A general biochemical kinetics data fitting algorithm for quasi-steady-state detection

Justinas V. Daugmaudis[1]**, Audrius Laurynėnas**[2]**,
Juozas Kulys**[2]**, Feliksas Ivanauskas**[1]

[1]*Faculty of Mathematics and Informatics, Vilnius University*
 Naugarduko str. 24, LT-03225 Vilnius, Lithuania
[2]*Institute of Biochemistry, Vilnius University*
 Mokslininkų str. 12, LT-102587 Vilnius, Lithuania
 E-mail: justinas.daugmaudis@mif.vu.lt, audrius.laurynenas@gmail.com
 E-mail: juozas.kulys@bchi.vu.lt, feliksas.ivanauskas@mif.vu.lt

Abstract. We develop a general algorithm for fitting the biochemical kinetics data. The developed algorithm searches and analyzes numerous minima. This approach allows us to analyze biochemical data without a priori quasi-steady-state assumptions. The algorithm allows us to treat the biochemical kinetics data that has varying degree of steadiness. We test the approach by analyzing experiment data from 4-nitrophenyl phosphate hydrolysis with alkaline phosphatase.

Keywords: biochemical kinetics, differential equations, curve fitting, reaction rate constants, phosphatase, 4-nitrophenylphosphate.

1 Introduction

Typically, in biochemical kinetics experiments, the concentration of substrates is much higher than that of enzymes. Therefore, the quasi-steady state assumption is frequently valid. This results in dimensional reduction of the number of reaction rate parameters, and integral curves can be approximated with less reaction rate parameters than defined in a reaction scheme.

The advent of computers and new algorithms has opened new possibilities to simulate and analyze data in biochemical kinetics. At first, integral curve analysis was limited only to the cases where an analytical solution was known. Nowadays, integral curve analysis can be applied for any reaction scheme of interest. Direct data fitting into numerical solutions of differential equations was implemented in various software tools a dozen of times [6, 4, 2, 1, 7]. However, this method also has pitfalls. Widespread belief (fundamental assumption in rate constant fitting software) that there exists one optimal set of rate constants is the most severe. This assumption leads to wrong fitting results in cases under quasi-steady state. However, it is desirable to describe a reaction scheme by simple chemical reactions and to have a method to prove a quasi-steady state instead of postulating it.

In this paper, we describe a novel algorithm that is able to fit data to a reaction scheme without assuming pre-steady or quasi-steady state. Neither we assume that there exists one optimal set of rate constants.

2 Mathematical model

2.1 A system of differential equations

A system of ordinary differential equations can be constructed from a reaction scheme, which will be represented as a set of simple unidirectional chemical reactions (bidirectional reactions are split into unidirectional reactions).

Suppose that that in a reaction scheme there are N different reagents and products. We can enumerate the reagents and products as $A_1 \ldots A_N$. Then all reactions in the reaction scheme can be written in the artificial matrix form

$$\sum_{n=1}^{N} \gamma_{i,n} A_n \xrightarrow{k_i} \sum_{n=1}^{N} \gamma'_{i,n} A_n, \quad i = 1 \ldots I, \tag{1}$$
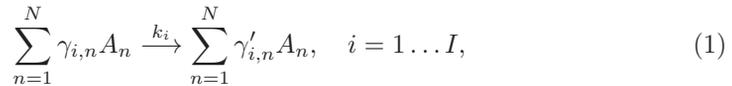
where I is the number of reactions, and $\gamma_{i,n} \geqslant 0$ and $\gamma'_{i,n} \geqslant 0$ are stoichiometric coefficients for reagents and reaction products. Note that if some reagent A_j does not participate in a reaction, its respective stoichiometric coefficients $\gamma_{i,j}$ and $\gamma'_{i,j}$ are simply zero.

We can then describe (1) as a system of N ordinary differential equations (one equation for each individual reagent):

$$\begin{cases} \frac{\partial [A_1]}{\partial t} &= \sum_{i=1}^{I} \tau_i (\gamma'_{i,1} - \gamma_{i,1}) k_i \prod_{n=1}^{N} [A_n]^{\gamma_{i,n}}, \\ \frac{\partial [A_2]}{\partial t} &= \sum_{i=1}^{I} \tau_i (\gamma'_{i,2} - \gamma_{i,2}) k_i \prod_{n=1}^{N} [A_n]^{\gamma_{i,n}}, \\ \quad \vdots \\ \frac{\partial [A_N]}{\partial t} &= \sum_{i=1}^{I} \tau_i (\gamma'_{i,N} - \gamma_{i,N}) k_i \prod_{n=1}^{N} [A_n]^{\gamma_{i,n}}. \end{cases} \tag{2}$$

Here $\tau_i \in \{1, -1\}$ indicates the direction of a reaction. If we have to split a bidirectional reaction, then the direction of the backward reaction is -1. Otherwise, it is simply 1.

