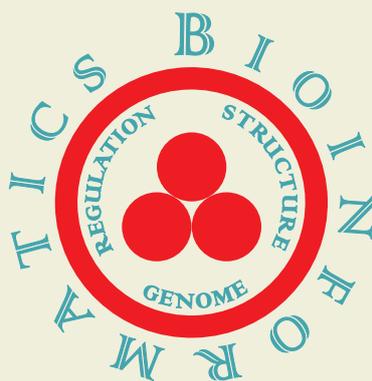


**PROCEEDINGS
OF THE FIFTH
INTERNATIONAL CONFERENCE
ON BIOINFORMATICS
OF GENOME REGULATION
AND STRUCTURE**



VOLUME 2

**BGRS' 2006
NOVOSIBIRSK, RUSSIA
JULY 16 - 22, 2006**

**RUSSIAN ACADEMY OF SCIENCES
SIBERIAN BRANCH**

INSTITUTE OF CYTOLOGY AND GENETICS

**PROCEEDINGS
OF THE FIFTH
INTERNATIONAL CONFERENCE
ON BIOINFORMATICS
OF GENOME REGULATION
AND STRUCTURE**

Edited by
N. Kolchanov, R. Hofestädt

Volume 2

**BGRS'2006
Novosibirsk, Russia
July 16–22, 2006**

**Novosibirsk
2006**

INTERNATIONAL PROGRAM COMMITTEE

Nikolay Kolchanov Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
(Chairman of the Conference)

Ralf Hofstadt University of Bielefeld, Germany
(Co-Chairman of the Conference)

Dagmara Furman Institute of Cytology and Genetics SB RAS, Novosibirsk,
(Conference Scientific Secretary)

Dmitry Afonnikov Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

Mikhail Gelfand GosNIIGenetika, Moscow, Russia

Vadim Govorun Institute of Physicochemical Medicine, RAMS, Moscow, Russia

Reinhart Heinrich Humboldt University Berlin, Berlin, Germany

Charlie Hodgman Multidisciplinary Centre for Integrative Biology,
School of Biosciences, University of Nottingham, UK

Alexey Kochetov Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

Eugene Koonin National Center for Biotechnology Information,
National Library of Medicine, National Institutes of Health, Bethesda, USA

Vassily Lyubetsky Institute for Informational Transmission Problems RAS,
Moscow, Russia

Luciano Milanesi National Research Council – Institute of Biomedical Technology, Italy

Viatcheslav Mordvinov Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

Yuriy Orlov Genome Institute of Singapore,
Laboratory of Information & Mathematical Sciences, Singapore

Igor Rogozin Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

Cenk Sahinalp Computing Science, Simon Fraser University, Burnaby, Canada

Maria Samsonova St.Petersburg State Polytechnic University, St.Petersburg, Russia

Akinori Sarai Kyushu Institute of Technology (KIT), Iizuka, Japan

Konstantin Skryabin Centre "Bioengineering" RAS, Moscow, Russia

Rustem Tchuraev Institute of Biology, Ufa Sci. Centre RAS, Ufa, Russia

Denis Thieffry ESIL-GBMA, Universite de la Mediterranee, Marseille, France

Jennifer Trelewicz IBM Almaden Research Center, San Jose, California, USA

Edgar Wingender UKG, University of Goettingen, Goettingen, Germany

Lev Zhivotovsky Institute of General Genetics RAS, Moscow, Russia

Jagath C. Rajapakse School of Computer Engineering,
Nanyang Technological University, Singapore

LOCAL ORGANIZING COMMITTEE

Sergey Lavryushev Institute of Cytology and Genetics SB RAS, Novosibirsk
(Chairperson)

Ekaterina Denisova Institute of Cytology and Genetics SB RAS, Novosibirsk

Andrey Kharkevich Institute of Cytology and Genetics SB RAS, Novosibirsk

Galina Kiseleva Institute of Cytology and Genetics SB RAS, Novosibirsk

Anna Onchukova Institute of Cytology and Genetics SB RAS, Novosibirsk

Yuri Orlov Institute of Cytology and Genetics, Novosibirsk

Natalia Sournina Institute of Cytology and Genetics SB RAS, Novosibirsk

Our sponsors

Organizers



Laboratory of Theoretical Genetics
of the Institute of Cytology and Genetics, SB RAS



Institute of Cytology and Genetics, SB RAS



Siberian Branch of the Russian Academy of Sciences



Vavilov Society of Geneticists and Breeders



Scientific Council on Bioinformatics,
Siberian Branch of the Russian Academy of Sciences
The Chair of Informational Biology
of the Department of Natural Sciences
of Novosibirsk State University

Grants



INTAS Grant



Russian Foundation for Basic Research

Contents

PART 3. COMPUTATIONAL SYSTEMS BIOLOGY

3.1. MODELLING OF MOLECULAR GENETIC SYSTEMS AND PROCESSES IN BACTERIAL CELL

<i>IN SILICO</i> CELL I. HIERARCHICAL APPROACH AND GENERALIZED HILL FUNCTIONS IN MODELING ENZYMATIC REACTIONS AND GENE EXPRESSION REGULATION <i>Likhoshvai V.A., Ratushny A.V.</i>	13
<i>IN SILICO</i> CELL II. INFORMATION SOURCES FOR MODELING <i>ESCHERICHIA COLI</i> METABOLIC PATHWAYS <i>Khlebodarova T.M., Ananko E.A., Nedosekina E.A., Podkolodnaya O.A., Poplavsky A.S., Smirnova O.G., Ibragimova S.S., Mishchenko E.L., Stepanenko I.L., Ignatieva E.V., Proskura A.L., Nishio Y., Imaizumi A., Usuda Y., Matsui K., Podkolodny N.L., Kolchanov N.A.</i>	19
MATHEMATICAL MODELING OF ELEMENTARY PROCESSES OF THE GENE NETWORK CONTROLLING HISTIDINE BIOSYNTHESIS IN <i>ESCHERICHIA COLI</i> <i>Ratushny A.V., Usuda Y., Matsui K., Podkolodnaya O.A.</i>	25
MODELING OF THE EFFECTS OF THREONINE, VALINE, ISOLEUCINE AND PYRIDOXAL 5'-MONOPHOSPHATE ON BIOSYNTHETIC THREONINE DEHYDRATASE REACTION <i>Nedosekina E.A., Usuda Y., Matsui K.</i>	30
REGULATION OF PYRIMIDINE BIOSYNTHESIS IN <i>ESCHERICHIA COLI</i> : GENE NETWORK RECONSTRUCTION AND MATHEMATICAL MODELING <i>Ratushny A.V., Nedosekina E.A.</i>	35
REGULATION OF THE PENTOSE PHOSPHATE PATHWAY IN <i>ESCHERICHIA COLI</i> : GENE NETWORK RECONSTRUCTION AND MATHEMATICAL MODELING OF METABOLIC REACTIONS <i>Ratushny A.V., Smirnova O.G., Usuda Y., Matsui K.</i>	40
AROMATIC AMINO ACID BIOSYNTHESIS IN <i>ESCHERICHIA COLI</i> : GENERALIZED HILL FUNCTION MODEL OF THE TRYPTOPHAN-SENSITIVE 3-DEOXY-D-ARABINO- HEPTULOSONATE-7-PHOSPHATE SYNTHASE REACTION DEMONSTRATE COMPLICATED MECHANISM <i>Ananko E.A., Ratushny A.V., Usuda Y., Matsui K.</i>	45
MATHEMATICAL MODELING OF REGULATION OF <i>cyoABCDE</i> OPERON EXPRESSION IN <i>ESCHERICHIA COLI</i> <i>Ratushny A.V., Khlebodarova T.M.</i>	49
GENE NETWORK RECONSTRUCTION AND MATHEMATICAL MODELING OF <i>E. COLI</i> RESPIRATION: REGULATION OF F ₀ F ₁ -ATP SYNTHASE BY METAL IONS <i>Khlebodarova T.M., Lashin S.A., Apasieva N.V.</i>	55

MATHEMATICAL MODELING OF REGULATION OF <i>ESCHERICHIA COLI</i> PURINE BIOSYNTHESIS PATHWAY ENZYMATIC REACTIONS <i>Nedosekina E.A.</i>	60
RECONSTRUCTION AND MATHEMATICAL MODELING OF THE GENE NETWORK CONTROLLING CYSTEINE BIOSYNTHESIS IN <i>ESCHERICHIA COLI</i> : REGULATION OF SERINE ACETYLTRANSFERASE ACTIVITY <i>Mishchenko E.L., Lashin S.A., Usuda Y., Matsui K.</i>	68
GENE NETWORK RECONSTRUCTION AND MATHEMATICAL MODELING OF SALVAGE PATHWAYS: REGULATION OF ADENINE PHOSPHORIBOSYLTRANSFERASE ACTIVITY BY STRUCTURALLY SIMILAR SUBSTRATES <i>Smirnova O.G., Ratushny A.V.</i>	73
MATHEMATICAL MODELING OF SERINE AND GLYCINE BIOSYNTHESIS REGULATION IN <i>ESCHERICHIA COLI</i> <i>Turnaev I.I., Ibragimova S.S., Usuda Y., Matsui K.</i>	78
REGULATION OF PYRUVATE BIOSYNTHESIS IN <i>ESCHERICHIA COLI</i> : GENE NETWORK RECONSTRUCTION AND MATHEMATICAL MODELING OF ENZYMATIC REACTIONS OF THE PATHWAY <i>Turnaev I.I., Smirnova O.G.</i>	84
BiotechPro: A DATABASE FOR MICROBIOLOGICALLY SYNTHESIZED PRODUCTS OF BIOTECHNOLOGICAL VALUE <i>Ibragimova S.S., Smirnova O.G., Grigorovich D.A., Khlebodarova T.M.</i>	89
DATABASE GenSensor AS INFORMATIONAL SOURCE FOR DESIGN OF BIOSENSORS. EXPERIMENTAL DEVELOPMENT OF BIOSENSOR BASED ON <i>yfiA</i> GENE <i>Khlebodarova T.M., Kachko A.V., Stepanenko I.L., Tikunova N.V.</i>	93
ABOUT RECONSTRUCTION OF REGULATORY MECHANISM OF GENE ELEMENTS' EXPRESSION <i>Peshkov I.M., Fadeev S.I.</i>	97

3.2. MODELLING OF MOLECULAR GENETIC SYSTEMS AND PROCESSES IN MULTICELLULAR ORGANISMS

A MATHEMATICAL MODEL FOR THE INFLUENZA VIRUS LIFE CYCLE <i>Akberdin I.R., Kashevarova N.A., Khlebodarova T.M., Bazhan S.I., Likhoshvai V.A.</i>	105
A MINIMUM MATHEMATICAL MODEL FOR SUPPRESSION HCV RNA REPLICATION IN CELL CULTURE <i>Bezmaternikh K.D., Mishchenko E.L., Likhoshvai V.A., Khlebodarova T.M., Ivanisenko V.A.</i>	109
A MATHEMATICAL MODEL OF IMMUNE RESPONSE IN INFECTION INDUCED BY MYCOBACTERIA TUBERCULOSIS. PREDICTION OF THE DISEASE COURSE AND OUTCOMES AT DIFFERENT TREATMENT REGIMENS <i>Bazhan S.I., Schwartz Ya.Sh., Gainova I.A., Ananko E.A.</i>	114
THE PACKAGE STEP+ FOR NUMERICAL STUDY OF AUTONOMOUS SYSTEMS ARISING WHEN MODELING DYNAMICS OF GENETIC-MOLECULAR SYSTEMS <i>Fadeev S.I., Korolev V.K., Gainova I.A., Medvedev A.E.</i>	118
MATRIX PROCESS MODELING: STUDY OF A MODEL OF SYNTHESIS OF LINEAR BIOMOLECULES WITH REGARD TO REVERSIBILITY OF PROCESSES <i>Fadeev S.I., Shtokalo D.N., Likhoshvai V.A.</i>	121
APPROXIMATE STATIONARY ATTRACTORS IN <i>DROSOPHILA</i> GAP GENE CIRCUITS IN THE LIMIT OF STEEP-SIGMOID INTERACTIONS <i>Gursky V.V., Samsonov A.M., Reinitz J.</i>	125
SIGNAL TRANSDUCTION PATHWAYS INVOLVED IN TRANSCRIPTIONAL REGULATION OF TYROSINE HYDROXYLASE <i>Kalashnikova E.V., Dymshits G.M., Kolpakov F.A.</i>	128

Cyclonet – AN INTEGRATED DATABASE ON CELL CYCLE REGULATION AND CARCINOGENESIS <i>Kolpakov F., Poroikov V., Sharipov R., Milanesi L., Kel A.</i>	132
A GENE NETWORK OF ADIPOCYTE LIPID METABOLISM: COORDINATED INTERACTIONS BETWEEN THE TRANSCRIPTION FACTORS SREBP, LXR α AND PPAR γ <i>Kouznetsova T.N., Ignatieva E.V., Oshchepkov D.Yu., Kondrakhin Y.V., Katokhin A.V., Shamanina M.Y., Mordvinov V.A.</i>	137
NEW MIGRATION SCHEME FOR PARALLEL DIFFERENTIAL EVOLUTION <i>Kozlov K.N., Samsonov A.M.</i>	141
HCCDB: DATA MINING SYSTEM FOR HUMAN CELL CYCLE GENE <i>Milanesi L., Alfieri R., Merelli I.</i>	145
HCV-KINET DATABASE: KINETIC PARAMETER REACTIONS AND REGULATORY PROCESSES OF THE LIFE CYCLE OF THE HEPATITIS C VIRUS <i>Mishchenko E.L., Korotkov R.O., Ivanisenko V.A.</i>	148
MATHEMATICAL MODELING OF RECEPTOR MEDIATED ENDOCYTOSIS OF LOW-DENSITY LIPOPROTEINS AND THEIR DEGRADATION IN LYSOSOMES <i>Ratushny A.V.</i>	151
CONSERVED PROPERTIES OF ENZYMATIC SYSTEMS: PRENYLTRANSFERASE KINETICS <i>Ratushny A.V., Bezzmaternikh K.D.</i>	156
MATHEMATICAL MODELING OF THE GENE NETWORK CONTROLLING HOMEOSTASIS OF INTRACELLULAR CHOLESTEROL <i>Ratushny A.V.</i>	160
CONSTRUCTION OF THE HCV-HEPATOCYTE SYSTEM MODEL <i>Rudenko V.M., Korotkov E.V.</i>	165
FORMAL DESCRIPTION OF NF- κ B PATHWAY, ITS ROLE IN INFLAMMATION, INHIBITION OF APOPTOSIS, CARCINOGENESIS, AND WAYS OF INACTIVATION FOR PREDICTION OF NEW TARGETS FOR ANTI-INFLAMMATORY AND ANTI-CANCER TREATMENT <i>Sharipov R.N., Kolpakova A.F.</i>	170
RECONSTRUCTION AND STRUCTURE ANALYSIS OF APOPTOSIS GENE NETWORK IN HEPATITIS C <i>Stepanenko I.L., Podkolodnaya N.N., Paramonova N.V., Podkolodny N.L.</i>	174
DYNAMIC FILTRATION OF VARIABILITY WITHIN EXPRESSION PATTERNS OF ZYGOTIC SEGMENTATION GENES IN <i>DROSOPHILA</i> <i>Surkova S.Yu., Samsonova M.G.</i>	178

3.3. MODELLING OF MORPHOGENESIS

A CELLULAR AUTOMATON TO MODEL THE DEVELOPMENT OF SHOOT MERISTEMS OF <i>ARABIDOPSIS THALIANA</i> <i>Akberdin I.R., Ozonov E.A., Mironova V.V., Komarov A.V., Omelyanchuk N.A.</i>	185
THE GENE NETWORK DETERMINING DEVELOPMENT OF <i>DROSOPHILA MELANOGASTER</i> MECHANORECEPTORS <i>Bukharina T.A., Katokhin A.V., Furman D.P.</i>	190
A MODEL FOR SENSING THE MORPHOGEN HEDGEHOG CONCENTRATION GRADIENT I: A NUMERICAL STUDY <i>Kogai V.V., Gunbin K.V., Omelyanchuk L.V., Fadeev S.I.</i>	194
A MODEL FOR SENSING THE HEDGEHOG CONCENTRATION GRADIENT II: A CHECK FOR ADEQUACY <i>Gunbin K.V., Kogai V.V., Fadeev S.I., Omelyanchuk L.V.</i>	199
METHODOLOGY FOR BUILDING OF COMPLEX WORKFLOWS WITH PROSTAK PACKAGE AND ISIMBIOS <i>Matveeva A., Kozlov K., Samsonova M.</i>	204

ONTOLOGY OF ARABIDOPSIS GENE NET SUPPLEMENTARY DATABASE (AGNS), CROSS DATABASE REFERENCES TO TAIR ONTOLOGY <i>Mironova V.V., Poplavsky A.S., Ponomarev D.K., Omelianchuk N.A.</i>	209
A ONE-DIMENSIONAL MODEL FOR THE REGULATION OF THE SIZE OF THE RENEWABLE ZONE IN BIOLOGICAL TISSUE <i>Nikolaev S.V., Fadeev S.I., Kogay V.V., Mjolsness E., Kolchanov N.A.</i>	213
A SYSTEM FOR SIMULATION OF 2D PLANT TISSUE GROWTH AND DEVELOPMENT <i>Nikolaev S.V., Penenko A.V., Belavskaya V.V., Mjolsness E., Kolchanov N.A.</i>	218
AGNS (ARABIDOPSIS GeneNet SUPPLEMENTARY DATABASE), RELEASE 3.0 <i>Omelianchuk N.A., Mironova V.V., Poplavsky A.S., Pavlov K.S., Savinskaya S.A., Podkolodny N.L., Mjolsness E.D., Meyerowitz E.M., Kolchanov N.A.</i>	223
SEMANTICALLY RICH ONTOLOGY OF ANATOMICAL STRUCTURE AND DEVELOPMENT FOR <i>ARABIDOPSIS THALIANA</i> L. <i>Ponomaryov D., Omelianchuk N., Kolchanov N., Mjolsness E., Meyerowitz E.</i>	227
A PROGRAM METHOD OF CONSTRUCTING ONTOLOGY OF PHENOTYPIC ABNORMALITIES FOR <i>ARABIDOPSIS THALIANA</i> <i>Ponomaryov D., Omelianchuk N., Mironova V., Kolchanov N., Mjolsness E., Meyerowitz E.</i>	231

Introduction

Three volumes of *Proceedings of the Fifth International Conference on Bioinformatics of Genome Regulation and Structure—BGRS'2006* (Akademgorodok, Novosibirsk, Russia, July 16–22, 2006) comprise about 200 peer-reviewed publications on the topical problems in bioinformatics of genome regulation and structure. Biology now is among the most dynamically developing scientific disciplines. The main factor of this progress is an unprecedented, both in the rate and volume, accumulation of new facts due to advent of novel state-of-the-art experimental technologies. The post-genome era in biology brought about a sharp up-scaling of the research in the fields of genomics, transcriptomics, and proteomics. We are the witnesses how new directions of experimental and computer molecular biology emerge and successfully advance, including sequencing and analysis of megagenomes of bacterial communities, regulation of gene expression by short RNAs, microarray analysis technique, construction of proteomic portraits of cells and tissues, metabolomics, high-throughput genotyping of human populations for biomedical purposes, and many others. However, the synthesis of these directions is developing to a lesser degree, while it is a primary need for creation of an orderly theory of development, function, and evolution of the living systems—systems biology (gene interaction, gene network functioning, signal transduction pathways, networks of protein–protein interactions, modeling of ontogenesis, molecular phylogeny, the theory of evolution, etc.). The reasons underlying this gap lie not only in the objective complexity of the living systems, but also in the specialization in various fields of biology, which is ever increasing with accumulation of new data and development of new methods. The holistic vision of the research object is disappearing. The goal of this Conference, similar to the preceding Conferences—BGRS'1998, BGRS'2000, BGRS'2002, and BGRS'2004, which were held in Novosibirsk in 1998, 2000, 2002, and 2004—is, first and foremost, to provide the possibility for a wide exchange of opinions for various experts in *in silico* biology and researchers involved in experimental studies who use computer methods in their work or have interest in applied or theoretical aspects of bioinformatics. BGRS'2006 provides a general forum for disseminating and facilitating the latest developments in bioinformatics in molecular biology. BGRS'2006 is a multidisciplinary conference. The scope covered by the Conference comprises (i) the issues of development of advanced methods for computational and theoretical analysis of structure–function genome organization, proteomics, transcriptomics microarray analysis, etc.; (ii) application of these methods in theoretical (various aspects of evolutionary biology) and applied (search for promising application points in biotechnology and medicine) fields; and (iii) the issues related to general informational support of biological research and education (creation and computer support of databases, retrieval systems, ontologies, etc.). Thus, the final goal of this Conference may be defined as a half the battle for the new synthesis in Biology, which is a long-standing need, via the dialogue between the experts in particular fields of biology. This is the reason why BGRS'2006,

along with the traditional sections (computational structural and functional genomics and transcriptomics, computational structural and functional proteomics, comparative and evolutionary genomics and proteomics, and bioinformatics and education), includes an essentially expanded section on *computational systems biology*, which contains the presentations on modeling of molecular genetic systems and processes in bacterial and multicellular organisms and modeling of morphogenesis. Moreover, as compared to the previous conferences, the presentations related to evolution and phylogeny are plentiful. Numerous interdisciplinary studies into various taxa performed by the methods of molecular phylogeny, computer genomics, proteomics, cytogenetics, etc., as well as comparison of these results with the data obtained by classical methods of evolutionary morphology, paleontology, and various directions of ecology revealed the basic differences between the rates and modes of evolution at different hierarchical levels of biological organization (genes, genomes, karyotypes, organisms, populations, and biocenoses). Thus, the actual evolutionary process cannot be reduced to the evolution on one of the listed levels and is, speaking in images, an interference pattern, which is the more complex, the more interacting blocks and hierarchical levels constitute a biological system and the more intricate are their interrelations. Deciphering of this interference pattern is one of the challenges for the biology of the XXI century, which is answerable only by the joint efforts of bioinformatics and experimental sciences. If BGRS'2006 succeeds in contributing to this to any degree, the organizers will reckon their goal fulfilled.

Among the main goals of BGRS is improvement in the quality of education in all its aspects. That is why the success and international acknowledgement of the preceding conferences and the 2005 BGRS Summer School "Evolution, Systems Biology and High Performance Computing Bioinformatics" has encouraged launching the 2006 BGRS Summer School "Evolution, Systems Biology and High Performance Computing Bioinformatics". This School being the co-event of the conference will precede BGRS'2006. This event will attract next generation of scientists to bioinformatics. The scientific scope of the school will include issues of the development and application of advanced methods of computational and theoretical analysis for structure-function genome organization, proteomics, evolutionary and systems biology. We hope that the School of Young Scientists will become a good BGRS tradition.

BGRS'2006 is organized by the Laboratory of Theoretical Genetics with the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, (Novosibirsk, Russia). The organizational sponsors of the Conference are the Institute of Cytology and Genetics and the Siberian Branch of the Russian Academy of Sciences. The financial sponsor is the Russian Foundation for Basic Research. The School of Young Scientists "Evolution, Systems Biology and High Performance Computing Bioinformatics" is sponsored by the Russian Foundation for Basic Research and INTAS. The organizational support for the School is provided by the Chair of the Informational Biology, Faculty of the Natural Sciences of the Novosibirsk State University and the Council of Young Scientists of the Institute of Cytology and Genetics, SB RAS.

Professor Nikolay Kolchanov
Head of the Laboratory of Theoretical Genetics
Institute of Cytology and Genetics SB RAS,
Novosibirsk, Russia
Chairman of the Conference

Professor Ralf Hofstaedt
Faculty of Technology
Bioinformatics Department
University of Bielefeld, Germany
Co-Chairman of the Conference



PART 3. COMPUTATIONAL SYSTEMS BIOLOGY

3.1. MODELLING OF MOLECULAR GENETIC SYSTEMS AND PROCESSES IN BACTERIAL CELL

***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:

$$\frac{dX}{dt} = F(X, Y, K), \quad (1)$$

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:

$$Y = G(X, Z, K), \quad (2)$$

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 $A = \{\alpha\} : \mathfrak{N} = \{X_\alpha : X_\alpha \subseteq X, \alpha \in A\}$. Null set \emptyset may enter into \mathfrak{N} . Let K , N , and Δ designate the sets of parameters called efficiency coefficients (EC), Hill coefficients (HC), and activity coefficients (AC), respectively: $\{k_\alpha : \alpha \in A\} = K$, $\{n_{\alpha,x} : x \in X_\alpha, \alpha \in A\} = N$, $\{\delta_\alpha : \alpha \in A\} = \Delta$. By implication of $k_\alpha, n_{\alpha,x} > 0, \delta_\alpha \geq 0$.

$$1. \quad h(x|x \in X) = R(X)/Q(X) = \sum_{\alpha} \delta_{\alpha} \prod_{x \in X_{\alpha}} (x/k_{\alpha})^{n_{\alpha,x}} / \sum_{\alpha} \prod_{x \in X_{\alpha}} (x/k_{\alpha})^{n_{\alpha,x}}$$

- This rational polynomial is a GHF (by convention, if $X_{\alpha} \emptyset$, then

$$\prod_{x \in X_{\alpha}} (x/k_{\alpha})^{n_{\alpha,x}} \equiv 1),$$

2. If $h(X, K, N, \Delta)$ is a GHF, then $\tilde{h}(X, K(X, K, N, \Delta), N(X, K, N, \Delta), \Delta)$, 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 $\mathbf{H1}(x_1)$. The result of the first stage is the generalized Hill function $\mathbf{H1}(x_1)$ 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 $\mathbf{H1}(x_1)$ 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 $\mathbf{H1}(x_1)$ 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).

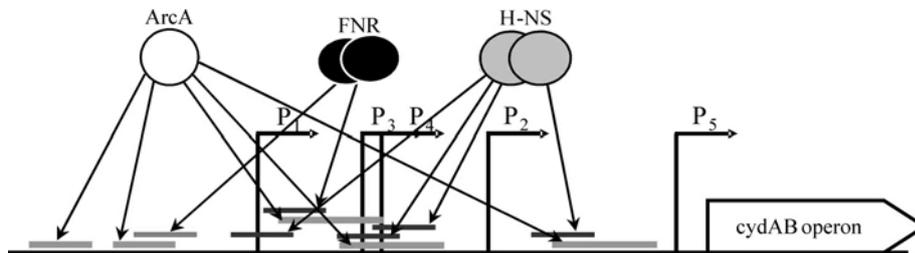


Figure 1. Structure of the promoter of the *cydAB* operon, coding for cytochrome bd oxidase. Retrieved from Ecocyc database (<http://ecocyc.org>). Factor ArcA is a transcription activator, H-NS is a repressor, and Fnr can be both an activator (left binding site of the second promoter) and a repressor.

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.

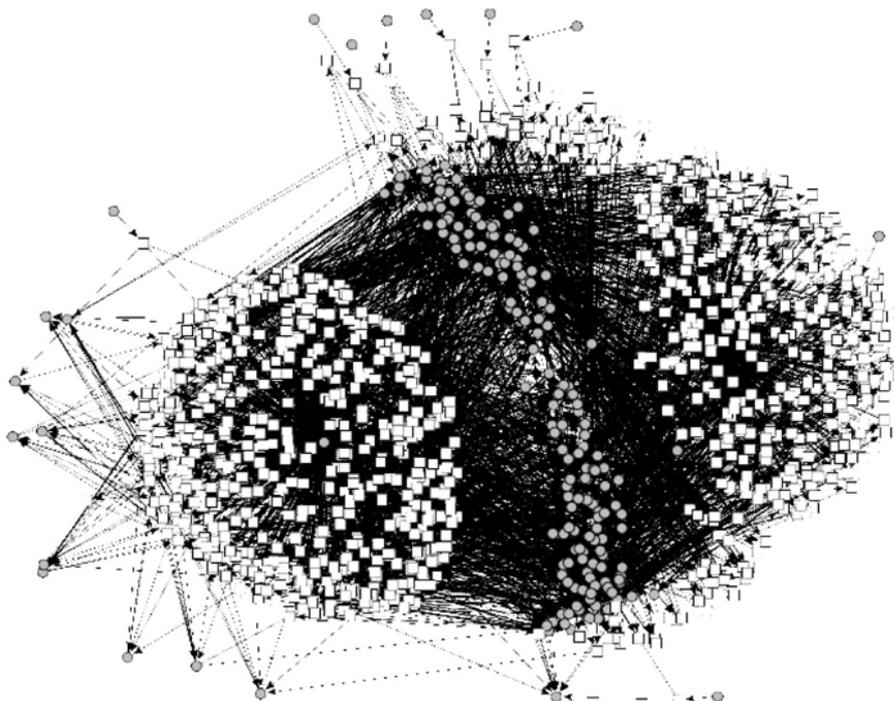


Figure 2. Bipartite graph of the mathematical model constructed in terms of mono- and bimolecular reactions to describe the regulation of transcription initiation in the promoter region of the *cydAB* operon. The graph contains 510 elementary processes (white square nodes) and 197 dynamic variables (gray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

Thus, we obtain the following model equation describing the level of *cydAB* expression (f_{Δ_a, Δ_f}) in the wild type ($\Delta_a = 1, \Delta_f = 1$) and knockout strains $\Delta arcA$ ($\Delta_a = 0, \Delta_f = 1$) and Δfnr ($\Delta_a = 1, \Delta_f = 0$) depending on oxygen concentration:

$$f_{\Delta_a, \Delta_f} = \frac{k_{\Delta_a, \Delta_f} + \Delta_a \left[\left(O_2 / k_{a11} \right)^{h_{a1}} + \left(O_2 / k_{a21} \right)^{h_{a2}} \right] + \Delta_f \left(O_2 / k_{f1} \right)^{h_{f1}} + \Delta_a \Delta_f \left(O_2 / k_{af1} \right)^{h_{af1}}}{1 + \Delta_a \left[\left(O_2 / k_{a12} \right)^{h_{a1}} + \left(O_2 / k_{a22} \right)^{h_{a2}} \right] + \Delta_f \left(O_2 / k_{f2} \right)^{h_{f2}} + \Delta_a \Delta_f \left(O_2 / k_{af2} \right)^{h_{af2}}}, \quad (3)$$

where O_2 is oxygen concentration in the medium; k_{Δ_a, Δ_f} , basal expression of the *cydAB* operon; k_{ai}, k_{fi}, k_{afi} are constants of the effect of ArcA, Fnr, and their combined effect on *cydAB* operon expression, respectively; h_{ai}, h_{fi}, h_{afi} , 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 ($f_{cyo(WT)}$) and for mutant strains ($f_{cyo(\Delta fnr)}$) and ($f_{cyo(\Delta arcA)}$) (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).

