

# BIOINFO 2005

Proceedings of the 2005 International Joint  
Conference of InCoB, AASBi and KSBI

September 22-24, 2005 Busan, KOREA

<http://www.ksbi.org/bioinfo2005>



## Editors:

Doheon Lee  
Limsoon Wong  
Sangsoo Kim  
Dae-Won Kim  
Cheolmin Kim  
Tin Wee Tan  
Kwang-Hyung Lee



## Hosted by

Association of Asian Societies for Bioinformatics  
Asia Pacific Bioinformatics Network  
Korean Society for Bioinformatics

## Organized by

Electronics and Telecommunications Research Institute  
Korea Basic Science Institute  
Korea Institute of Science and Technology Information  
Korea National Institute of Health  
Korea Research Institute of Bioscience and Biotechnology  
National Institute of Agricultural Biotechnology  
National Livestock Research Institute  
National Core Research Center for Systems Bio-Dynamics  
Busan Genome Center  
Chungbuk BIT Research-Oriented University Consortium

*Shigeo Orie*

# BIOINFO 2005

**Proceedings of the 2005 International Joint  
Conference of InCoB, AASBi and KSBI**

September 22-24, 2005, Busan, KOREA

<http://www.ksbi.org/bioinfo2005>

## **Hosted by**

Association of Asian Societies for Bioinformatics,  
Asia Pacific Bioinformatics Network,  
Korean Society for Bioinformatics

## **Organized by**

Electronics and Telecommunications Research Institute  
Korea Basic Science Institute  
Korea Institute of Science and Technology Information  
Korean National Institute of Health  
Korea Research Institute of Bioscience and Biotechnology  
National Institute of Agricultural Biotechnology  
National Livestock Research Institute  
National Core Research Center for Systems Bio-Dynamics  
Busan Genome Center  
Chungbuk BIT Research-Oriented University Consortium

## **Sponsored by**

Apple Computer Korea  
Mirae Corporation  
IDR Tech. Inc.  
CT & D  
Silicon Graphics Inc.  
ISTECH Inc.  
Macrogen Inc.  
Busan Convention & Visitors Bureau  
Korea National Tourism Organization

# Symbolic-numeric Estimation of Parameters in Biochemical Models by Quantifier Elimination

Shigeo Orii<sup>1\*</sup> Hirokazu Anai<sup>2</sup> Katsuhisa Horimoto<sup>3\*</sup>

<sup>1</sup>Science Solutions Group, FUJITSU LTD., Chiba, Japan

<sup>2</sup>IT Core Laboratories, FUJITSU LABORATORIES LTD./JST, CREST, Kawasaki, Japan

<sup>3</sup>Laboratory of Biostatistics, Institute of Medical Science, University of Tokyo, Tokyo, Japan

Email : orii@strad.ssg.fujitsu.com, anai@jp.fujitsu.com, khorimoto@ims.u-tokyo.ac.jp

**ABSTRACT:** We introduce a new approach to optimize the parameters in biological kinetic models by quantifier elimination (QE), in combination with numerical simulation methods. The optimization method was applied to a model for the inhibition kinetics of HIV proteinase with ten parameters and nine variables, and attained the goodness of fit to 300 points of observed data with the same magnitude as that obtained by the previous optimization methods, remarkably by using only one or two points of data. Furthermore, the utilization of QE demonstrated the feasibility of the present method for elucidating the behavior of the parameters in the analyzed model. The present symbolic-numeric method is therefore a powerful approach to reveal the fundamental mechanisms of kinetic models, in addition to being a computational engine.

## 1 INTRODUCTION

The availability of genomic sequence information is facilitating the fundamental understanding of biological phenomena on the molecular level. By utilizing complete genomic sequences, for example, the expression of many genes can be simultaneously monitored to reveal the regulatory network on a genomic scale [1]. As a further challenging investigation, systems biology is aiming at modeling and simulating the global network of biological molecules in a cell [2]. Interestingly, the investigations in these studies share one procedure in the numerical analyses; in the kinetic model constructed for the corresponding biological phenomena, the kinetic parameters are estimated by various optimization methods, based on the observed data. Subsequently, the parameter optimizations emerge as mathematical subjects in a wide variety of issues in molecular biology.

