Chapter 3

A Tutorial on Mathematical Modeling of Biological Signaling Pathways

Zhike Zi

Abstract

Mathematical models have been widely used in the studies of biological signaling pathways. Among these studies, two systems biology approaches have been applied: top-down and bottom-up systems biology. The former approach focuses on X-omics researches involving the measurement of experimental data in a large scale, for example proteomics, metabolomics, or fluxomics and transcriptomics. In contrast, the bottom-up approach studies the interaction of the network components and employs mathematical models to gain some insights about the mechanisms and dynamics of biological systems. This chapter introduces how to use the bottom-up approach to establish mathematical models for cell signaling studies.

Key words: Mathematical modeling, Signaling pathway, Reaction kinetics, Systems biology

1. Introduction

Traditional cell signaling studies have focused on understanding the functions of individual signaling components. It is now realized that cell behaviors are not only determined by the characteristics of individual system components, but also by the interactions of such components acting together as a system. Conventional biochemistry and cell biology approaches are useful to describe and discover the components of cellular signaling pathways. However, most of the descriptions are carton-type hypotheses that give us the static and qualitative information about the system. The nonlinear interactions of the components might exhibit some emergent properties, such as bistability, oscillation, and robustness, which are difficult to be discovered by experimental analyses but are easier to be investigated by kinetic analysis of mathematical models.

On the other hand, mathematical modeling of signaling pathways requires experimental data and prior knowledge of the pathways. Therefore, systems biology research could not be

Xuedong Liu and Meredith D. Betterton (eds.), *Computational Modeling of Signaling Networks*, Methods in Molecular Biology, vol. 880, DOI 10.1007/978-1-61779-833-7_3,
© Springer Science+Business Media, LLC 2012

done without the contribution of traditional biochemical and cell biology. However, classical cell signaling studies face the limitations, such as inaccuracy, inability to deal with emergent properties, and so on (1). The bottom-up systems biology approaches could fetch up these limitations by merging model simulations and quantitative experimental analyses. Therefore, systems biology approach with mathematical models has been emerged as a powerful tool in studying signaling pathways because it provides a way to study the signaling dynamics and the underlying principles.

Many sources of uncertainty, including errors, noise, incomplete information, and poor experimental design, impose limitations on our confidence of the experimental data. Some experimental data might lead to a misunderstanding of the true property of biological systems due to the limited number of experimental data. For example, we might get a time course data for protein X as shown in Fig. 1a. The amount of protein X oscillates with the increase of time. The experimentalists may judge the number of periods by counting the number of peaks appearing in the data by connecting the discrete data. In this case, the number of peaks will be counted as 3 (Fig. 1b). However, if we perform model simulation for this system, we will find that the number of the oscillations during the investigated time might be 7, rather than 3 (Fig. 1c). Inspired by the model prediction, we could design a new experiment to verify the actual number of oscillations by taking more samplings at proper time points.

2. Materials

The material requirement for mathematical modeling of cellular signaling pathways is not demanding. However, getting high quality of quantitative experimental data sets is still one of the bottlenecks for the development of the mathematical models. Normally, computer and software packages are required to perform model simulations. The specific configuration of the computers and software depends on the modelers' preference. Till now, more than 200 software tools have been developed with the support of the Systems Biology Markup Language (SBML) format (2). A summary of these software packages is available in the official SBML Web site (http://sbml.org).

3. Method

The mathematical format for modeling signaling pathways depends on the properties of the studied system and the specific questions that are going to be answered. The most commonly used mathematical forms are ordinary and partial differential equation systems, which can be in either deterministic or stochastic format (3).