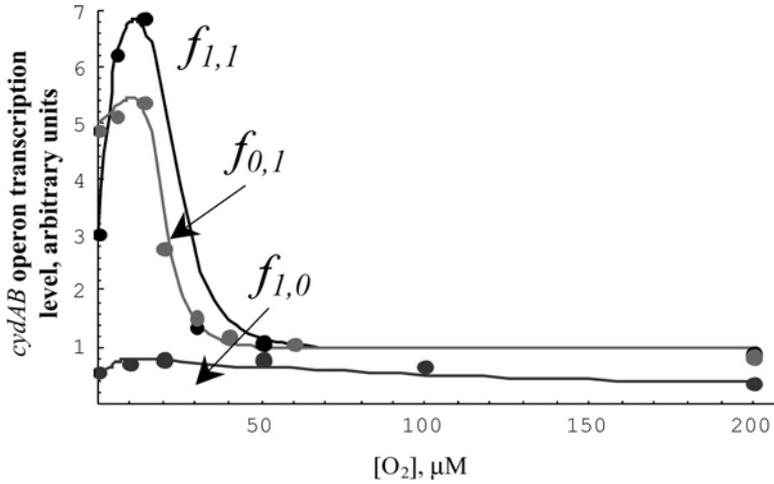


Figure 3. Dependence of the *cydAB* operon transcription on oxygen concentration in wild-type *E. coli* cells ($f_{1,1}$) and mutant strains $\Delta arcA$ ($f_{0,1}$) and Δfnr ($f_{1,0}$). Dots indicate experimental data (Tseng *et al.*, 1996). Curves indicate calculations from Eq. (3). The curves were calculated with the following parameters: $k_{1,1} = 3$; $k_{0,1} = 5$; $k_{a11} = 20 \mu\text{M}$; $k_{a12} > 500 \mu\text{M}$; $h_{a1} = 1$; $k_{a21} = 20 \mu\text{M}$; $h_{a2} = 6$; $k_{a22} = 20 \mu\text{M}$; $k_{1,0} = 0.4$; $k_{f1} = 9 \mu\text{M}$; $h_{f1} = 1$; $k_{f2} = 11 \mu\text{M}$; $h_{f2} = 1.4$; $k_{af1} = 1.2 \mu\text{M}$; $h_{af1} = 1$; $k_{af2} = 100 \mu\text{M}$; $h_{af2} = 6$.

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

- Ananko E.A., Ratushny A.V., Usuda Y., Matsui K. (2006) Aromatic amino acid biosynthesis in *Escherichia coli*: generalized hill function model of the tryptophan-sensitive 3-deoxy-d-arabino-heptulosonate-7-phosphate synthase reaction demonstrate complicated mechanism. *This issue*.
- Cornish-Bowden A. (1977) An automatic method for deriving steady-state rate equations. *Biochem. J.*, **165**, 55–59.
- Khlebodarova T.M. *et al.* (2006) *In silico* cell II. Information sources for modeling *Escherichia coli* metabolic pathways. *This issue*.
- King E.L., Altman C. (1956) A Schematic Method of Deriving the Rate Laws for Enzyme-Catalyzed Reactions. *J. Phys. Chem.*, **60**, 1375–1378.
- Nedosekina E.A., Usuda Y., Matsui K. (2006) Modeling of the effects of threonine, valine, isoleucine and pyridoxal 5'-monophosphate on biosynthetic threonine dehydratase reaction. *This issue*.
- Ratushny A.V., Likhoshvai V.A., Anan'ko E.A., Vladimirov N.V., Gunbin K.V., Lashin S.A., Nedosekina E.A., Nikolaev S.V., Omel'yanchuk L.V., Matushkin Yu.G., Kolchanov N.A. (2005) Novosibirsk school of computational systems biology: history and future trends. *Informatsionnyi vestnik VOGiS*, **9**(2), 232–261.
- Ratushny A.V., Usuda Y., Matsui K., Podkolodnaya O.A. (2006a) Mathematical modeling of elementary processes of the gene network controlling histidine biosynthesis in *Escherichia coli*. *This issue*.
- Ratushny A.V., Nedosekina E.A. (2006b) Regulation of pyrimidine biosynthesis in *Escherichia coli*: gene network reconstruction and mathematical modeling. *This issue*.
- Ratushny A.V., Smirnova O.G., Usuda Y., Matsui K. (2006c) Regulation of the pentose phosphate pathway in *Escherichia coli*: gene network reconstruction and mathematical modeling of metabolic reactions. *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*.
- Tseng C.P. *et al.* (1996) Effect of microaerophilic cell growth conditions on expression of the aerobic (cyoABCDE and cydAB) and anaerobic (narGHJI, frdABCD, and dmsABC) respiratory pathway genes in *Escherichia coli*. *J. Bacteriol.*, **178**, 1094–1098.

IN SILICO CELL II. INFORMATION SOURCES FOR MODELING *ESCHERICHIA COLI* METABOLIC PATHWAYS

**Khlebodarova T.M.¹, Ananko E.A.^{*1}, Nedosekina E.A.¹, Podkolodnaya O.A.¹,
Poplavsky A.S.¹, Smirnova O.G.¹, Ibragimova S.S.¹, Mishchenko E.L.¹,
Stepanenko I.L.¹, Ignatieva E.V.¹, Proskura A.L.¹, Nishio Y.², Imaizumi A.²,
Usuda Y.², Matsui K.², Podkolodny N.L.¹, Kolchanov N.A.¹**

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ² Institute of Life Sciences, Ajinomoto, Co., Inc., Kawasaki-city, Japan

* Corresponding author: e-mail: eananko@bionet.nsc.ru

Key words: database, transcription, expression, gene networks, metabolism, constants, dynamics, kinetics, modeling, functional parameters, *Escherichia coli*

SUMMARY

Motivation: Development of an *in silico* cell, a computer resource for modeling and analysis of physiological processes is an urgent task of systems biology and computational biology. This field of inquiry demands development of software for accumulation and analysis of information on operation of the cell as an integrated system.

Results: We developed three information sources for computer modeling of manifold processes of *E. coli*: 1) EcoTRRD database designed for description of structure and function of regulatory regions of *E. coli* genes; 2) EcoGeneNet database was developed for description of metabolic networks in *E. coli*; 3) EcoKiNET database allows to accumulate the quantitative information about the dynamics on biosynthesis of amino acids, nucleotides, and sugars in *E. coli*.

Availability: The diagrams of *E. coli* metabolic networks are available through the GeneNet viewer at <http://wwwmgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/>; the EcoKiNET database is available an authorized access to the web recourse is provided or the data exported into XML files may be granted on the basis of collaboration; the EcoTRRD database is available as a flat-file that could be accessible on the basis of collaboration via kol@bionet.nsc.ru.

INTRODUCTION

Modern biology deals with large volumes of experimental data, which cannot be comprehended without up-to-date computational technologies and efficient mathematical methods for data processing and modeling of biological systems and processes. In recent times, increasing attention is given to systems biology, considering the whole set of processes in a living body. Computer simulation of the life of certain types of cells (Nakayama *et al.*, 2005) or a unicellular organism (Takahashi *et al.*, 2003; Ishii *et al.*, 2004) was repeatedly undertaken. However, simulation of cell life demands computer analysis of kinetic and actual experimental data, which should be structured and formalized.

The TRRD database was being designed for accumulation of experimental data on extended regulatory regions of eukaryotic genes (Kolchanov *et al.*, 2002), the GeneNet database was being developed for the accumulation of heterogenous data on gene networks

of multicellular organisms (Ananko *et al.*, 2005). Then, the TRRD and GeneNet formats were adapted to collect information on the basal metabolism of *E. coli* and applying this information to computer simulation.

The EcoKiNET database is a new resource and intended for accumulation of the information about the constants of biochemical reactions and processes, the concentrations of various compounds in the cell and the organism, and, which is of the utmost importance, the dynamics of these reactions.

The main goal of this work was to develop a comprehensive databases on the regulation of major metabolic processes in *E. coli*.

METHODS AND ALGORITHMS

The EcoTRRD database was developed with the use of the TRRD format (Kolchanov *et al.*, 2002) modified as described below.

The EcoGeneNet database was developed with the use of the GeneNet format (Ananko *et al.*, 2005) modified as described below.

Reconstruction of metabolic networks in *E. coli* was performed with the GenED gene network editor (Loktev *et al.*, 2002).

IMPLEMENTATION AND RESULTS

The EcoTRRD as an information source for description of regulatory regions *Escherichia coli* genes.

The EcoTRRD database has been designed for accumulation of experimental data on *E. coli* genes, promoters, transcription factor binding sites, and gene expression patterns. A database entry corresponds to a single gene. EcoTRRD is a novel database developed on the base of the TRRD. TRRD format was adapted to the *E. coli* specificity. Some fields were abolished, because they cannot be used for description of prokaryotic organisms gene expression regulation. Some new fields were added for a general gene description (operon name, B number, open reading frame, point of the transcription termination); for a binding site description (positions of a site on the *E. coli* chromosome), and for a transcription factor description (active form, a metabolite operating in a complex with active transcription factor). The designed format of the TRRD database gives possibility to describe any essential feature of structural and functional organization of prokaryotic genes regulatory regions providing for an adequate level of gene expression under particular living conditions of bacterium.

The structure of EcoTRRD entry reflects the modular organization of the gene regulatory regions and the hierarchy of constituent regulatory units. Each line of an entry begins with a two-character line code indicating the type of information contained in the line and denoting some information field. EcoTRRD contains 76 various line types. All information fields of EcoTRRD may be divided into the 6 sections (Table 1).

Currently, EcoTRRD comprises the description of 363 *E. coli* genes, 553 regulatory units, 1298 transcription factor binding sites, and 2565 expression patterns. More than 2200 scientific publications were annotated.

The EcoGeneNet database as an information source for description of Escherichia coli metabolic pathways. For description of gene networks of the *E. coli* the format of the GeneNet database was modified so that the features of the organization and regulation of metabolic ways of this organism should be taken into account. The designed format of the EcoGeneNet database allows description of any essential feature of structure, regulation, and functional organization of bacterial genetic and metabolic networks. For example, the new table for an operon description was added. Some new fields were added for a general

gene description (link to the operon containing this gene, standard B number of the gene). The field “half-life time” was added for substance and RNA descriptions. Also, some new possibilities were added for a protein description, namely, subunit content and half-life time. New fields for a reaction description are: link to the process in KEGG database, cofactor, enzyme, catalyzing this reaction, kinetic constants.

Table 1. Information sections of the EcoTRRD database

The section name	Number of information fields
General gene description	18
Information section describing expression patterns of gene	14
Description of promoters	7
Description of transcription factor binding sites	17
Description of transcription factors regulating the gene expression	11
Bibliography	9

The EcoGeneNet was used for describing the regulation of basal metabolism in *E. coli*. At present, EcoGeneNet contains descriptions of 20 *E. coli* metabolic networks, including 739 proteins, 258 RNAs, 389 genes, 227 operons, 2443 relationships between entities, 265 small molecules, 873 links to SWISS-PROT, 600 links to EMBL/GenBank, and 921 links to EcoCyc. The database was filled with information from 2667 publications from 131 journals. The information content of EcoGeneNet is given in Table 2.

Table 2. Information content of the EcoGeneNet database

Diagram name	Number of		
	operons	proteins	interrelations
Alanine biosynthesis	4	10	33
Anaplerotic reactions	9	23	133
Arginine, putrescine, and spermidine biosynthesis	23	56	205
Aromatic amino acids	15	30	130
Aspartate and asparagine biosynthesis	7	18	89
Branched chain amino acid biosynthesis	11	38	124
Cysteine biosynthesis	6	26	99
Glutamate and glutamine biosynthesis	5	34	204
Histidine biosynthesis	2	21	63
Membrane transport	3	17	58
Methionine biosynthesis	9	21	75
One carbon metabolism	6	25	68
Pentose phosphate pathway	12	35	104
Proline biosynthesis	5	17	57
Purine biosynthesis	11	33	126
Pyrimidine biosynthesis	7	20	61
Pyruvate metabolism	19	52	178
Respiration	22	111	313
Salvage pathways	30	85	476
Serine and glycine biosynthesis	10	32	107

EcoKiNET database as an information source for description of kinetic data.

EcoKiNET is a novel database, which allows for systematizing and accumulating the quantitative information about the dynamics of various processes of both enzymatic and molecular genetic natures. The structure of EcoKiNET database reflects the specificity of experimental approach to studying the kinetics of any process and is independent of its type. The structure and format of this database allow for describing both the molecular genetic processes and biochemical reactions.

The unit entry of the database is process – a regulatory event or an enzymatic reaction. Its overall description comprises, first, a general description of the components involved in

the process (the “process” module) with indication of their particular roles, namely, cofactors, enzymes, substrates, products, inducers, or repressors. A possibility of playing several parts is provided, for example, a reaction product can also be an inhibitor of the reaction.

Second, the overall description includes the description of experiment with indication of the research object, method used, and experimental conditions (the “experiment” module). The characteristics of the process (constants) obtained in a particular experiment are described in the block “functional parameters”. This block comprises five tables allowing for description of the parameters of an enzymatic reaction or a molecular genetic process, parameters of formation of protein–protein or DNA–protein complexes, and the half-life times of the components of the process. The database format now allows 15 types of various parameters to be described. The page of experiment description in the EcoKiNET database is shown in Fig. 1.

KiNET database editor

[Save as draft](#)
[Up](#)

[Save](#)
[Add annotation source](#)

[Home](#)

Kinet
7BE20272- 6CFA-
328A- 814D-
B4A976C5B0B7
experiment
6AC2B8CF- 8912-
04E7- A9EB-
87A3DC065FF5

process 1
experiment 1
experiment 2
diagram 1
diagram 2
diagram 3
diagram 4
experiment 3
diagram 1
diagram 2
diagram 3
diagram 4
diagram 5
experiment 4 >

[bib ref](#) 1 unrelated
[bib ref](#) 2 unrelated
[bib ref](#) 3 related

"experiment" module

[Clone "experiment" module for lesser changes](#)
[New "diagram" module](#)

Object

Species

Organ/Tissue

Strain id
[Find strain](#)

Strain description

Plasmid

Object status

Culture density

Growth medium

Oxygen status

Method

Name

Description

Conditions

Parameter	Value	Unit
Buffer	<input type="text" value="70 mM HEPES buffer, 0.5 mM EDTA, 10 mM MgCl2"/>	
Temperature	<input type="text" value="25"/>	<input type="text" value="°C"/>
pH	<input type="text" value="8.0"/>	

Functional parameters

[Enzymatic reaction parameters](#)
[Regulation parameters](#)
[Transcription factor characterization](#)
[Half-life value](#)
[Enzyme complex formation parameters](#)

Figure 1. The page of experiment description in the EcoKiNET database.

The overall description of process dynamics in each particular experiment is provided in the “diagram” module. The information about the process dynamics is extracted from

graphical data, tables, and diagrams given in the experimental papers. The information inputted can be visualized as plots including one or several curves. Fig. 2 shows the structure of this information block.

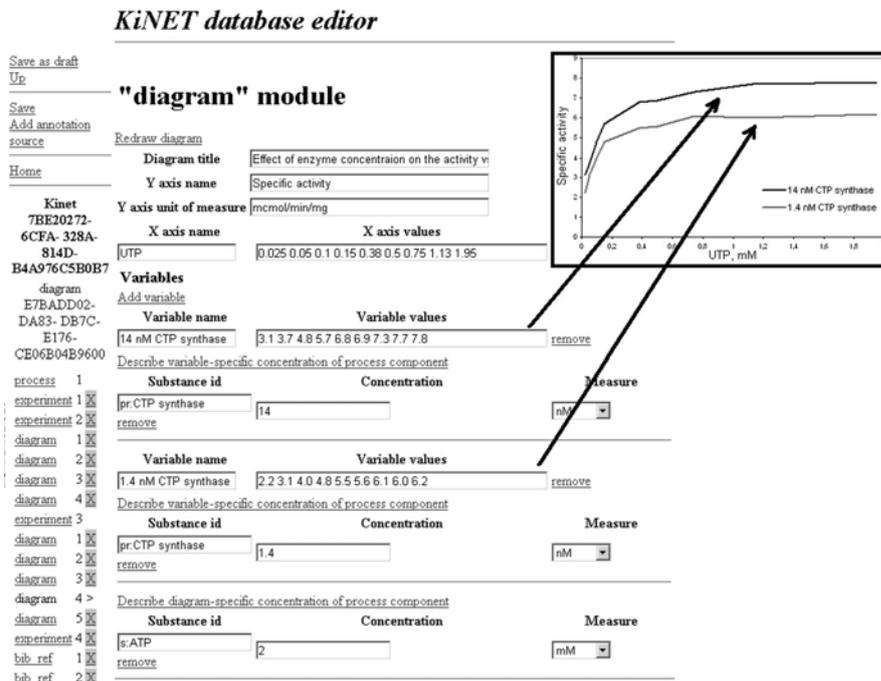


Figure 2. The page of the EcoKiNET database that describes dynamic data and the example of plot, visualizing kinetic data.

The last but not the least information block of the “experiment” module and related to particular experiments is the block containing bibliographic information. Note that the database format allows also for accumulation of the information generated by modeling; however, the database now stores only the information extracted from experimental papers.

Table 3. The block of constants “Enzymatic reaction parameters” (the data on K_m include the K_m values of both substrates and cofactors)

Class	Dynamic	Enzymatic reaction parameters						
	Profile	Michaelis constant K_m	Constant of catalytic activity k_{cat}	Maximal velocity V_{max}	Association constant K_a	Dissociation constant K_d	Hill coefficient n^H	Specific activity
No.	3077	1789	438	469	1	155	265	401

The data in the EcoKiNET database are represented in a specially designed XML format. The complete formal description of the database format and individual fields is made in the XML Schema language, supported by the W3 consortium. Application of XML provides a high data portability and equal simplicity of data processing for both analytical and presentational aims.

Currently, the database EcoKiNET stores the information related to the description of specific features of the *E. coli* metabolism and genetic regulation extracted from 794 original experimental papers published in 68 journals. The content of the database is illustrated in Tables 3 and 4.

As is evident, the database contains descriptions of 4486 constants and 3077 dynamic profiles involved in kinetics of the processes taking place in 23 gene networks of *E. coli*.

Table 4. The blocks of constants “Regulation parameters”, “Enzyme complex formation parameters”, and “Half-life value”

Class	Regulation parameters			Half-life value	Enzyme complex formation parameters			
	Constant of inhibition K_i	Constant of activation K_{act}	Hill coefficient n^H	Half-life $t_{1/2}$	First order velocity constant	Second order velocity constant	Association constant K_a	Dissociation constant K_d
Number	552	144	74	38	11	7	20	128

DISCUSSION

We developed three information sources for accumulation of the information about the regulation and dynamics of metabolic pathways expression in *E. coli*. The data collected in the EcoTRRD, EcoGeneNet and EcoKiNET databases were implemented for construction of more than 300 mathematical models of elementary processes describing the metabolism and genetic regulation genes of *E. coli* (Likhoshvai, Ratushny, 2006). Being assembled into an integral system, these models allow predicting the complex and dynamic behavior of living cell. With this system, one can simulate the results of gene modification or different growth conditions.

ACKNOWLEDGEMENTS

This work was supported by the Russian Government (Contract No. 02.467.11.1005), Siberian Branch of the Russian Academy of Sciences (the project “Evolution of molecular-genetical systems: computational analysis and simulation” and integration projects Nos 24 and 115), and the Federal Agency of Science and innovation (innovation project IT-CP.5/001). The authors are grateful to I.V. Lkhova for bibliographical support, K.A. Loktev for the assistance at initial step of EcoGeneNet development, and to V.V. Gulevich and G.B. Chirikova for translating the manuscript from Russian into English.

REFERENCES

- Ananko E.A. *et al.* (2005) GeneNet in 2005. *Nucl. Acids Res.*, **33**, D425–427.
- Ishii N. *et al.* (2004) Toward large-scale modeling of the microbial cell for computer simulation. *J. Biotechnol.*, **113**, 281–294.
- Kolchanov N.A. *et al.* (2002) Transcription Regulatory Regions Database (TRRD): its status in 2002. *Nucl. Acids Res.*, **30**, 312–317.
- Likhoshvai V.A., Ratushny A.V. (2006) In silico cell. I. Hierarchical approach and generalized hill functions in modeling enzymatic reactions and gene expression regulation. *This issue*.
- Loktev K.A. *et al.* (2002) A system for visual modelling of gene networks’ structural and functional organization. *Proc. Third International Conference on Bioinformatics of Genome Regulation and Structure (BGRS’2002)*. ICG, Novosibirsk, **2**, 132–135.
- Nakayama Y. *et al.* (2005) Dynamic simulation of red blood cell metabolism and its application to the analysis of a pathological condition. *Theor. Biol. Med. Model.*, **2**, 18.
- Takahashi K. *et al.* (2003) E-Cell 2: multi-platform E-Cell simulation system. *Bioinformatics*, **19**, 1727–1729.

MATHEMATICAL MODELING OF ELEMENTARY PROCESSES OF THE GENE NETWORK CONTROLLING HISTIDINE BIOSYNTHESIS IN *ESCHERICHIA COLI*

Ratushny A.V.^{*1}, Usuda Y.², Matsui K.², Podkolodnaya O.A.¹

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ² Institute of Life Sciences, Ajinomoto, Co., Inc., Kawasaki, Japan

* Corresponding author: e-mail: ratushny@bionet.nsc.ru

Key words: mathematical modeling, gene network, regulation, histidine biosynthesis, *Escherichia coli*

SUMMARY

Motivation: Development of an *in silico* cell as a computer resource for simulation and analysis of processes within living cells is an urgent task of systems biology and computational biology. Mathematical modeling of the genetic regulation of a basic metabolic process, histidine biosynthesis, is an important problem to be solved as part of this line of work.

Results: By using the GeneNet technology, we reconstructed the gene network of the regulation of histidine biosynthesis in *E. coli*. Mathematical models were constructed by the method of generalized Hill functions to describe the efficiency of enzyme systems and regulation of expression of genes coding for these enzymes.

Availability: Models are available on request; the gene network is available at <http://wwwmgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/index.shtml>.

INTRODUCTION

The genes of the histidine biosynthesis pathway are grouped into two operons: *prsA* and *hisGDCBHAFI*. Transcription of the *prsA* operon, coding for the enzyme phosphoribosyl pyrophosphate synthase (RPPS), is repressed by the PurR protein. The repressive action of PurR is enhanced by guanine (He *et al.*, 1993). Two *prsA* transcripts have been found. The transcripts originating from promoter P1 constitute only 5 % or less of the total *prs*-encoding mRNA under all conditions examined; promoter P2 appears to be the sole site of regulation of *prsA* expression by pyrimidines (Post *et al.*, 1993). Phosphoribosyl pyrophosphate (PRPP), the product of the reaction catalyzed by RPPS, is utilized by microorganisms as a precursor of purines, pyrimidines, tryptophan, and histidine. The enzyme demands MgATP as a substrate and free Mg²⁺ as a cofactor.

Eight genes of the *hisGDCBHAFI* operon code for ten histidine biosynthesis pathway enzymes. Three of the eight genes, *hisD*, *hisB*, and *hisI*, code for bifunctional enzymes; two, *hisH* and *hisF*, for polypeptides forming a heterodimeric enzyme. Transcription of polycistronic mRNAs of the *his* operon, differing in length, starts from different promoters. Promoter P1 gives rise to transcript *hisGDCBHAFI*, promoter P2, to *hisBHAFI*. Transcription from P1 is repressed by histidine by the attenuation mechanism at the level of the leader region. Promoter P2 is not histidine-regulatable. Both promoters are metabolically regulatable.

The key enzyme of the histidine biosynthesis pathway is ATP phosphoribosyltransferase. It is encoded by the first gene of the *hisG* operon. The enzyme activity is regulated by the end product of the pathway, histidine. This regulation is a typical example of enzymatic feedback inhibition.

We reconstructed the gene network of regulation of histidine biosynthesis in *E. coli* and developed mathematical models describing the operation of some individual enzyme systems and regulation of expression of genes coding for enzymes and their subunits. We also constructed a database storing experimental data on the behavior of the components of this network (Khlebodarova *et al.*, this issue). Parameters of the models were determined by numerical experiments. The results of calculations of steady-state and dynamic parameters of the gene network deduced from the models are in agreement with experimental evidence.

METHODS AND ALGORITHMS

The gene network was reconstructed with the use of the GeneNet technology (Ananko *et al.*, 2005). Mathematical models were constructed by the method of generalized Hill functions (Likhoshvai, Ratushny, 2006).

RESULTS AND DISCUSSION

The gene network of the regulation of histidine biosynthesis stores information on 21 proteins, 4 RNAs, 2 operons, 25 small molecules, and 64 reactions or regulatory events (Fig. 1, Table 1). The information was obtained by annotating 55 scientific publications. The genes, enzymes, and enzymatic reactions of the histidine biosynthesis pathway are listed in Table 1.

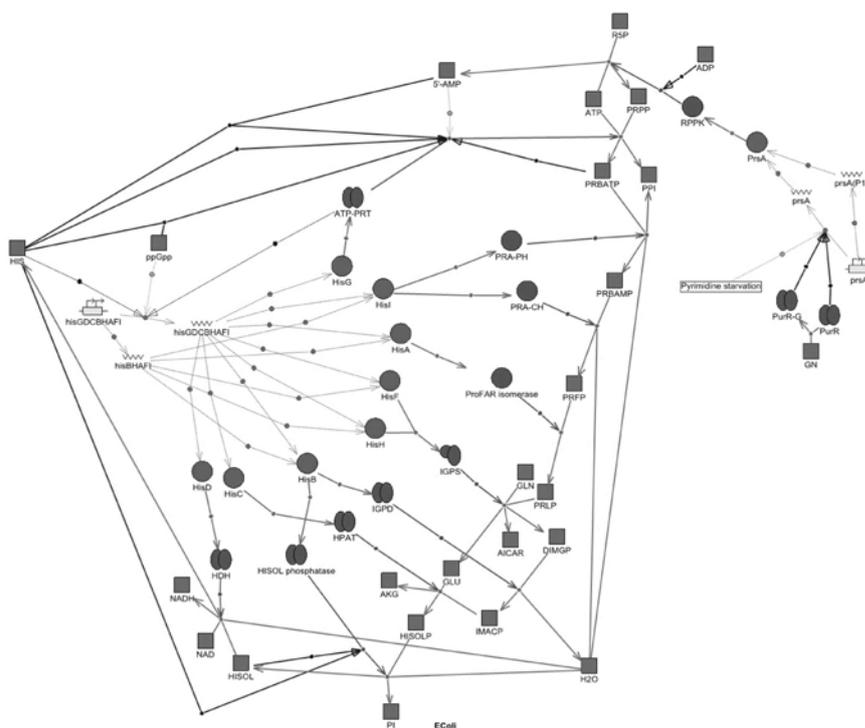


Figure 1. Histidine biosynthesis regulation gene network reconstruction in the GeneNet.

Table 1. Enzymatic reactions present in the histidine biosynthesis pathway network

Genes	Enzyme names	Reactions
<i>prsA</i>	Phosphoribosyl pyrophosphate synthase	R5P + ATP \leftrightarrow PRPP + AMP
<i>hisG</i>	ATP phosphoribosyltransferase	PRPP + ATP \rightarrow PPI + PRBATP
<i>hisIE</i>	Phosphoribosyl-ATP pyrophosphatase	PRBATP \rightarrow PPI + PRBAMP
<i>hisIE</i>	Phosphoribosyl-AMP cyclohydrolase	PRBAMP \rightarrow PRFP
<i>hisA</i>	Phosphoribosylformimino-5-amino-1-phosphoribosyl-4-imidazole carboxamide isomerase	PRFP \rightarrow PRLP
<i>hisFH</i>	Imidazoleglycerol phosphate synthase	PRLP + GLN \rightarrow GLU + AICAR + DIMGP
<i>hisB</i>	Imidazoleglycerol phosphate dehydratase	DIMGP \rightarrow IMACP
<i>hisC</i>	L-Histidinol phosphate aminotransferase	IMACP + GLU \rightarrow AKG + HISOLP
<i>hisB</i>	Histidinol phosphatase	HISOLP \rightarrow PI + HISOL
<i>hisD</i>	Histidinol dehydrogenase	HISOL + 3 NAD \rightarrow HIS + 3 NADH

Consider the expression of the *prsA* gene in *E. coli* as an example of application of generalized Hill functions to modeling molecular processes in histidine biosynthesis. This gene is regulated at the transcriptional level by the PurR protein. It is known that PurR more efficiently inhibits *prsA* transcription when bound to guanine (He *et al.*, 1993). An equation is proposed to describe the level of *prsA* transcription (f_{prsA}):

$$f_{prsA} = \frac{1}{1 + \left(\frac{d_{PurR,b}}{k_{PurR,b}} \right)^{h_b} + \left(\frac{d_{PurR,f}}{k_{PurR,f}} \right)^{h_f}}, \quad d_{PurR,b} = \frac{PurR \cdot G}{K + PurR + G}, \quad d_{PurR,f} = PurR - d_{PurR,b} = \frac{(PurR + K) \cdot PurR}{K + PurR + G}. \quad (1)$$

Here $PurR$ and G are concentrations of PurR and guanine, respectively; $d_{PurR,b}$ is the proportion of PurR bound to guanine; $d_{PurR,f}$ is the proportion of free PurR; K is the generalized PurR–guanine binding constant; $k_{PurR,b}$ and $k_{PurR,f}$ are constants of the efficiency of the effect of guanine-bound and free PurR on *prsA* transcription, respectively; and h_b , h_f are constants characterizing the nonlinearity of the effect of guanine-bound and free PurR on *prsA* transcription.