Many methods for local and global optimization have been developed [3, 4], and some simulators based on various optimization methods have also been designed (e.g. [5]). In the optimization methods, the estimation of kinetic parameters plays a key role in the development of kinetic models, which, in turn, promotes functional understanding at the system level, for example, in several biological pathways [6, 7]. In addition to the development of optimization methodology, the high performance of computers for numerical calculations also supports the optimization of the kinetic parameters in the complex dynamics within a reasonable amount of computational time.

The high computer performance supports not only the numerical calculations based on calculus, but also the symbolic computations based on computer algebra (CA). Indeed, symbolic computation is popular in software platforms such as Maple [8] and Mathematica [9], and the use of symbolic computation is increasing rapidly in biology [10]. The quantifier eliminator (QE) is one of the subjects in CA [11]. The QE was originally described in the mathematical proof by Tarski [12], which stated that the elementary theory for reals is decidable. Although the original algorithm was too complex to solve the problem, a new method by Collins [13], cylindrical algebraic decomposition (CAD), allowed efficient implementation. Subsequent improvements in the algorithm provided feasible software platforms for symbolic computation (e.g., [14]). Although some applications of QE to biological issues by symbolic computation were reported [15], an amalgam of symbolic computation by QE and numerical calculation has not been designed.

In this paper, we developed a novel optimization method in combination with symbolic computation and numerical simulation. A procedure for parameter optimization was designed to solve differential equations by QE in combination with numerical simulation. The performance of our procedure is illustrated by optimizing ten parameters between nine variables in a model for the inhibition kinetics of HIV proteinase [16]. As for the optimization performance, the goodness of fit to the observed data and the optimized parameters are compared with those from the previous studies [16, 3]. Furthermore, some characteristics of the symbolic-numeric method are discussed with the behaviors of the parameters in the model.

## 2 MATERIALS AND METHODS

### 2.1 Main Tool

#### 2.1.1 Quantifier elimination

Our key tool for realizing symbolic-numeric optimization is an algebraic method, quantifier elimination (QE) over the reals [11]. QE deals with first-order formulas, which consist of polynomial equations, inequalities, quantifiers ( $\exists, \forall$ ) and Boolean operators ( $\wedge, \vee, \Rightarrow, \neg$  etc). The QE procedure is an algorithm to compute an equivalent quantifier-free formula for a given first-order formula over the real closed field. For example, for the input  $\forall x(x^2 + bx + c > 0)$ , QE outputs the

\* To whom correspondence should be addressed.

equivalent quantifier-free formula  $b^2 - 4c < 0$ . This implies that we can obtain a condition for unquantified free variables that makes the input formula true by QE. Moreover, if all of the variables are quantified, then QE decides whether the formula is true or false.

### 2.1.2 QE for uncertain input data

When we find feasible model parameters according to the observed data, we can solve some constraints derived by substituting observed data for the corresponding variables/parameters in the original constraints. Such constraints have some uncertainty, due to the inexact input data. Hence, if we apply QE directly to the constraints, then there is a real danger of arriving at an incorrect answer. Actually, we often obtain a "false" result for feasible cases. In order to extract the nontrivial information of feasible parameters, even for the incorrect cases, we propose the introduction of new variables into the constraints (see 2.2). We call them "error variables", which play a role in absorbing the uncertainty due to inaccurate input data. If we apply QE for the constraints including error variables, then we obtain possible ranges of error variables, so that the constraints are feasible. Then we obtain feasible regions of the model parameters by applying QE again to the constraints, where the error variables are substituted with the minimum value of their feasible ranges.

## 2.2 Mathematical Framework

### 2.2.1 Problem

In this paper, we consider the following fitting problem: the biological kinetic model analyzed here is of the form:

$$\dot{x}_i = v_i(\mathbf{X}, \mathbf{K}) \quad (1)$$

where  $\mathbf{X} = \{x_1, \dots, x_{n_x}\}$  is a set of variables, and

$\mathbf{K} = \{k_1, \dots, k_{n_p}\}$  is a set of parameters. The problem is to fit the parameters  $\mathbf{K}$  of the model to the observed data  $\hat{\mathbf{X}} = \{\hat{x}_i^t\}$  for  $i = 1, \dots, n_x$ ,  $t = 0, 1, \dots, n_t$ , under the following additional conditions:

- (i) Conservation laws:  $h_t(\mathbf{X}) = 0$ ,
- (ii) Variable ranges:  $x_i \in D_i$ , where  $D_i = [a, b]$ ,  $a, b \in \mathbf{R} \cup \{\infty\}$ .

