New computer technologies for construction and numerical analysis of mathematical models of molecular genetic systems

I. R. Akberdin[2], S.I. Fadeev[1], I. A. Gainova[1*], F. V. Kazantsev[2], V. K. Korolev[1], V. A. Likhoshvai[2] and A. E. Medvedev[3]

[1] *Sobolev Institute of Mathematics,*
Siberian Branch of the Russian Academy of Sciences,
Koptyug Avenue, 4, Novosibirsk, 630090, Russia
[*] *E-mail: gajnova@math.nsc.ru; fadeev@math.nsc.ru; korolev@math.nsc.ru*

[2] *Institute of Cytology and Genetics,*
Siberian Branch of the Russian Academy of Sciences,
Lavrentyev Avenue, 10, Novosibirsk, 630090, Russia
E-mail: akberdin@bionet.nsc.ru; kazfdr@bionet.nsc.ru; likho@bionet.nsc.ru

[3] *Khristianovich Institute of Theoretical and Applied Mechanics,*
Siberian Branch of the Russian Academy of Sciences,
Institutskaya Street, 4/1, Novosibirsk, 630090, Russia
E-mail: medvedev@itam.nsc.ru

We have represented a new computer system for generation and analysis of mathematical models describing the dynamics of the molecular genetic systems functioning in pro- and eukaryotes. The system consists of two program modules: MGSgenerator and STEP+.

Keywords: Molecular-genetic systems (MGS); Numerical analysis; Mathematical model; Nonlinear system; Program module.

1. Introduction

Mathematical models of molecular-genetic systems are based on the information about the structural and functional organization of gene networks and their dynamic properties disseminated over hundreds and thousands of scientific papers. The problem arises of data comparison and analysis of non-uniformed experimental data, analysis of cause-and-effect relations between molecular structure, dynamics and phenotypic features of MGS, and software development for automatic generation of mathematical models, storage of creating models in the database and their numerical analysis.

In the context of solving some of the above mentioned problems, a group of researchers from the Institute of Cytology and Genetics and Sobolev Institute of Mathematics is developing an integrated computer system that does not only render automatically the process of reconstruction of mathematical models based on the structural and functional organization of gene networks but also implements original approaches and algorithms to modeling and studying molecular-genetic systems. Part of this system is the computer system consisting of two interlinked program modules (MGSgenerator [1] and STEP+ [2]), presented in the given paper.

2. Description of the MGSgenerator program module

The MGSgenerator program module generates automatically a mathematical model from the structural model of a gene network extracted from the GeneNet database [3] and delivers it in an appropriate computer format. The model can be exported to SiBML [4], SBML [5], and Step+ formats.

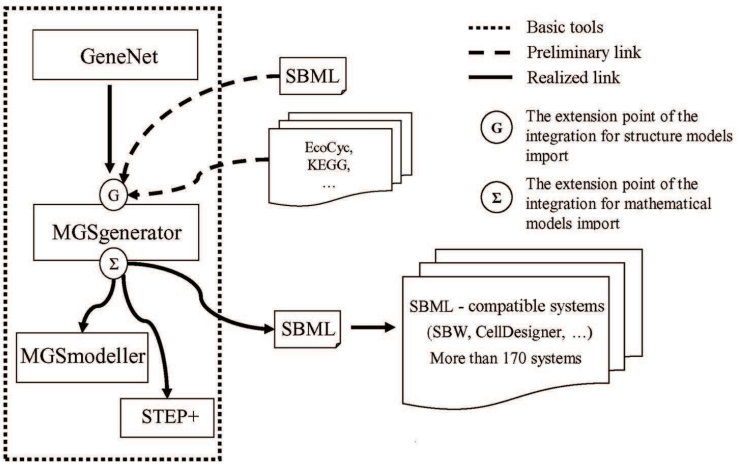

Fig. 1. Diagram of the MGSgenerator software.

The construction of a mathematical model is based on the generalized chemical-kinetic method [6] that allows describing the molecular-genetic processes linking the components of the gene network, by a nonlinear system of differential equations. The MGSgenerator matches every biological process to some template function and, based on this information, generates a mathematical model. To describe the biological processes, template

models of three various levels of complexity are used that form the functions determining the right-hand sides of the system of ordinary differential equations.