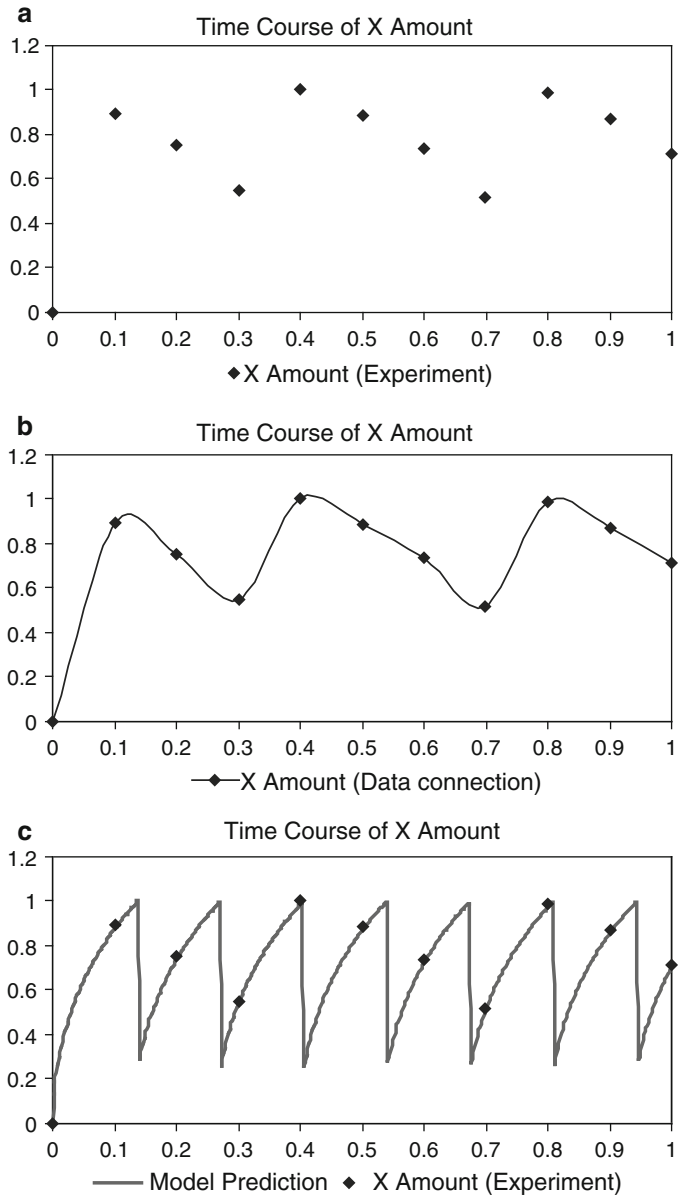

Fig. 1. Illustration of the role of model prediction. (**a**) Time course data for the amount of protein X. (**b**) The intuitive data connection predicts three oscillation periods. (**c**) Computational model simulation of the underlying network suggests that protein X has oscillated seven times.

Stochastic equations are useful to investigate the properties of the system arising from random fluctuations or noise. Here, we introduce how to establish mathematical models for the signaling pathways with a system of ordinary differential equations (ODE). The ODE modeling approach can be applied to the signaling studies when the components of the signaling pathway can be assumed

to homogenously distribute in the cell and the stochastic effects could be ignored.

3.1. The Choice of Reaction Kinetics

There are different ways to model the reaction kinetics in signaling pathways. Typical reaction kinetics include the law of mass action, Hill equation, and Michaelis–Menten kinetics. Here, we use the example of ligand receptor interaction to illustrate how to use these kinetics to describe signaling reactions.

(1) *The law of mass action*: In the nineteenth century, Guldberg and Waage introduced the law of mass action to describe the biochemical kinetics (4). The law of mass action states that the reaction rate is proportional to the probability of the collision of the reactants. This probability is also proportional to the concentration of reactants to the power of their molecularity, the number of them entering the specific reaction (5).

The ligand and receptor interaction is a signaling event that happens in most signaling pathways. With the binding of the ligand (L), the receptor (R) forms a ligand–receptor complex (LRC) with a characteristic rate constant k_{on}. On the other hand, the LRC can dissociate to ligand and receptor with a rate constant k_{off}. The ligand–receptor interaction can be described with the following reaction:

$$\mathrm{L} + \mathrm{R} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\Leftrightarrow}} \mathrm{LRC}. \tag{1}$$

By the definition of mass action law, we can derive the concentration change over time of ligand, receptor, and LRC by the following ODE system:

$$\frac{\mathrm{d}[L]}{\mathrm{d}t} = k_{\mathrm{off}}[\mathrm{LRC}] - k_{\mathrm{on}}[L][R], \tag{2}$$

$$\frac{\mathrm{d}[R]}{\mathrm{d}t} = k_{\mathrm{off}}[\mathrm{LRC}] - k_{\mathrm{on}}[L][R], \tag{3}$$

$$\frac{\mathrm{d}[\mathrm{LRC}]}{\mathrm{d}t} = k_{\mathrm{on}}[L][R] - k_{\mathrm{off}}[\mathrm{LRC}]. \tag{4}$$