2.2 Relation between the observed quantities and solutions of the ODE system

A response function r is a linear combination of reagent concentrations at some time t.

In order to define the response function, we need to compute the concentration at time t. We do that by integrating each equation in (2):

$$[A_m(t)] = [A_m(0)] + \int_0^t \sum_{i=1}^{I} \tau_i (\gamma'_{i,m} - \gamma_{i,m}) k_i \prod_{n=1}^{N} [A_n]^{\gamma_{i,n}} \, \mathrm{dt}, \quad m = 1, \ldots, N. \tag{3}$$

Note that $[A_m(0)]$ here is the initial concentration at time $t = 0$.

Once we have computed the concentrations for each reagent, we can define the response function

$$r_j(\mathbf{A}, t) = \sum_{m=1}^{N} c_{j,m} [A_m(t)], \quad j = 1, \ldots, K. \tag{4}$$

Here $c_{j,m}$, $m = 1, \ldots, N$, are scaling constants (for example, molar absorptivities) and $j = 1, \ldots, K$ are the indices of independent experiments (where K is the number of

experiments with different initial concentrations). A linear combination is sufficient to express the measurement in biochemical kinetics. Note that the coefficients $c_{j,m}$ can act as a "comb" on the concentrations of reactants. We can filter out uninteresting concentrations by setting an appropriate coefficient to 0. Moreover, if all $c_{j,m} = 1$, then a well-behaving system yields $r_j(\mathbf{A}, t) = \text{const}$.

A response function, tracked for the duration of the experiment, produces the kinetic curve

$$\Gamma_j = r_j(\mathbf{A}, t), \quad t \in [0, t_j]. \tag{5}$$

2.2.1 Goodness of fit

Let c and $\bar{c}$ be curves with M points. The similarity metric between those curves is defined as

$$d(c, \bar{c}) = \sqrt{\frac{1}{M} \sum_{i=1}^{M} (c_i - \bar{c}_i)^2}, \tag{6}$$

where $M > 0$ is the number of samples in a curve. The metric d gives an average distance between the samples in curves c and $\bar{c}$.

The goodness of fit for K experiment curves is then defined as the sum of similarity scores for each individual experiment:

$$q = \sum_{j=1}^{K} d(\Gamma_j, \bar{\Gamma}_j), \tag{7}$$

where Γ_j is a simulated kinetic curve, and $\bar{\Gamma}_j$ is an experiment curve. Ideal fits l have the score $q = 0$.

2.2.2 Fitting algorithm

During grid search, a set of rate constants with minimal scores q is constructed. The Powell optimizer [5] is applied on this set, which refines the rate constants even further. Finally, the contour of likelihood region of the fit scores is estimated, and the rate constants are filtered using the equation [3]

$$c = \frac{q_{\min}}{1 - I_{1-\alpha}^{-1}\left(\frac{P}{2}, \frac{U-P}{2}\right)} \left[\frac{\sum_{i=1}^{K} \left(\frac{n_i}{n_i - 1}\right)}{K} \right]. \tag{8}$$

Here $K > 0$ is the number of experiments, $U = \sum_{i=1}^{K} n_i$ is the total number of measurements in all experiments, n_i is the number of measurements during a single ith experiment, P is the number of rate constants in a reaction scheme, α is the significance value (the value $\alpha = 0.05$ is used in the fitting of data in Section 3), I^{-1} is the inverse beta regularized function, and $q_{\min}$ is the smallest fit score found during the optimization.

The inverse beta regularized function is the solution to the equation $I_{\frac{xn}{xn+m}}(n, m) = 1 - \alpha$ for x, where I is the beta regularized function that arises from the Fisher–Snedecor cumulative distribution [8].