Equation (1) was tested by applying it to experimental data on the efficiency of binding between PurR and the regulatory region of the *prsA* operon with and without guanine reported in (He *et al.*, 1993). We assumed that, if RNA polymerase concentration and other factors were neglected, then, as a first approximation, it may be admitted that the portion of the operator region bound to PurR was in proportion to the efficiency of *prsA* expression.

The estimated values of parameters of Eq. (1) characterize the increase of the efficiency of inhibition of *prsA* by PurR owing to guanine binding ($k_{PurR,b} = 0.15 \mu\text{M}$, $k_{PurR,f} = 0.9 \mu\text{M}$) and the nonlinearity of the effect of free and guanine-bound PurR ($h_f = 2$, $h_b = 2$) on *prsA* transcription.

The *prsA* gene codes for phosphoribosyl pyrophosphate synthase (PRPPS) (Table 1). It is known that the rate of the corresponding reaction decreases dramatically at high riboso-5-phosphate (R5P) concentrations with the presence of ADP (Hove-Jensen *et al.*, 1986). We propose an equation describing the steady-state rate of the reaction:

$$V = \frac{k_{cat} \cdot e_0 \cdot R5P \cdot ATP}{(K_{m,R5P} + R5P) \cdot (K_{m,ATP} + ATP)} \cdot \frac{1}{1 + \frac{ADP^{h_{ADP}} \cdot R5P^{h_{R5P}}}{k_{ADP,R5P}^{h_{ADP} + h_{R5P}}}}, \quad (2)$$

where e_0 is PRPPS concentration; R5P, ATP, PRPP, AMP, ADP are concentrations of corresponding low-molecular-weight substances; f_{Mg} is the efficiency of the effect of

Mg^{2+} ions on PRPPS activity; k_{cat} , catalytic constant; $K_{m,R5P}$, $K_{m,ATP}$ are Michaelis–Menten constants for corresponding substrates; $k_{ADP,R5P}$ is the constant describing the efficiency of the combined effect of ADP and R5P on PRPPS activity; and h_{ADP} and h_{R5P} are constants describing the nonlinearity of the effect of ADP and R5P on PRPPS activity.

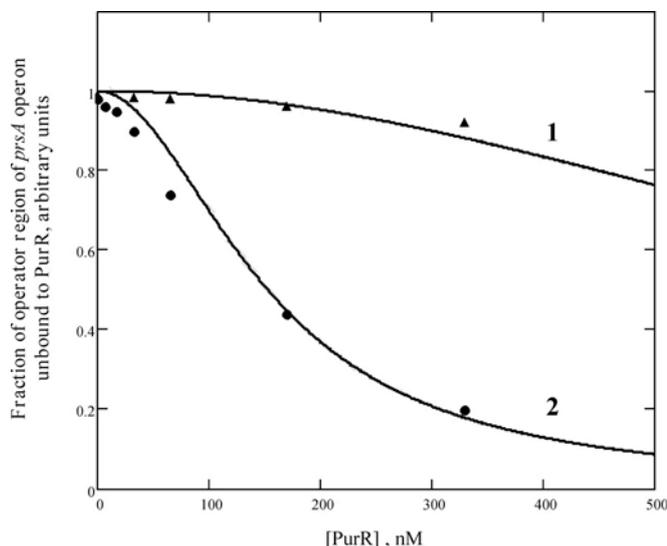


Figure 2. Efficiency of binding between PurR and the operator region of the *prsA* operon *in vitro*: comparison of predictions (curves) according to Eq. (1) with experimental data (dots) reported in (He *et al.*, 1993); 1, without guanine; 2, with 50 mM guanine. The curves were constructed for the following parameters of Eq. (1): $K = 1 \mu\text{M}$, $k_{PurR,b} = 0.15 \mu\text{M}$, $k_{PurR,f} = 0.9 \mu\text{M}$, $h_b = 2$, and $h_f = 2$.

The equation describing the regulation of PRPPS activity was tested by applying it to experimental data on the effect of R5P and ATP on PRPPS activity at various ADP concentrations. The results of calculation according to Eq. 2 and experimental data on PRPPS inhibition by ADP at various R5P concentrations are shown in Fig. 3.

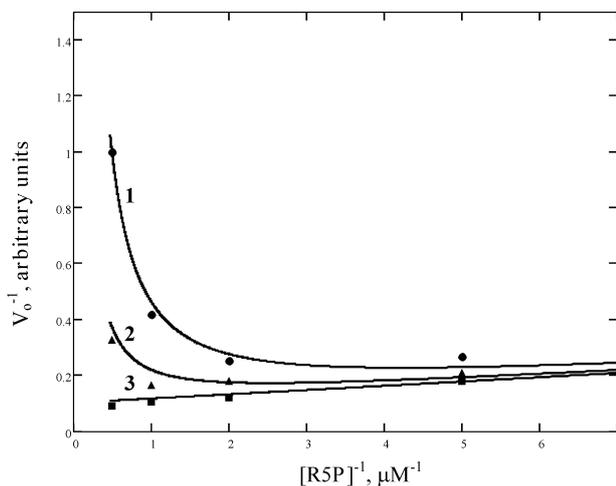


Figure 3. Inhibition of PRPPS by ADP. The initial rate was measured with the presence of 2 mM ATP. ADP concentration: 1, 0.4 mM; 2, 0.2 mM; 3, absent. Dots indicate experimental data reported in (Hove-Jensen *et al.*, 1986); curves are the results of calculation according to Eq. (2) at the following parameter values: $k_{cat} \cdot e_0 = 220 \text{ nmol}/(\text{ml} \cdot \text{min})$, $K_{m,R5P} = 150 \mu\text{M}$, $K_{m,ATP} = 300 \mu\text{M}$, $k_{ADP,R5P} = 430 \mu\text{M}$, $h_{ADP} = 1.75$, and $h_{R5P} = 1.45$.

The proposed mathematical model, constructed with the use of generalized Hill functions, is simple in comparison with the intricate mechanisms of molecular processes. Nevertheless, it provides good agreement between experimental and predicted results (Fig. 3).

Reconstruction of the gene network and development of mathematical models describing the efficiency of operation of enzyme systems and regulation of genes coding for these enzymes must be the initial stage of the construction of an integral kinetic model of the gene network of histidine biosynthesis. Such a model would allow prediction of the course of processes in the system, understanding of their mechanisms, determination of the key links of the gene network, and analysis of the effects of mutations on its operation. It will be an inextricable part of the “*in silico* cell” computer resource.

ACKNOWLEDGEMENTS

The authors are grateful to Vitaly Likhoshvai for valuable discussions, to Irina Lokhova for bibliographical support, and to Victor Gulevich for translating the manuscript from Russian into English. This work was supported in part by the Russian Government (Contract No. 02.467.11.1005), by Siberian Branch of the Russian Academy of Sciences (the project “Evolution of molecular-genetical systems: computational analysis and simulation” and integration projects Nos 24 and 115), and by the Federal Agency of Science and Innovation (innovation project No. IT-CP.5/001).

REFERENCES

- Ananko E.A., Podkolodny N.L., Stepanenko I.L., Podkolodnaya O.A., Rasskazov D.A., Miginsky D.S., Likhoshvai V.A., Ratushny A.V., Podkolodnaya N.N., Kolchanov N.A. (2005) GeneNet in 2005. *Nucl. Acids Res.*, **33**, D425–D427.
- He B., Choi K.Y., Zalkin H. (1993) Regulation of *Escherichia coli* *glnB*, *prsA*, and *speA* by the purine repressor. *J. Bacteriol.*, **175**(11), 3598–606.
- Hove-Jensen B., Harlow K.W., King C.J., Switzer R.L. (1986) Phosphoribosylpyrophosphate synthetase of *Escherichia coli*. Properties of the purified enzyme and primary structure of the *prs* gene. *J. Biol. Chem.*, **261**(15), 6765–6771.
- Khlebodarova T.M. *et al.* (2006) *In silico* cell II. Information sources for modeling *Escherichia coli* metabolic pathways. *This issue*.
- Likhoshvai V.A., Ratushny A.V. (2006) *In silico* cell I. Hierarchical approach and generalized hill functions in modeling enzymatic reactions and gene expression regulation. *This issue*.
- Post D.A., Hove-Jensen B., Switzer R.L. (1993) Characterization of the *hemA-prs* region of the *Escherichia coli* and *Salmonella typhimurium* chromosomes: identification of two open reading frames and implications for *prs* expression. *J. Gen. Microbiol.*, **139**(2), 259–266.

MODELING OF THE EFFECTS OF THREONINE, VALINE, ISOLEUCINE AND PYRIDOXAL 5'-MONOPHOSPHATE ON BIOSYNTHETIC THREONINE DEHYDRATASE REACTION

Nedosekina E.A.^{*1}, *Usuda Y.*², *Matsui K.*²

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ² Institute of Life Sciences, Ajinomoto, Co., Inc., Kawasaki, Japan

* Corresponding author: e-mail: nzhenia@bionet.nsc.ru

Key words: mathematical modeling, gene network, branched chain amino acid.

SUMMARY

Motivation: Development of mathematical models that adequately describe molecular-genetic mechanisms is one of the main tasks in modern bioinformatics. Not only should a model be sensitive to the particulars of enzyme reactions underlying cell life activity, but also to regulation exerted on enzyme activity by metabolites and specialized cell proteins.

Results: The gene network for branched-chain amino acid biosynthesis in *E. coli* has been reconstructed using the GeneNet system. On the basis of available information on this gene network, elementary mathematical models for enzyme reactions have been developed using generalized Hill functions. We herein give a detailed description of the mathematical model for the biosynthetic threonine dehydratase-catalyzed reaction that yields OBUT¹; the model includes enzyme activity regulation.

Availability: The model is available on request. Gene network is available at <http://www.mgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/index.shtml>.

INTRODUCTION

Branched-chain amino acids biosynthesis (BCAAB) in *Escherichia coli* runs down parallel pathways, largely encoded by dually functional enzymes. For example, five BCAAB operons: *ilvBN*, *ilvGMEDA*, *ilvYC*, *ilvIH* and *avtA*, encode enzymes that possess dual enzyme activity. The operons *ilvBN* and *ilvIH* encode the subunits of AHAS-I and AHAS-III, respectively (Table 1). The operon *ilvGMEDA* encode four enzymes involved in the biosynthesis of L-isoleucine and L-valine, of which: *ilvGM* encodes the subunits of AHAS-II; *ilvD*, dihydroxy-acid dehydratase; *ilvE* genes, transaminase B (the only enzyme that has three functions to perform); *ilvA*, biosynthetic

¹ The abbreviations used are: ABUT, 2-Aceto-2-hydroxy butyrate; ACCOA, Acetyl-CoA; ACLAC, Acetolactate; AKG, alpha-ketoglutarate; ALA, Alanine; CBHCAP, 3-Carboxy-3-hydroxy-isocaproate; CO₂, Carbon dioxide; COA, Coenzyme A; DHMVA, 2,3-Dihydroxy-3-methyl-valerate; DHVAL, Dihydroxy-isovalerate; GLU, Glutamate; ILE, Isoleucine; IPPMAL, 3-Isopropylmalate; LEU, Leucine; NAD, Nicotinamide adenine dinucleotide; NADH, NAD reduced; NADP, NAD phosphate; NADPH, Dihydronicotinamide adenine dinucleotide phosphate reduced; NH₃, Ammonia; OBUT, alpha-ketobutyrate; OICAP, 2-Oxoisocaproate; OIVAL, Oxoisovalerate; OMVAL, Oxomethylvalerate; PLP, pyridoxal 5'-monophosphate; PYR, Pyruvate; THR, threonine; VAL, Valine.

L-threonine deaminase (enzyme specific for L-isoleucine biosynthesis). The other members of this gene network are biodegradative L-threonine deaminase (*tdcB*), aromatic-amino-acid aminotransferase (*tyrB*) and the enzymes encoded by the operon *leuABCD* (leucine-biosynthetic operon).

Table 1. Biochemical reactions and genes encoding corresponding enzymes of the gene network for branched-chain amino acid biosynthesis

№	Enzyme names	Genes	Reaction
1	TD-bs; Threonine dehydratase, biosynthetic	<i>ilvA</i>	THR → NH ₃ + OBUT
2	TD-cat; Threonine dehydratase, catabolic	<i>tdcB</i>	THR → NH ₃ + OBUT
3	AHAS-I; Acetohydroxybutanoate synthase I	<i>ilvBN</i>	OBUT + PYR → ABUT + CO ₂
4	AHAS-II; Acetohydroxybutanoate synthase II	<i>ilvGM</i>	OBUT + PYR → ABUT + CO ₂
5	AHAS-III; Acetohydroxybutanoate synthase III	<i>ilvIH</i>	OBUT + PYR → ABUT + CO ₂
6	AHAIR; Acetohydroxy acid isomeroreductase	<i>ilvC</i>	ABUT + NADPH → NADP + DHMVA
7	DAD; Dihydroxy acid dehydratase	<i>ilvD</i>	DHMVA → OMVAL
8	BCAT; Branched chain amino acid aminotransferase	<i>ilvE</i>	OMVAL + GLU ↔ AKG + ILE
9	AHAS-I; Acetolactate synthase I	<i>ilvBN</i>	2 PYR → CO ₂ + ACLAC
10	AHAS-II; Acetolactate synthase II	<i>ilvGM</i>	2 PYR → CO ₂ + ACLAC
11	AHAS-III; Acetolactate synthase III	<i>ilvIH</i>	2 PYR → CO ₂ + ACLAC
12	AHAIR; Acetohydroxy acid isomeroreductase	<i>ilvC</i>	ACLAC + NADPH → NADP + DHVAL
13	DAD; Dihydroxy acid dehydratase	<i>ilvD</i>	DHVAL → OIVAL
14	BCAT; Branched chain amino acid aminotransferase	<i>ilvE</i>	OIVAL + GLU → AKG + VAL
15	TrC; Valine-pyruvate aminotransferase	<i>avtA</i>	OIVAL + ALA → PYR + VAL
16	IMS; Isopropylmalate synthase	<i>leuA</i>	ACCOA + OIVAL → COA + CBHCAP
17	IMI; Isopropylmalate isomerase	<i>leuCD</i>	CBHCAP ↔ IPPMAL
18	IMDH; 3-Isopropylmalate dehydrogenase	<i>leuB</i>	IPPMAL + NAD → NADH + OICAP + CO ₂
19	BCAT; Branched chain amino acid aminotransferase	<i>ilvE</i>	OICAP + GLU → AKG + LEU
20	ArAT; Aromatic amino acid transaminase	<i>tyrB</i>	OICAP + GLU → AKG + LEU

Biosynthetic L-threonine deaminase (TD-bs) of *E. coli K-12* was the first enzyme, on which inhibition by the end product of the biosynthetic pathway, in which it participates, was demonstrated (Umberger *et al.*, 1956). TD-bs too was in the spotlight while the concept of allostery as a regulatory mechanism was in the making (Monod *et al.*, 1965). The catalytically active form of TD-bs is tetrameric; the presence of the cofactor PLP is required to prevent the formation of non-active apodimers (Calhoun *et al.*, 1973). The allosteric repressor ILE strongly inhibits the enzyme, while VAL is rather a powerful activator of enzyme activity (Fig. 1). In the absence of ILE, the rate of the reaction catalyzed by TD-bs depends, in a sigmoid manner, on the initial concentration of THR, which is indicative of a substrate-dependent switch from non-active to active status of the enzyme (Eisenstein *et al.*, 1991).

We have reconstructed the BCAAB gene network and developed elementary mathematical models for enzyme reactions using Hill's generalized functions. We herein give a detailed description of the model for the biochemical reaction catalyzed by TD-bs. Unlike the model previously proposed by Yang *et al.* (2005), ours captures the effects of the cofactor of the enzyme (PLP).

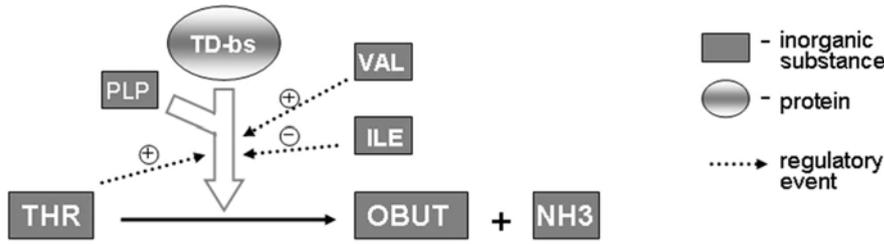


Figure 1. Regulation exerted on enzyme activity by TD-bs.

METHODS AND ALGORITHMS

The reconstruction of the BCAAB gene network was performed using the GeneNet system (Ananko *et al.*, 2005). The biochemical reactions were modeled using generalized Hill functions (Likhoshvai, Ratushny, 2006).

RESULTS

The BCAAB gene network has been reconstructed using the GeneNet system, a previous development (Khlebodarova *et al.*, 2006). The numbers of BCAAB components, of which GeneNet is aware, are presented in Table 2.

Table 2. The numbers of components of the branched-chain amino acid biosynthesis gene network, reconstructed in the GeneNet

Item	Protein	RNA	Gene	Operon	Reaction, regulation	Inorganic substance	Protein-repressor	Literature source
Amount	38	11	21	7	124	32	8	234

Using Hill's generalized functions, mathematical models have been developed for the enzyme reactions listed in Table 1. Verification of the model's parameters was performed using data in the Kinet database containing constants and dynamic data from published experimental works (Khlebodarova *et al.*, 2006). The steady-state and dynamic characteristics of the biochemical reactions studied using the model were compared with the experimental data and a good agreement was obtained.

Knowing how VAL and ILE affect the OBUT biosynthesis rate (VAL enhances enzyme activity, ILE reduces it) and because enzyme activation is substrate-dependent, the OBUT synthesis rate can be written in a generalized form as follows:

$$V = V_{\max} \frac{\left[\frac{THR}{K_m \frac{1 + ILE/k_1}{1 + VAL/k_3}} \right]^{N_h \cdot \frac{1 + ILE/k_2}{1 + VAL/k_4}} \cdot \left(\frac{PLP}{K_L} \right)}{\left(1 + \frac{THR}{K_m \frac{1 + ILE/k_1}{1 + VAL/k_3}} \right)^{N_h \cdot \frac{1 + ILE/k_2}{1 + VAL/k_4}} \left(1 + \frac{PLP}{K_L} \right) \cdot \left(1 + \frac{ILE}{K_I} \right)}$$

where V , V_{\max} are the current and maximum OBUT synthesis rates under standard conditions, respectively; THR , VAL , ILE and PLP are the concentrations of THR, VAL, ILE and PLP, respectively; K_m is the Michaelis-Menten constant for the substrate THR; k_1 , k_2 , K_I are the constants of ILE affecting OBUT synthesis rates; k_3 , k_4 are the constants of VAL affecting OBUT synthesis rates (they both equal to 1 mM); K_L is the

constant of PLP affecting OBUT synthesis rates; Nh is the constant of the non-linearity of THP influence on enzyme activity.

How the model was fit to experimental data on the effect of the THR, VAL, ILE and PLP concentrations on the OBUT biosynthesis rates is presented in Fig. 2. Fitting resulted in the following values for the parameters: $V_{max} = 210 \mu\text{mol}/\text{mg}/\text{min}$, $K_m = 8 \text{ mM}$, $k_1 = 0.006 \text{ mM}$, $k_2 = 0.06 \text{ mM}$, $k_3 = 1 \text{ mM}$, $k_4 = 1 \text{ mM}$, $K_I = 0.37 \text{ mM}$, $K_L = 0.0007 \text{ mM}$, $Nh = 2$.

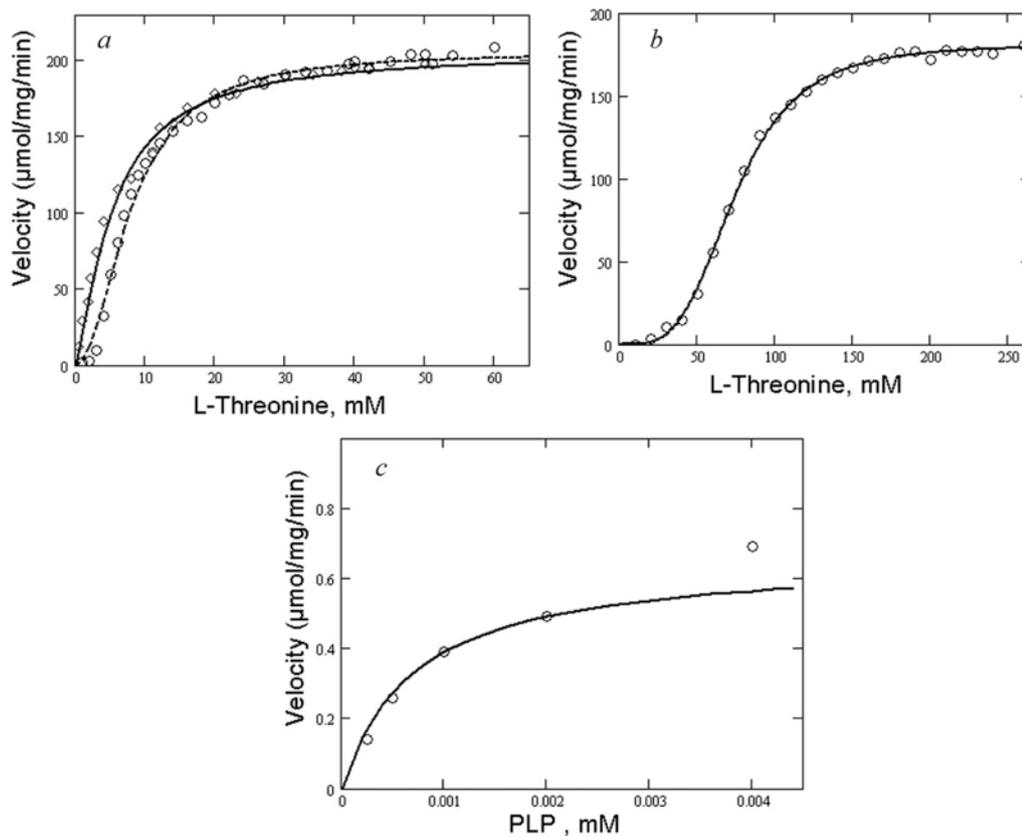


Figure 2. OBUT synthesis rates depending on concentrations of the effectors. A, depending on THR concentration (solid lines and diamonds: 0.5mM VAL, 0.03mM PLP, 0 ILE; broken line and circles: 0 VAL, 0.03mM PLP, 0 ILE). B, depending on THR concentration (0.05mM ILE, 0.03mM PLP, 0 VAL). C, depending on PLP concentration (0.45 THR, 0 VAL and ILE). Circles and diamonds indicate experimental data; a, b: from Eisenstein *et al.*, 1991, c – Calhoun *et al.*, 1973; curves indicate results obtained from the model.

DISCUSSION

The proposed model for OBUT synthesis rates captures the effects of the substrate THR, the reaction inhibitor ILE, the activator VAL and, unlike the model previously published by Yang (2005), that of the cofactor PLP (in the absence of which the enzyme is non-active). The model respects the fact that VAL and ILE affect the Michaelis-Menten constant (K_m): the former decreases it, the latter increases it. In the absence of the effectors, K_m is equal to 8 mM. Furthermore, VAL and ILE affect the degree of non-linearity of the dependence of reaction rates on THR concentrations: VAL makes the sigmoid quality of the dependence curve less pronounced; ILE, more pronounced. In the

absence of these effectors, Hill's non-linearity constant Nh is equal to 2. Under our model, isoleucine, too, affects the reaction rate by reducing it.

The development of mathematical models for metabolic processes allows their behavior to be predicted and the mechanisms underlying these processes to be studied. This model forms part of the computer resource known as Electronic Cell, which is currently under construction.

ACKNOWLEDGEMENTS

The authors are grateful to Vitaly Likhoshvai for valuable discussions, to Irina Lokhova for bibliographical support, and to Vladimir Filonenko for translating the manuscript from Russian into English. This work was supported in part by the Russian Government (Contract No. 02.467.11.1005), by Siberian Branch of the Russian Academy of Sciences (the project "Evolution of molecular-genetical systems: computational analysis and simulation" and integration projects Nos 24 and 115), and by the Federal Agency of Science and Innovation (innovation project No. IT-CP.5/001).

REFERENCES

- Ananko E.A. *et al.* (2005) GeneNet in 2005. *Nucl. Acids Res.*, **33**, D425–D427.
- Calhoun D.H., Rimerman R.A., Hatfield G.W. (1973) Threonine deaminase from *Escherichia coli*. I. Purification and properties. *J. Biol. Chem.*, **248**, 3511–3516.
- Eisenstein E. (1991) Cloning, expression, purification, and characterization of biosynthetic threonine deaminase from *Escherichia coli*. *J. Biol. Chem.*, **266**, 5801–5807.
- Khlebodarova T.M., *et al.* (2006) *In silico* cell II. Information sources for modeling *Escherichia coli* metabolic pathways. *This issue*.
- Likhoshvai V.A., Ratushny A.V. (2006) *In silico* cell I. Hierarchical approach and generalized hill functions in modeling enzymatic reactions and gene expression regulation. *This issue*.
- Monod J., Wyman J., Changeux J.-P. (1965) On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.*, **12**, 88–118.
- Umbarger H.E. (1956) Evidence for a negative-feedback mechanism in the biosynthesis of isoleucine. *Science*, **123**, 848.
- Yang C.R., Shapiro B.E., Hung S.P., Mjolsness E.D., Hatfield G.W. (2005) A mathematical model for the branched chain amino acid biosynthetic pathways of *Escherichia coli* K12. *J. Biol. Chem.*, **280**, 11224–11232.

REGULATION OF PYRIMIDINE BIOSYNTHESIS IN *ESCHERICHIA COLI*: GENE NETWORK RECONSTRUCTION AND MATHEMATICAL MODELING

Ratushny A.V.*, **Nedosekina E.A.**

Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia

* Corresponding author: e-mail: ratushny@bionet.nsc.ru

Key words: mathematical modeling, gene network, regulation, pyrimidine biosynthesis, *Escherichia coli*

SUMMARY

Motivation: Development of an *in silico* cell as a computer resource for the modeling and analysis of intracellular processes is a topical problem of the systems biology and bioinformatics. Within this direction, it is necessary to develop mathematical models of the genetic regulation of cell metabolic pathways, in particular, the regulation of pyrimidine biosynthesis.

Results: This work utilizes the GeneNet technology to reconstruct the gene network of pyrimidine biosynthesis in *E. coli* cell. In the context of generalized Hill functions, the mathematical models describing the functioning efficiencies of enzymatic systems and expression regulation of the genes coding for enzymes and their subunits are constructed.

Availability: Models are available on request; the gene network is available at <http://www.mgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/index.shtml>.

INTRODUCTION

A de novo synthesis of CTP² in *E. coli* involves eight consecutive enzymatic reactions catalyzed by aspartate transcarbamoylase (ATCase), dihydroorotase (DHOase), dihydroorotate oxidase (DHODase), orotate phosphoribosyltransferase (OPT), orotidine monophosphate decarboxylase (OMPase), uridylylate kinase (UMP kinase), nucleoside-diphosphate kinase (NDP kinase), and cytidine-triphosphate synthetase (CTP synthase). The corresponding genes are *pyrB* and *pyrI* (encoding ATCase), *pyrC*, *pyrD*, *pyrE*, *pyrF*, *pyrH*, *ndk*, and *pyrG*. The expression of the *pyr* genes is controlled in a complex manner by intracellular nucleotide pools. Thus, *pyrBI* operon appears to be repressed by a uridine nucleotide; *pyrC*, by a cytidine nucleotide; and *pyrE*, by uridine and guanine nucleotides. The nucleotide pools also regulate other stages of enzyme syntheses and activities. In addition, regulation of pyrimidine biosynthesis gene expression involves the protein substances PurR, NusA, and PyrS.

² The abbreviations used are: CTP, cytidine triphosphate; CAP, carbamoyl phosphate; ASP, L-aspartate; PI, phosphate; PPI, pyrophosphate; CAASP, carbamoyl aspartate; DOROA, dihydroorotic acid; Q, ubiquinone; QH₂, ubiquinol; OROA, orotic acid; PRPP, phosphoribosylpyrophosphate; OMP, orotidylate; CO₂, carbon dioxide; UMP, uridine monophosphate; UDP, uridine diphosphate; ATP, adenosine triphosphate; UTP, uridine triphosphate; GLN, glutamine; ADP, adenosine diphosphate; and GLU, glutamate.

The gene network regulating pyrimidine biosynthesis in the *E. coli* cell is reconstructed in the work. The mathematical models describing the functioning efficiencies of enzymatic systems and expression regulation of the genes coding for enzymes and their subunits are constructed.

METHODS AND ALGORITHMS

The gene network of pyrimidine biosynthesis was reconstructed using the GeneNet system (Ananko *et al.*, 2005). The method of generalized Hill functions (Likhoshvai, Ratushny, 2006) was used to model the regulation of gene expression and the functioning efficiency of enzymatic systems.

RESULTS AND DISCUSSION

This work utilizes the GeneNet technology (Ananko *et al.*, 2005) to reconstruct the gene network of pyrimidine biosynthesis in the *E. coli* cell (Fig. 1).