(2) *Hill equation*: The signaling receptors might exist not only as monomers, but also form dimers or oligomers with two or more identical binding sites for the ligand. The bound subunit has a cooperative effect on the later binding subunits, which means that the affinity of the receptor to the later bound ligand is significantly increased by the already bound ligand. A typical example that has such a property is the binding of oxygen to hemoglobin (Hb). The early experimental research found that the fractional saturation of Hb with oxygen had a sigmoid response to the oxygen partial pressure. Hill explained the interaction between the oxygen

binding sites of Hb subunits by the following fractional saturation (6):

$$\Upsilon = \frac{[L_n R]}{[R_t]} = \frac{[L]^n}{(K_a)^n + [L]^n}, \tag{5}$$

where n is known as Hill coefficient and K_a corresponds to the ligand concentration occupying half of the receptor-binding sites.

In general, for a receptor with n subunits, the reaction rate of the receptor to the ligand can be approximated as:

$$v = V_{max}\Upsilon = \frac{V_{max}[L]^n}{(K_a)^n + [L]^n}. \tag{6}$$

The Hill coefficient (n) can be calculated by

$$n = \frac{\log(81)}{\log(S_{90}/S_{10})}. \tag{7}$$

Here, S_{10} and S_{90} correspond to the ligand (stimulus) levels that are required to achieve 10% and 90% of the maximum activation, respectively (7).

One can use Hill equation kinetics with a larger value of Hill coefficient (n) to approximately model some switch-like processes, for example the turn on of ion channels.

(3) *Michaelis–Menten kinetics*: When an enzyme (kinase or phosphatase) catalyzes a biochemical reaction, it is not consumed or produced by this reaction, but forms a temporary complex with the substance in the reaction. For such a reaction, we can use Michaelis–Menten kinetics to describe the reaction rate under the key assumption of quasi-steady-state approximation, which is valid when the enzyme concentration is much lower than the substrate concentrations and when the enzyme is not allosteric. Michaelis–Menten kinetics is named after Leonor Michaelis and Maud Menten and has the following formulation:

$$v = \frac{V_{max} S}{K_m + S}, \tag{8}$$

where K_m is the Michaelis constant and is equal to the substance concentration that causes the half-maximal reaction rate V_{max}.

3.2. How to Construct the Systems of Ordinary Differential Equations?

The ODE modeling approach describes the concentration change of a component over time by a system of ordinary differential equations. For a certain component's concentration $[C_i]$ evolved over time, we calculate both the sum of the reaction rates producing C_i (the reactions coming towards component C_i, v^+) and the sum of the rates consuming C_i (the reaction going outwards component

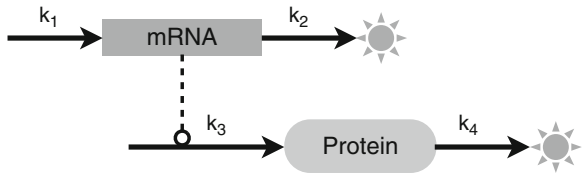

Fig. 2. A simple network for the production and degradation of mRNA and protein.

C_i, v^-). The system of ODEs is determined by the subtraction of the former to the latter item as the following equation:

$$\frac{\mathrm{d}[C_i]}{\mathrm{d}t} = \sum v^+ - \sum v^-. \tag{9}$$

More examples about how to construct the ODE system will be illustrated in the following section.

3.3. How to Model Different Signal Transduction Steps?

Some signal transduction steps are shared in different signaling pathways. In order to establish a complete model for a certain signal pathway, it would be useful to understand how to model representative signal transduction steps (8).

3.3.1. Production and Degradation of mRNA and Protein

The production and degradation of mRNA and protein are typical processes involved in signaling pathways (Fig. 2). Here, we can assume that there is a constitutive mRNA synthesis with a production rate constant k_1. The degradation of mRNA follows first-order mass-action kinetics with degradation rate constant k_2. The protein is produced with a rate constant k_3 and is proportional to the concentration of mRNA. Similarly, the protein is degraded with rate constant k_4.