So, the processes of replication, transcription, and translation with ignored regulatory interactions are described by the simple equation

$$V = xk_0\,,$$

where V is the rate of the process, x is the number of genes (concentrations), mRNA in a cell, etc., and k_0 is the reaction rate constant.

If molecular-genetic regulation of these processes is taken into account, another template model is used:

$$V = xk_0\left[\frac{\delta + \sum\limits_i \left(\frac{a_i}{k_{1,ai}}\right)^{h_i}}{1 + \sum\limits_i \left(\frac{a_i}{k_{2,ai}}\right)^{hd_i} + \sum\limits_i \left(\frac{r_i}{k_{ri}}\right)^{g_i}}\right].$$

Here x is a chosen element (gene/RNA/enzyme), a_i is the activator concentration, r_i is the inhibitor concentration, k_0 is the reaction rate constant, k_{ai} is the activator constant, k_{ri} is the inhibitor constant, δ is the basal activity, and h_i, g_i and hd_i are the Hill coefficients. By default, the interaction of regulators with each other is assumed to be competitive, and their interaction with the chosen element (x) is assumed to be non-competitive.

The mathematical description of enzymatic synthesis with allowance for molecular-genetic regulation and reaction reversibility is even more complicated:

$$V = x\frac{k_{s,0}\prod\limits_i \frac{S_i}{k_{mS,i}} - k_{p,0}\prod\limits_i \frac{P_i}{k_{mP,i}}}{\prod\limits_i \left(1 + \frac{S_i}{k_{mS,i}} + \frac{P_i}{k_{mP,i}}\right)} \times \frac{\delta + \sum\limits_i \left(\frac{a_i}{k_{1,ai}}\right)^{h_i}}{1 + \sum\limits_i \left(\frac{a_i}{k_{2,ai}}\right)^{hd_i} + \sum\limits_i \left(\frac{r_i}{k_{ri}}\right)^{g_i}}\,.$$

Here x is the enzyme (regulator), S is the substrate (input), P is the product (output), a_i is the activator concentration, r_i is the inhibitor concentration, k_0 is the reversal constant, $k_{s,0}$ is the constant of the direct reaction, $k_{p,0}$ is the constant of the reverse reaction, k_{mS} is the Michaelis constant for the substrate, k_{mP} is the Michaelis constant for the product, k_{ai} is the activator constant, k_{ri} is the inhibitor constant, δ is the basal activity, and h_i, g_i and hd_i are the Hill coefficients.

The MGSgenerator software was created with the use of plug-in technology and has a module architecture. Such a structure of the generator allows for the attachment of new data sources, export models from other

databases, such as KEGG [7], EcoCyc [8], etc., change or add to existing program modules, and integrate the given module into other computer systems.

3. Description of the STEP+ program module

The STEP+ program module allows a comprehensive study of the mathematical model generated by the MGSgenerator in the STEP+ input format in the form of an autonomous system composed of n differential equations with the vector of parameters p:

$$\dot{x} = f(x, p). \tag{1}$$

Here $f(x, p)$ is a sufficiently smooth vector-function of the vector arguments $x \in R^n$ and $p \in R^m$ in the domain of its definition.

The STEP+ program module includes algorithms for numerical analysis of the solution of the autonomous system, depending on the model parameters: multistep Gear method with a variable order of accuracy for integration of stiff systems [9]; method of solution continuation with respect to a parameter for constructing the stationary solutions diagram of the autonomous system (1) [10, 11]; Godunov-Bulgakov numerical criterion for determining the guaranteed asymptotic stability of stationary solutions, depending on the model parameter $\alpha \in p$, [12].

3.1. *Continuation method and parametrization*

The continuation method with respect to a parameter is also used to study the dependence of the solution of a system of nonlinear equations (not related to the autonomous system)

$$f(x, \alpha) = 0\,, \tag{2}$$

on a scalar parameter α, $\alpha \in p \in R^m$. The plots of one or several components of the vector-function $x = x(\alpha)$ which is the solution of Eq. (2) will be referred to as the diagram of stationary solutions. The method is based on the implicit function theorem. According to this theorem, the plot of the solution of Eq. (2) in the $(n + 1)$-dimensional space will be a smooth space curve if in a neighborhood of the space curve the rank of matrix A $(A = [f_x, f_\alpha])$ of the derivatives of the right-hand sides is always equal to n, regardless of α. Notice that the smooth space curve may intersect the hyperplane $\alpha = \alpha^*$, $\alpha^* \in [\alpha_0, \alpha_1]$ several times, meaning that there is a multiplicity of solutions of (2) when $\alpha = \alpha^*$.