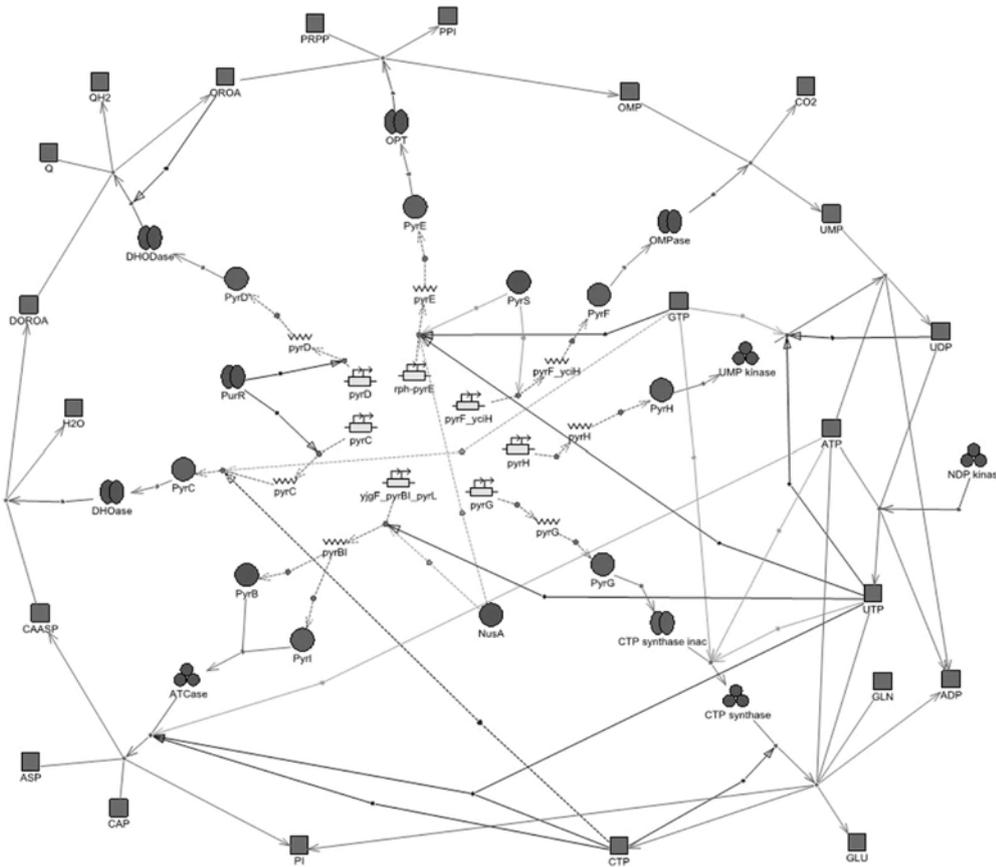


Figure 1. Reconstruction of pyrimidine biosynthesis gene network using GeneNet.

Table 1 shows the number of components of the gene network of pyrimidine biosynthesis.

Table 1. List of the components of gene network of pyrimidine biosynthesis

Item	Protein	RNA	Gene	Operon	Reaction	Inorganic substance	Repressor protein	Literature source
Number	20	7	8	7	61	22	3	180

Table 2 lists the enzymatic reactions involved in the gene network reconstruction, names of the enzymes catalyzing the corresponding reactions, and names of the genes encoding the corresponding enzymes. The mathematical models of enzymatic reactions (Table 2, column Reaction) and the models describing expression regulation of the genes encoding the corresponding enzymes and their subunits (Table 2, column Gene) were constructed using the method of generalized Hill functions. A database compiling the experimental information about dynamics of the components of this gene network (Khlebodarova *et al.*, 2006) was created. The parameters of these models were determined by numerical experiments. It was demonstrated that the equilibrium and dynamic characteristics of the gene network in question calculated using the models developed fit the experimental data.

Table 2. The enzymatic reactions involved in the pyrimidine biosynthesis gene network

No.	Enzyme name	Gene	Reaction
1	ATCase; aspartate-carbamoyltransferase	<i>pyrBI</i>	$CAP + ASP \rightarrow CAASP + PI$
2	DHOase; dihydroorotase	<i>pyrC</i>	$CAASP \leftrightarrow DOROA$
3	DHODase; dihydroorotate dehydrogenase	<i>pyrD</i>	$DOROA + Q \leftrightarrow QH2 + OROA$
4	OPT; orotate phosphoribosyl transferase	<i>pyrE</i>	$OROA + PRPP \leftrightarrow PPI + OMP$
5	OMPase; OMP decarboxylase	<i>pyrF</i>	$OMP \rightarrow CO2 + UMP$
6	UMP kinase; uridylylate kinase	<i>pyrH</i>	$UMP + ATP \leftrightarrow UDP + ADP$
7	NDP kinase; nucleoside-diphosphate kinase	<i>ndk</i>	$UDP + ATP \leftrightarrow UTP + ADP$
8	CTP synthetase	<i>pyrG</i>	$UTP + GLN + ATP \rightarrow GLU + CTP + ADP + PI$

Let us consider the expression regulation of *pyrC* gene in the *E. coli* cell as an example of application of the generalized Hill functions to modeling the molecular processes within the pyrimidine biosynthesis gene network. It is known that expression of this gene, coding for dihydroorotase, is regulated mainly at the level of translation via a nucleotide-dependent choosing of transcription start site (Wilson *et al.*, 1990). The transcription initiation at *pyrC* promoter is possible from four adjacent nucleotides located downstream of Pribnow box, namely, U-6, C-7, C-8, and G-9 (Wilson, Turnbough, 1987). It is demonstrated that C-7 transcripts are synthesized when the intracellular CTP level is high. A stable hairpin, preventing the binding to ribosome, is formed in these transcripts, which causes a low level of *pyrC* expression. When the CTP level is low and the GTP level is high (the cells with a decreased level of pyrimidines), G-9 transcripts are predominantly synthesized. These transcripts are shorter and incapable of forming the hairpin; consequently, the ribosome binds freely to RNA, thereby increasing the expression of dihydroorotase (Liu, Turnbough, 1994; Fig. 2).

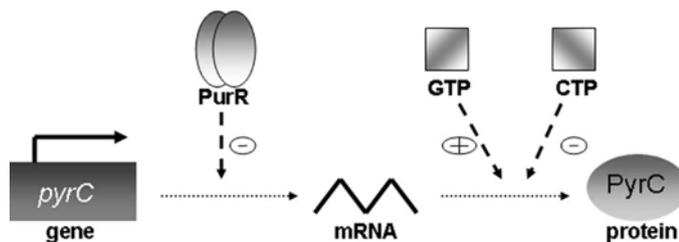


Figure 2. Expression regulation of *pyrC*.

In addition to the translation mechanisms, *pyrC* gene is also regulated at the transcription level: PurR repressor protein binds to the conservative sequence of *pyrC* gene, thereby inhibiting transcription initiation (Wilson *et al.*, 1990; Choi, Zalkin, 1990; Fig. 2).

Without going into details of the genetic mechanism of *pyrC* gene expression regulation, which as is evident from the above description, is very complex, note that the method of generalized Hill functions allows for obtaining a very compact description of the overall processes considered. Taking into account the type of PurR, GTP, and CTP effects on the expression of *pyrC* gene (PurR decreases the transcription level of *pyrC* gene, GTP increases the fraction of translated *pyrC* transcripts, and CTP decreases this fraction; in addition, GTP and CTP act independently of one another), the biosynthesis rate of PyrC protein may be represented in a generalized form by the following equation:

$$V = d \cdot V_{max}, \quad d = \frac{\left(\frac{GTP}{k1_{GTP}}\right)^{h_g}}{1 + \left(\frac{GTP}{k2_{GTP}}\right)^{h_g} + \left(\frac{CTP}{k_{CTP}}\right)^{h_c} + \frac{GTP^{h_g} \cdot CTP^{h_c}}{k_{GTP,CTP}^{h_g+h_c}} + \frac{1}{1 + \left(\frac{PurR}{k_{PurR}}\right)^{h_p}}}, \quad (1)$$

where V and V_{max} are the current and maximal biosynthesis rates of PyrC protein under standard conditions, respectively; d , the fraction of the current rate with respect to the maximal rate; $PurR$, GTP , and CTP are concentrations of PurR, GTP, and CTP; $k1_{GTP}$, $k2_{GTP}$, and k_{CTP} are the efficiency constants of the GTP and CTP effects on PyrC protein biosynthesis; $k_{GTP,CTP}$, efficiency constant of the joint GTP and CTP effect on PyrC protein biosynthesis; h_g and h_c , the constants characterizing the nonlinearity degree of the GTP and CTP effects on PyrC protein biosynthesis; k_{PurR} , the constant of PurR inhibition of PyrC synthesis; and h_p are the constants characterizing the nonlinearity degree of the PurR effect on PyrC synthesis.

Note that characteristic of the mathematical model proposed, constructed using a generalized Hill function, is a considerable simplicity (compared with the molecular biological processes simulated); nonetheless, it provides a very good fit of the experimental data and numerical calculations.

Fig. 3a, b shows the results of calculations using model (1) and their comparison with the experimental data on the effects of various CTP and GTP concentrations on the synthesis of *pyrC* translated transcripts. This comparison demonstrates a high adequacy of the model of genetic expression regulation of the gene considered.

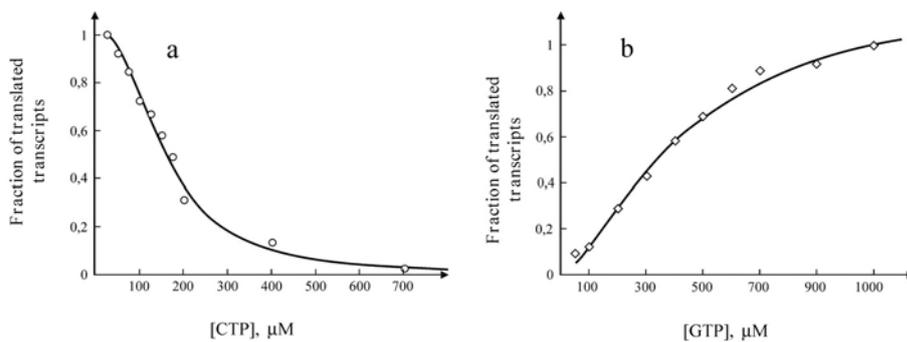


Figure 3. The effects of (a) CTP and (b) GTP on the synthesis of translated transcripts of *pyrC* (Wilson *et al.*, 1992). Experimental conditions: the transcripts were synthesized in the presence of ATP (2.7 mM) and (a) GTP (1.1 mM) or (b) CTP (0.1 mM). The abscissa, the concentration of (a) CTP and (b) GTP, μM ; the ordinate, the fraction of translated transcripts. (a) Circles and (b) diamonds represent experiment data of Wilson *et al.* (1992); the curves, calculations using model (1) with the following values of parameters ($k1_{GTP} = 30 \mu\text{M}$, $k2_{GTP} = 50 \mu\text{M}$, $k_{CTP} = 25 \mu\text{M}$, $k_{GTP,CTP} = 150 \mu\text{M}$, $h_c = 2.4$, and $h_g = 1.5$).

Gene network reconstruction and construction of the mathematical models describing the functioning efficiency of enzymatic systems and expression regulation of the genes encoding the corresponding enzymes is a necessary initial stage in constructing the general kinetic model of pyrimidine biosynthesis gene network. This model will allow for predicting dynamics of the processes going on in the system considered, studying their mechanisms, detecting key components of the gene network, and analyzing the effects of mutations on its function and will be an integral component of the computer resource under development—an *in silico* cell.

ACKNOWLEDGEMENTS

The authors are grateful to Vitaly Likhoshvai for valuable discussions, to Irina Lokhova for bibliographical support 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), 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

- Ananko E.A., Podkolodny N.L., Stepanenko I.L., Podkolodnaya O.A., Rasskazov D.A., Miginsky D.S., Likhoshvai V.A., Ratushny A.V., Podkolodnaya N.N., Kolchanov N.A. (2005) GeneNet in 2005. *Nucl. Acids Res.*, **33**, D425–D427.
- Choi K.Y., Zalkin H. (1990) Regulation of *Escherichia coli* pyrC by the purine regulon repressor protein. *J. Bacteriol.*, **172**, 3201–3207.
- Khlebodarova T.M. *et al.* (2006) *In silico* cell II. Information sources for modeling *Escherichia coli* metabolic pathways. *This issue*.
- Likhoshvai V.A., Ratushny A.V. (2006) *In silico* cell I. Hierarchical approach and generalized hill functions in modeling enzymatic reactions and gene expression regulation. *This issue*.
- Liu J., Turnbough C.L.Jr. (1994) Effects of transcriptional start site sequence and position on nucleotide-sensitive selection of alternative start sites at the pyrC promoter in *Escherichia coli*. *J. Bacteriol.*, **176**, 2938–2945.
- Wilson H.R., Archer C.D., Liu J.K., Turnbough C.L.Jr. (1992) Translational control of pyrC expression mediated by nucleotide-sensitive selection of transcriptional start sites in *Escherichia coli*. *J. Bacteriol.*, **174**, 514–524.
- Wilson H.R., Chan P.T., Turnbough C.L. (1987) Nucleotide sequence and expression of the pyrC gene of *Escherichia coli* K-12. *J. Bacteriol.*, **169**, 3051–3058.
- Wilson H.R., Turnbough C.L.Jr. (1990) Role of the purine repressor in the regulation of pyrimidine gene expression in *Escherichia coli* K-12. *J. Bacteriol.*, **172**, 3208–3213.

REGULATION OF THE PENTOSE PHOSPHATE PATHWAY IN *ESCHERICHIA COLI*: GENE NETWORK RECONSTRUCTION AND MATHEMATICAL MODELING OF METABOLIC REACTIONS

Ratushny A.V.^{*1}, Smirnova O.G.¹, Usuda Y.², Matsui K.²

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ² Institute of Life Sciences, Ajinomoto, Co., Inc., Kawasaki, Japan

* Corresponding author: e-mail: ratushny@bionet.nsc.ru

Key words: mathematical modeling, gene network, regulation, pentose phosphate pathway, glucose 6-phosphate-1-dehydrogenase, *Escherichia coli*

SUMMARY

Motivation: Development of an *in silico* cell, a computer resource for modeling and analysis of physiological processes, is an urgent task of systems biology and computational biology. Mathematical modeling of the genetic regulation of one of the basic metabolic processes, the pentose phosphate pathway, is an important problem to be solved as part of this line of work.

Results: By using the GeneNet technology, we reproduced the gene network of the regulation of the pentose phosphate pathway in the *E. coli* cell. Mathematical models were constructed by the method of generalized Hill functions to describe the efficiency of enzyme systems.

Availability: Models are available on request. Gene network is available at <http://www.mgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/index.shtml>.

INTRODUCTION

Apart from glycolysis, the pentose-phosphate pathway is a major route of intermediary carbohydrate metabolism. In addition to its role as a route for the breakdown of sugars such as glucose or pentoses, the pentose phosphate pathway is involved in the generation of reducing power (NADPH) for biosynthesis and provides the cell with intermediates for the anabolism of amino acids, vitamins, nucleotides, and cell wall constituents. In *E. coli*, two branches of the pentose-phosphate pathway occur: (1) the oxidative branch, in which pentose phosphate is formed from glucose-6-P, and (2) a non-oxidative branch in which fructose-6-P and glyceraldehydes-3-P, participants of the Embden-Meyerhof-Parnas pathway, are formed. Moreover, the pentose-phosphate pathway is the only pathway that allows *E. coli* to utilize sugars such as D-xylose, D-ribose, or L-arabinose, which cannot be catabolized by other routes (Sprenger, 1995).

The gene network of regulation of the pentose phosphate pathway in the *E. coli* cell was reconstructed. Mathematical models of the efficiency of enzyme systems were constructed. A database storing experimental data on the behavior of components of this gene network was developed (Khlebodarova *et al.*, this issue). Parameters of the models were determined by numerical simulation. The results of calculation of steady-state

properties and behavior of the components of the molecular system derived from the models are in agreement with experimental evidence.

METHODS AND ALGORITHMS

The gene network was reconstructed using the GeneNet system (Ananko *et al.*, 2005).

The method of generalized Hill functions (Likhoshvai, Ratushny, 2006) was used to model the regulation of the functioning efficiency of enzymatic systems.

RESULTS AND DISCUSSION

The pentose phosphate pathway of *Escherichia coli* involves 10 enzymatic reactions (Fig. 1, Table 1, 2). Most of the corresponding enzymes are repressed by the final products of the reactions: glucose 6-phosphate (G6P), 2-keto-3-deoxy-6-phosphogluconate (2KD6PG), fructose 6-phosphate (F6P), ribose 5-phosphate (R5P), glyceraldehyde 3-phosphate (T3P1), and ribulose 5-phosphate (RL5P). Also, fructose 1,6-diphosphate (FDP), phosphoenolpyruvate (PEP), arabinose 5-phosphate (A5P) and inorganic phosphate (PI) are downregulators of the pentose phosphate pathway. The activities of glucose 6-phosphate-1-dehydrogenase and 6-phosphogluconate dehydrogenase depend on the energy and redox potential of the cell. They are repressed by ATP, NADH and NADPH. The following proteins control the expression of the genes of the pentose phosphate pathway: SoxS, MarA, Rob (*zwf*), GadE (*gnd*), FruR (*edd-eda*), CreB (*talA-ktkB*), RpiR (*rpiB*), and LipB. Moreover, the pentose phosphate pathway is superoxide-sensitive. An iron-sulfur cluster of 6-phosphogluconate dehydratase coded by the *edd* gene is the superoxide-sensitive target site, which is readily destroyed by oxidation.

Table 1 summarizes the components of the network. Table 2 shows the enzymatic reactions present in the pentose phosphate pathway gene network, names of enzymes catalyzing corresponding reactions, and names of genes coding for the enzymes.

Table 1. Components of the pentose phosphate pathway gene network

Operon	RNA	Protein	Reaction	Inorganic substance	Repressor	Transcription factor	Reference
11	11	35	104	31	19	5	109

Table 2. Enzymatic reactions constituting the pentose phosphate pathway gene network

Enzyme	Gene	Reaction	EC
Glucose 6-phosphate-1-dehydrogenase	<i>zwf</i>	$G6P + NADP \leftrightarrow D6PGL + NADPH$	1.1.1.49
6-Phosphogluconolactonase	<i>pgl</i>	$D6PGL \rightarrow D6PGC$	3.1.1.31
6-Phosphogluconate dehydrogenase (decarboxylating)	<i>gnd</i>	$D6PGC + NADP \rightarrow NADPH + CO_2 + RL5P$	1.1.1.44
Ribose-5-phosphate isomerase A, B	<i>rpiA, rpiB</i>	$RL5P \leftrightarrow R5P$	5.3.1.6
Ribulose phosphate 3-epimerase	<i>rpe</i>	$RL5P \leftrightarrow X5P$	5.1.3.1
Transketolase I, II	<i>tktA, tktB</i>	$R5P + X5P \leftrightarrow T3P1 + S7P$	2.2.1.1
Transketolase I, II	<i>tktA, tktB</i>	$X5P + E4P \leftrightarrow F6P + T3P1$	2.2.1.1
Transaldolase A, B	<i>talA, talB</i>	$T3P1 + S7P \leftrightarrow E4P + F6P$	2.2.1.2
Phosphogluconate dehydratase	<i>edd</i>	$D6PGC \rightarrow 2KD6PG$	4.2.1.12
2-Keto-3-deoxy-6-phosphogluconate aldolase	<i>eda</i>	$2KD6PG \rightarrow T3P1 + PYR$	4.1.2.14

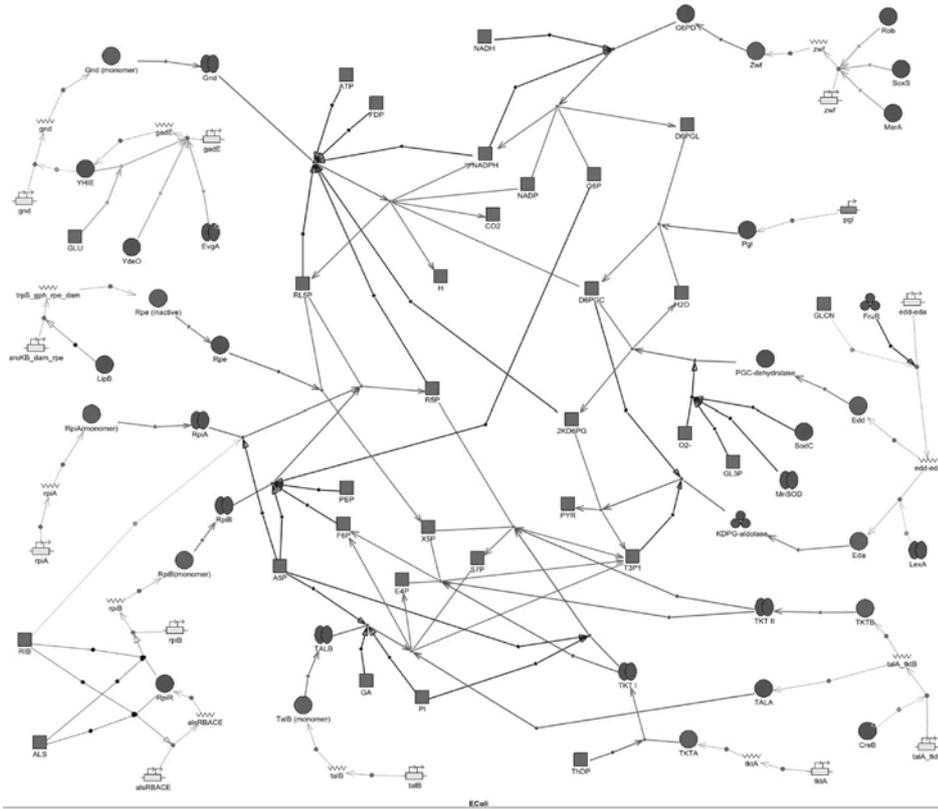


Figure 1. Pentose phosphate pathway gene network reconstruction in the GeneNet.

Application of the method of generalized Hill functions to modeling the molecular processes of the pentose phosphate pathway can be exemplified by regulation of the activity of the enzyme glucose 6-phosphate-1-dehydrogenase (G6PD) in the *E. coli* cell (Table 1). The mechanism of this process is very intricate. NADH inhibits the reaction and intricately affects the G6PD activity. The reaction rate sigmoidally depends on NADP content with the presence of NADH. With the absence of NADH, this dependence assumes a hyperbolic form. No sigmoidal dependence is observed at variable G6P concentrations under the same conditions either (Sanwal, 1970). The reaction rate is affected by NADP, which reduces G6P affinity to the enzyme. The reaction product, NADPH, inhibits the G6PD activity. A model for the steady-state rate of the reaction is proposed:

$$V = \frac{k_{cat} \cdot e_0 \cdot G6P}{K_{m,G6P} \cdot \left(1 + k_{ign} \cdot \frac{NADP^{htg}}{k_{ig}^{htg} + NADP^{htg}} \right) + G6P} \cdot \frac{\left(\frac{NADP}{K_{m,NADP}} \right)^{1+k_{dm} \cdot \frac{NADH^{hdt}}{k_{dt}^{hdt} + NADH^{hdt}}}}{1 + \left(\frac{NADP}{K_{m,NADP}} \right)^{1+k_{dm} \cdot \frac{NADH^{hdt}}{k_{dt}^{hdt} + NADH^{hdt}}} + \left(\frac{NADPH}{k_{NADPH}} \right)^{h_{NADPH}} + 1 + \left(\frac{NADH}{k_{NADH}} \right)^{h_{NADH}}}$$

(1)

where e_0 is the concentration of the enzyme glucose 6-phosphate-1-dehydrogenase; G6P, NADP, D6PGL, NADPH are concentrations of the corresponding low-molecular-weight substances; k_{cat} , the catalytic constant; $K_{m,G6P}$, $K_{m,NADP}$, Michaelis constants for corresponding substrates; k_{NADPH} , constant of inhibition by NADPH; h_{NADPH} , constant determining the nonlinearity of the effect of NADPH on the reaction rate; k_{dm} , k_{dt} , k_{NADH} , constants of the efficiency of the effect of NADH on the reaction rate; h_{hd} , h_{NADH} , constants determining the nonlinearity of the effect of NADH on the reaction rate; k_{ign} , k_{ig} , constants of the efficiency of the effect of NADP on the reaction rate; and h_{ig} , constant determining the nonlinearity of the effect of NADP on the reaction rate.

Experimental data obtained by Sanwal (1970) were used for testing the model of regulation of G6P1D activity. These data illustrate the effects of the substrates G6P and NADP on G6P1D activity with various concentrations of NADH and NADPH (Fig. 2).

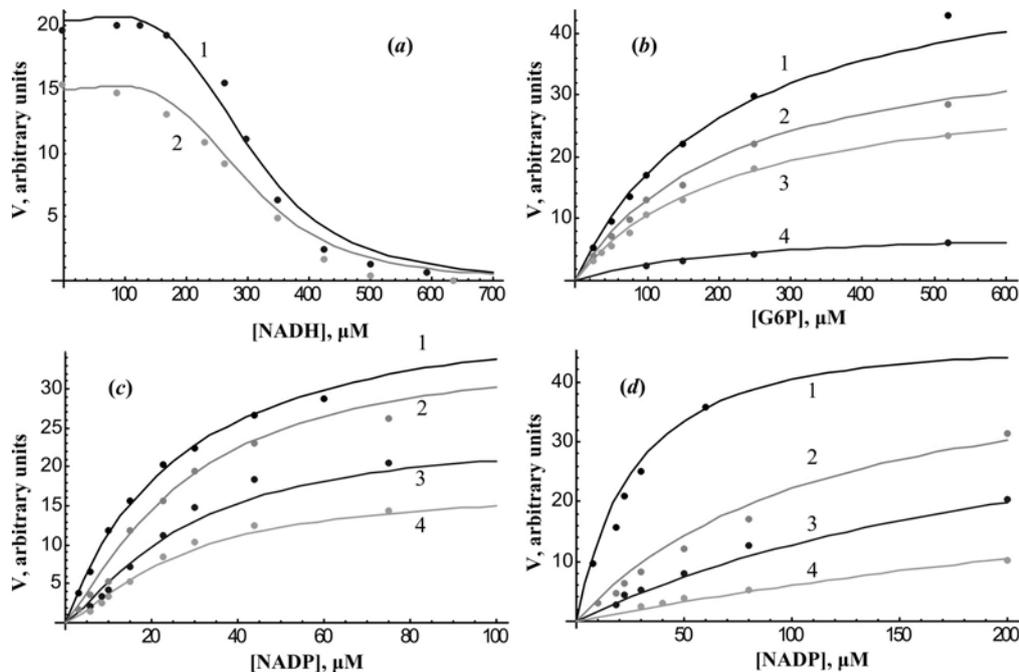


Figure 2. (a) Effect of NADH on the rate of the reaction catalyzed by G6P1D at various G6P concentrations; (b) Effect of G6P on the rate of the reaction catalyzed by G6P1D at various NADH concentrations; (c, d) Effect of NADP on the rate of the reaction catalyzed by G6P1D at various concentrations of (c) NADH and (d) NADPH. The enzyme activity was measured (a) with 50 μM NADP at pH = 7.5, G6P concentrations equaling: 1, 520 μM ; 2, 210 μM ; (b) with 0.3 mM NADP and 5 mM MgCl_2 , pH=7.5 at NADH concentrations: 1, null; 2, 176 μM ; 3, 264 μM ; 4, 44⁰ μM ; (c) with 0.42 mM G6P, 5 mM MgCl_2 , pH=7.5 at NADPH concentrations: 1, null; 2, 37 μM ; 3, 74 μM ; 4, 148 μM . Dots indicate experimental data from (Sanwal, 1970), and curves are the results of simulation according to model (1) with the following parameters: $K_{m,G6P} = 150 \mu\text{M}$, $K_{m,NADP} = 30 \mu\text{M}$, $k_{ign} = 1$, $k_{ig} = 400$, $h_{ig} = 1$, $k_{dm} = 0.3$, $k_{dt} = 100$, $h_{dt} = 2$, $k_{NADPH} = 15$, $h_{NADPH} = 1.4$, $k_{NADH} = 300$, and $h_{NADH} = 4$.

The model takes into account the intricate nonlinear mechanism of the reaction and dependence of the enzyme activity on various low-molecular-weight components. For example, the mode of the NADP effect on the reaction rate changes from hyperbolic without NADH to sigmoid with NADH. The Hill coefficient in the model functionally depends on NADH, equaling unity with its absence and increasing in a threshold manner with its presence, which also allows proper description of experimental evidence (Fig. 2c). Moreover, NADH itself extremely nonlinearly influences the reaction rate by reducing it (Fig. 2a). In model (1), inhibition by NADH is described by the product including the last fraction, the Hill coefficient, estimated to be $h_{NADH} = 4$. The model also

describes the effect of NADP on G6P affinity to the enzyme, a nonlinear twofold increase. Model (1) takes into account the nontrivial competition between NADPH and NADP, which allows proper description of experimental evidence (Fig. 2*d*).

Thus, the modeling method used in this work allows construction of proper mathematical models with relatively simple description of the processes modeled with the lack of knowledge on their fine mechanisms, when, for example, the King and Altman algorithm (King, Altman, 1956; Cornish-Bowden, 1977) cannot be applied.

Reconstruction of the gene network of regulation of the pentose phosphate pathway and development of mathematical models describing the efficiency of operation of enzyme systems and regulation of genes coding for these enzymes is essential for construction of an overall kinetic model of the network. Such a model will allow determination of key links of the gene network, prediction of the course of processes accompanying carbohydrate conversion in the pentose phosphate pathway, and analysis of the effects of mutations on its operation. The model of the pentose phosphate pathway will be an inextricable part of the “*in silico* cell” computer resource.

