IN SILICO CELL I. HIERARCHICAL APPROACH AND GENERALIZED HILL FUNCTIONS IN MODELING ENZYMATIC REACTIONS AND GENE EXPRESSION REGULATION

Likhoshvai V.A.*, Ratushny A.V.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia
* Corresponding author: e-mail: likho@bionet.nsc.ru

Key words: mathematical modeling, gene network, regulation, enzyme reaction, Hill function

SUMMARY

Motivation: Development of an in silico cell is an urgent task of systems biology. The core of this cell should consist of mathematical models of intracellular events, including enzymatic reactions and control of gene expression.

Results: In this study, we put forward a concept of hierarchical modeling of intracellular processes. A class of generalized Hill functions is determined, and a method for modeling enzymatic reactions and gene expression events with the use of this class is proposed.

INTRODUCTION

Development of an in silico cell is certainly a challenge of the postgenomic epoch in systems biology. Such large-scale projects demand huge volumes of cross-disciplinary studies involving biology, mathematics, and computer science. The cores of such systems must include knowledge bases storing comprehensive experimental and theoretical information on the cell and its processes, as well as mathematical and computer models allowing experiments in silico. The compatibility between models is essential for developing the in silico cell. Formerly, we developed a generalized chemokinetic modeling method (GCKMM) for molecular-genetic systems (MGS), which took into account genetic maps, multiple alleles, and multiple compartmentalization (Ratushny et al., 2005). On its base, we developed a unified standard of model specification SiBML (Podkolodny et al., 2006). In this report, we supplement the hierarchy-based method of elementary subsystem description used in GCCMM/SiBML with a method of modeling with the use of generalized Hill functions (GHF). Examples of modeling enzymatic reactions and genetic events are presented.

RESULTS

One of the mandatory stages of the GCKMM/SiBML approach is the construction of models of elementary subsystems (ES) of the object to be modeled. Such elementary subsystems may be: (i) enzymatic reactions; (ii) formation of complexes with biologically significant functions (active enzyme species, transcription factors, RNA polymerases, ribosomes, etc.); (iii) molecular-genetic events, including initiation and termination of replication, transcription, and translation, with regard to their regulation; (iv)
macromolecule degradation processes; and so on. Elementary models describing the rates of changes of substance concentrations are written in the differential form:

$$\frac{dX}{dt} = F(X, Y, K),$$

(1)

where $X$ is the list of dynamic variables; $K$, list of parameters; $Y$, list of functional activities; $F(X, Y, K)$, monitor rule.

For calculating quasi-steady state concentrations of functional substances, the model is presented as the function:

$$Y = G(X, Z, K),$$

(2)

where $Z$ and $Y$ are lists of input and output functional activities, and $G(X, Z, K)$ is the rule for generating functional activities from list $Y$.

Target object models (TOMs) are constructed on the grounds of a set of elementary models. The algorithm includes two stages. At the first stage, a combined set of differential equations (SDE) is constructed. The dynamic variables of the SDE are variables entering the union of all lists $X$ for all selected elementary models (1). The right side for each variable is determined as the sum of the right sides $F$ of all elementary models containing this variable in their lists $X$. The second stage involves repetitive substitution of equations $G$ from elementary models of view (2) for functional activities of list $Y$ in the right sides of the SDE. The resulting TOM is used for solving corresponding tasks.

Let us dwell on methods for defining the types of functions $F$ and $G$.

We are working out three approaches to modeling intracellular processes: nonequilibrium, quasi-steady, and GHF-based. An approach is chosen depending on the level of knowledge of processes under consideration.

The first approach involves construction of elementary subsystems. For this purpose, a biochemical model of an elementary subsystem is constructed and transformed by a standard procedure to an elementary model of view (1), where the right sides are polynomial functions of variables. Construction of portrait elementary models requires good understanding of the structure of the subsystem under study and a large volume of kinetic data.

The second approach is based on construction of a biochemical model of an elementary process and its consideration in a quasi-steady approximation. Therefore, we call such models steady-state. This approach is broadly used in enzymatic kinetics (King, Altman, 1956; Cornish-Bowden, 1977). It can be applied as is for modeling other intracellular processes, including genetic ones. This approach also demands much structural and dynamic data on the elementary process. Nevertheless, steady-state elementary models generally contain fewer variables and parameters than in the previous approach. Here, functions $F$ and $G$ are rational polynomials.

The third approach involves construction of approximating functions in the form of GHF. These functions are natural generalization of rational polynomials obtained during steady-state consideration of biochemical processes. Generally, this approach does not require any information on the mechanisms of the process under consideration. A model can be constructed directly from kinetic curves, and variables and parameters of the model correspond only to experimentally measured values. The resulting models involve the least numbers of variables.

With regard to the level of knowledge of the E. coli subsystems considered by us, we most often choose the third method for modeling mechanisms of enzymatic reactions and
regulation of genetic element expression. Therefore, we describe the GHF-based modeling method in more detail.