### 2.2.2 Basic Formula

Here we set up the leading formula of this paper. As mentioned above, we have the following constraints  $F$  with error variables  $e_i$  from kinetic models:

$$\Psi = \bigwedge_i \psi_i,$$

where

$$\psi_i = \hat{x}_i - v_i(\mathbf{X}, \mathbf{K}) + e_i = \theta.$$

For the error variables we introduce a new variable,  $e_{\max}$ , which means the magnitude of the error variables:  $|e_i| \leq e_{\max}$ . Moreover, for the variables whose observed data is given, we consider the following objective conditions:

$$\hat{x}_j^{(i)} - x_j^{(i)} = 0,$$

to achieve fitting. Then the "basic formula" is given as

$$F(\hat{\mathbf{X}}, \mathbf{X}, \mathbf{K}, e_{\max}, e_i) = (\Psi \wedge h_t(\mathbf{X}) = 0 \wedge x_i \in D_i \wedge |e_i| \leq e_{\max} \wedge \hat{x}_j^{(i)} - x_j^{(i)} = 0). \quad (2)$$

We apply our symbolic-numeric approach to formulas derived by slightly modifying the basic formula according to various purposes.

## 2.3 Optimization Procedure

We explain the concrete procedure of symbolic-numeric optimization, which consists of six parts as illustrated in Figure 1.

### Part (1): Numerical simulation

First we prepare simulation data for  $\hat{\mathbf{X}}$  and  $x_i$ , for which we lack observed data, by performing a numerical simulation of the kinetic models.

1. Set initial conditions  $\tilde{\mathbf{X}}^{(0)}$  and starting values for unknown parameters  $\tilde{\mathbf{K}}^{(0)}$  as follows:

$$\tilde{\mathbf{X}}^{(0)} = \{\tilde{x}_i^{(0)} \mid i = 1, \dots, n_x\} \text{ and } \tilde{\mathbf{K}}^{(0)} = \mathbf{K}_1^{(0)} \cup \mathbf{K}_2^{(0)},$$

where  $\mathbf{K}_1^{(0)} = \{k_1^{(0)}, \dots, k_{n_p}^{(0)}\}$  are starting values, and

$\mathbf{K}_2^{(0)} = \{k_{j+1}^{(0)}, \dots, k_{n_p}^{(0)}\}$  are given fixed parameters.

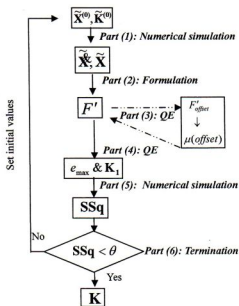


Figure 1: Flowchart of symbolic-numeric optimization. The variables and values are enclosed by the boxes, and the procedures are numbered corresponding to the description in the text.

2. By numerical simulation of the kinetic model (1), we obtain a time series for  $x_i$  and  $\tilde{x}_i$ :

$$\tilde{X} = \{\tilde{x}_i^{(t)} \mid i = 1, \dots, n_1, t = 0, 1, \dots, n_t\} \text{ and}$$

$$\tilde{X} = \{\tilde{x}_i^{(t)} \mid i = 1, \dots, n_1, t = 0, 1, \dots, n_t\}.$$

#### Part (2): Formulation

After choosing some variables from  $X$ , we call them "focusing variables",  $Y$ , and substitute observed/simulated data into the remaining variables:

1. Choose a subset  $Y$  of  $X$ :  $Y \subseteq X$ .

2. Substitute  $\tilde{X}$ ,  $X \setminus Y$  by the values of  $\tilde{X}$ ,  $\tilde{X}$  at a time point  $t$ :  $\tilde{x}_i \leftarrow \tilde{x}_i^{(t)}$ ,  $x_j \leftarrow \tilde{x}_j^{(t)}$ , where  $x_j \in X \setminus Y$ .

Then we denote the new formula as  $F(Y, K_1, e_{\max}, e)$ . We note that by performing a QE computation for the formula,  $\exists Y \exists K_1 \exists e_{\max} \exists e (F)$ , we can check the feasibility (true or false) of  $F$ .

