is composed mainly from phenylalanine.

Two negative feed back loops exist: phenyl pyruvate is inhibited by phenyl lactate; and phenylalanine is inhibited by phenyl pyruvate. Although it can be readily imagined that inhibition of feedback reactions raises the production of phenyl lactate, the main inhibitory effect to phenyl pyruvate is by shikimic acid. Disrupting one or some genes of reactions to phenyl pyruvate from shikimic acid might improve phenyl lactate production.

Phenylalanine decreases gradually (Fig. 3), but it shows no natural decomposition, represented as $A \exp(-Bt)$, that might be caused by the incoming link from phenyl pyruvate.

In conclusion, results show that the reliability of the estimated parameter values might not the best or even very high because the reconstructed pathway and the ODE system are simplified. These values suggest that target genes can be modified for industrial improvement of production by microorganisms. This case study presents several suggestions that are useful when constrained by few samples or low observation costs.

Acknowledgments Stock of the microorganism that is used in this study is provided by professor Naoki Takaya at Graduate School of Life and Environmental Science, Tsukuba University.

This work was supported by New energy and Industrial Technology Development Organization, Japan.

References

- Bisswanger, H.: Enzyme Kinetics, Principles and Methods (2nd ed.), Wiley-VCH, Weinheim, Germany (2008).
- [2] Cleland, W. W.: Enzyme kinetics, Ann. Rev. Biochem., Vol.36, pp.77-112 (1967).
- [3] Davidi, D., Noor, E., Liebermeister, W., Bar-Even., A., Flamholz, A., Tummler, K., Barenholz, U., Goldenfelda, M., Shlomi, T. and Milo, R.: Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro k_{cat} measurements, *Proc. Natl. Acad. Sci.* USA, Vol.113, pp.3401–3406 (2016).
- [4] Eunen, K. v., Kiewiet, J. A. L., Westerhoff, H. V., Bakker, B. M.: Testing Biochemistry Revisited: How *In Vivo* Metabolism Can Be Understood from *In Vitro* Enzyme Kinetics, *PLoS Comput. Biol.*, Vol. 8, e1002483 (2012).
- [5] Gavalas, G. R.: Nonlinear Differential Equation of Chemically Reacting Systems. Springer-Verlag, Berlin, Germany (1968).

tems. Structure, control, and optimality. *BioSystems*, Vol.47, pp.61-77 (1998).

- [7] Kawaguchi, H., Teramura, H., Uematsu, K., Hara, Y. K., Hasunuma, T., Hirano, K., Sazuka, T., Kitano, H., Tsuge, Y., Kahar, P., Niimi-Nakamura, S., Oinuma, K-I., Takaya, N., Kasuga, S., Ogino, C. and Kondo, A.: Phenyllactic acid production by simultaneous saccharification and fermentation of pretreated sorghum bagasse, *Biores. Technol.*, Vol.182, pp.169–178 (2015).
- [8] Kawaguchi, H., Uematsu, K., Ogino, C., Teramura, H., Niimi-Nakamura, S., Tsuge, Y., Hasunuma, T., Oinuma, K., Takaya, N. and Kondo, A.: Simultaneous saccharification and fermentation of kraft pulp by recombinant Escherichia coli for phenyllactic acid production, *Biochem. Eng. J.*, Vol.88, pp.188–194 (2014).
- [9] Koma, D., Yamanaka, H., Moriyoshi, K., Ohmoto, T. and Sakai, K.: Production of Aromatic Compounds by Metabolically Engineered *Escherichia coli* with an Expanded Shikimate Pathway, *App. Envoron. Microbiol.*, Vol.78, No.17, pp.6203–6216 (2012).
- [10] Liu, S-P., Xiao, M-R., Zhang, L., Xu, J., Ding, Z-Y., Gu, Z-H. and Shi G-Y.: Production of l-phenylalanine from glucose by metabolic engineering of wild type Escherichia coli W3110, *Process Biochem.*, Vol.48, pp.413–419 (2013).
- [11] Placzek, S., Schomburg, I., Chang, A., Jeske, L., Ulbrich, M., Tillack, J. and Schomburg, D.: BRENDA in 2017: new perspectives and new tools in BRENDA, *Nucleic Acids Res.*, Vol.45(D1), pp.D380–D388 (2017).
- [12] Savageau, M. A.: Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology, Addison-Wesley Educational Publishers, Boston, U. S. A.(1977).
- [13] Storn, R. and Price, K.V.: Differential Evolution -a simple and efficient heuristic for global optimization over continuous spaces-, J. Global Optim., Vol.11, pp.341–359 (1997).
- [14] Tominaga, D. and Horton, P.: Inference of Scale-free Networks From Gene Expression Time Series, J. Bioinfor. Comp. Biol., Vol.4, No.2, pp.503–514 (2006).
- [15] Voit, E.O.: Computational Analysis of Biochemical Systems, Cambridge University Press, Cambridge, UK (2000).