Definition of generalized Hill functions
Let \( X \) designate a set of variables. Denote the set of indices for elements of a subset of the set of all subsets of set \( X \) as \( A = \{ \alpha \} : N = \{ X_\alpha : X_\alpha \subseteq X, \alpha \in A \} \). Null set \( \emptyset \) may enter into \( \mathbb{N} \). Let \( K, N, \text{ and } \Delta \) designate the sets of parameters called efficiency coefficients (EC), Hill coefficients (HC), and activity coefficients (AC), respectively: \( \{ k_\alpha : \alpha \in A \} = K \), \( \{ n_{\alpha,a} : x \in X_\alpha, \alpha \in A \} = N \), \( \{ \delta_\alpha : \alpha \in A \} = \Delta \). By implication of \( k_\alpha, n_{\alpha,a} > 0, \delta_\alpha \geq 0 \).

1. \( h \left( x \mid x \in X \right) = R \left( X \right)/Q \left( X \right) = \sum_{\alpha} \delta_\alpha \prod_{x \in X_\alpha} \left( x/k_\alpha \right)^{n_{\alpha,a}}/\sum_{\alpha} \prod_{x \in X_\alpha} \left( x/k_\alpha \right)^{n_{\alpha,a}} \) - This rational polynomial is a GHF (by convention, if \( X_\alpha \emptyset \), then \( \prod_{x \in X_\alpha} \left( x/k_\alpha \right)^{n_{\alpha,a}} = 1 \)).

2. If \( h \left( X, K, N, \Delta \right) \) is a GHF, then \( \tilde{h} \left( X, K \left( X, K, N, \Delta \right), N \left( X, K, N, \Delta \right), \Delta \right) \), obtained by substitution of \( K \) and \( N \) parameters to the GHF, is also a GHF.

Description of the GHF model constructing procedure.
Experimental data are reported in terms of the (discrete) function \( E : S \subseteq DR^n \rightarrow R \), where \( n \) is the dimension of vector \( X \). Point \( S_i \) of \( S \) is associated with the corresponding value of the function \( E(S) \) and determines the rate of an enzymatic reaction or gene expression efficacy. The problem is to construct the continuous (on all arguments) GHF having the least deviation from function \( E \) at corresponding points (on \( S \)). Note that this problem is soluble by virtue of the validity of theorems about generalized Hill function completeness (Bazaikin et al., Private report).

The process of construction begins with choice of the vector \( X \) variable to be considered at the first stage of the process. As a rule, the first variable has the most informative scripts (experimental data). The script is understood as a discrete function constructed from experimental data for variation of the chosen substance under identical conditions with regard to all other substances. It is obvious that the more points are there in one script, the more information such function contains. Without loss of generality, we may select the first variable of vector \( X \). Then we construct the common structured GHF of only one selected variable which allows reproducing each script, but it is possible with different sets of kinetic and Hill parameters. We designate the constructed function as \( H_1(x_1) \). The result of the first stage is the generalized Hill function \( H_1(x_1) \) of a universal form and tables of all its parameters obtained by varying variables \( X_1 = X \setminus \{ x_1 \} \). Then we consider the coefficients of the function \( H_1(x_1) \) as generalized Hill functions of \( n-1 \) variables and the tables of parameters as an analogue of experimental data (it should be noted that not all of \( H_1(x_1) \) parameters vary at different conditions. Some of them can be constant, and for such parameters we finish the process).

The next stage is to construct the GHF for each parameter. The procedure of its definition is the same as described above. For each parameter we select the most informative variable, prepare the scripts and construct the universal function of one variable and the corresponding table of parameters. After definition of the GHF for each parameter we pass to the following stage. We repeat the above described procedure until all variables are exhausted.
As a result we get set of generalized Hill functions of view (1) with certain coefficients and the order of their composition to uniform generalized Hill function of view (2).

**Regulation of the expression of the cydAB operon**

As an example, consider transcription initiation in the *cydAB* operon (Fig. 1).

![Figure 1. Structure of the promoter of the cydAB operon, coding for cytochrome bd oxidase. Retrieved from Ecocyc database (http://ecocyc.org). Factor ArcA is a transcription activator, H-NS is a repressor, and Fnr can be both an activator (left binding site of the second promoter) and a repressor.]

The transcription initiation region has a complex structure: 5 transcription starts and 10 binding sites within transcription sites ArcA (5 sites), H-NS (3 sites), and Fnr (2 sites). Therefore, the portrait of this subsystem involves 510 bimolecular reactions with 197 dynamic variables (Fig. 2). Verification of its parameters demands detailed dynamic information on the operation of the subsystem depending on transcription factor concentrations. At present, such data are absent from available sources. However, this subsystem has been studied in the context of oxygen concentration variation (Tseng *et al.*, 1996). Therefore, we model it with generalized Hill functions.