#### Part (3): Computation of offset by QE

Observed data often contain an offset. Therefore, we must first determine the offset value. Here we consider the case that the offset appears linearly. For the sake of simplicity, we assume that only  $x_1$  has an offset. Let  $F'_{\text{offset}}$  be the

formula obtained by putting  $x_1 - \text{offset}$  into  $x_1^{(t)}$  of  $F'$ , where  $\text{offset}$  is a variable for offset. By performing QE for  $\exists X \exists K_1 \exists e_{\max} \exists e (F'_{\text{offset}})$ , we obtain the quantifier-free formula  $\mu(\text{offset})$ , which stands for the feasible ranges of  $\text{offset}$ . Then we substitute the minimum value of the offset for the variable  $\text{offset}$  in  $F'$ , and we denote it again by  $F(Y, K_1, e_{\max}, e)$ .

#### Part (4): Estimation of $e_{\max}$ and $K_1$ by QE

First, we use QE to make the magnitude of  $e_i$  as small as possible, and then we estimate the parameters  $K_1$  by QE:

1. Compute the feasible range of  $e_{\max}$ : by computing QE for  $\exists Y \exists K_1 \exists e (F)$ , we obtain a quantifier-free formula  $\pi(e_{\max})$  describing the feasible ranges of  $e_{\max}$ . Next, we put the minimum value of  $e_{\max}$  into  $e_{\max}$  in  $F'$ , and denote the resulting formula as  $F''(Y, K_1, e)$ .

2. Compute  $K_1$ : by computing QE for  $\exists Y \exists e (F'')$ , we obtain a quantifier-free formula  $\pi(K_1)$  describing the feasible ranges of  $K_1$ .

Actually, the feasible ranges of  $K_1$  are usually sufficiently narrow intervals (e.g., about  $10^7$ ) to choose an appropriate specific value of  $K_1$ .

#### Part (5): Computation of sum of squares (SSQ)

We estimate the goodness-of-fit for the obtained parameter values  $K_1$  from the feasible ranges of  $K_1$  in terms of SSQ.

1. Set initial conditions  $\tilde{X}^{(0)}$  and  $K_1$ .

2. Perform numerical simulation of kinetic model (1).

3. Compute SSQ:  $SSQ = \sum_i (\tilde{x}_i^{(t)} - \tilde{x}_i^{(t)})^2$ .

#### Part (6): Termination

If SSQ is smaller than a specific level  $\theta$ , then output  $K_1$ . Otherwise, set new initial values and go to (1).

## 2.4 Kinetic Model

We analyzed a model for the inhibition kinetics of HIV proteinase [16], as shown in Figure 2. The proteinase monomer ( $M$ ) is inactive, but the enzyme ( $E$ ) is active in the dimeric form. The dimer catalyzes the conversion of the substrate ( $S$ ) to the product ( $P$ ). The inhibitor ( $I$ ) is competitive for the substrate and the product, and the inhibitor-binding enzyme is irreversibly deactivated ( $EI$ ). In the model, there are ten parameters and nine variables. According to the previous studies [16, 3], five parameters ( $k_{11}$ ,  $k_{12}$ ,  $k_{21}$ ,  $k_{41}$ ,  $k_{51}$ ) are given, and the remaining five unknown parameters ( $k_{22}$ ,  $k_3$ ,  $k_{42}$ ,  $k_{52}$ ,  $k_6$ ), two initial values ( $E_{\text{init}}$ ,  $S_{\text{init}}$ ) and the offset of the fluorimeter are estimated by the present method. The experimental data of the product  $[P]$ , which are composed of 300 data points measured from 0 to 3600 seconds, were downloaded from a web site (<http://www.gepasi.org/tutorials/opt/hivfit.html>).

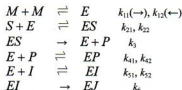


Figure 2: Kinetic model for the inhibitor of HIV proteinase.

The start values for ten parameters and the initial values for nine variables [3] are as follows:  $k_{11}=0.1$ ,  $k_{12}=1E-4$ ,  $k_{21}=100$ ,  $k_{22}=300$ ,  $k_3=10$ ,  $k_{41}=100$ ,  $k_{42}=500$ ,  $k_{51}=100$ ,  $k_{52}=0.1$ , and  $k_6=0.1$ ;  $\tilde{x}_1=0$ ,  $\tilde{x}_2=0.004$ ,  $\tilde{x}_3=25.0$ ,  $\tilde{x}_4=0$ ,  $\tilde{x}_5=0$ ,  $\tilde{x}_6=0$ ,  $\tilde{x}_7=0.003$ ,  $\tilde{x}_8=0$ , and  $\tilde{x}_9=0$ .