For this simple network, we can use the following system of ODEs to describe its dynamics.

$$\frac{\mathrm{d}[\mathrm{mRNA}]}{\mathrm{d}t} = k_1 - k_2[\mathrm{mRNA}], \tag{10}$$

$$\frac{\mathrm{d}[\mathrm{Protein}]}{\mathrm{d}t} = k_3[\mathrm{mRNA}] - k_4[\mathrm{Protein}]. \tag{11}$$

It is worth noting that mRNA has appeared in Eq. 11 for the protein production, but it does not mean that mRNA is a reactant or substrate for protein production. Therefore, parameter k_3 should not appear in Eq. 10. mRNA is used as a template during protein translation and mRNA itself is not converted to protein. In SBML, such molecules are defined as "modifiers."

For Eq. 10, we can derive an analytic solution to calculate the mRNA amount at time t with the following equation:

$$[\mathrm{mRNA}]_t = \frac{k_1}{k_2} - \left(\frac{k_1}{k_2} - [\mathrm{mRNA}]_0\right)\mathrm{e}^{-k_2 t}, \tag{12}$$

where $[\mathrm{mRNA}]_0$ is the concentration of mRNA at initial time 0.

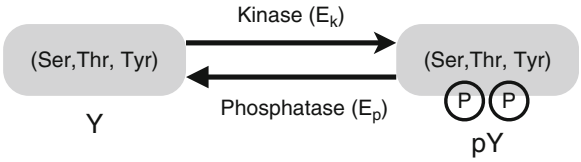

Fig. 3. Scheme of protein phosphorylation and dephosphorylation.

3.3.2. Phosphorylation and Dephosphorylation

Protein phosphorylation and dephosphorylation take place very frequently in cellular signaling transduction. In Fig. 3, the protein Υ is phosphorylated to pΥ by a kinase and then the phosphorylated pΥ is dephosphorylated to Υ by a phosphatase. Here, we assume that the total amount of Υ and pΥ is constant and equals to Υ_t.

For this process, there are two different ways to model it. We can choose mass-action kinetics or Michaelis–Menten kinetics depending on the context of the signaling pathway. If the concentrations of kinase (E_k) and phosphatase (E_p) are much lower than their substrates, we can model this simple system with Michaelis–Menten kinetics:

$$\frac{d[\Upsilon]}{dt} = -\frac{V_{max\,1}[\Upsilon]}{K_{m1} + [\Upsilon]} + \frac{V_{max\,2}(\Upsilon_t - [\Upsilon])}{K_{m2} + (\Upsilon_t - [\Upsilon])}, \tag{13}$$

$$[pY] = \Upsilon_t - [\Upsilon]. \tag{14}$$

In some cases, the kinase or phosphatase can be one component of the signaling pathway and its concentration is in a similar range as the substrate. The assumption for Michaelis–Menten kinetics is not valid anymore. In this case, mass-action kinetics might be better to model the system. It might be necessary to include the intermediate steps. For simplicity, we can model this system with mass-action kinetics in the following way.

$$\frac{d[\Upsilon]}{dt} = -k_1[\Upsilon] + k_2(\Upsilon_t - [\Upsilon]), \tag{15}$$

$$[pY] = \Upsilon_t - [\Upsilon]. \tag{16}$$

3.3.3. Feedbacks

Feedback regulations are important strategies for cells to control the signaling pathways. For example, positive feedback may allow signaling pathways to read continuously graded signal input and generate digital signaling output (9). Moreover, cells can adapt to some types of signals by using negative-feedback regulation (10). In some cases, a signaling pathway with delayed negative feedback could generate an oscillatory response of the signaling output.

Figure 4 shows a simple example of feedback network, in which the activation of molecule A leads to the activation of molecule B and activated B (B *) has a positive or negative feedback on the activation of A. We can use mass-action kinetics to model the activation of A and

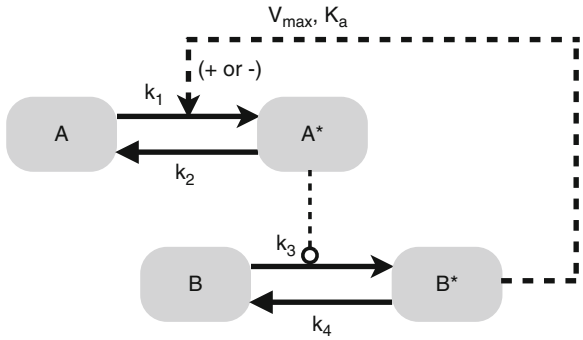

Fig. 4. Scheme of network feedback.