ACKNOWLEDGEMENTS

The authors are grateful to Vitaly Likhoshvai for valuable discussions, to Irina Lokhova for bibliographical support, and to Victor Gulevich for translating the manuscript from Russian into English. This work was supported in part by the Russian Government (Contract No. 02.467.11.1005), by Siberian Branch of the Russian Academy of Sciences (the project “Evolution of molecular-genetical systems: computational analysis and simulation” and integration projects Nos 24 and 115), and by the Federal Agency of Science and Innovation (innovation project No. IT-CP.5/001).

REFERENCES

- Ananko E.A. *et al.* (2005) GeneNet in 2005. *Nucl. Acids Res.*, **33**, D425–D427.
- Khlebodarova T.M. *et al.* (2006) *In silico* cell II. The kinet database as an information source for description of kinetic data. *This issue*.
- Likhoshvai V.A., Ratushny A.V. (2006) *In silico* cell I. Hierarchical approach and generalized hill functions in modeling enzymatic reactions and gene expression regulation. *This issue*.
- Cornish-Bowden A. (1977) An automatic method for deriving steady-state rate equations. *Biochem. J.*, **165**, 55–59.
- King E.L., Altman C. (1956) A schematic method of deriving the rate laws for enzyme-catalyzed reactions. *J. Phys. Chem.*, **60**, 1375–1378.
- Sanwal B.D. (1970) Regulatory mechanisms involving nicotinamide adenine nucleotides as allosteric effectors. 3. Control of glucose 6-phosphate dehydrogenase. *J. Biol. Chem.*, **245**(7), 1626–1631.
- Sprenger G.A. (1995) Genetics of pentose-phosphate pathway enzymes of *Escherichia coli* K-12. *Arch. Microbiol.*, **164**(5), 324–330.

AROMATIC AMINO ACID BIOSYNTHESIS IN *ESCHERICHIA COLI*: GENERALIZED HILL FUNCTION MODEL OF THE TRYPTOPHAN- SENSITIVE 3-DEOXY-D-ARABINO- HEPTULOSONATE-7-PHOSPHATE SYNTHASE REACTION DEMONSTRATE COMPLICATED MECHANISM

Ananko E.A.^{*}, Ratushny A.V., Usuda Y.¹, Matsui K.¹

Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ¹Institute of Life Sciences, Ajinomoto, Co., Inc., Kawasaki, Japan

^{*} Corresponding author: e-mail: eananko@bionet.nsc.ru

Key words: mathematical modeling, gene network, aromatic amino acid biosynthesis, *Escherichia coli*

SUMMARY

Motivation: Development of an *in silico* cell as a computer resource for simulation and analysis of processes within living cells is an urgent task of systems biology and computational biology.

Results: By using the GeneNet technology, we reproduced the gene network of the regulation of aromatic amino acid biosynthesis in the *E. coli* cell. Mathematical models were constructed by the method of generalized Hill functions. The models describe the efficiency of enzymatic systems and regulation of expression of related genes. Mathematical model of the enzyme tryptophan-sensitive 3-deoxy-d-arabino-heptulosonate-7-phosphate synthase reaction demonstrate complicated mechanism.

Availability: Models are available on request. The diagram of the gene network regulating aromatic amino acid biosynthesis in *E. coli* is available through the GeneNet viewer at <http://www.mgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/index.shtml>.

INTRODUCTION

The pathway of aromatic amino acid biosynthesis shows the common pathway (shikimate pathway) and three terminal pathways in which chorismate is converted to phenylalanine, tyrosine, and tryptophan, respectively (Pittard, 1996). In *E. coli*, these pathways involve 23 enzymes coded by the genes *aroF*, *aroG*, *aroH*, *aroB*, *aroD*, *aroE*, *aroK*, *aroL*, *aroA*, *aroC*, *tyrA*, *tyrB*, *aspC*, *pheA*, *trpE*, *trpD*, *trpC*, *trpA*, and *trpB*. Many of these genes, as well as the activity of the corresponding enzymes are controlled by end products, namely, tyrosine, phenylalanine, and tryptophan.

We reconstructed the gene network regulating aromatic amino acid biosynthesis in *E. coli* and developed mathematical models describing the operation of some individual enzymatic systems³.

³ The abbreviations used are: 3DDAH7P, 3-deoxy-D-arabino-heptulosonate-7-phosphate; DAHPS, 3-deoxy-d-arabino-heptulosonate-7-phosphate synthase; Trp, L-tryptophan.

regulated isozymes that catalyze the first step of aromatic biosynthesis, the condensation of phosphoenolpyruvate (PEP) and erythrose-4-phosphate (E4P) to form 3-deoxy-D-arabino-heptulosonate-7-phosphate (3DDAH7P). Here is the reaction of synthesis of 3DDAH7P from E4P and PEP, catalyzed by DAHPS: $E4P + PEP + H_2O \rightarrow 3DDAH7P$.

The enzyme is homodimeric and has two independent inhibitor binding sites. DAHPS(Trp) displays sigmoid kinetics with respect to both substrates, E4P and PEP. Both catalytic activity and substrate affinity of the DAHPS(Trp) are dependent on the species of activating metal ion. L-Tryptophan (Trp) binding decreases k_{cat} , decreases positive homotropic cooperativity for both substrates and activates the enzyme at low concentrations of E4P (Akowski, Bauerle, 1997).

The enzymatic reaction has a very intricate mechanism of DAHPS(Trp) regulation by Trp. With regard to the effect of Trp on various parameters of the molecular system under consideration, the rate of the enzymatic reaction can be expressed as:

$$V = \frac{k_{cat} \cdot e_0 \cdot \left(S_1 / (K_{m,S_1} \cdot f_2) \right)^{h_{S_1} \cdot f_3} \cdot \left(S_2 / K_{m,S_2} \right)^{h_{S_2} \cdot f_4}}{P_2 / K_{i,P_2} + \left[1 + \left(S_1 / (K_{m,S_1} \cdot f_2) \right)^{h_{S_1} \cdot f_3} \right] \cdot \left[1 + \left(S_2 / K_{m,S_2} \right)^{h_{S_2} \cdot f_4} \right]} \cdot f_1, \quad (1)$$

$$f_1 = \frac{1}{1 + kl_{R,V_{max}} \cdot \frac{R}{k_{R,V_{max}} + R}}, \quad f_2 = \frac{1}{1 + kl_{R,K_{m,S_1}} \cdot \frac{R^{h_{R,K_{m,S_1}}}}{k_{R,K_{m,S_1}} \cdot R^{h_{R,K_{m,S_1}}} + R^{h_{R,K_{m,S_1}}}}}, \quad (1)$$

$$f_3 = \frac{1}{1 + kl_{R,h_{S_1}} \cdot \frac{R^{h_{R,h_{S_1}}}}{k_{R,h_{S_1}} \cdot R^{h_{R,h_{S_1}}} + R^{h_{R,h_{S_1}}}}}, \quad f_4 = \frac{1}{1 + kl_{R,h_{S_2}} \cdot \frac{R^{h_{R,h_{S_2}}}}{k_{R,h_{S_2}} \cdot R^{h_{R,h_{S_2}}} + R^{h_{R,h_{S_2}}}}},$$

where V is the rate of the reaction; e_0 is the concentration of the enzyme DAHPS(Trp); S_1 , S_2 , P_1 , P_2 , and R are the concentrations of E4P, PEP, PI, 3DDAH7P, and Trp, respectively; k_f is the catalytic constant; K_{m,S_1} and K_{m,S_2} are the Michaelis–Menten constants for the substrates E4P and PEP, respectively; K_{i,P_2} is the constant of inhibition by the 3DDAH7P product; h_{S_1} and h_{S_2} are constants determining the nonlinearity of the effect of the substrates E4P and PEP on the reaction rate, respectively; $kl_{R,V_{max}}$ is the constant determining the maximum degree of reaction rate inhibition by Trp; $k_{R,V_{max}}$ is the constant of efficiency of the effect of Trp on the maximum rate of the reaction; $kl_{R,k}$ is the constant determining the maximum effect of Trp on the constant designated by k , where k assumes a character value from the set $\{K_{m,S_1}, h_{S_1}, h_{S_2}\}$; $k_{R,k}$ is the constant determining the efficiency of the effect of Trp on constant k ; and $h_{R,k}$ is the constant determining the nonlinearity of the effect of Trp on constant k .

Although simple, the model provides good agreement between experimental data and simulation results. Fig. 2a, b present the results of calculations according to model (1) compared with experimental data on the effect of various Trp concentrations on the rate of the reaction catalyzed by the enzyme DAHPS(Trp). Apparently, the proposed model of the genetic regulation of the gene under study is fairly precise.

DISCUSSION

Reconstruction of the gene network and development of mathematical models describing the efficiency of operation of enzymatic systems and regulation of genes coding for these enzymes must be the initial stage of the development of an overall

kinetic model of the gene network of aromatic amino acid biosynthesis. The advantage of the method suggested is that it enables to construct appropriate mathematical models with the minimal complexity of description of the modelled processes in conditions of the lack of knowledge about fine mechanisms of reactional process. It was demonstrated by the example of the model describing regulation of DAHPS(Trp) enzyme activity by Trp. The model will be an inextricable part of the “*in silico* cell” computer resource.

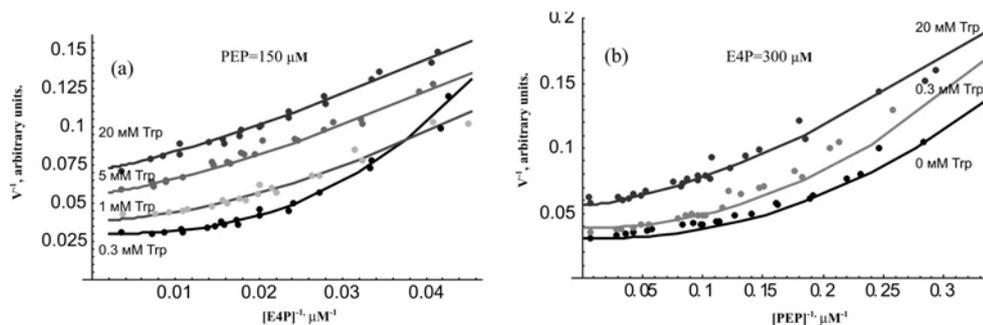


Figure 2. Effect of various Trp concentrations on the rate of the enzymatic reaction catalyzed by the DAHPS(Trp). Experimental conditions: *a* – PEP = 150 mM, *b* – E4P = 300 mM. On the X axis the reverse concentrations of substrates are plotted; on the Y axis, the reverse rates on the enzymatic reactions. Dots indicate experimental data according to (Akowski, Bauerle, 1997); curves are the results of simulation according to model (1); parameter values: $k_f = 20.6$ s⁻¹; $K_{m,S1} = 35$ mM; $K_{m,S2} = 5.3$ mM; $h_{S1} = 2.6$; $h_{S2} = 2.2$; $K_{i,P1} = 1$ mM; $kl_{R,Vmax} = 1.7$; $k_{R,Vmax} = 5$ mM; $kl_{R,KmS1} = 0.85$; $k_{R,KmS1} = 25$ mM; $h_{R,KmS1} = 0.6$; $kl_{R,hS1} = 1.1$; $k_{R,hS1} = 1$ mM; $h_{R,hS1} = 1$; $kl_{R,hS2} = 0.47$; $k_{R,hS2} = 1$ mM; $h_{R,hS2} = 2$.

ACKNOWLEDGEMENTS

The authors are grateful to Vitaly Likhoshvai for valuable discussions, to Irina Lokhova for bibliographical support, and to Victor Gulevich for translating the manuscript from Russian into English. This work was supported in part by the Russian Government (Contract No. 02.467.11.1005), by Siberian Branch of the Russian Academy of Sciences (the project “Evolution of molecular-genetical systems: computational analysis and simulation” and integration projects No. 24 and 115), and by the Federal Agency of Science and Innovation (innovation project No. IT-CP.5/001).

REFERENCES

- Akowski J.P., Bauerle R. (1997) Steady-state kinetics and inhibitor binding of 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase (tryptophan sensitive) from *Escherichia coli*. *Biochemistry*, **36**, 15817–15822.
- Ananko E.A. *et al.* (2005) GeneNet in 2005. *Nucl. Acids Res.*, **33**, D425–D427.
- Khlebodarova T.M. *et al.* (2006) *In silico* cell II. Information sources for modeling *Escherichia coli* metabolic pathways. *This issue*.
- Likhoshvai V.A., Ratushny A.V. (2006) *In silico* cell I. Hierarchical approach and generalized hill functions in modeling enzymatic reactions and gene expression regulation. *This issue*.
- Pittard J.A. (1996) Biosynthesis of the aromatic amino acids. In Neidhardt F.C. (ed.), *Escherichia coli and Salmonella*. In *Cellular and Molecular Biology*. Second Edition. Washington, DC: American Society for Microbiology Press., **1**, 458–484.

MATHEMATICAL MODELING OF REGULATION OF *cyoABCDE* OPERON EXPRESSION IN *ESCHERICHIA COLI*

Ratushny A.V.* , Khlebodarova T.M.

Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia

* Corresponding author: e-mail: ratushny@bionet.nsc.ru

Key words: mathematical modeling, regulation, gene, operon, *cyoABCDE*, *Escherichia coli*

SUMMARY

Motivation: Development of an *in silico* cell is an urgent task of systems biology and computational biology. Mathematical modeling of the genetic regulation of cell respiration is an important problem to be solved as part of this line of work.

Results: The mathematical model was constructed by the method of generalized Hill functions to describe the *cyoABCDE* operon expression regulation in wild-type *E. coli* and mutants lacking the ArcA ($\Delta arcA$) and Fnr (Δfnr) regulators. It was found that synergism between transcription factors ArcA and Fnr is important for the function of this operon.

INTRODUCTION

Aerobic metabolism is characterized by a high rate and efficiency of NAD⁺ production. This feature determines the choice of the preferential dehydrogenase/terminal oxidase couple under aerobic conditions. Depending on oxygen concentration, such couples are: NADH dehydrogenase II/cytochrome *bd* oxidase or NADH dehydrogenase II/cytochrome *bo* oxidase (Unden, Bongaerts, 1997). It is known that oxygen is essential for the function of cytochrome oxidases, being their substrate. However, expression of the *cydAB* operon, coding for cytochrome *bd* oxidase, is observed even under anoxic conditions. Owing to high affinity to oxygen, this enzyme can utilize residual oxygen even at anaerobiosis (Iuchi *et al.*, 1990). This suggests fine regulation mechanisms for components of the respiratory chain, which helps the cell survive under abrupt environmental changes.

We analyze expression of the *cyoABCDE* operon in *E. coli* by computer modeling depending on oxygen concentration and expression of global respiration regulators ArcA and Fnr.

METHODS AND ALGORITHMS

The mathematical model of the *cyoABCDE* operon expression regulation was constructed by the method of generalized Hill functions (Likhoshvai, Ratushny, 2006). The following generalized model was proposed for describing the dependence of gene (operon) expression regulated at the transcription initiation level on transcription factor concentrations:

$$\frac{dmRNA_g}{dt} = k \cdot \frac{k_0 + \sum_{s_{i_1}}^{C_{As,1}} \left(\frac{r_{s_{i_1}}}{k_{s_{i_1}}} \right)^{h_{s_{i_1}}} + \sum_{s_{i_1}, s_{i_2}}^{C_{As,2}} \frac{r_{s_{i_1}}^{h_{s_{i_1}}} r_{s_{i_2}}^{h_{s_{i_2}}}}{k_{s_{i_1,2}}} + \dots + \sum_{s_{i_1}, \dots, s_{i_M}}^{C_{As,M}} \frac{\prod_{v=1}^M r_{s_{i_v}}^{h_{s_{i_v}}}}{k_{s_{i_1, \dots, i_M}}^{\sum_{v=1}^M h_{s_{i_v}}}}}{1 + \sum_{s_{j_1}}^{C_{Is,As,1}} \left(\frac{r_{s_{j_1}}}{k_{s_{j_1}}} \right)^{h_{s_{j_1}}} + \sum_{s_{j_1}, s_{j_2}}^{C_{Is,As,2}} \frac{r_{s_{j_1}}^{h_{s_{j_1}}} r_{s_{j_2}}^{h_{s_{j_2}}}}{k_{s_{j_1,2}}} + \dots + \sum_{s_{j_1}, \dots, s_{j_N}}^{C_{Is,As,N}} \frac{\prod_{w=1}^N r_{s_{j_w}}^{h_{s_{j_w}}}}{k_{s_{j_1, \dots, j_N}}^{\sum_{w=1}^N h_{s_{j_w}}}}}, \quad (1)$$

where $mRNA_g$ is the concentration of the mRNA coded by the g gene; r_{s_i} , concentration of the transcription factor interacting with the s_i regulatory site; k , transcription rate constant of the g gene; k_0 , basal g transcription rate; k_{s_i} are constants of the efficiency of regulators r_{s_i} on the rate of g transcription; $k_{s_{i_1}, \dots, s_{i_z}}$ is the constant of the efficiency of the combined effect of regulators $(r_{s_{i_1}}, \dots, r_{s_{i_z}})$ on the rate of g transcription; $h_{s_{i_z}}$, constant describing the nonlinearity of the effect of regulator $r_{s_{i_z}}$ on g transcription; $C_{As,m}$, number of combinations of m nonoverlapping transcription activation sites ($m = \overline{1..M}$); and $C_{Is,As,n}$, number of combinations of nonoverlapping transcription activation and repression sites ($n = \overline{1..N}$).

RESULTS AND DISCUSSION

On the grounds of Eq. (1), taking into account the functional organization of the *cyoABCDE* promoter region (Fig. 1) and mechanisms of the action of transcription factors (Fnr and ArcA repress the transcription; GadE activates the transcription) (<http://ecocyc.org>), we propose the following equation for the efficiency of *cyoABCDE* operon expression:

$$f_{cyo} = \frac{k_0 + \left(\frac{G}{k_{G,1}} \right)^{h_G}}{1 + \left(\frac{A}{k_A} \right)^{h_A} + \left(\frac{G}{k_{G,2}} \right)^{h_G} + \left(\frac{Fn}{k_{Fn}} \right)^{h_{Fn}} + \frac{Fn^{h_{Fn1}} \cdot A^{h_{A1}}}{k_{Fn,A}^{h_{Fn1}+h_{A1}}} + \frac{A^{h_{A2}} \cdot G^{h_{G1}}}{k_{A,G}^{h_{A2}+h_{G1}}}}, \quad (2)$$

where f_{cyo} is the rate of *cyoABCDE* expression; A , G , Fn are concentrations of transcription factors ArcA, GadE, and Fnr, respectively; k_0 is the basal rate of *cyoABCDE* expression; $k_{G,i}$, k_A , k_{Fn} , are constants of the efficiency of regulators GadE, ArcA и Fnr, respectively, on the rate of *cyoABCDE* expression; $k_{Fn,A}$, $k_{A,G}$ constants of the efficiency of mutual action of regulators Fnr and ArcA, ArcA and GadE, respectively; h_i constant describing the nonlinearity of the effect of regulator i , where i assumes a character value from the set $\{A, G, Fn\}$.

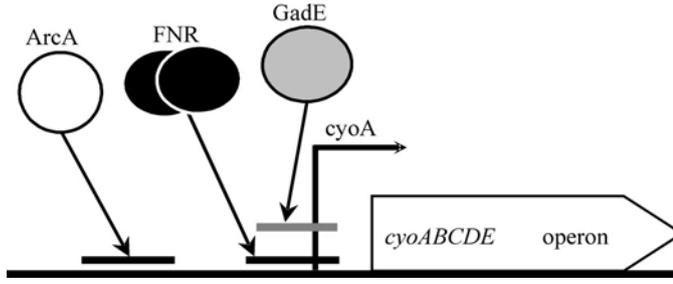


Figure 1. Structure of the promoter of the *cyoABCDE* operon, coding for cytochrome *bo* oxidase. Retrieved from Ecocyc database (<http://ecocyc.org>).

The activity of regulators Fnr and ArcA is known to depend on oxygen concentration (Tseng *et al.*, 1996). To take this fact into account, we propose generalized models describing the dependence of the relative activities of Fnr and ArcA in the cell on oxygen concentration – Eqs. (3, 4), respectively:

$$f_{Fnr} = \frac{k_{f0,O_2} + \left(\frac{O_2}{k_{f1,O_2}}\right)^{h_{f1,O_2}}}{1 + \left(\frac{O_2}{k_{f2,O_2}}\right)^{h_{f2,O_2}}} \quad (3) \quad f_A = \frac{k_{a0,O_2} + \left(\frac{O_2}{k_{a1,O_2}}\right)^{h_{a1,O_2}}}{1 + \left(\frac{O_2}{k_{a2,O_2}}\right)^{h_{a2,O_2}}}, \quad (4)$$

where k_{f0,O_2} and k_{a0,O_2} are activities of Fnr and ArcA without oxygen, respectively; k_{f1,O_2} and k_{a1,O_2} , constants of the effect of oxygen on the activities of Fnr and ArcA, respectively; and h_{f1,O_2} and h_{a1,O_2} , constants describing the nonlinearity of the effect of oxygen on the activities of Fnr and ArcA, respectively.

From Eqs. (3, 4) and the model of *cyoABCDE* expression (2) constructed in terms of regulator concentrations, we deduce the equation for regulation of *cyoABCDE* operon transcription in terms of oxygen concentration:

$$f_{cyo(WT)} = \frac{k_0 + \left(\frac{O_2}{k_{a1}}\right)^{h_{a1}} + \left(\frac{O_2}{k_{f1}}\right)^{h_{f1}} + \left(\frac{O_2}{k_{af1}}\right)^{h_{af1}}}{1 + \left(\frac{O_2}{k_{a2}}\right)^{h_{a2}} + \left(\frac{O_2}{k_{f2}}\right)^{h_{f2}} + \left(\frac{O_2}{k_{af2}}\right)^{h_{af2}}}, \quad (5)$$

where $f_{cyo(WT)}$ is the dependence of *cyoABCDE* operon expression; O_2 , oxygen concentration in the medium; k_0 , basal expression of the *cyoABCDE* operon; k_{a1} , k_{f1} , k_{af1} are constants of the effect of ArcA, Fnr, and their combined effect on *cyoABCDE* operon expression, respectively; h_{a1} , h_{f1} , h_{af1} , constants describing the nonlinearity of the effect of ArcA, Fnr, and their combined effect on *cyoABCDE* expression, respectively, expressed in terms of oxygen concentration.

Equation (5) takes into account the functional organization of the expression regulation system of the *cyoABCDE* operon. By elimination of components considering the Fnr- and ArcA-mediated effect of oxygen, we obtain corresponding models of mutant strains Δfnr (6) and $\Delta arcA$ (7), respectively:

$$f_{\text{cyo}(\Delta\text{fnr})} = \frac{k_{a0} + \left(\frac{O_2}{k_{a1}}\right)^{h_{a1}}}{1 + \left(\frac{O_2}{k_{a2}}\right)^{h_{a2}}} \quad (6)$$

$$f_{\text{cyo}(\Delta\text{arcA})} = \frac{k_{f0} + \left(\frac{O_2}{k_{f1}}\right)^{h_{f1}}}{1 + \left(\frac{O_2}{k_{f2}}\right)^{h_{f2}}} \quad (7)$$

Fig. 2 shows the results of predictions from models represented by Eqs. (5–7) in comparison with experimental data on the oxygen concentration dependence of *cyoABCDE* operon transcription for wild-type *E. coli* cells ($f_{\text{cyo}(WT)}$) and mutant strains ($f_{\text{cyo}(\Delta\text{fnr})}$ and $f_{\text{cyo}(\Delta\text{arcA})}$) (Tseng *et al.*, 1996). Calculations from Eqs. (6, 7), referring to corresponding mutant strains, are in good agreement with experimental data shown by dots in the figures. The description of the life of wild-type cells under microaerobic conditions provided by Equation (5) is not quite perfect.

The data indicate that the combined regulating effect of two transcription factors, Fnr and ArcA is very complex. The degree of cooperativity of Fnr and ArcA is very high under microaerobic conditions. To correct for this fact in model (5), we increased the Hill coefficients for terms in the numerator and denominator corresponding to the combined effect of Fnr and ArcA. This yielded the desired pattern of *cyoABCDE* expression in wild-type cells but increased the total transcription level dramatically.

Analysis of the system led us to the conclusion that wild-type cells use a mechanism restraining *cyoABCDE* transcription. It was found that such a mechanism can involve complex effects of Fnr and ArcA on each other's expression (Compan, Touati, 1994; Shalel-Levanon *et al.*, 2005).

With regard to these data, model (5) was modified so that the equation assumed the form:

$$f_{\text{cyo}(WT)} = \frac{k_0 + \left(\frac{O_2}{k_{a1}}\right)^{h_{a1}} + \left(\frac{O_2}{k_{f1}}\right)^{h_{f1}} + \left(\frac{O_2}{k_{af1}}\right)^{h_{af1}}}{1 + \left(\frac{O_2}{k_{a2}}\right)^{h_{a2}} + \left(\frac{O_2}{k_{f2}}\right)^{h_{f2}} + \left(\frac{O_2}{k_{af2}}\right)^{h_{af2}}} \cdot \frac{I + \frac{O_2}{k_{fa1}}}{I + \frac{O_2}{k_{fa2}}}, \quad (8)$$

where k_{fa1} , k_{fa2} are constants of the effect of Fnr on ArcA activity expressed in terms of oxygen concentration. Other designations as in Eq. (5).

The following parameter values were chosen by numerical fitting: $k_0 = 0.4$; $k_{a0} = 1.4$; $k_{a1} = 2.8 \mu\text{M}$; $h_{a1} = 1$; $k_{a2} = 40 \mu\text{M}$; $h_{a2} = 1.7$; $k_{f0} = 1.2$; $k_{f1} = 1.8 \mu\text{M}$; $h_{f1} = 1$; $k_{f2} = 19 \mu\text{M}$; $h_{f2} = 1.5$; $k_{af1} = 10 \mu\text{M}$; $h_{af1} = 5.5$; $k_{af2} = 20 \mu\text{M}$; $h_{af2} = 5.7$; $k_{fa1} = 0.54 \mu\text{M}$; $k_{fa2} = 0.1 \mu\text{M}$. These values give the best fit to the experimental data reported in (Tseng *et al.*, 1996) (Fig. 3). As shown in Fig. 3, curves predicted from equation (8) perfectly describe the experimental data, indicated by dots. The estimated Hill coefficient values ($h_{af1} = 5.5$, $h_{af2} = 5.7$) point to a complex mechanism of *cyoABCDE* expression and an important role of synergism between the ArcA and Fnr proteins.

Generally, we consider this study to be a reconstruction of a small fragment of a complex gene network describing respiration regulation in *E. coli*. It is the first indispensable step in the construction of the general kinetic model of this molecular system. This model will be applied to determination of key links of the network, analysis of the effect of mutations on its operation, and study of mechanisms of transition between respiration modes. The mathematical model of regulation of the *cydAB* (Likhoshvai, Ratushny, 2006) and *cyoABCDE* operons and the general model of the respiration regulation gene network in the *E. coli* cell will be inextricable parts of the “*in silico* cell” computer resource.

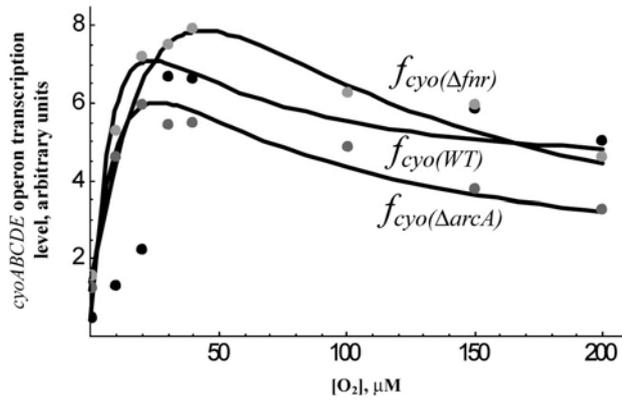


Figure 2. Dependence of the *cyoABCDE* operon transcription on oxygen concentration in wild-type *E. coli* cells ($f_{cyo(WT)}$) and mutant strains ($f_{cyo(\Delta arcA)}$ and $f_{cyo(\Delta fnr)}$). Dots indicate experimental data (Tseng *et al.*, 1996). Curves indicate calculations from Eqs. (5–7). The curves were calculated with the following parameters: $k_0 = 0.4$; $k_{a0} = 1.4$; $k_{a1} = 2.8 \mu\text{M}$; $h_{a1} = 1$; $k_{a2} = 40 \mu\text{M}$; $h_{a2} = 1.7$; $k_{f0} = 1.2$; $k_{f1} = 1.8 \mu\text{M}$; $h_{f1} = 1$; $k_{f2} = 19 \mu\text{M}$; $h_{f2} = 1.5$; $k_{af1} = 20 \mu\text{M}$; $h_{af1} = 2$; $k_{af2} = 40 \mu\text{M}$; $h_{af2} = 1.3$.

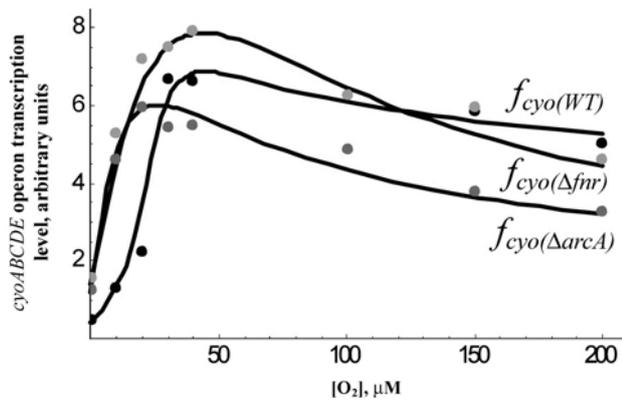


Figure 3. Dependence of the *cyoABCDE* operon transcription on oxygen concentration in wild-type *E. coli* cells ($f_{cyo(WT)}$) and mutant strains ($f_{cyo(\Delta arcA)}$ and $f_{cyo(\Delta fnr)}$). Dots indicate experimental data (Tseng *et al.*, 1996). Curves indicate calculations from Eqs. (8–7).