## 3 RESULTS

First, we will describe the practical procedure for parameter optimization in the kinetic model for HIV proteinase, and then we will evaluate the optimized parameters by the goodness of fit to the observed data.

### 3.1 Optimization Procedure in HIV Inhibition Proteinase Kinetics Model

To perform the numerical simulation (Part 1),  $K_1$  and  $K_2$  are defined as the five unknown parameters and the nine variables parameters, and the nine variables  $X = \{x_1, x_2, \dots, x_9\}$  are allocated to  $[P]$ ,  $[E]$ ,  $[S]$ ,  $[ES]$ ,  $[M]$ ,

$[EP]$ ,  $[I]$ ,  $[EI]$ , and  $[EJ]$ . Then we set the start value  $\tilde{K}^{(0)}$  and the initial value  $\tilde{X}^{(0)}$ . The start values for ten parameters and the initial values for nine variables are cited from the previous study [3] (see the legend in Figure 2). Also, the two initial values,  $E_{\text{init}}$  and  $S_{\text{init}}$ , are changed within a limited range with reference to the previous studies [16, 3]: 31 discrete values for  $\tilde{x}_2^{(0)}$  ( $[E]=0.00350, 0.00355, \dots, 0.00500$ ) and 9 values for  $\tilde{x}_3^{(0)}$  ( $[S]=24.0, 24.5, \dots, 28.0$ ).

The focusing variables  $Y$  (Part 2) are simply obtained by the symbolic computation with QE from the relationship

between  $X$  and  $K_1$  in the model. In the inequality  $v_i(X, K) \Delta t + x_i' \geq 0$ , the elimination of  $\Delta t$  by QE outputs five inequalities including five parameters:  $100*[E]*[I] - k_{22}*[E]$ ,  $[E] - k_6*[E] > 0$ ,  $100*[E]*[I] - k_{22}*[E] > 0$ ,  $100*[E]*[P] - k_{22}*[E] - k_3*[E] < 0$ ,  $100*[E]*[P] - k_{22}*[E] > 0$ , and  $100*[E]*[S] - k_{22}*[E] - k_3*[E] > 0$ . Among the seven variables in the above five inequalities,  $[P]$  is included in the objective function, and  $[S]$  is a large value relative to the other variables in the reaction molecules. Except for the last three inequalities including  $[P]$  and  $[S]$ , only  $[E]$  appears in the terms related to the unknown parameters in the first two inequalities. Thus, the focusing variables  $Y$  are defined as  $[P]$ ,  $[S]$ , and  $[E]$  in the present model. In addition, the conservation laws in the present model are obtained by Gepasi [5], a tool for estimating the kinetic flux in a given model, as follows:  $h_1(X) = [S] + [ES] + [P] + [EP] - S_{init} = 0$  and  $h_2(X) = [M] + 2[E] - 2[S] - 2[P] + 2[E] + 2[E] - (2E_{init} - 2S_{init}) = 0$ . The computation of *offset* by QE (Part 3) is realized by eliminating all of the variables by QE, except for *offset* in  $F^{offset}$ . Then we obtain  $\mu(offset)$ , depending on  $\tilde{x}_1^{(i)}$ , and the values of six variables and five parameters, as the following inequalities:  $\mu(offset) = c_1 + c_2 * offset > 0$  and  $c_3 + c_4 * offset < 0$ , where  $c_1$ ,  $c_2$ ,  $c_3$ , and  $c_4$  are constants. The constants in the above equations are estimated in each optimization.