![Figure 2. Bipartite graph of the mathematical model constructed in terms of mono- and bimolecular reactions to describe the regulation of transcription initiation in the promoter region of the cydAB operon. The graph contains 510 elementary processes (white square nodes and 197 dynamic variables (gray circular nodes).]
Modelling of molecular genetic systems in bacterial cell

Thus, we obtain the following model equation describing the level of cydAB expression \( f_{a_f \Delta_f} \) in the wild type \( (\Delta_u = 1, \Delta_f = 1) \) and knockout strains \( \Delta \text{arcA} \) \( (\Delta_u = 0, \Delta_f = 1) \) and \( \Delta \text{fnr} \) \( (\Delta_u = 1, \Delta_f = 0) \) depending on oxygen concentration:

\[
f_{a_f \Delta_f} = \frac{k_{a_f \Delta_f} + \Delta_u \left[ \left( \frac{O_2}{k_{a1}} \right)^{h_{a1}} + \left( \frac{O_2}{k_{a21}} \right)^{h_{a21}} \right] + \Delta_f \left[ \left( \frac{O_2}{k_{f1}} \right)^{h_{f1}} + \frac{O_2}{k_{f21}} \right] + \Delta_u \Delta_f \left( \frac{O_2}{k_{a2}} \right)^{h_{a2}}}{1 + \Delta_u \left[ \left( \frac{O_2}{k_{a21}} \right)^{h_{a21}} + \left( \frac{O_2}{k_{a22}} \right)^{h_{a22}} \right] + \Delta_f \left[ \left( \frac{O_2}{k_{f2}} \right)^{h_{f21}} + \frac{O_2}{k_{f22}} \right] + \Delta_u \Delta_f \left( \frac{O_2}{k_{f2}} \right)^{h_{f22}}}, \tag{3}
\]

where \( O_2 \) is oxygen concentration in the medium; \( k_{a_f \Delta_f} \), basal expression of the cydAB operon; \( k_{a1}, k_{a2}, k_{a21} \), are constants of the effect of ArcA, Fnr, and their combined effect on cydAB operon expression, respectively; \( h_{a1}, h_{f1}, h_{a2}, \) constants describing the nonlinearity of the effect of ArcA, Fnr, and their combined effect on cydAB expression, respectively; expressed in terms of oxygen concentration.

Fig. 3 shows the results of comparison between calculations according to model (3) and experimental data describing the dependence of the rate of cydAB transcription on oxygen concentration for wild-type \( E. \ coli \) cells \((f_{cydABWT})\) and for mutant strains \((f_{cydAB\DeltaarcA})\) and \((f_{cydAB\Deltafnr})\) (Tseng et al., 1996).

By now, the accumulated knowledge (Khlebodarova et al., 2006) has allowed construction of about 300 elementary enzymatic reaction models and about 20 elementary models describing the regulation of expression of various genes according to the method presented here. More examples of constructing mathematical models of enzymatic reactions and gene expression regulation using generalized Hill functions are provided in (Ananko et al., 2006; Nedosekina et al., 2006; Ratushny et al., 2006a, b, c).

![Figure 3](image)

**Figure 3.** Dependence of the cydAB operon transcription on oxygen concentration in wild-type \( E. \ coli \) cells \((f_{1,1})\) and mutant strains \( \Delta \text{arcA} \) \((f_{0,1})\) and \( \Delta \text{fnr} \) \((f_{1,0})\). Dots indicate experimental data (Tseng et al., 1996). Curves indicate calculations from Eq. (3). The curves were calculated with the following parameters: \( k_{a1} = 3; k_{a2} = 5; k_{a21} = 20 \ \mu M; k_{a22} > 500 \ \mu M; h_{a1} = 1; k_{f1} = 20 \ \mu M; k_{f2} = 6; k_{f21} = 20 \ \mu M; k_{f22} = 20 \ \mu M; h_{f1} = 0.4; k_{f2} = 9 \ \mu M; h_{f2} = 1; k_{f21} = 11 \ \mu M; h_{f22} = 1.4; k_{a21} = 1.2 \ \mu M; h_{a2} = 1; k_{a22} = 100 \ \mu M; h_{a22} = 6. \)
ACKNOWLEDGEMENTS

The authors are grateful to Eric Mjolsness, Tamara Khlebodarova, Evgeniya Nedosekina, Sergey Lashin and Nikolay Kolchanov for valuable discussion and to Victor Gulevich for translating the manuscript from Russian into English. Work was supported in part by the Russian Foundation for Basic Research (Grant No. 04-01-00458), NSF:FIBR (Grant No. EF–0330786), by the Russian Government (Contract No. 02.467.11.1005), and by the Federal Agency of Science and Innovation (innovation project No. IT-CP.5/001).

REFERENCES


Khlebodarova T.M. et al. (2006) In silico cell II. Information sources for modeling Escherichia coli metabolic pathways. This issue.


Podkolodny N.L. et al. (2006) An integration of the descriptions of gene networks and their models presented in sigmoid (cellerator) and genenet. This issue.