B. For the feedback of activated B on the activation of A, we can choose Hill equation to model it. Here, the total amount of A and A* is A_t, and the total amount of B and B* is B_t. Therefore, for the network shown in Fig. 4, it is possible to model it with the following system of ODE:

$$\frac{\mathrm{d}[A^*]}{\mathrm{d}t} = k_1(A_t - [A^*]) - k_2[A^*] + \frac{V_{\max}(B^*)^n}{(K_\mathrm{a})^n + (B^*)^n}(A_t - [A^*]), \tag{17}$$

$$\frac{\mathrm{d}[B^*]}{\mathrm{d}t} = k_3[A^*](B_t - [B^*]) - k_4[B^*]. \tag{18}$$

In Eq. 17, parameter $V_{\max}$ denotes how strong the feedback is. Parameter K_a and Hill coefficient n indicate how sensitive the feedback is to the activated B. If n is negative, then the feedback is negative regulation. Otherwise, it represents a positive feedback.

3.3.4. How to Model Signaling Inputs?

A variety of signaling sources trigger different types of signaling pathway, which includes environment stimuli and signaling molecules. According to the specific signaling input profile, different mathematical approximations can be used to model the signaling input change over time. Two typical approximations are employed to model the step change and the depletion of signaling input (S) change.

For step change of signal, the signal profile is constant, for example the osmotic change for the stress signaling pathways in bacterial and yeast cells. In this case, the ODE for the signal input is defined as:

$$\frac{\mathrm{d}[S]}{\mathrm{d}t} = 0. \tag{19}$$

Therefore,

$$[S_t] = [S_0]. \tag{20}$$

In mammalian cells, the signaling molecules (or ligands) can be internalized into the cells and then degraded. The signaling molecules in the extracellular medium will be depleted over time. In this case, for a certain dose of signaling molecules, we can assume that the degradation of the signaling molecules is proportional to their concentration, which reads as

$$\frac{\mathrm{d}[S]}{\mathrm{d}t} = -\lambda[S]. \tag{21}$$

The signaling molecules' time course profile is a decaying function.

$$[S_t] = [S_0]\mathrm{e}^{-\lambda t} \tag{22}$$

Here, λ characterizes the speed of ligand depletion.

It is worth noting that the value of parameter λ may depend on the ligand dose and cell density (11). For a very high dose of ligand, λ is approximate to 0. Therefore, the depletion of the signaling molecules can be ignored and it is approximately constant. However, when the ligand dose is relatively low, the depletion of signaling molecules plays an important role and λ cannot be ignored. When the signaling responses to different ligand doses are modeled, it would be better to develop a more complicated mathematical model, including the ligand–receptor interaction, and internalization and degradation of the ligand.

3.4. How to Run Simulation for the Mathematical Model?

Two pieces of important information are essential for implementing simulation for the ODE mathematical model, which are (a) the initial condition of the variables and (b) the values for the reaction rate constants.

The initial condition of the ODE system corresponds to the abundance of the proteins involved in the signaling pathway. Normally, it is difficult to know the complete information about the absolute number of molecules per cell and it is still a challenge to directly measure the rate constants by in vivo experiments. Some model parameter values have to be estimated based on quantitative experimental data. The estimation of the model parameter values is called parameter estimation. The goal of parameter estimation for the ODE models is to find the possible parameter sets that minimize the difference between the model simulation data and experimental data, which is formalized as:

$$F_{\mathrm{obj}} = \sum_{i=1}^{n} \left(\frac{y_i^{\mathrm{sim}} - y_i^{\mathrm{data}}}{\sigma_i} \right)^2. \tag{23}$$

Here, F_{obj} is called the objective function (cost function) or sum of squared errors. y_i^{sim} is computer simulation data and y_i^{data} is the corresponding experimental data. σ_i represents the noise or standard deviation of the experimental data. Different optimization algorithms can be used for parameter estimation. The performance of

the algorithms usually depends on the specific models (12). Different tools are available for parameter estimations, for example Copasi (13), SBML-PET-MPI (14, 15), PottersWheel (16), and so on.

When we know the initial conditions and parameter values for the ODE model, it is possible to run computer simulation for the model with software tools. Model simulation is usually implemented with some ODE solver algorithms. The user needs to specify the ODE solver with the initial condition, parameter values, and ODE system. Some SBML software tools are able to automatically generate the ODE equations if the modeler can define all reactions involved in the pathway and set the reaction kinetics. Copasi (18) and CellDesigner (17) are two popular model construction and simulation tools. Details about the software tools are available in the user guide of these tools or other resources.