Using  $F^*$  of  $Y$  and the *offset* obtained above, we can estimate  $e_{max}$  and  $K_1$  by QE (Part 4). Note that 279 sets of  $e_{max}$  and  $K_1$  are obtained by the corresponding sets of  $E_{init}$  and  $S_{init}$ . Since the fitting of simulated data strongly depends on the initial values, we further simulate numerically  $E_{init}$  and  $S_{init}$  within the above ranges of  $E_{init}$  and  $S_{init}$  by a standard technique of the bisection method,  $E_{init}$  and  $S_{init}$  for each set of  $e_{max}$  and  $K_1$  are estimated to minimize the  $SSq$  that is calculated for 300 values of  $[P]$  (Part 5). Finally, we obtain a set of  $e_{max}$ ,  $K_1$ ,  $E_{init}$  and  $S_{init}$  by selecting a minimum  $SSq$  among the 279  $SSq$ 's.

To judge whether the loop in Figure 1 terminates or not (Part 6), the minimum of  $SSq$ 's is compared with the threshold  $\theta$ . In the present study, the threshold is set to 0.01 to attain the same magnitude as that in the previous study [3]. If the  $SSq$  is smaller than  $\theta$ , then we terminate the optimization process. If the  $SSq$  is larger than  $\theta$ , then we start the loop by substituting 279 sets of  $K_1$  into  $\tilde{K}^{(0)}$  with

the same initial value sets of  $\tilde{x}_2^{(0)}$  ( $[E_{init}]$ ) and  $\tilde{x}_3^{(0)}$  ( $[S_{init}]$ ).

Although the number of starting values  $\tilde{K}^{(0)}$  increases as 279<sup>n</sup> with the  $n$ -th iteration, the restriction of the parameter and the variable spaces prevents multiple iterations. Indeed, only one or two iterations were sufficient to attain the threshold in the present model.

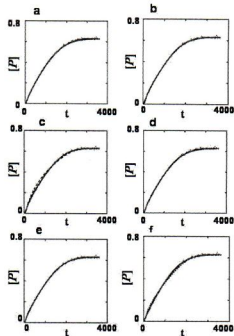


Figure 3: Fitting to observed data with optimized parameters. The amount of product  $[P]$  is multiplied by a coefficient (0.024), according to [3]. The experimental data are denoted by the jagged curve. The simulated curves are denoted by the solid curves (finally optimized) and if the loop iterates twice, by the broken curves (first optimized): a,  $r=336$ ; b,  $r=984$ ; c,  $r=1848$ ; d,  $r=336$  and 984; e,  $r=336$  and 1848; f,  $r=984$  and 1848.

time	Iteration	$k_{22}$	$k_3$	$k_{21}$	$k_{23}$	$k_6$	$SSq$	$E_{max}$	$S_{max}$	<i>offset</i>
336	1	250.2	9.776	1306	0.103	0.0969	0.00962	0.00495	27.5	-0.028591
984	1	155.5	9.982	1127	0.102	0.0980	0.00825	0.00470	28.0	-0.040591
1848	1	40.58	9.990	1211	0.102	0.0983	0.06380	0.00385	28.0	-0.040591
	2	134.5	9.817	1128	0.110	0.0900	0.00963	0.00460	28.0	-0.040591
336, 984	1	225.9	9.970	1239	0.103	0.0969	0.00856	0.00490	27.5	-0.028591
336, 1848	1	191.9	9.980	1304	0.103	0.0969	0.00666	0.00475	27.5	-0.028591
984, 1848	1	99.6	9.990	1254	0.100	1.00	0.03334	0.00430	27.5	-0.028591
	2	211.6	9.719	1140	0.101	1.00E+08	0.00659	0.00495	27.0	-0.016591
Mendes & Kell	-	201.1	7.352	1171	1.31E+04	3.00E+04	0.00513	0.00547	26.79	-0.008962
Kuzmic	-	179.7	9.46	1117	0.0831	0.1224	-	0.00387	24.63	-0.01

Table 1: Goodness of fit with optimized parameters by symbolic-numeric method. For reference, the values related to the present optimization are also cited from previous studies [16, 3].

### 3.2 Fitting to Observed Data with the Optimized Parameters

The optimized parameters with the six sets of observed data are listed in Table 1, together with the iteration number, the goodness of fit measured by  $SSq$ , the initial values of  $E_{init}$  and  $S_{init}$ , and the offset. In addition, the fittings of simulated values to the observed data in six cases are described in Figure 3.