Let the solution of Eq. (2) be known for a certain value of α and A be the matrix formed by this solution. As the rank of A is n, there is a nondegenerate square matrix $B = [f_z]$ that is obtained from A by deleting some k-th column. Denote by z the vector whose components are equal to those of the vector (x, α) obtained by deleting the component with the index k, which will be denoted by μ. By defining the derivative vector z_μ as a solution of the system with the matrix B and the right-hand side f_μ, let us find the maximal by module component of vector $(z_\mu, 1)$ with the index j. The vector consisting of the components of vector (x, α) after deleting the j-th component is denoted by u, while the component deleted is denoted by λ.

Vector u can be considered as a solution of Eq. (2) at the same value of α, but with parameter λ. In this case, the derivative vector of the solution with respect to λ can be found by normalization of the components of vector $(z_\mu, 1)$, using its maximal by module component for the normalization. As a result, this maximal by module component of the derivative vector will be equal to 1. This means that λ can be seen as a parameter of Eq. (2) in the neighborhood of the solution under consideration.

The procedure of determining λ will be referred to as parametrization, and λ itself as a current parameter of Eq. (2) as it is defined regularly for one $\Delta\lambda$ step in the continuation method. The role of the current parameter can be equitably taken by any component of vector (x, α). To sum, the solution continuation with respect to the parameter, together with the parametrization and adaptation of the current step, allows for plotting a smooth space curve defined by Eq. (2) that may have turning points and, consequently, have several solutions within some ranges of α.

3.2. *Godunov-Bulgakov method*

The STEP+ program module contains a numeric criterion of the guaranteed asymptotic stability of the stationary solution developed under the guidance of academician S.K.Godunov. The method, which does not require calculations of eigenvalues of the Jacobi matrix ($J = [f_x]$) of Eq. (1), is based on estimating the norm of the solution H of the Lyapunov matrix equation

$$J^*H + HJ = -E\,,$$

where E is a unit matrix, J^* is a matrix adjoint to J. Matrix J is a Hurwitz matrix (i.e. the considered stationary solution is asymptotically stable) if H is a Hermitian positive-definite matrix. In this case, H has an integral

presentation in which the matrix exponents of J and J^* are used:

$$H = \int_0^\infty \exp(tJ^T)\exp(tJ)\,dt\ .$$

The numeric characteristic $\kappa(J)$ of asymptotic stability when J is a Hurwitz matrix has the form

$$\kappa(J) = 2\|J\| \sup_{v(0)\neq 0} \frac{\int_0^\infty \|v(t)\|^2\,dt}{\|v(0)\|^2}\ ,\ v(t) = [\exp(tJ)]v(0)\,.$$

At each iteration in the process of calculating $\kappa(J)$, the verification for the inequality is carried out: $\bar{\kappa}(J) < k_0$, where $\bar{\kappa}(J)$ is the approximation of $\kappa(J)$ at the iteration, and k_0 is a constant whose value depends only on the precision of number representation by the computer. The asymptotic stability is guaranteed if the limiting value $\kappa(J)$ also satisfy this inequality. If, at some iteration, the inequality is not satisfied, then the guaranteed conclusion about 'practical' instability of the considered stationary solution is given.

The interface of STEP+ module automatically constructs the Jacobi matrix and the matrix of partial derivatives of the right-hand sides of the system with respect to the parameters. The analytic representation of the matrix elements allows us to perform a numerical study of the system solution depending on an arbitrary model parameter of interest.

4. Conclusion

The system has been tested on the gene network that regulates auxin metabolism in the meristem of shoots of *Arabidopsis thaliana L* [13,14]. Not only has testing the system of mathematical models automatic generation with MGSgenerator module and numerical analysis of the model by STEP+ module demonstrated its validity, but it has also led to interesting biological results on the dynamics of auxin metabolism functioning as well as the evolution of its transformations.