(a)

(b)

(c)

(c)
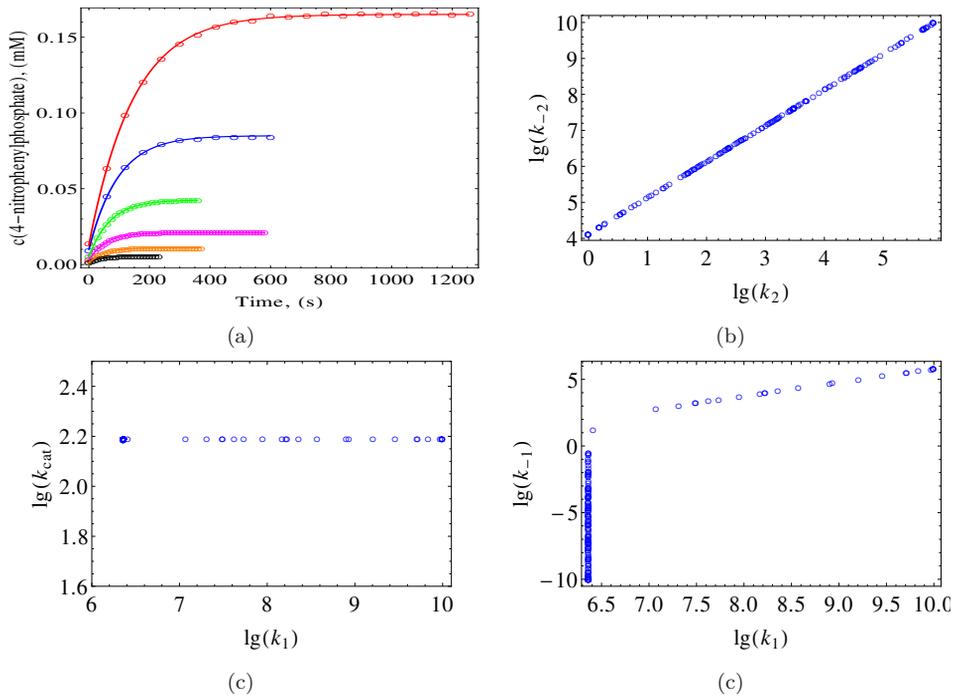
Fig. 1. (a) Kinetic curves fitted to experiment data; relationship between rate constants (b), (c) and initial rate dependence on substrate concentration (d).

3 Results

Let us consider hydrolysis of para-nitrophenilphosphate catalysed by alkaline phosphatase. The following reaction scheme, which describes reaction kinetics is relevant in our case:

$$
\begin{aligned}
E + S &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES, \\
ES &\overset{k_{cat}}{\rightarrow} EP, \\
E + P &\underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E_{inh}.
\end{aligned}
\tag{9}
$$
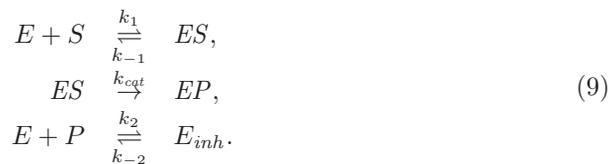
Here S is a substrate, P is a product, E, E_{inh} and ES are enzyme and enzyme/substrate complex respectively, k_1, k_{-1}, k_2, k_{-2}, and k_{cat} are the rate constants for reactions. Without any unnecessary experiments, the reaction scheme (9) results in an ideal fit (Fig. 1(a)) with fitted parameters $k_1 = 2.21 \times 10^6$ M^{-1} s^{-1}, $k_{-1} = 3.18 \times 10^{-7}$ s^{-1}, $k_2 = 1$ M^{-1} s^{-1}, $k_{cat} = 1.54 \times 10^2$ s^{-1}. This result is valid under the assumption that only one optimal solution exists. This assumption is a typical error in biochemical kinetics data fitting software. If we discard this erroneous assumption our algorihtm can find numerous local minima (see Fig. 1(b), (c), (d)), all of which are valid fits for this dataset, and indicate a quasi-steady state.