ACKNOWLEDGEMENTS

The authors are grateful to Vitaly Likhoshvai for valuable discussions, to Irina Likhova for bibliographical support, and to Victor Gulevich for translating the manuscript from Russian into English. This work was supported in part by the Russian Government (Contract No. 02.467.11.1005), by Siberian Branch of the Russian Academy of Sciences (the project “Evolution of molecular-genetical systems: computational analysis and simulation” and integration projects Nos 24 and 115), and by the Federal Agency of Science and Innovation (innovation project No. IT-CP.5/001).

REFERENCES

- Compan I., Touati D. (1994) Anaerobic activation of *arcA* transcription in *Escherichia coli*: roles of Fnr and ArcA. *Mol. Microbiol.*, **11**(5), 955–964.
- Iuchi S. *et al.* (1990) Requirement for terminal cytochromes in generation of the aerobic signal for the *arc* regulatory system in *Escherichia coli*: study utilizing deletions and lac fusions of *cyo* and *cyd*. *J. Bacteriol.*, **172**, 6020–6025.
- Likhoshvai V.A., Ratushny A.V. (2006) *In silico* cell I. Hierarchical approach and generalized hill functions in modeling enzymatic reactions and gene expression regulation. *This issue*.
- Shalel-Levanon S. *et al.* (2005) Effect of oxygen, and ArcA and FNR regulators on the expression of genes related to the electron transfer chain and the TCA cycle in *Escherichia coli*. *Metab. Eng.*, **7**, 364–374.
- Tseng C.P. *et al.* (1996) Effect of microaerophilic cell growth conditions on expression of the aerobic (*cyoABCDE* and *cydAB*) and anaerobic (*narGHJI*, *frdABCD*, and *dmsABC*) respiratory pathway genes in *Escherichia coli*. *J. Bacteriol.*, **178**, 1094–1098.
- Uden G., Bongaerts J. (1997) Alternative respiratory pathways of *Escherichia coli*: energetics and transcriptional regulation in response to electron acceptors. *Biochim. Biophys. Acta.*, **1320**, 217–234.

GENE NETWORK RECONSTRUCTION AND MATHEMATICAL MODELING OF *E. COLI* RESPIRATION: REGULATION OF F0F1-ATP SYNTHASE BY METAL IONS

Khlebodarova T.M.^{*}, *Lashin S.A.*, *Apasieva N.V.*

Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia

^{*} Corresponding author: e-mail: tamara@bionet.nsc.ru

Key words: mathematical modeling, gene network, respiration, *Escherichia coli*

SUMMARY

Motivation: Development of an *in silico* cell as a computer resource for simulation and analysis of processes within living cells is an urgent task of systems biology and computational biology. Within this direction, it is necessary to develop mathematical models of the genetic regulation of cell metabolic pathways, in particular, the regulation of *E. coli* respiration enzymes.

Results: By using the GeneNet technology, we reproduced the gene network of the regulation of respiration in the *E. coli* cell. Mathematical models were constructed by the method of generalized Hill functions. The models describe the velocities of enzymatic reactions.

Availability: The diagram of the gene network “Respiration” is available through the GeneNet viewer at <http://wwwmgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/index.shtml>.

INTRODUCTION

Fine mechanisms of expression regulation of the respiration enzymes in *E. coli* allow to cells well living in various conditions. More than 20 various enzymes provide this diversity forming respiration chain, which consist of pairs of interacting enzymes: oxidizer-deoxidizer. Optimal pair (or set of pairs) choice depends on a lot of parameters, but the significative ones are the oxygen concentration and presence of some substrate in the environment. Complex analysis of such complicated processes as respiration is very difficult without computer modeling. Stating the problem of *in silico* cell model development we have used the gene network reconstruction methodology and modeling in the terms of elementary processes. On this evidence the gene network of respiration in *E. coli* was reconstructed and the enzymatic processes database was developed.

METHODS AND ALGORITHMS

The gene network of regulation respiration was reconstructed with the GeneNet technology (Ananko *et al.*, 2005). Mathematical models were constructed by the method of generalized Hill functions (Likhoshvai, Ratushny, 2006).

RESULTS AND DISCUSSION

The GeneNet technology (Ananko *et al.*, 2005) was applied to reconstruction of the gene network regulating the respiration of *E. coli* cells. The gene network of respiration contains description of genetic regulation of operons, controlling enzymes synthesis of the respiration chain and metabolic processes, providing respiration of *E. coli* cells. The fragment of this network is shown on the Fig. 1. Section “Respiration” of the GeneNet contains description of 22 operons, 27 mRNA, 111 proteins, 64 different metabolites and others small molecules, and 313 interrelations between components. The information has been extracted from 241 scientific papers.

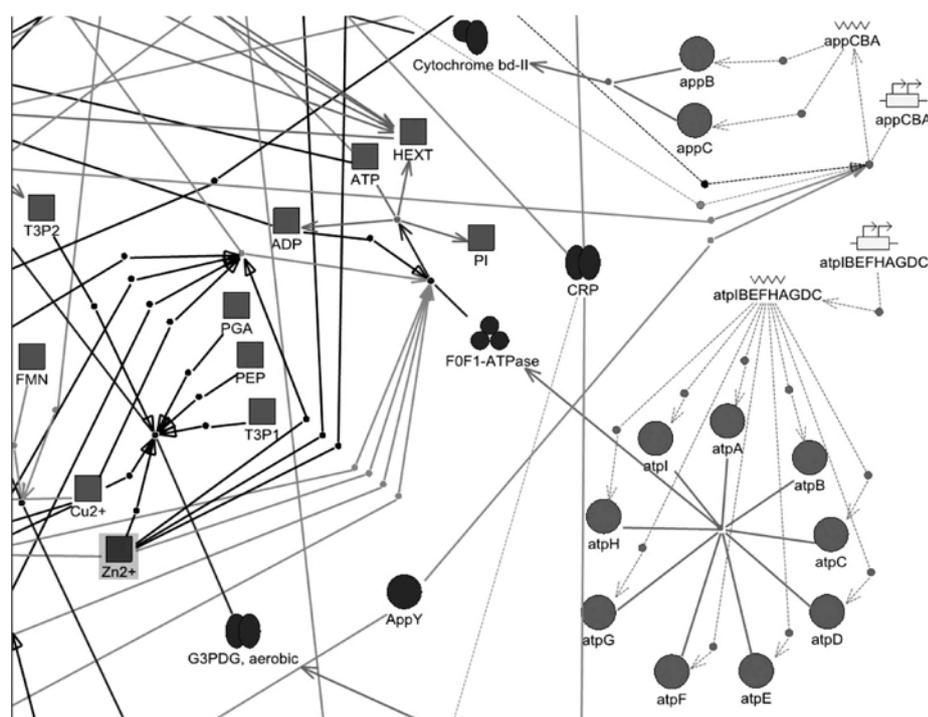


Figure 1. Fragment of the “Respiration” GeneNet diagram.

We also developed a database storing experimental data on the dynamic behaviour of the gene network components. Its content is illustrated in Table 1. As seen, at the present day, the database accumulates descriptions of ~340 constants and dynamic profiles involved in kinetics of the processes taking place in gene network of respiration.

Table 1. The parameters of enzymatic reactions, providing respiration of *E. coli* cells (the data K_m include the values of both K_m of substrates and cofactors)

Class	Dynamic	Enzymatic reaction parameters						
	Profile	Michaelis constant K_m	Constant of catalytic activity k_{cat}	Maximal velocity V_{max}	Dissociation constant K_d	Hill coefficient n^H	Constant of activation K_{act}	Constant of inhibition K_i
Amount	92	122	22	47	16	11	2	26

Mathematical models of 19 enzymatic reactions were constructed. Parameters of the models were determined by numerical experiments. The results of calculation of steady-state and dynamic parameters deduced from the models were in agreement with experimental data. Now, we consider regulation of the reaction catalyzed by the membrane-

bound F_0F_1 -ATPase, coded by the *atpABCDEFGHI* operon, as an example of simulation of molecular processes in the gene network of in *E. coli* (Fig. 1). This enzyme catalyzes the reaction of ATP synthesis and hydrolysis in accordance of the following equation.



Moreover the ATP synthesis carries in aerobic conditions and is provided of those respiration enzymes, which are capable of forming the proton gradient in the chain of electrons transport. ATP hydrolysis carries in anaerobic conditions and enzyme's catalytic domain (F_1) contains three binding sites of ATP whose affinity depends on the Mg^{2+} presence (Weber *et al.*, 1996). It determines the complex cooperative effect of ATP and Mg^{2+} on the ATPase activity. In the proposed model we also take into account ATP and ADP effects of substrate and competitive inhibition respectively. There is the activator effect of Na^+ on the ATPase activity at the concentrations lower than 15–20 mM and inhibitor effect at the higher concentrations and Na^+ influence on the inhibitory effect of K^+ (Koebmann *et al.*, 2002). Without going into details of the enzymatic activity mechanism which as is evident from the above description, is very complex, note that the method of generalized Hill functions allows for obtaining a very compact description of the overall processes considered. Taking into account effects of Mg^{2+} , Na^+ , and K^+ described above on the enzymatic activity, the rate of reaction, catalyzing by the ATPase enzyme may be represented in a generalized form by the following equation:

$$V = V_{\max} \cdot \frac{\text{ATP} / K_{m,\text{ATP}}}{1 + \frac{\text{ATP}}{K_{m,\text{ATP}}} + \frac{\text{ADP}}{K_{i,\text{ADP}}}} \cdot \frac{1 + \left(\frac{\text{ATP}}{k_{i,\text{ATP}}} \right)^{h_{i,\text{ATP}}}}{1 + \left(l_{i,\text{ATP}} \cdot \frac{\text{ATP}}{k_{i,\text{ATP}}} \right)^{h_{i,\text{ATP}}}} \cdot f_{\text{Mg}} \cdot f_{\text{Na}} \cdot f_{\text{K}}, \quad (1)$$

where $K_{m,\text{ATP}}$ – Michaelis constant for ATP, $K_{i,\text{ADP}}$ – inhibitor constant for ADP, $k_{i,\text{ATP}}$, $l_{i,\text{ATP}}$ – constants describing efficiency of ATP substrate inhibition, $h_{i,\text{ATP}}$ – constant of substrate inhibition nonlinearity. The subformulas f_{Mg} , f_{Na} , f_{K} , describing effects of magnesium, sodium and potassium respectively on the enzyme activity are the generalized Hill functions:

$$f_{\text{Mg}} = \frac{\frac{k_{0,\text{Mg}}}{1 + \left(\text{ATP} / k_{i,\text{Mg}0\text{ATP}} \right)} + \frac{l_{a,\text{Mg}}(\text{ATP}) \cdot \text{Mg}}{k_{a,\text{Mg}}(\text{ATP})}}{1 + \frac{\text{Mg}}{k_{a,\text{Mg}}(\text{ATP})} \cdot \left(1 + \left(\frac{\text{Mg}}{k_{i,\text{Mg}}(\text{ATP})} \right)^{n_{i,\text{Mg}}(\text{ATP})} \right)},$$

where $k_{0,\text{Mg}}$, $k_{i,\text{Mg}0\text{ATP}}$ – describing ATP effect at the low magnesium concentrations, the subformulas $k_{a,\text{Mg}}$, $k_{i,\text{Mg}}$, $l_{a,\text{Mg}}$ and $n_{i,\text{Mg}}$, describing the efficiency and nonlinearity of magnesium influence depending on ATP concentration are also generalized Hill functions (not presented).

$$f_{\text{Na}} = \frac{1 + \left(\frac{l_{a,\text{Na}} \cdot \text{Na}}{k_{a,\text{Na}}} \right)^{n_{a,\text{Na}}} \cdot \left(1 + \left(\frac{\text{Na}}{k_{i,\text{Na}}} \right)^{n_{i,\text{Na}}} \right)}{1 + \left(\frac{\text{Na}}{k_{a,\text{Na}}} \right)^{n_{a,\text{Na}}} \cdot \left(1 + \left(\frac{l_{i,\text{Na}} \cdot \text{Na}}{k_{i,\text{Na}}} \right)^{n_{i,\text{Na}}} \right)},$$

where k_{aNa} , l_{aNa} , n_{aNa} – constants of efficiency and nonlinearity of sodium activator effects at the low concentrations, k_{iNa} , l_{iNa} , n_{iNa} – constants of efficiency and nonlinearity of sodium inhibitor effects at the higher concentrations.

$$f_K = \frac{1 + \left(\frac{K}{k_{iK}(Na)} \right)^{n_{iK}}}{1 + \left(\frac{l_{iK}(Na) \cdot K}{k_{iK}(Na)} \right)^{n_{iK}}},$$

where subformulas k_{iK} and l_{iK} are the generalized Hill functions (not presented) and describe the efficiency of potassium influence depending on sodium concentration and $n_{i,Mg}$ – constant of nonlinearity of potassium concentration.

Note that characteristic of the mathematical model proposed, constructed using a generalized Hill function, is a considerable simplicity (compared with the molecular biological processes simulated); nonetheless, it provides a very good fit of the experimental data and numerical calculations.

Fig. 2 shows the results of calculations using model (1) and their comparison with the experimental data on the effects of various Mg^{2+} , Na^+ , and K^+ and ATP concentrations on the ATPase enzymatic activity. This comparison demonstrates a high adequacy of the model of genetic expression regulation of the gene considered.

The example given above demonstrates a strategy of metabolic reactions modeling, which is also used for modeling and description of complex molecular-genetic processes (Likhoshvai *et al.*, 2006, this issue). In general, the work shows the reconstruction of fragment being the component of complex gene network of respiration regulation in *E.coli* cell. It is the first and essential step in constructing of complete kinetic model of any molecular-genetic system. It is necessary to say that the proposed methodology of constructing the elementary kinetic models allows us to reduce the model dimension keeping the complex effects and description adequacy. The models of the all enzymatic reactions of gene network “Respiration”, including the model described above will be an inextricable part of the “*in silico* cell” computer resource.

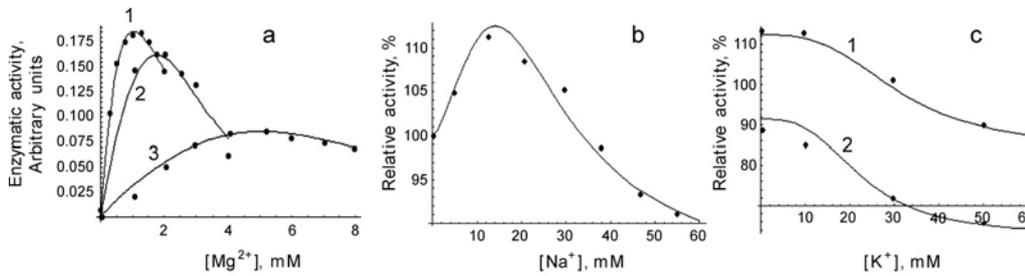


Figure 2. The effects of a – Mg^{2+} concentration at various ATP concentrations, b – Na^+ concentration, c – K^+ concentration at various Na^+ concentrations on the ATPase enzymatic activity. The enzyme activity was measured (a) with the ATP concentration of 2.5 mM, 5 mM and 10 mM (1,2,3 respectively); (b) ATP (2.5 mM), Mg^{2+} (1 mM); (c) ATP (2.5 mM), Mg^{2+} (1 mM), Na (1, 15 mM, 2, 55 mM). (a) Dots indicate experimental data from (Koebmann *et al.*, 2002), and curves are the results of simulation according to model (1) with the estimated parameters.

ACKNOWLEDGEMENTS

The authors are grateful to Vitaly Likhoshvai and Aleksandr Ratushny for valuable discussions, to Irina Lokhova for bibliographical support, and to Victor Gulevich for translating the manuscript from Russian into English. This work was supported in part by the Russian Government (Contract No. 02.467.11.1005), by Siberian Branch of the Russian Academy of Sciences (the project “Evolution of molecular-genetical systems: computational analysis and simulation” and integration projects Nos 24 and 115), and by the Federal Agency of Science and Innovation (innovation project No. IT-CP.5/001).

REFERENCES

- Ananko E.A. *et al.* (2005) GeneNet in 2005. *Nucl. Acids Res.*, **33**, D425–D427.
- Koebmann B.J. *et al.* (2002) The glycolytic flux in *Escherichia coli* is controlled by the demand for ATP. *J. Bacteriol.* **184**, 3909–3916.
- Likhoshvai V.A., Ratushny A.V. (2006) *In silico* cell I. Hierarchical approach and generalized hill functions in modeling enzymatic reactions and gene expression regulation. *This issue*.
- Weber J. *et al.* (1996) Specific tryptophan substitution in catalytic sites of *Escherichia coli* F1-ATPase allows differentiation between bound substrate ATP and product ADP in steady-state catalysis. *J. Biol. Chem.*, **271**, 18711–18718.

MATHEMATICAL MODELING OF REGULATION OF *ESCHERICHIA COLI* PURINE BIOSYNTHESIS PATHWAY ENZYMATIC REACTIONS

*Nedosekina E.A.**

Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia

* Corresponding author: e-mail: nzhenia@bionet.nsc.ru

Key words: mathematical modeling, gene network, purine biosynthesis, *E. coli*

SUMMARY

Motivation: Development of mathematical models that adequately describe molecular-genetic mechanisms is one of the main tasks in modern bioinformatics. Not only should a model be sensitive to the particulars of enzyme reactions underlying cell life activity, but also to regulation exerted on enzyme activity by metabolites and specialized cell proteins.

Results: The gene network for purine biosynthesis in *E. coli* has been reconstructed using the GeneNet system. The elementary mathematical models for enzyme reactions have been developed using Hill's generalized functions. We herein give a detailed description of the mathematical models (1) for the biosynthetic IMP⁴ dehydrogenase-catalyzed reaction and (2) for the reaction catalyzed by Adenylosuccinate synthetase (AdSS); the models include enzyme activity regulation.

Availability: The models are available on request; the gene network is available at <http://www.mgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/index.shtml>.

INTRODUCTION

Biosynthesis of the purines AMP and GMP in *E. coli* is a many-staged process supported by a complex network of enzymes (Table 1). Some of the genes that encode these enzymes, are arranged into operons (purF, purHD, purMN, purEK, guaBA, purB), others are located in single cistrons (purT, purI, purC, purA, guaC). Many of the genes in

⁴ The abbreviations used are: ADP, Adenosine diphosphate; AdSS, Adenylosuccinate synthetase; AICAR, 5-Phosphate-ribose-5-amino-4-imidazole carboxamide; AIR, Aminoimidazole ribotide; AMP, Adenosine monophosphate; ASP, Aspartate; ASUC, Adenylosuccinate; ATP, Adenosine triphosphate; CAIR, 5-Phosphoribosyl-5-aminoimidazole-4-carboxylate; CMP, Cytidine monophosphate; CO₂, Carbon dioxide; dAMP, DeoxyAMP; dGMP, DeoxyGMP; GAR, Glycinamide ribonucleotide; GDP, Guanosine diphosphate; GLN, glutamine; GMP, Guanosine monophosphate; GTP, Guanosine triphosphate; FGAM, 5-Phosphoribosyl-n-formylglycineamide; FGAR, N-Formyl-GAR; FOR, Formate; FTHF, 10-Formyltetrahydrofolate; FUM, Fumarate; GLN, Glutamine; GLU, Glutamate; GLY, Glycine; IMP, Inosine monophosphate; NAD, Nicotinamide adenine dinucleotide; NADH, NAD reduced; NADP, NAD phosphate; NADPH, NAD phosphate reduced; NCAIR, N⁵-carboxyaminoimidazole ribonucleotide; NH₃, Ammonia; PI, Phosphate; PPI, Pyrophosphate; PRAM, Phosphate-ribose amine; PRFICA, Phosphoribosyl-formamido-carboxamide; PRPP, Phosphoribosylpyrophosphate; SAICAR, 5-Phosphoribosyl-4-(N-succinocarboxamide)-5-amino-imidazole; SUCC, Succinate; THF, Tetrahydrofolate; UMP, uridine monophosphate; XMP, Xantosine monophosphate.

the network for purine biosynthesis are controlled by the protein PurR (which is encoded by the gene *purR*) and its co-repressors hypoxanthine and guanine. Besides, the operon *guaBA* is repressed by GMP and the protein DnaA, and is activated by CRP (the cAMP receptor protein) and AMP. AMP and GMP have opposing effects on the operon *guaC*. The operon *guaC*, too, is repressed by GLN. Some other low-molecular-weight items participate in the regulation of gene expression and enzyme activity in this gene network.

Table 1. Enzymes, genes and reactions involved in the *de novo* purine biosynthesis

№	Enzyme names	Genes	Reaction
1	Amidophosphoribosyl transferase	<i>purF</i>	PRPP + GLN → PPI + GLU + PRAM
2	Phosphoribosylamine-glycine ligase	<i>purD</i>	PRAM + ATP + GLY ↔ ADP + PI + GAR
3	Phosphoribosylglycinamide formyltransferase	<i>purN</i>	GAR + FTHF → THF + FGAR
4	GAR transformylase T	<i>purT</i>	GAR + FOR + ATP → ADP + PI + FGAR
5	Phosphoribosylformylglycinamide synthetase	<i>purL</i>	FGAR + ATP + GLN → GLU + ADP + PI + FGAM
6	Phosphoribosylformylglycinamide cyclo-ligase	<i>purM</i>	FGAM + ATP → ADP + PI + AIR
7	Phosphoribosylaminoimidazole carboxylase 1	<i>purK</i>	AIR + CO ₂ + ATP ↔ NCAIR + ADP + PI
8	Phosphoribosylaminoimidazole carboxylase 2	<i>purE</i>	NCAIR ↔ CAIR
9	Phosphoribosylaminoimidazole-succinocarboxamide synthetase	<i>purC</i>	CAIR + ATP + ASP ↔ ADP + PI + SAICAR
10	5'-Phosphoribosyl-4-(N-succinocarboxamide)-5-aminoimidazole lyase	<i>purB</i>	SAICAR ↔ FUM + AICAR
11	AICAR transformylase	<i>purH</i>	AICAR + FTHF ↔ THF + PRFICA
12	IMP cyclohydrolase	<i>purH</i>	PRFICA ↔ IMP
13	Adenylosuccinate synthetase	<i>purA</i>	IMP + GTP + ASP → GDP + PI + ASUC
14	Adenylosuccinate lyase	<i>purB</i>	ASUC ↔ FUM + AMP
15	IMPD; IMP dehydrogenase	<i>guaB</i>	IMP + NAD → NADH + XMP
16	GMP synthase	<i>guaA</i>	XMP + ATP + GLN → GLU + AMP + PPI + GMP
17	GMP reductase	<i>guaC</i>	GMP + NADPH → NADP + IMP + NH ₃

IMP dehydrogenase of *E. coli* (IMPD; EC 1.2.1.14) catalyzes the NAD⁺-dependent conversion of IMP into XMP. IMPD attracts scientists as a potential target in developing antimicrob, antitumor and immune suppressing drugs. This is a tetramer, which consists of identical subunits and is competitively inhibited by GMP. The enzyme activity of IMPD is enhanced by K⁺: the mechanism of this enhancement is not yet absolutely clear, however, it is proposed that K⁺ facilitates the binding of the enzyme to NAD (Gilbert *et al.*, 1979; Kerr *et al.*, 2000).

The enzyme adenylosuccinate synthetase (AdSS; GDP-forming IMP: L-aspartate ligase, EC 6.3.4.4) catalyzes the reaction of conversion of IMP to ASUC in the presence of Mg²⁺. There are many nucleotides that inhibit AdSS: AMP is an inhibitor competitive with respect to IMP; ASUC - to IMP; dGMP - to IMP; GMP - to GTP; GDP - with respect to GTP, which in part explains a gradual decrease in the rates of formation ASUC in the solutions in which GTP is not reduced. Weak inhibitory effects are also produced by dAMP, CMP and UMP (Wyngaarden *et al.*, 1963). Mathematical models for the reaction catalyzed by AdSS were proposed in a variety of works: in 1969, Rudolph proposed an equation which included the effect of one inhibitor – that equation contained 11 parameters; in 1979, Stayton proposed a slightly different equation for one inhibitor – that equation contained 12 parameters (Stayton, Fromm, 1979); in 1995, Kang proposed an equation which related the reaction rate and ASP, which is one of the substrates, and Mg²⁺ ions – that equation contained 4 parameters.

We have reconstructed the gene network for purine biosynthesis and developed elementary mathematical models for enzyme reactions using generalized Hill functions. Herein we give a detailed description of the models (1) for the biochemical reaction catalyzed by IMPD and regulated by GMP and K^+ ions, and (2) propose rather a simple model for the reaction catalyzed by AdSS, considering the effects of five inhibitors and Mg^{2+} ions and containing 10 parameters.

METHODS AND ALGORITHMS

The reconstruction of the gene network for purine biosynthesis was performed using the GeneNet system (Ananko *et al.*, 2005). The biochemical reactions were modeled using Hill's generalized functions (Likhoshvai *et al.*, 2006).

RESULTS

The gene network for purine biosynthesis has been reconstructed using the GeneNet system (Fig. 1) and is accessible via the Internet. The number of the reconstructed gene network components, of which GeneNet is aware, is presented in Table 2.

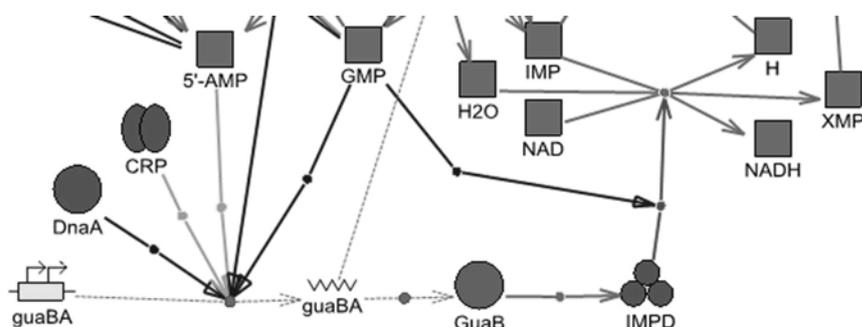


Figure 1. A fragment of the gene network for purine biosynthesis reconstructed using the GeneNet system: regulation of the expression of the *guaB* operon and the enzyme activity of IMPD. Squares denote low-molecular-weight compounds; circles, monomeric proteins; double oval, the dimeric protein; triple oval, the multimer; broken line, RNA; arrowed rectangle, the operon; arrowed dash lines, transcription and translation; other arrows, reactions and regulatory actions.

Table 2. The number of items collected as data on the gene network for purine biosynthesis

Item	Protein	RNA	Gene	Operon	Reaction, regulation	Inorganic substance	Protein-repressor	Literature source
Quantity	33	12	15	11	126	42	3	140

Using generalized Hill functions, mathematical models have been developed for the enzyme reactions listed in Table 1. Verification of the model's parameters was performed using data in the Kinet database containing constants and dynamic data from published experimental works (Khlebodarova *et al.*, this issue). The steady-state and dynamic characteristics of the biochemical reactions studied using the model were compared with the experimental data and a good agreement was obtained. For example, knowing how GMP and K^+ ions affect the IMPD-catalyzed reaction (K^+ ions enhance enzyme activity, GMP inhibits it competitively with respect to IMP), the reaction rate can be written in a generalized form as follows:

$$V = E_0 \cdot k_0 \cdot \frac{(IMP/K_{mIMP}) \cdot (NAD/K_{mNAD})}{(1 + IMP/K_{mIMP} + GMP/K_{GMP}) \cdot (1 + NAD/K_{mNAD})} \cdot \frac{1 + K^+/k_1}{1 + K^+/k_2}, \quad (1)$$

where V is the XMP synthesis rate; E_0 is the IMP concentration; k_0 is the rate constant of enzyme; IMP , NAD , GMP and K^+ are IMP , NAD , GMP and K^+ concentrations; K_{mIMP} , K_{mNAD} are the Michaelis-Menten constants for the respective substrates; K_{GMP} is the constant of GMP inhibiting the reaction; k_1 , k_2 are the constants of K^+ ions affecting enzyme activity.

How the model (1) was fit to experimental data on the effect of the IMP , NAD , GMP and K^+ concentrations on the XMP biosynthesis rates is presented in Fig. 2. Fitting resulted in the following values for the parameters: $k_0 = 0.18$ 1/sec, $K_{mIMP} = 0.0115$ mM, $K_{mNAD} = 0.334$ mM, $K_{GMP} = 0.08$ mM, $k_1 = 0.18$ mM, $k_2 = 5$ mM.

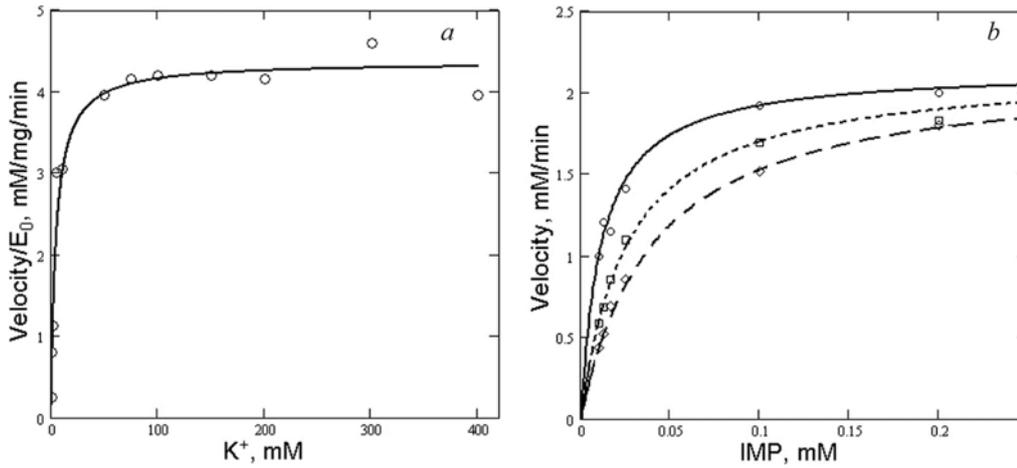


Figure 2. XMP synthesis rates: a – depending on IMP concentrations with 10mM NAD and without K^+ (solid line and \circ : 0 GMP ; dotted line and \square : 0.1mM GMP ; broken lines and \diamond : 0.2mM GMP); b – depending on the concentration of K^+ ions (1mM IMP , 2.5 mM NAD , 0 GMP). \circ , \square , \diamond indicate experimental data from Gilbert *et al.*, 1979; B, Kerr *et al.*, 2000); curves indicate results obtained from the model (1).