Acknowledgments

The work has been conducted during the Special Semester on Quantitative Biology Analyzed by Mathematical Methods, October 1^{st}, 2007 – January 27^{th}, 2008, organized by RICAM, Austrian Academy of Sciences.

This work has been partially supported by the Russian Foundation for Basic Research (Grant Nos. 09-01-08109, 08-01-91204-YaF_a), by the Grant NSh-2447.2008.4, and by the SB RAS Interdisciplinary project Nos. 91, 107, and 119, and also by the RAS projects 21.26, 22.8, 23.29.

References

1. F. V. Kazantsev, I. R. Akberdin, K. D. Bezmaternykh, and V. A. Likhoshvai, *System of automated generation of mathematical models of gene networks*, VOGiS Herald journal, **13** 1, (2009).
2. S. I. Fadeev, V. K. Korolev, I. A. Gainova, A. E. Medvedev, *The package Step+ for numerical study of autonomous systems arising when modeling dynamics of genetic-molecular systems*, in *Proc. of the 6th Intern. Conf. on Bioinformatics of Genome Regulation and Structure (BGRS-06)*, **2** (2006).
3. E. A. Ananko, N. L. Podkolodny, I. L. Stepanenko, O. A. Podkolodnaya, D. A. Rasskazov, D. S. Miginsky, V. A. Likhoshvai, A. V. Ratushny, N. N. Podkolodnaya, and N. A. Kolchanov, *GeneNet in 2005*, Nucleic Acids Res. Vol., **33** (2005).
4. V. A. Likhoshvai, Yu. G. Matushkin, A. V. Ratushny, E. A. Ananko, E. V. Ignatyeva, O. V. Podkolodnaya, *Generalized chemical-kinetic approach for modeling gene networks*, Molecular biology, **3**, 6 (2001).
5. M. Hucka, A. Finney, H. Sauro, H. Bolouri, J. Doyle, H. Kitano, et al., *The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models*, Bioinformatics. March (2003).
6. V. Likhoshvai and A. Ratushny, *Generalized Hill function method for modeling molecular processes*, J. of Bioinformatics and Comp. Biology, **5**, 2 (2007).
7. M. Kanehisa and S. Goto, *KEGG: Kyoto Encyclopedia of Genes and Genomes*, Nucleic Acids Res., **28** (2000).
8. I. Keseler, J. Collado-Vides, S. Gama-Castro, J. Ingraham, S. Paley, I. Paulsen, M. Peralta-Gil and P. Karp, *EcoCyc: a comprehensive database resource for Escherichia coli*, Nucleic Acids Res. Vol., **33** (2005).
9. G. W. Gear, *The automatic integration of ordinary differential equations*, Comm. ACM.V., **14**, 3 (1971).
10. S. I. Fadeev, S. A. Pokrovskaya, A. Yu. Berezin, and I. A. Gainova, STEP program package for a numerical study of systems of nonlinear equations and autonomous systems of the general form. Description of STEP package operation by examples of the tasks from the educational course "Engineering chemistry of catalytic processes". Tutorial. (Izd. Novosib. Gos. Univ., Novosibirsk, Russia, 1998).
11. S. I. Fadeev and V. V. Kogai, *Using parameter continuation based on the multiple shooting method for numerical research of nonlinear boundary value problems*, Int. J. of Pure and Applied Mathematics, **14**, 4 (2004).
12. S. K. Godunov, Ordinary Differential Equations with Constant Coefficients. (Izd. Novosib. Gos. Univ., Novosibirsk, 1994).
13. I. R. Akberdin, N. A. Omelyanchuk, S. I. Fadeev, V. M. Efimov, I. A. Gainova and V. A. Likhoshvai, *Mathematical model of auxin metabolism in shoots of Arabidopsis thaliana L*, Proc. of the 6th Intern. Conf. on Bioinformatics of Genome Regulation and Structure (BGRS'2008), (2008).
14. I. R. Akberdin, F. V. Kazantsev, N. A. Omeliyanchuk and V. A. Likhoshvai, *Mathematical model of auxin metabolism in meristem cells of plant shoots*, VOGiS Herald journal, **13**, 1 (2009).