Other data fitting tools based on the uniqueness of optimal fit [6, 4, 2, 1, 7] give one solution with values for all rate constants in the reaction scheme (9). Analytical derivation of initial rate equation, based on quasi-steady approximation, for the

reaction scheme (9) yields

$$v_0 = \frac{[E_0]k_{cat}[S_0]}{K_M\left(1 + \frac{[P_0]}{K_{inh}}\right) + S}, \tag{10}$$

where $K_M = \frac{k_{-1}+k_{cat}}{k_1}$ and $K_{inh} = \frac{k_2}{k_{-2}}$. Using these relations, from the multiple minima we calculated K_M and K_{inh}. This result should be interpreted as *impossibility* to fit real values of separate rate constants k_1, k_{-1}, k_2, k_{-2} if an experiment is carried out in quasi-steady state. Only values of k_{cat}, K_M, and K_{inh} could be recovered in this case. Indeed, a linear relation of k_2 and k_{-2} in logarithmic coordinates (Fig. 1(b)) is the result of quasi-equilibrium in the third reaction (9). The same applies to the logarithmic relation for constants $k_1$and k_{-1} in a bit complicated way. Here we have two cases:

- the first, where $k_{-1} \gg k_{cat}$, and $K_M = \frac{k_{-1}+k_{cat}}{k_1}$ simplifies to $K_M \approx \frac{k_{-1}}{k_1}$;

- the second, where $k_{cat} \gg k_{-1}$, and $K_M = \frac{k_{-1}+k_{cat}}{k_1}$ simplifies to $K_M \approx \frac{k_{cat}}{k_1}$.

Indeed, such a relation is found in this case (Fig. 1(d)). The only definitely fitted rate constant, which shows no dependence on other rate constants, is k_{cat}. The values of K_M are calculated using the relations $K_M = \frac{k_{-1}+k_{cat}}{k_1} = 66.9 \pm 0.5$ µM, $v_{\max} = [E_0]k_{cat} = 156 \pm 1 \text{nMs}^{-1}\text{mg}^{-1}$, $K_{inh} = \frac{k_2}{k_{-2}} = 72.3 \pm 0.1$ µM from fitted values shown in Fig. 1.

4   Conclusions

We demonstrated that the assumption about the uniqueness of optimal solution is not necessary. Without it, a unified pre-steady and quasi-steady biochemical kinetics data analysis with extended statistical fit evaluation is possible. Applicability of the proposed algorithm was demonstrated with experiment data from 4-nitrophenyl phosphate hydrolysis with alkaline phosphatase.

References

[1] P. Kuzmič. Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.*, **237**(2):260–273, 1996.

[2] P. Kuzmič. DynaFit – a software package for enzymology. *Meth. Enzymol.*, **467**:247–280, 2009.

[3] A. Laurynenas and J. Kulys. An exhaustive search approach for chemical kinetics experimental data fitting, rate constants optimization and confidence interval estimation. *Nonlinear Anal., Model. Control*, **20**(1):145–157, 2015.

[4] P. Mendes. Biochemistry by numbers: simulation of biochemical pathways with Gepasi 3. *Trends Biochem. Sci.*, **22**(9):361–363, 1997.

[5] M.J.D. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Comput. J.*, **7**(2):155–162, 1964.

[6] G.E. Rovati, R. Shrager, S. Nicosia and P.J. Munson. KINFIT II: a nonlinear least-squares program for analysis of kinetic binding data. *Mol. Pharmacol.*, **50**(1):86–95, 1996.

[7] I.B. Svir, O.V. Klymenko and M.S. Platz. Kinfitsim software to fit kinetic data to a user selected mechanism. *Comput. Chem.*, **26**(4):379–386, 2002.

[8] Wolfram Research, Inc. *Inverse of the Regularized Incomplete Beta Function.* 2001–2008.

REZIUMĖ

Apibendrintas biocheminės kinetikos duomenų analizės algoritmas kvazi-stacionarių būsenų aptikimui
J.V. Daugmaudis, A. Laurynėnas, J. Kulys, F. Ivanauskas

Šiame darbe pasiūlėme ir realizavome metodą, aptinkantį biocheminių reakcijų kvazi-stacionarias būsenas. Aprašomas algoritmas buvo pritaikytas analizuojant 4-nitrofenilfosfato hidrolizę su šarmine fosfataze. Atsisakius prielaidos, jog tėra tik vienas greičio konstantų rinkinys, kuris optimaliai sugludintų cheminės kinetikos kreives su eksperimentinių duomenų kreivėmis, galime sudaryti ir ištirti reakcijos greičio konstantų rinkinius. Analizuodami tų greičio konstantų tarpusavio priklausomybes, galime nustatyti ar reakcija yra kvazi-stacionarioje būsenoje.

Raktiniai žodžiai: biocheminė kinetika, reakcijų greičio konstantos, diferencialinės lygtys, kreivių gludinimas, fosfatazė, 4-nitrofenilfosfatas.