Second example – is the model of reaction, catalyzed by AdSS. This enzyme is inhibited by GMP , GDP , AMP , $ASUC$ and $SUCC$, and requires the presence of Mg^{2+} ions. Knowing how these effectors work, the reaction rate can be written in a generalized form as follows:

$$V = V_{max} \cdot \frac{\frac{GTP}{K_{mGTP}} \cdot \frac{IMP}{K_{mIMP}} \cdot \frac{ASP}{K_{mASP}} \cdot \frac{Mg^{2+}}{K_{mMg}}}{\left(1 + \frac{GTP}{K_{mGTP}} + \frac{GMP}{K_{GMP}} + \frac{GDP}{K_{GDP}}\right) \cdot \left(1 + \frac{IMP}{K_{mIMP}} + \frac{AMP}{K_{AMP}} + \frac{ASUC}{K_{ASUC}}\right) \cdot \left(1 + \frac{ASP}{K_{mASP}} + \frac{SUCC}{K_{SUCC}}\right) + \frac{Mg^{2+}}{K_{mMg}}}, \quad (2)$$

where V_{max} is the maximum reaction rate; GTP , IMP , ASP are the concentrations of the corresponding substrates; GMP , GDP , AMP , $ASUC$, $SUCC$ are the concentrations of the corresponding inhibitors; Mg^{2+} is the concentration of Mg^{2+} ions; K_{mGTP} , K_{mIMP} , K_{mASP} are the Michaelis-Menten constants for the respective substrates; K_{mMg} is the Michaelis-Menten constant for Mg^{2+} ions; K_{GMP} , K_{GDP} , K_{AMP} , K_{ASUC} , K_{SUCC} are the constants of the efficiency of reaction inhibition by corresponding substances.

The model's parameters were verified against 61 curves from literature (Wyngaarden *et al.*, 1963; Rudolph, Fromm, 1969; Kang, Fromm, 1995). Rate constant of the enzyme

AdSS is 15600 1/s (Dong *et al.*, 1991); however, since no enzyme concentrations were indicated, we calculated the value for V_{max} using our model.

The results of calculations using our model and the experimental data (Rudolph, Fromm, 1969), to which the fitting of the model was performed, are presented in Fig. 3. The parameter values inferred from curves were as follows (Fig. 3a, b, c): $V_{max} = 1.35 \cdot 10^{-3}$ mM/min, $K_{mGTP} = 0.023$ mM, $K_{mIMP} = 0.02$ mM, $K_{mASP} = 0.3$ mM.

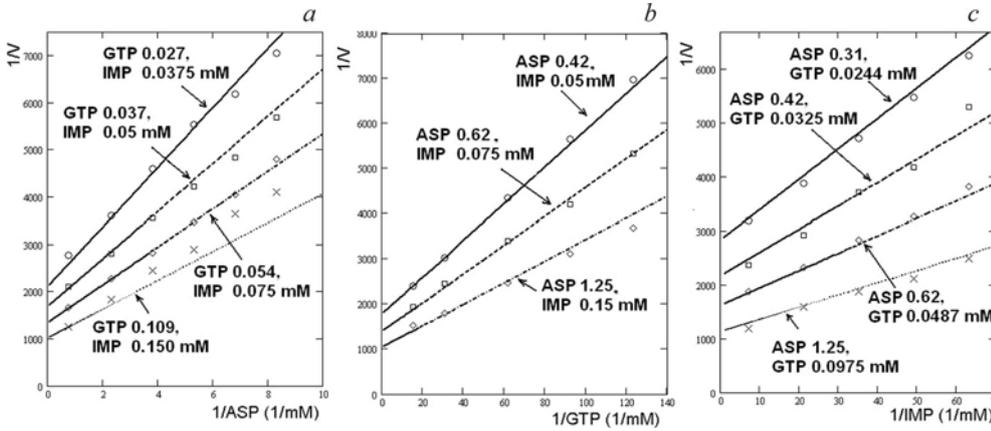


Figure 3 Relationships between the reaction rate (V) and the concentrations of GTP (a), IMP (b) and ASP (c), respectively. O, □, ◇ and × represent experimental data from Rudolph *et al.* (1969); lines are the results of calculations using our mathematical model (2).

The effect of SUCC has been studied in detail (Rudolph, Fromm, 1969); we examined its effect on all the three substrates. The modeling results and experimental data are presented in Fig. 4. Using the model, the value of the constant K_{SUCC} , is 8 mM. As can be seen from Fig. 4a, there is an inconsistency between modeling results and experimental evidence. To do away with this inconsistency, however, is easy, by introducing to the model a correction factor (a multiplier equal to 0.74), which in fact implies a reduced concentration of the enzyme compared to the experimental data shown in the other curves (calculations not shown).

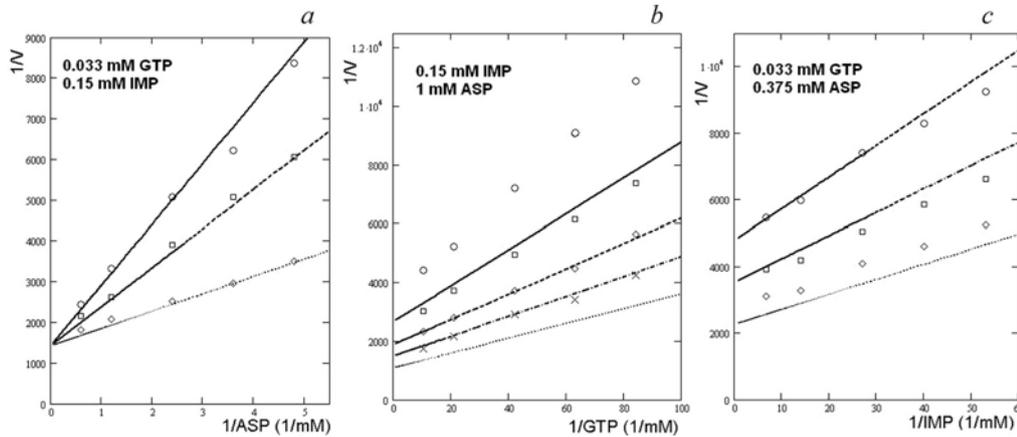


Figure 4. Relationships between the reaction rate (V) and the concentrations of GTP (a), IMP (b) and ASP (c), respectively. a – SUCC concentrations were (solid line and O) 50 mM; (dashed line and □) 25 mM; (dotted line and ◇) 12.5 mM; (lowest line and ×) 0. b – SUCC concentrations were (solid line and O) 20 mM; (dashed line and □) 10 mM; (dotted line and ◇) 0. c – SUCC concentrations were (solid line and O) 20 mM; (dashed line and □) 10 mM; (dotted line and ◇) 0. O, □, ◇, × represent experimental data from Rudolph, Fromm (1969); lines are the results of calculations using our mathematical model (2).

The modeling results and experimental data on GDP effects are presented in Fig. 5a. Using the model, the value of the constant K_{GDP} is $8 \cdot 10^{-3}$ mM. The effects of ASUC on the reaction substrates IMP and ASP have been examined; the results, both experimental and modeling, are presented in Fig. 5b, c. Using the model, the value of the constant K_{ASUC} is $7.5 \cdot 10^{-3}$ mM.

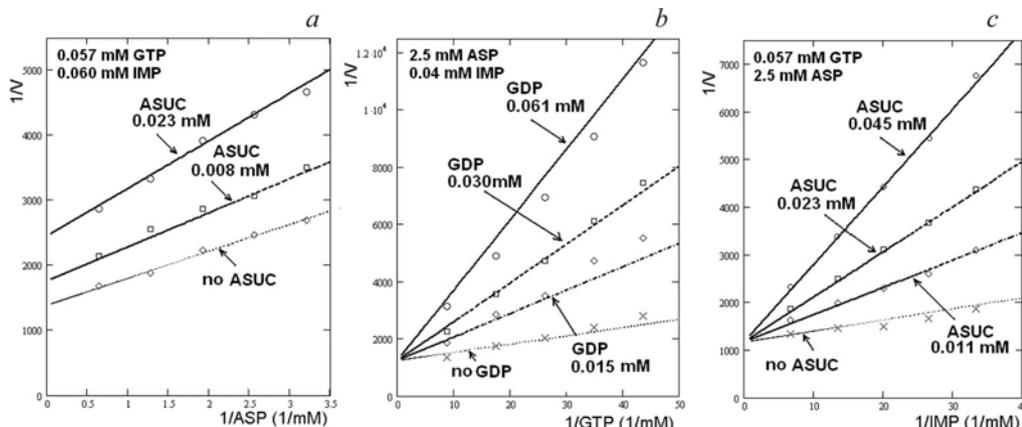


Figure 5. Relationships between the reaction rate (V) and the concentrations of GTP (a), IMP (b) and ASP (c), respectively. O, □, ◇, × represent experimental data from Rudolph, Fromm (1969); lines are the results of calculations using our mathematical model (2).

The constant K_{mMg} was verified against data published by Kang, Fromm (1995) and is 0.08 mM (Fig. 6a). However, in so doing, we had to reduce K_{mASP} two-fold, to 0.17 mM and introduced to the model a correction factor (a multiplier equal to 5000), which increases the amount of the enzyme. The other parameters were absolutely consistent with those experimental data.

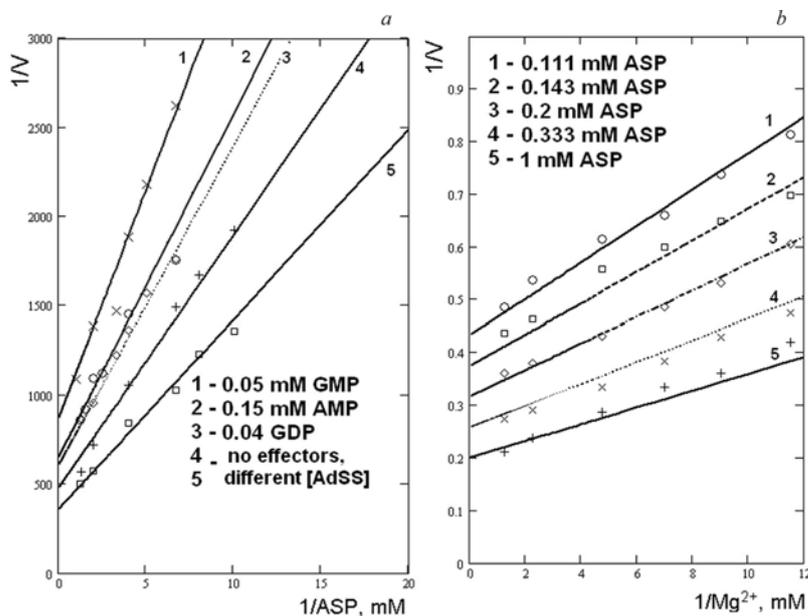


Figure 6. Relationships between the reaction rate (V) and: concentrations of Mg^{2+} ions at varying concentrations of ASP (a) or concentrations of ASP under different conditions (b). O, □, ◇, × represent experimental data (a – Kang, Fromm, 1995; b – Wyngaarden *et al.*, 1963); lines are the results of calculations using our mathematical model.

Data from the work by Wyngaarden *et al.* (1963) were used as the control experiments. Analyzing the curves in this paper, an apparent inconsistency revealed. It is possible that the authors were dealing with different amounts of the enzyme in different experiments. This assumption was supported by introduction of a coefficient like the ones for the curves in Fig. 4a. A comparison of modeling results and experimental data (Wyngaarden *et al.*, 1963) is presented in Fig. 6b, lines 4 and 5. Based on the work by Wyngaarden *et al.* (1963), we calculated the constants of GMP and AMP effects (Fig. 6b, lines 1 and 2): $K_{GMP} = 0.024$ mM, $K_{AMP} = 0.01$ mM. However, these estimates may suffer due to lack of concordance between the experimental data.

Note that the discrepancies revealed at fitting the parameters could in part be explained by different temperatures at which the different authors conducted their experiments: Rudolph, Fromm (1969), at 28 °C, Wyngaarden *et al.* (1963), at 25 °C, and Kang, Fromm (1995), at 22 °C. However, in the present work we did not look at temperature as a factor.

DISCUSSION

Using Hill's generalized functions, mathematical models have been developed for the reactions involved in purine biosynthesis (listed in Table 1). The models capture the effects of all the regulatory proteins and low-molecular-weight compounds. The parameters were introduced to the models using published experimental data.

We herein present the models (1) for the dependence of XMP biosynthesis rates on the concentration of the reaction substrates IMP, NAD, inhibitor GMP and the activator K^+ ; (2) for the reaction catalyzed by the AdSS, which includes relationships between the reaction rate, the concentrations of three substrates (GTP, IMP and ASP), the effects of five inhibitors (GMP, GDP, AMP, ASUC and SUCC) and Mg^{2+} . Fitting of the model's parameters was performed on the basis of experimental data in published literature. A methodical nuisance that came from the lack of concordance between data in different publications published was dealt with by introduction of correction coefficients. The adequacy of the model was ensured by comparing the theoretical calculations and the experimental data from the literature sources that were not being used while the fitting procedure was under way.

ACKNOWLEDGEMENTS

The authors are grateful to Vitaly Likhoshvai for valuable discussions, to Irina Lokhova for bibliographical support, and to Vladimir Filonenko for translating the manuscript from Russian into English. This work was supported in part by the Russian Government (Contract No. 02.467.11.1005), by Siberian Branch of the Russian Academy of Sciences (the project "Evolution of molecular-genetical systems: computational analysis and simulation" and integration projects Nos. 24 and 115), and by the Federal Agency of Science and Innovation (innovation project No. IT-CP.5/001).

REFERENCES

- Ananko E.A. *et al.* (2005) GeneNet in 2005. *Nucl. Acids Res.*, **33**, D425–D427.
Dong Q., Liu F., Myers A.M., Fromm H.J. (1991) Evidence for an arginine residue at the substrate binding site of *Escherichia coli* adenylosuccinate synthetase as studied by chemical modification and site-directed mutagenesis. *J. Biol. Chem.*, **266**, 12228–12233.

- Gilbert H.J., Lowe C.R., Drabble W.T. (1979) Inosine 5'-monophosphate dehydrogenase of *Escherichia coli*. Purification by affinity chromatography, subunit structure and inhibition by guanosine 5'-monophosphate. *Biochem J.*, **183**, 481–494.
- Kang C., Fromm H.J. (1995) Identification of an essential second metal ion in the reaction mechanism of *Escherichia coli* adenylosuccinate synthetase. *J. Biol. Chem.*, **270**, 15539–15544.
- Kerr K.M. *et al.* (2000) Monovalent cation activation in *Escherichia coli* inosine 5'-monophosphate dehydrogenase. *Arch. Biochem. Biophys.*, **375**, 131–137.
- Khlebodarova T.M. *et al.* (2006) *In silico* cell II. Information sources for modeling *Escherichia coli* metabolic pathways. *This issue*.
- Likhoshvai V.A., Ratushny A.V. (2006) *In silico* cell I. Hierarchical approach and generalized hill functions in modeling enzymatic reactions and gene expression regulation. *This issue*.
- Rudolph F.B., Fromm H.J. (1969) Initial rate studies of adenylosuccinate synthetase with product and competitive inhibitors. *J. Biol. Chem.*, **244**, 3832–3839.
- Stayton M.M., Fromm H.J. (1979) Guanosine 5'-diphosphate-3'-diphosphate inhibition of adenylosuccinate synthetase. *J. Biol. Chem.*, **254**, 2579–2581.
- Wynngaarden J.B., Greenland R.A. (1963) The inhibition of succinoadenylate kinosynthetase of *Escherichia coli* by adenosine and guanosine 5'-monophosphates. *J. Biol. Chem.*, **238**, 1054–1057.

RECONSTRUCTION AND MATHEMATICAL MODELING OF THE GENE NETWORK CONTROLLING CYSTEINE BIOSYNTHESIS IN *ESCHERICHIA COLI*: REGULATION OF SERINE ACETYLTRANSFERASE ACTIVITY

Mishchenko E.L.^{*1}, Lashin S.A.¹, Usuda Y.², Matsui K.²

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ² Institute of Life Sciences, Ajinomoto, Co., Inc., Kawasaki, Japan

* Corresponding author: e-mail: elmish@bionet.nsc.ru

Key words: mathematical modeling, bioinformatics, cysteine biosynthesis, regulation, *Escherichia coli*

SUMMARY

Motivation: Development of an *in silico* cell as a computer resource for the modeling and analysis of intracellular processes is a topical problem of the systems biology.

Results: This work reconstructed the gene network of cysteine biosynthesis and degradation in *E. coli* cell. In the context of generalized Hill functions, the mathematical models describing the functioning efficiencies of enzymatic systems and expression regulation of the genes coding for enzymes and their subunits are constructed.

Availability: The model is available on request; the gene network is available at <http://www.mgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/index.shtml>

INTRODUCTION

The synthesis of L-cysteine from inorganic sulfur is the predominant mechanism by which reduced sulfur is incorporated into organic compounds; it produces significant quantities of L-cysteine only in plants and microorganisms, including *E. coli*. In this process, inorganic sulfate⁵ is taken up and reduced to sulfide which is then incorporated into L-cysteine by a relatively simple two-step process requiring conversion of L-serine to O-acetyl-L-serine, which then reacts with sulfide. The reduction of sulfate requires its prior activation to a phosphosulfate mixed anhydride. This activation is achieved by the ATP sulfurylase-catalyzed reaction of sulfate with ATP to give adenosine 5'-phosphosulfate (APS) and PPi. Then APS kinase phosphorylates APS with another ATP to give PAPS which is then reduced by PAPS reductase to sulfite. The reduction of sulfite to sulfide is catalyzed by NADPH-sulfite reductase. Serine transacetylase catalyzes the acetylation of L-serine by acetyl-CoA to give O-acetyl-L-serine, the direct precursor of L-cysteine. The enzyme is feedback inhibited by L-cysteine and L-serine, thus providing kinetic regulation of this short branch of the pathway. Synthesis of L-cysteine from O-

⁵ The abbreviations used are: SLF, Sulfate; APS, adenosine 5'-phosphosulfate; PPi, pyrophosphate; Pi phosphate; PAPS, 3'-phosphoadenosine 5'-phosphosulfate; PAP, 3'-phosphoadenosine 5'-phosphate; RTHIO, reduced thioredoxin; OTHIO, oxidated thioredoxin; ACSER, O-acetyl-L-serine; AC, acetic acid; ACCOA, acetyl-CoA; COA, CoA; PYR, pyruvate; ACC, acetylcysteine; CSA, L-cysteic acid; SAT, serine acetyltransferase.

acetyl-L-serine and sulfide is catalyzed by two distinct O-acetylserine(thiol)-lyase isozymes (Kredich, 1996).

METHODS AND ALGORITHMS

The gene network of regulation respiration was reconstructed with the GeneNet technology (Ananko *et al.*, 2005). Mathematical models were constructed by the method of generalized Hill functions (Likhoshvai, Ratushny, 2006).

RESULTS AND DISCUSSION

The reconstructed gene network of Cysteine biosynthesis and degradation in *E. coli* cell is shown on the Fig. 1. Table 1 shows the number of components of the gene network of Cysteine biosynthesis and degradation.

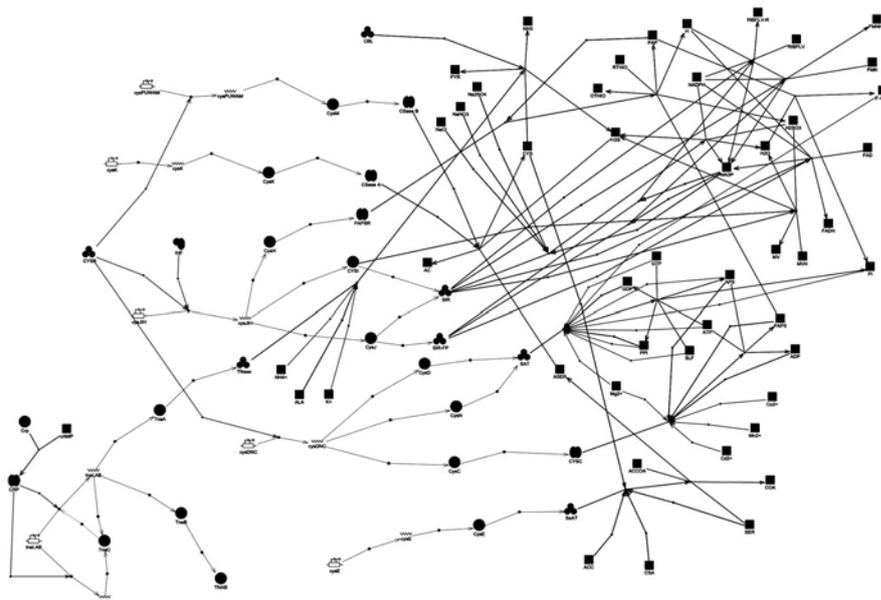


Figure 1. Cysteine biosynthesis and degradation gene network reconstruction in the GeneNet.

Table 1. List of the components of gene network of Cysteine biosynthesis and degradation

Item	Protein	RNA	Operon	Gene	Reaction	Substance	Literature source
Amount	27	7	6	15	100	48	112

Table 2 lists the enzymatic reactions involved in the gene network reconstruction, names of the enzymes catalyzing the corresponding reactions, and names of the genes encoding the corresponding enzymes. The mathematical models of enzymatic reactions (Table 2, column Reaction) and the models describing expression regulation of the genes encoding the corresponding enzymes and their subunits (Table 2, column Gene) were constructed using the method of generalized Hill functions. Table 3 shows the number of kinetic parameters for the enzymatic reactions and the regulatory processes (Michaelis constants, K_m ; constants of catalytic activity, k_{cat} ; constants of inhibition, K_i ; constants of activation, K_a ; the equilibrium constants of the dissociation, K_d ; Hill coefficients, n^H) used

upon modeling of functioning efficiencies of enzymes. The set of other parameters of these models were determined by numerical experiments. It was demonstrated that the equilibrium and dynamic characteristics of the gene network in question calculated using the models developed fit the experimental data.

Table 2. Genes, enzymes and reactions of the gene network of Cysteine biosynthesis and degradation

Enzyme	Gene	Reaction
Sulfate adenylyltransferase	cysDN	SLF+ATP+GTP →
PPi+APS+GDP+Pi		
Adenylylsulfate kinase	cysC	APS +ATP →ADP +PAPS
3'-Phospho-adenylylsulfate reductase	cysH	PAPS +RTHIO
→OTHIO+H2SO3+PAP		
Sulfite reductase	cys IJ	H2SO3 +3NADPH↔H2S+3NADP
Serine acetyltransferase	cysE	SER+ACCOA↔COA+ACSER
O-Acetylserine (thiol)-lyase A	cysK	ACSER +H2S→AC+CYS
O-Acetylserine (thiol)-lyase B	cysM	ACSER +H2S→AC+CYS
Tryptophanase	tnaA	CYS →PYR+NH3+H2S
L-Cysteine desulphydrase		CYS →PYR+NH3+H2S

Table 3. Kinetic parameters for the enzymatic reactions and the regulatory processes involved in the gene network reconstruction

	Kinetic parameters						
	Michaelis-Menten constant K_m	Constant of catalytic activity k_{cat}	Maximal velocity V_{max}	Dissociation constant K_d	Hill coefficient h	Constant of activation K_{act}	Constant of inhibition K_i
Amount	49	12	1	25	6	4	16

Let us consider the model of enzyme serine acetyltransferase (Kredich, Tomkins, 1966) activity regulation as an example of modeling method. In this reaction, L-cysteine is competitive inhibitor, L-serine, L-cysteic acid and acetylcysteine are also inhibit the enzymatic activity at the concentrations of 1–4 mM.

A model for the steady-state rate of the reaction is proposed (where e_0 is the concentration of the enzyme serine acetyltransferase; SER, ACCOA, ACSER, COA, ACC, CYS, CSA are concentrations of the corresponding low-molecular-weight substances; k_f , the catalytic constant of direct reaction, k_r , the catalytic constant reverse reaction; $K_{m,SER}$, $K_{m,ACCOA}$, $K_{m,ACSER}$, $K_{m,COA}$, Michaelis constants for corresponding substrates and products; $K_{i,ACC}$, $K_{i,CYS}$, corresponding inhibition constants; k_{SER} , $l_{i,SER}$, constants of substrate inhibition efficiency (SER), h_{SER} constant determining the nonlinearity of the effect of SER on the reaction rate; k_{CSA} , constant of the efficiency of the effect of CSA on the reaction rate).

$$V = \frac{e_0 \cdot \left(k_f \cdot \frac{SER}{K_{m,SER}} \cdot \frac{ACCOA}{K_{m,ACCOA}} - k_r \cdot \frac{ACSER}{K_{m,ACSER}} \cdot \frac{COA}{K_{m,COA}} \right) \cdot \left(1 + \left(\frac{SER}{k_{SER}} \right)^{h_{SER}} \right) \cdot \frac{1}{1 + \left(\frac{CSA}{k_{CSA}} \right)}}{\left(1 + \frac{SER}{K_{m,SER}} + \frac{ACSER}{K_{m,ACSER}} + \frac{ACC}{K_{i,ACC}} \right) \left(1 + \frac{ACCOA}{K_{m,ACCOA}} + \frac{COA}{K_{m,COA}} + \frac{CYS}{K_{i,CYS}} \right) \cdot \left(1 + \left(l_{i,SER} \cdot \frac{SER}{k_{SER}} \right)^{h_{SER}} \right) \cdot \left(1 + \left(\frac{CSA}{k_{CSA}} \right) \right)}$$

The experimental data (Kredich *et al.*, 1966) were used for model testing. These data illustrate the effects of the substrates SER and ACCOA on SAT activity with various concentrations of CYS and COA (Fig. 2).

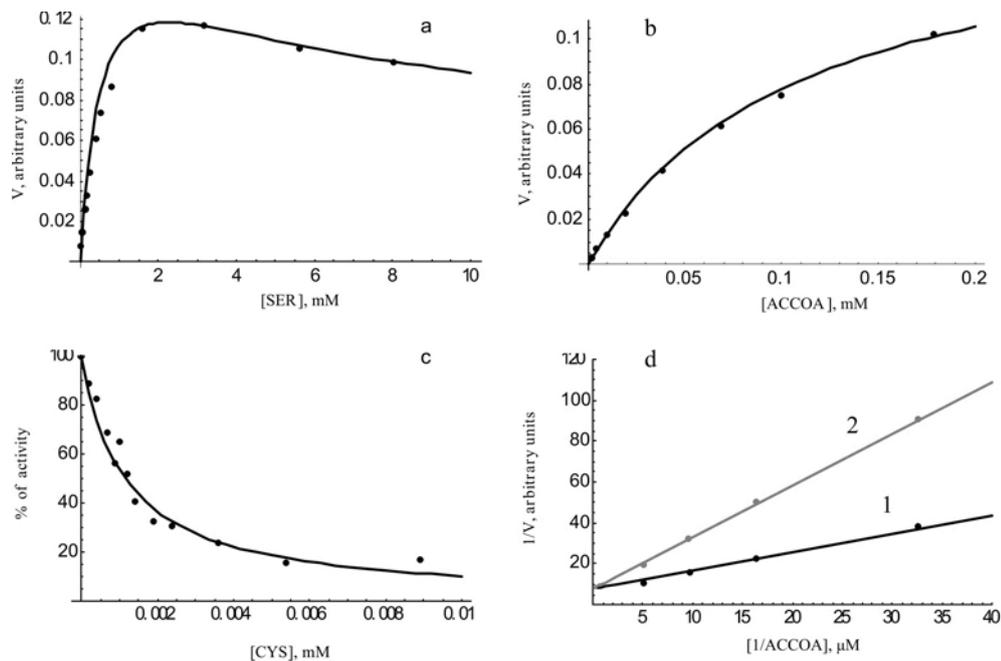


Figure 2. Effect of SER on the rate of the reaction catalyzed by SAT (a); effect of ACCOA on the rate of the reaction catalyzed by SAT (b); effect of CYS on the rate of the reaction catalyzed by SAT (c, d). The enzyme activity was measured (a) with 0.1 mM ACCOA, SAT concentration equaling 0.4 mM; (b) with 1 mM SER, SAT concentration equaling 0.29 mM; (c) with 1 mM SER, 0.1 mM ACCOA, SAT concentration equaling 0.4 mM; (d) 1 mM SER, at CYS concentrations: 1, null; 2, 1.1 μ M. Dots indicate experimental data (Kredich *et al.*, 1966), and curves are the results of simulation according to model (1) with the following parameters: $K_{m,SER} = 0.56$ mM, $K_{m,ACCOA} = 0.11$ mM, $K_{i,ACC} = 1$ mM, $K_{i,CYS} = 0.6$ μ M, $k_{SER} = 18$ mM, $h_{SER} = 1.1$, $i_{i,SER} = 3.5$, $k_f = 10$ u/s $k_r = 1$ u/s.

The model parameters are estimated according to the correspondence of experimental data. The proposed mathematical model and the other eight models of enzymatic reactions developed in the framework of the study are the components of more generalized model of the cysteine biosynthesis in the *E. coli* cell. On the next stage we plan to describe genetic regulation of expression of enzymes and their subunits. Then we will be able to analyze functioning of the cysteine biosynthesis pathway taking into account genetic processes. Furthermore, the developed models will be the significant component of the computer resource – an *in silico* cell.