One of the remarkable features of the present fitting is that only one or two points of the observed data are sufficient to fit 300 data points with an  $SSq$  value of less than 0.01. The data point for the optimization is randomly chosen from 300 points of data, and all fittings attain the threshold by one or two iterations of the loop. In two of the six cases, two rounds of iterations were required, but the first fitting in each case agreed well with the observed data. This is partly because QE powerfully restricts the possible ranges of the parameters and the variables, and partly because the present model is simpler than that expected from the complex kinetics of ten parameters and nine variables. These points will be discussed in the following section.

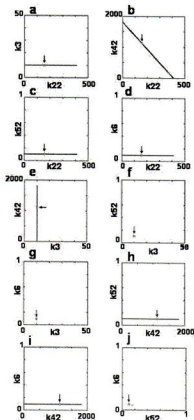


Figure 4: Relationships between the five optimized parameters. The parameter ranges were estimated by the given values at  $t=984$ . The open circles indicated by arrows are the optimized parameter values at  $t=984$ .

Another feature is that the values of the parameters agree well with those in the previous studies [16, 3]. In particular, the highlighted parameters in this model, the inhibitor binding constant ( $k_{52}$ ) and the deactivation rate constant ( $k_6$ ), are about 0.10 and 0.097 in three of the six cases, which are similar values to the constants in one previous study [16]. In contrast, the constants in the three remaining cases are enormously large, except for  $k_{52}$  at  $t=984$  and 1848, which are also similar to the magnitude of the rate obtained in the other previous study [3]. In comparison with both cases, the value in the latter case is unreasonably large for the analysis to be acceptable. Thus, the large dissociation and deactivation rate constants suggest that the potency of the inhibitor is overestimated in terms of the inhibitor reaction.

Two problems in the present optimization remain: one is the choice of the observed data for the simulation, and the other is the confidence intervals for the parameters. As for the data choice, the data showing a flat slope in the kinetic curve seem intuitively inadequate for the simulation. Indeed, by using the data of more than  $t=2500$  in Figure 3, QE frequently outputs 'false'; this means no parameter and variable spaces for the initial conditions in  $F'$ . Any data, except for those in the steady states, may possibly output 'true' for the optimization by QE. As for the confidence intervals for the parameters, we will discuss them in terms of the parameters space in the following section.

## 4 DISCUSSION

### 4.1 Elucidation of Parameter Space by QE

One of the merits of the present method is that some relationships between the parameters in the optimization process can be easily estimated by QE. By utilizing this merit, we can further elucidate the optimization process in the present model.

#### 4.1.1 Relationship between parameters by QE

The relationship between the parameters can be easily estimated by QE. Indeed, the parameter relationship in arbitrary ranges of variables is obtained by only the exclusion of the object function in  $F'$ , and is useful to elucidate visually the parameter optimization by QE in terms of the confidence intervals of the optimized parameters.

Figure 4 shows the relationships between five optimized parameters. Given  $E_{init}$  and  $S_{init}$ , the  $F'$  obtained by excluding the object function consists of five known and five unknown parameters and three focusing and six remaining variables. In  $F'$ , the three focusing variables and three of the five unknown parameters can be eliminated by QE, given numerical values at  $t=984$  for six known variables, and then we can obtain the relationship between the remaining pair of parameters (a quantifier-free formula for the parameters).

One of the striking features of the elimination by QE is that the parameter spaces are highly restricted in all parameter pairs. The parameter space of each pair emerges in a very narrow range between two boundaries, which is a visual representation of the uncertainty of the parameters. For example, the narrow range for the parameter space between  $k_{52}$  and  $k_{42}$  is obtained with the following equations:  $155.49657522 < k_{22} < 155.49657525$  and  $1127.2733904 < k_{42} <$

1127.2733905. Furthermore, the total parameter ranges are shown visually in the restricted ranges; the maximum ranges of the parameter pairs are  $0 \leq k_{21} \leq 411$  and  $0 \leq k_{12} \leq 1813$ , and the dissociation constant of  $[EP]$  ( $k_{e2}$ ) varies in a wider range than that of  $[ES]$  ( $k_{e1}$ ).

As seen in the figure, the relationships between the two parameters are divided into independent and dependent relationships. The parameter range with a slope indicates that the two parameters change mutually, and are dependent on each other, and the range with no slope indicates that the parameters are independent of each other. Only the relationship between  $k_{21}$  and  $k_{12}$  ( $b$  in Figure 4) is dependent, and the remaining relationships are independent. Thus, the present method reveals the mutual dependency of the parameters with their feasible ranges.