3.5. A General Workflow for Modeling of Cellular Signaling Pathways

The procedure of building a mathematical model for biological signaling pathways can be summarized as the following steps.

(a) Draw a complete reaction interaction map of the signaling pathway based on the existing biological knowledge. Occasionally, the signaling pathway might be disconnected; new reactions can be added to bridge the disconnected subsystems together. Such a map is a schematic representation of the mathematical model. The modeler can define the reactions with some software tools in different ways, for example write the reaction in text format (for example, Copasi) or draw the graphic reaction map (for example, CellDesigner).

(b) Assign appropriate reaction kinetics for all the reactions that are involved in the signaling pathway.

(c) Generate the ODE equations for the signaling pathway. For a simple system, it is possible to write it down manually. However, for large signaling network, it is more convenient to generate the ODE system by SBML software packages, which avoid typo mistakes as well.

(d) Set the initial concentration (or initial amount) of the molecules in the signaling pathway. The initial concentrations of the molecules might be unknown and need to be estimated with quantitative experimental data sets.

(e) Set the values of the parameter appeared in the reaction kinetics. If some parameter values are unknown, perform parameter estimation based on quantitative experimental data sets.

(f) Run simulations for the mathematical model. With the mathematical model, one can implement additional kinetic analysis of the signaling pathway, for example steady-state analysis, sensitivity analysis, and other perturbation analysis such as over-expression, down-regulation, knockout, and so on.

Acknowledgments

This work was supported by the Excellence Initiative of the German Federal and State Governments (EXC 294).

References

1. Boogerd FC, Bruggeman FJ, Hofmeyr J-HS, Westerhoff HV (2007) Systems biology: philosophical foundations, 1st edn. Elsevier, Amsterdam
2. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H et al (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19:524–531
3. Aldridge BB, Burke JM, Lauffenburger DA, Sorger PK (2006) Physicochemical modelling of cell signalling pathways. Nat Cell Biol 8:1195–1203
4. Guldberg CM, Waage P (1879) Concerning chemical affinity. Erdmann's J Pract Chem 127:69–114
5. Klipp E, Herwig R, Kowald A, Wierling C, Lehrach H (2005) Systems biology in practice: concepts, implementation and application, 1st edn. Wiley, Berlin
6. Hill AV (1910) The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. J Physiol 40:4–7
7. Legewie S, Bluthgen N, Herzel H (2005) Quantitative analysis of ultrasensitive responses. FEBS J 272:4071–4079
8. Szallasi Z, Stelling J, Periwal V (2006) System modeling in cellular biology: from concepts to nuts and bolts, 1st edn. The MIT, Boston, MA
9. Ferrell JE Jr (2008) Feedback regulation of opposing enzymes generates robust, all-or-none bistable responses. Curr Biol 18:R244–R245
10. Ma W, Trusina A, El-Samad H, Lim WA, Tang C (2009) Defining network topologies that can achieve biochemical adaptation. Cell 138:760–773
11. Zi Z, Klipp E (2007) Cellular signaling is potentially regulated by cell density in receptor trafficking networks. FEBS Lett 581:4589–4595
12. Moles CG, Mendes P, Banga JR (2003) Parameter estimation in biochemical pathways: a comparison of global optimization methods. Genome Res 13:2467–2474
13. Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N et al (2006) COPASI—a COmplex PAthway SImulator. Bioinformatics 22:3067–3074
14. Zi Z (2011) SBML-PET-MPI: a parallel parameter estimation tool for Systems Biology Markup Language based models. Bioinformatics 27:1028–1029
15. Zi Z, Klipp E (2006) SBML-PET: a Systems Biology Markup Language-based parameter estimation tool. Bioinformatics 22:2704–2705
16. Maiwald T, Timmer J (2008) Dynamical modeling and multi-experiment fitting with PottersWheel. Bioinformatics 24:2037–2043
17. Tanimura AFMMHKN (2003) CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. Biosilico 1:159–162
18. Mendes P, Hoops S, Sahle S, Gauges R, Dada J, Kummer U (2009) Computational modeling of biochemical networks using COPASI. Methods Mol Biol 500:17–59