## 4.2 Computational Performance

The present analyses were performed on a PC with a Pentium XEON 3.2 GHz CPU and 4GB of memory, using the WINDOWS operating system. The computational time was about 4.1 minutes for the total optimization of the first loop, and was about 19.5 hours for the second loop.

## 4.3 Conclusions

The present study is the first application of QE to the parameter optimization problem in conjunction with a numerical simulation. Our symbolic-numeric method by QE shows the same magnitude of goodness of fit as the previous numerical optimization. Furthermore, the present method has the distinct potential to elucidate the relationships between the parameters and the variables in the kinetic model. Thus, our method provides a new direction for the analysis of kinetic models in the field of computational biology.

## ACKNOWLEDGEMENTS

We would like to express our gratitude to Mr. Taku Takeshima for his kind assistance. One of the authors (K. H.) was partly supported by a Grant-in-Aid for Scientific Research on Priority Areas "Genome Information Science" (grant 15014208) and for Scientific Research (B) (grant 15310134), from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

## REFERENCES

[1] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. In Proc. Natl. Acad. Sci. USA, 92:6479–6483, 1995.  
 [2] H. Kitano. Computational systems biology. Nature, 421:573, 2002.  
 [3] P. Mendes, and D. B. Kell. Non-linear optimization of biochemical pathways: applications to metabolic engineering, and parameter estimation. Bioinformatics, 14:869–883, 1998.  
 [4] C. G. Moles, P. Mendes, and J. R. Banga. Parameter estimation in biochemical pathways: a comparison of global optimization methods. Genome Res., 13:2467–2474, 2003.

[5] P. Mendes, and D. B. Kell. MEG (Model Extender for Gepasi): a program for the modelling of complex, heterogeneous, cellular systems. Bioinformatics, 17:288–289, 2001.  
 [6] M. H. Hoefnagel, M. J. C. Starrenburg, D. E. Martens, J. Hugenholtz, M. Kleerebezem, I. I. Van Swam, R. Bongers, H. V. Westerhoff, and J. Snoep. Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modeling, metabolic control and experimental analysis. Microbiology, 148:1003–1013, 2002.  
 [7] I. Swameye, T. G. Muller, J. Timmer, O. Sandra, and U. Lingmuller. Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. In Proc. Natl. Acad. Sci. USA, 100:1028–1033, 2003.  
 [8] F. Garvan. The Maple Book. Chapman and Hall, London, 2001.  
 [9] S. Wolfram. The Mathematica Book. 4th ed. Cambridge University Press, 1999.  
 [10] M. Barnett. Computer algebra in the life sciences. ACM SIGSAM Bulletin, 36:5–32, 2002.  
 [11] B. Caviness, and J. Johnson (ed.). Quantifier Elimination and Cylindrical Algebraic Decomposition. Springer-Verlag, 1998.  
 [12] A. Tarski. Decision Methods for Elementary Algebra and Geometry. Berkeley University of California Press, 1951.  
 [13] G. E. Collins. Quantifier elimination for the elementary theory of real closed fields by cylindrical algebraic decomposition. In Lecture Notes in Computer Science, 33:134–183. Springer-Verlag, Berlin, 1975.  
 [14] H. Anai, and H. Yanami. SYNAC: A maple-package for solving real algebraic constraints. In Sloot, P.M.A. et al. (eds.), Proceedings of the International Workshop on Computer Algebra Systems and Their Applications. CASA'2003, volume 2657 of LNCS. Springer-Verlag, 2003.  
 [15] C. Chauvin, M. Muller, and A. Weber. An application of quantifier elimination to mathematical biology. In J. Fleischer, J. Grabmeier, F. W. Hehl, and W. Kuchlin (eds.), Computer Algebra in Science and Engineering. Pages 287–298, World Scientific, 1995.  
 [16] P. Kuzmic. Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. Anal. Biochem., 237:260–273, 1996.



# *Honorable Mention Paper Award*

*September 24, 2005*

*Shigeo Orit, Japan*

*We are honored to extend our sincere appreciation for your honorable mention paper presentation in BIOINFO2005, International Joint Conference of InCoB, AASBi, and KSBI.*

*Kwang H. Lee, Ph.D.* 

*General Chair, BIOINFO2005*