ACKNOWLEDGEMENTS

The authors are grateful to Vitaly Likhoshvai for valuable discussions and to Irina Lokhova for bibliographical support. This work was supported in part by the Russian Government (Contract No. 02.467.11.1005), by Siberian Branch of the Russian Academy of Sciences (the project “Evolution of molecular-genetical systems: computational analysis and simulation” and integration projects Nos. 24 and 115), and by the Federal Agency of Science and Innovation (innovation project No. IT-CP.5/001).

REFERENCES

- Ananko E.A., Podkolodny N.L., Stepanenko I.L., Podkolodnaya O.A., Rasskazov D.A., Miginsky D.S., Likhoshvai V.A., Ratushny A.V., Podkolodnaya N.N., Kolchanov N.A. (2005) GeneNet in 2005. *Nucl. Acids Res.*, **33**, D425–D427.
- Kredich N.M. (1996) Biosynthesis of Cysteine. In Neidhardt F.C. (ed.), *Escherichia coli and Salmonella typhimurium*. Washington, 514–527.
- Kredich N.M., Tomkins G.M. (1966) The enzymic synthesis of L-cysteine in *Escherichia coli* and *Salmonella typhimurium*. *J. Biol. Chem.*, **241**, 4955–4965.
- Likhoshvai V.A., Ratushny A.V. (2006) *In silico* cell I. Hierarchical approach and generalized hill functions in modeling enzymatic reactions and gene expression regulation. *This issue*.

GENE NETWORK RECONSTRUCTION AND MATHEMATICAL MODELING OF SALVAGE PATHWAYS: REGULATION OF ADENINE PHOSPHORIBOSYLTRANSFERASE ACTIVITY BY STRUCTURALLY SIMILAR SUBSTRATES

*Smirnova O.G.**, *Ratushny A.V.*

Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia

* Corresponding author: e-mail: planta@bionet.nsc.ru

Key words: mathematical modeling, gene network, regulation, salvage pathways, adenine phosphoribosyltransferase, *Escherichia coli*

SUMMARY

Motivation: Development of an *in silico* cell, a computer resource for modeling and analysis of physiological processes is an urgent task of systems biology and computational biology. Mathematical modeling of the genetic regulation of cell metabolism pathways, in particular, salvage pathways, is an important problem to be solved as part of this line of work.

Results: By using the GeneNet technology, we reproduced the gene network of the regulation of salvage pathways in the *E. coli* cell. Mathematical models were constructed by the method of generalized Hill functions to describe the efficiency of enzyme systems and regulation of expression of genes coding for these enzymes.

Availability: The diagram of the gene network is available through the GeneNet viewer at <http://www.mgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/index.shtml>. Models are available on request.

INTRODUCTION

Salvage pathways are the metabolic pathways used by *Escherichia coli* for synthesis and conversion of adenine, hypoxanthine, guanine, xanthine, and their nucleosides and pyrimidine ribo- and deoxyribonucleotides.

The gene network of regulation of salvage pathways in the *E. coli* cell was reconstructed. Mathematical models of enzymatic reactions were constructed. A database storing experimental data on the behavior of components of this gene network was developed (Khlebodarova *et al.*, 2006). Parameters of the models were determined by numerical simulation. The results of calculation of steady-state properties and behavior of the components of the molecular system derived from the models are in agreement with experimental evidence.

METHODS AND ALGORITHMS

The gene network of salvage pathways was reconstructed with the use of the GeneNet technology (Ananko *et al.*, 2005), allowing accumulation and presentation of data on the

structure and function of molecular systems. Mathematical models of the regulation of gene expression and efficiency of enzyme systems were constructed by the method of generalized Hill functions (Likhoshvai, Ratushny, 2006).

RESULTS

The gene network of regulation of salvage pathways was reconstructed with the use of the GeneNet technology (Ananko *et al.*, 2005). In addition to *de novo* synthesis, purines, pyrimidines and their nucleosides can be formed in the cells via salvage reactions. The salvage pathways of *E. coli* involve 33 enzymes catalyzing 85 reactions (Table 1).

Table 1. Enzymatic reactions present in the salvage pathway network

Enzyme	Gene	Reaction	Alternative substrate	EC
Adenylate kinase	<i>adk</i>	GTP + AMP \leftrightarrow ADP + GDP	ATP, ITP, DAMP	2.7.4.3
Guanylate kinase	<i>gmk</i>	GMP + ATP \leftrightarrow GDP + ADP	DGMP	2.7.4.8
Nucleoside-diphosphate kinase	<i>ndk</i>	GDP + ATP \leftrightarrow GTP + ADP	UDP, CDP, DGDP, DUDP, DCDP, DADP, DTDP	2.7.4.6
AMP Nucleosidase	<i>amn</i>	AMP \rightarrow AD + R5P		3.2.2.4
Adenosine deaminase	<i>add</i>	ADN \rightarrow INS + NH ₃	DA	3.5.4.4
Adenine deaminase	<i>yicP</i>	AD \rightarrow NH ₃ + HYXN		3.5.4.2
Inosine/ Guanosine kinase	<i>gsk</i>	INS + ATP \rightarrow IMP + ADP	GSN	2.7.1.73
Adenine phosphoribosyltransferase	<i>apt</i>	AD + PRPP \rightarrow PPI + AMP		2.4.2.7
Xanthine-guanine phosphoribosyltransferase	<i>gpt</i>	XAN + PRPP \rightarrow XMP + PPI	GN, HYXN	2.4.2.22
Hypoxanthine phosphoribosyltransferase	<i>hpt</i>	HYXN + PRPP \rightarrow PPI + IMP	GN	2.4.2.8
Xanthosine phosphorylase	<i>xapA</i>	DIN + PI \leftrightarrow HYXN + DR1P	DA, DG, INS, ADN, GSN, XTSN	2.4.2.1
Purine nucleotide phosphorylase	<i>deoD</i>	DIN + PI \leftrightarrow HYXN + DR1P	DA, DG, INS, ADN, GSN	2.4.2.1
Uridine phosphorylase	<i>udp</i>	URI + PI \leftrightarrow URA + R1P		2.4.2.3
Thymidine/deoxyuridine phosphorylase	<i>deoA</i>	DU + PI \leftrightarrow URA + DR1P	DT	2.4.2.4
Cytidylate kinase	<i>cmk</i>	CMP + ATP \leftrightarrow ADP + CDP	UMP, DCMP	2.7.4.14
dTMP kinase	<i>tmk</i>	DTMP + ATP \leftrightarrow ADP + DTD		2.7.4.9
Uridylate kinase	<i>pyrH</i>	UMP + ATP \leftrightarrow UDP + ADP	DUMP	2.1.4.-
Uracil phosphoribosyltransferase	<i>upp</i>	URA + PRPP \rightarrow UMP + PPI		2.4.2.9
Cytosine deaminase	<i>codA</i>	CYTS \rightarrow URA + NH ₃		3.5.4.1
Uridine/Cytidine kinase	<i>udk</i>	URI + GTP \rightarrow GDP + UMP	CYTD	2.7.1.48
Thymidine (deoxyuridine) kinase	<i>tdk</i>	DT + ATP \rightarrow ADP + DTMP	DU	2.7.1.21
dCTP deaminase	<i>dcd</i>	DCTP \rightarrow DUTP + NH ₃		3.5.4.13
Cytidine deaminase	<i>cdd</i>	DC \rightarrow NH ₃ + DU	CYTD	3.5.4.5

Enzyme	Gene	Reaction	Alternative substrate	EC
5'-Nucleotidase	<i>ushA</i>	AMP → PI + ADN	GMP, IMP, XMP, UMP, CMP, DCMP, DGMP, DAMP, DTMP, DUMP	3.1.3.5
Ribonucleoside-diphosphate reductase	<i>nrdAB</i>	ADP + RTHIO → DADP + OTHIO	GDP, CDP, UDP	1.17.4.1
Ribonucleoside-triphosphate reductase	<i>nrdD</i>	ATP + RTHIO → DATP + OTHIO	GTP, CTP, UTP	1.17.4.2
Ribonucleoside-diphosphate reductase II	<i>nrdEF</i>	CDP + RTHIO → DCDP + OTHIO		
dUTP pyrophosphatase	<i>dut</i>	DUTP → PPI + DUMP		3.6.1.23
Thymidylate synthetase	<i>thyA</i>	DUMP + METTHF → DHF + DTMP		2.1.1.45
Nucleoside triphosphatase	<i>mutT</i>	GTP → GMP + PPI	DGTP	3.6.1.-
Deoxyguanosinetriphosphate triphosphohydrolase	<i>dgt</i>	DGTP → DG + PPP	GTP	3.1.5.1

Table 1 shows the enzymatic reactions present in the gene network under consideration, names of enzymes catalyzing corresponding reactions, and names of genes coding for the enzymes. Table 2 summarizes the components of the salvage pathway network.

Table 2. Components of the salvage pathway network

Operon	RNA	Enzyme	Reaction	Inorganic substance	Repressor	Transcription factor	Reference
30	30	32	476	82	58	10	390

Escherichia coli possesses the ability to take up purine and pyrimidine nucleosides from the growth medium and use them as sources of nitrogen and carbon. Nucleoside phosphorylases catalyze the phosphorolytic cleavage of the nucleoside, thereby forming the free nucleotide base and (deoxy)ribose-1-phosphate. The base can be utilized by the purine or the pyrimidine salvage pathways, and the ribose-1-phosphate and the deoxyribose-1-phosphate can be converted to intermediates of the pentose phosphate shunt and of glycolysis, respectively. Of the four different nucleoside phosphorylases in *E. coli*, uridine phosphorylase (*udp*) and thymidine phosphorylase (*deoA*) are specific for pyrimidine nucleosides whereas purine nucleoside phosphorylase (*deoD*) and xanthosine phosphorylase (*xapA*) are specific for purine nucleosides. Purine nucleoside phosphorylase is important for the breakdown of all purine nucleosides and deoxynucleosides except xanthosine. Xanthosine phosphorylase (XapA), on the other hand, has specificity toward xanthosine and all other purine nucleosides and deoxynucleosides except adenosine and deoxyadenosine. Purine nucleoside phosphorylase (DeoD) is encoded by the last gene of the *deoCABD* operon. The regulation of these genes is complex and involves two repressors (CytR and DeoR) and an activator (cyclic AMP [cAMP] receptor protein-cAMP complex). Despite the action of two repressors, the *deo* genes are always expressed at a low basal level to ensure a rapid metabolism of purine nucleosides taken up from the medium. In contrast, *xapA* is expressed only if the inducer xanthosine is present in the growth medium. The xanthosine-induced activation of *xapA* expression is mediated by the regulatory protein XapR (Jorgensen, Dandanell, 1999). Apart of XapR, the following transcription factors control the expression of the genes of the salvage pathways: FNR (*nrdDG*), FUR (*nrdHIEF*), Fis (*nrdAB*), Nac (*codBA*), CRP (*hpt*), CRP, CytR (*udp*, *cdd*, *cytR*, *deoCABD*), DeoR (*deoCABD*), IHF (*hpt*). In addition to the transcription level, salvage pathway genes are regulated at the translation level. Expression of the *upp* gene of *E. coli*, which encodes the pyrimidine salvage enzyme uracil phosphoribosyltransferase, is negatively regulated by pyrimidine availability. The regulation occurs mainly by UTP-sensitive selection of alternative transcriptional start sites, which produces transcripts that differ in the ability to be productively elongated (Tu, Turnbough, 1997).

Application of the method of generalized Hill functions to modeling the molecular processes of salvage pathways can be exemplified by regulation of the activity of adenine phosphoribosyltransferase (APRT, coded by *apt* gene) in *E. coli*. The enzyme catalyzes a salvage reaction yielding AMP (Table 1). It is known that all acyclic nucleoside-5'-phosphates considerably inhibit APRT activity by competition for the substrate PRPP (Hochstadt-Ozer and Stadtman, 1971). An equation for the steady-state rate of the reaction is proposed:

$$V = \frac{k_{cat} \cdot e_0 \cdot \frac{S_1}{K_{m,S_1}} \cdot \frac{S_2}{K_{m,S_2}}}{\left(I + \frac{S_1}{K_{m,S_1}} \right) \cdot \left(I + \frac{S_2}{K_{m,S_2}} + \frac{P_1}{K_{i,P_1}} + \frac{P_2}{K_{i,P_2}} + \sum_{j=1}^5 \frac{R_j}{k_{i,R_j,S_2}} + \sum_{j=6}^{10} \left(\frac{R_j}{k_{i,R_j,S_2}} \right)^2 + \frac{R_{11}}{k_{i,R_{11},S_2}} \right) \cdot \left(I + kl_{R_{12}} \cdot \frac{R_{12}}{k_{R_{12}} + R_{12}} \right)}, \quad (1)$$

where e_0 is APRT concentration; S_1 , S_2 , P_1 , P_2 , R_1 , R_2 , R_3 , R_4 , R_5 , R_6 , R_7 , R_8 , R_9 , R_{10} , R_{11} , R_{12} are concentrations of low-molecular-weight substances AD, PRPP, PPI, AMP, ADP, dADP, ATP, dATP, dAMP, GTP, ITP, XTP, UTP, GDP, Mg^{2+} , and cAMP, respectively; k_{cat} is the catalytic constant; K_{m,S_i} are the Michaelis constants for corresponding substrates; K_{i,P_i} , constants of inhibition by corresponding products; k_{i,R_i,S_2} , constants of inhibition by the corresponding regulator competing for the substrate PRPP; and $kl_{R_{12}}$, $k_{R_{12}}$, constants determining the efficiency of the effect of cAMP on the reaction rate.

Experimental data reported in (Hochstadt-Ozer and Stadtman, 1971) were used for testing the model of regulation of APRT activity. These data illustrate the effects of various low-molecular-weight substances (see comments on Eq. (1)) on APRT activity at various concentration combinations (Fig. 1).

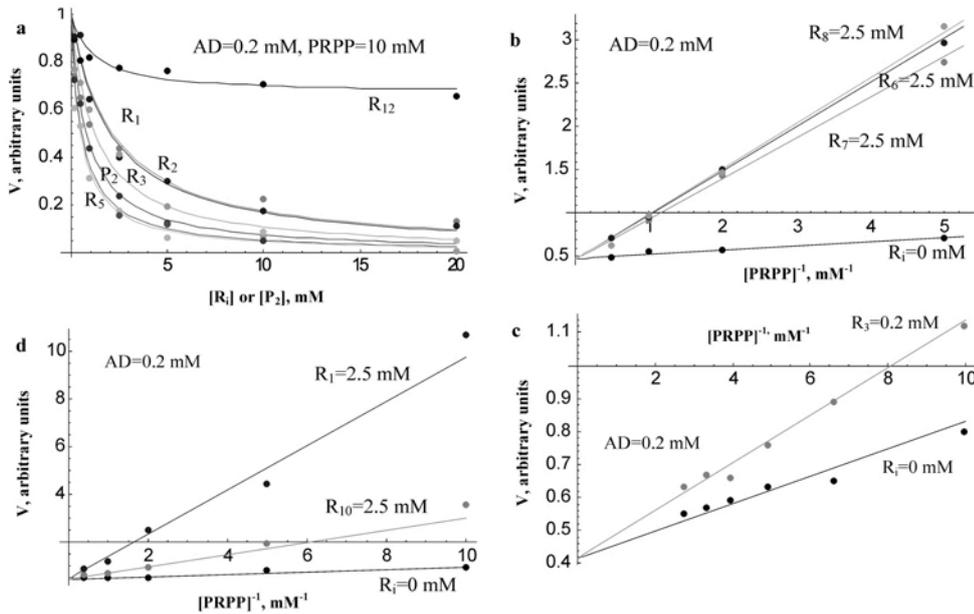


Figure 1. Effect of various regulators (for designations see text, comments on Eq. (1)) on the rate of the reaction catalyzed by APRT (a). Effect of PRPP of the rate of the reaction catalyzed by G6P1D at various concentrations of regulators R_i . (for each predicted curve $R_j = 0$ mM, at $j \neq i$) and AD = 0.2 mM (b, c, d). Dots indicate experimental values reported in (Hochstadt-Ozer, Stadtman, 1971). Curves are results of calculation according to Eq. (1) at the following parameter values: $k_{cat} = 560 \text{ min}^{-1}$; $K_{m,S_1} = 0.011 \text{ mM}$; $K_{m,S_2} = 0.1 \text{ mM}$; $K_{i,P_1} = 0.8 \text{ mM}$; $K_{i,P_2} = 0.03 \text{ mM}$; $k_{i,R_1,S_2} = 0.13 \text{ mM}$; $k_{i,R_2,S_2} = 0.02 \text{ mM}$; $k_{i,R_3,S_2} = 0.27 \text{ mM}$; $k_{i,R_4,S_2} = 0.008 \text{ mM}$; $k_{i,R_5,S_2} = 0.0055 \text{ mM}$; $k_{i,R_6,S_2} = 0.79 \text{ mM}$; $k_{i,R_7,S_2} = 0.84 \text{ mM}$; $k_{i,R_7,S_2} = 2.0$; $k_{i,R_8,S_2} = 0.8 \text{ mM}$; $k_{i,R_9,S_2} = 10 \text{ mM}$; $k_{i,R_{10},S_2} = 1.7 \text{ mM}$; $kl_{R_{12}} = 0.5$; $k_{i,R_9,S_2} = 0.27 \text{ mM}$; and $k_{i,R_{10},S_2} = 0.27 \text{ mM}$.

DISCUSSION

The reaction catalyzed by APRT is an example of multicomponent and complex regulation of a molecular system involving numerous structurally similar components. When this feature of the cellular system is taken into account, the connectivity of the graph representing the system increases dramatically. Molecular systems can also acquire this property because of a vast number of nonspecific interactions in the cell. Thus, the enzymatic system under discussion concerns a basic property of the cell. Consideration of such features in mathematical modeling of molecular systems is of paramount importance for proper description of actual molecular processes in a living cell.

Reconstruction of the gene network and development of mathematical models describing the efficiency of operation of enzymatic systems are essential for constructing a general kinetic model of salvage pathways. Such a model would allow predicting the progress of processes in the system, understanding their mechanisms, determining key links of the gene network, and analyzing effects of mutations on its operation. It will be an inextricable part of the “*in silico* cell” computer resource.

ACKNOWLEDGEMENTS

The authors are grateful to Vitaly Likhoshvai for valuable discussions, to Irina Lokhova for bibliographical support, and to Victor Gulevich for translating the manuscript from Russian into English. This work was supported in part by the Russian Government (Contract No. 02.467.11.1005), by Siberian Branch of the Russian Academy of Sciences (the project “Evolution of molecular-genetical systems: computational analysis and simulation” and integration projects Nos 24 and 115), and by the Federal Agency of Science and Innovation (innovation project No. IT-CP.5/001).

REFERENCES

- Ananko E.A., Podkolodny N.L., Stepanenko I.L., Podkolodnaya O.A., Rasskazov D.A., Miginsky D.S., Likhoshvai V.A., Ratushny A.V., Podkolodnaya N.N., Kolchanov N.A. (2005) GeneNet in 2005. *Nucl. Acids Res.*, **33**, D425–D427.
- Hochstadt-Ozer J., Stadtman E.R. (1971) The regulation of purine utilization in bacteria. I. Purification of adenine phosphoribosyltransferase from *Escherichia coli* K12 and control of activity by nucleotides. *J. Biol. Chem.*, **246**, 5294–5303.
- Jorgensen C., Dandanell G. (1999) Isolation and characterization of mutations in the *Escherichia coli* regulatory protein XapR. *J. Bacteriol.*, **181**, 4397–4403.
- Khlebodarova T.M. *et al.* (2006) *In silico* cell II. Information sources for modeling *Escherichia coli* metabolic pathways. *This issue*.
- Likhoshvai V.A., Ratushny A.V. (2006) *In silico* cell I. Hierarchical approach and generalized hill functions in modeling enzymatic reactions and gene expression regulation. *This issue*.
- Tu A.H., Turnbough C.L.Jr. (1997) Regulation of upp expression in *Escherichia coli* by UTP-sensitive selection of transcriptional start sites coupled with UTP-dependent reiterative transcription. *J. Bacteriol.*, **179**, 6665–6673.

MATHEMATICAL MODELING OF SERINE AND GLYCINE BIOSYNTHESIS REGULATION IN *ESCHERICHIA COLI*

Turnaev I.I.^{*1}, Ibragimova S.S.¹, Usuda Y.², Matsui K.²

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ² Institute of Life Sciences, Ajinomoto, Co., Inc., Kawasaki, Japan

* Corresponding author: e-mail: turn@bionet.nsc.ru

Key words: mathematical modeling, gene network, regulation, serine and glycine biosynthesis, *Escherichia coli*

SUMMARY

Motivation: Development of an electronic cell as a resource for use on computers for modeling and analysis of intracellular processes is important to further advancements in systems biology and bioinformatics. To contribute to the progress, it is of importance that mathematical models of cellular metabolic pathway regulation, in particular, regulation of serine and glycine biosynthesis be developed.

Results: The gene network for regulation of serine and glycine biosynthesis in *E. coli* has been reconstructed using the GeneNet technology. Using Hill's generalized functions, mathematical models have been developed for the enzyme reactions. Based on these models, a mathematical model for serine and glycine has been developed.

Availability: The model is available on request, the gene network is accessible at <http://www.mgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/index.shtml>

INTRODUCTION

SER⁶ is the most important cell metabolite. In cells grown on minimum medium with glucose as a source of carbon, about 15 % cell carbon follows through the PGDH reaction and onwards to the synthesis of SER and its derivatives (Pizer, Potochny, 1964). SER being a growth factor, cell division starts at the SER concentration of 0.05 mM (Prub, Matzumura, 1996). *E. coli* depends on the SER biosynthetic pathway, because mutations, which result in the loss of PGDH, lead to a growth requirement for SER (Pizer, Potochny, 1964). High levels of SER in growth medium lead to growth inhibition. The intracellular SER pool is regulated by feedback inhibition of PGDH, the product of the *serA* gene, by SER through conformational change in the enzyme. SER serves as a precursor for GLY, TRY and CYS. SER is also incorporated into phospholipids. SER is converted to GLY by SHMT, the *glyA* gene product. This reaction is the only source of GLY and the major source of one-carbon units to the cell (Stauffer, 1996).

We report herein the reconstruction of the gene network for serine and glycine biosynthesis and a mathematical model of this process.

⁶ The abbreviations used are: SER, L-serine, PGDH, 3-phosphoglycerate dehydrogenase, GLY, glycine, TRY, tryptophan, CYS, cysteine, SHMT, serine hydroxymethyltransferase.

A listing of enzyme reactions in the reconstructed gene network, the names of enzymes catalyzing the corresponding reactions and the names of genes encoding the corresponding enzymes are presented in Table 2⁷.

Table 2. A listing of enzyme reactions in the gene network for serine and glycine biosynthesis and degradation

No.	Enzyme name	Gene	Reaction	Low-molecular regulators
1	3-Phosphoglycerate dehydrogenase (PGDH)	<i>serA</i>	3PG + NAD → NADH + PHP	SER, H ⁺ NADH, PHP
2	Phosphoserine transaminase (PSAT)	<i>serC</i>	PHP + GLU → AKG + 3PSER	AKG, 3PSER
3	Phosphoserine phosphatase (PSP)	<i>serB</i>	3PSER → PI + SER	PI, SER
4	Glycine hydroxymethyltransferase (SHMT)	<i>glyA</i>	THF + SER → GLY + MTHF	GLY, MTHF

A database containing experimental data on the dynamics of the gene network components has been developed (Khlebodarova *et al.*, 2006). Using Hill's generalized functions, mathematical models have been developed for the enzyme reactions listed in Table 2, and the models' parameters, evaluated.

Based on the models of separate enzyme reactions, a dynamic model for serine and glycine biosynthesis in *E. coli* has been developed, a general schematic of which is presented in Fig. 1. The mathematical model comprises 9 differential equations.

Enzyme reaction rates:

$$V1 \text{ (PGDH activity)} = \frac{k_{15} \cdot [\text{PGDH}] \cdot [\text{3PG}] \cdot [\text{NAD}] \cdot \frac{(1+k_{28} \cdot [\text{SER}]/k_{26})(1+k_{29} \cdot [\text{SER}]/k_{27})}{(1+[\text{SER}]/k_{26})(1+[\text{SER}]/k_{27})} \cdot \left(\frac{1}{1+([\text{SER}]/k_{20})^{k_{33}} \cdot (k_{22}/[\text{H}^+] + [\text{H}^+]/k_{21})} \right)}{k_{16} \cdot k_{17} \cdot \left(1 + [\text{3PG}] / \left(\frac{1+k_{28} \cdot [\text{SER}]/k_{26}}{1+[\text{SER}]/k_{26}} + [\text{PHP}]/k_{19} \right) \right) \cdot \left(1 + [\text{NAD}] / \left(\frac{1+k_{29} \cdot [\text{SER}]/k_{27}}{1+[\text{SER}]/k_{27}} + [\text{NADH}]/k_{18} \right) \right)}$$

$$V2 \text{ (PSAT activity)} = \frac{k_{30} \cdot [\text{PSAT}] \cdot [\text{GLU}] \cdot [\text{PHP}]}{k_{31} \cdot k_{32} \cdot (1 + [\text{AKG}]/k_{33} + [\text{PHP}]/k_{31}) \cdot (1 + [\text{GLU}]/k_{32} + [3\text{PSER}]/k_{34})}$$

$$V3 \text{ (PSP activity)} = \frac{k_{35} \cdot [\text{PSP}] \cdot (1 + [\text{PI}]/k_{38}) \cdot [3\text{PSER}]}{k_{36} \cdot \frac{1 + [\text{PI}]/k_{39}}{1 + k_{41} \cdot [\text{PI}]/k_{39}} \cdot (1 + k_{40} \cdot [\text{PI}]/k_{38}) \cdot \left(1 + [3\text{PSER}] / \left(\frac{1 + [\text{PI}]/k_{39}}{k_{36} \cdot \frac{1 + [\text{PI}]/k_{39}}{1 + k_{41} \cdot [\text{PI}]/k_{39}} + [\text{SER}]/k_{37}} \right) \right)}$$

$$V4 \text{ (SHMT activity)} = \frac{k_{42} \cdot [\text{SHMT}] \cdot [\text{THF}] \cdot [\text{SER}]}{k_{43} \cdot k_{44} \cdot (1 + [\text{MTHF}]/k_{46} + [\text{THF}] \cdot (1 + [\text{MTHF}]/k_{47})/k_{44}) \cdot (1 + [\text{GLY}]/k_{45} + [\text{SER}]/k_{43})}$$

Differential equations:

$$\frac{d[\text{PDGH}]}{dt} = k_1 - k_6 \cdot [\text{PDGH}]; \quad \frac{d[\text{PSAT}]}{dt} = k_2 - k_7 \cdot [\text{PSAT}]; \quad \frac{d[\text{PSP}]}{dt} = k_3 - k_8 \cdot [\text{PSP}]; \quad \frac{d[\text{SHMT}]}{dt} = k_5 - k_9 \cdot [\text{SHMT}];$$

$$\frac{d[\text{3PG}]}{dt} = k_4 - v_1 - k_{10} \cdot [\text{3PG}]; \quad \frac{d[\text{PHP}]}{dt} = v_1 - v_2; \quad \frac{d[\text{P3SER}]}{dt} = v_2 - v_3; \quad \frac{d[\text{SER}]}{dt} = v_3 - v_4 - k_{13} \cdot [\text{SER}]; \quad \frac{d[\text{GLY}]}{dt} = v_4 - k_{14} \cdot [\text{GLY}].$$

Constant values: $k_1 = 1.15 \cdot 10^{-3}$ mM/s, $k_2 = 10^{-4}$ mM/s, $k_3 = 2.5 \cdot 10^{-4}$ mM/s, $k_4 = 0.4$ mM/s, $k_5 = 1.15 \cdot 10^{-4}$ mM/s, $k_6, k_7, k_8, k_9 = 10^{-3}$ s⁻¹, $k_{10} = 0.4$ s⁻¹, $k_{13} = 0.0103$ s⁻¹, $k_{14} = 8.03 \cdot 10^{-3}$ s⁻¹, $k_{15} = 0.55$ s⁻¹, $k_{16} = 1.1$ mM, $k_{17} = 7,8 \cdot 10^{-3}$ mM, $k_{18} = 0.78$ mM,

⁷ The abbreviations used are: 3PG, 3-phosphoglycerate, PHP, 3-hosphohydroxypyruvate, PSAT, 3-phosphoserine aminotransferase, 3PSER, 3-phosphoserine, PSP, 3-phosphoserine phosphatase, , LEU, Leucine, Lrp, Leucine-responsive regulatory protein, MTHF, 5,10-methylenetetrahydrofolate, MET, methionine, MetR, Homocysteine transcriptional activator, HCYS, homocysteine, PurR, Hypoxanthine transcriptional repressor, THF, tetrahydrofolate, PI, phosphate, Crp, cAMP receptor protein.

$k_{19} = 110$ mM, $k_{20} = 0.047$ mM, $k_{21} = 0.074$ mM, $k_{22} = 3.5 \cdot 10^{-3}$ mM, $k_{23} = 2.2$, $k_{26} = 1.8$ mM, $k_{27} = 3 \cdot 10^{-4}$ mM, $k_{28} = 40$, $k_{29} = 0.6$, $k_{30} = 1.75$ s⁻¹, $k_{31} = 0.15$ mM, $k_{32} = 20$ mM, $k_{33} = 100$ mM, $k_{34} = 1.5$ mM, $k_{35} = 1.43$ s⁻¹, $k_{36} = 1.5 \cdot 10^{-3}$ mM, $k_{37} = 0.15$ mM, $k_{38} = 500$ mM, $k_{39} = 1$ mM, $k_{40} = 36$, $k_{41} = 1.5$, $k_{42} = 1.83$ s⁻¹, $k_{43} = 0.8$ mM, $k_{44} = 0.04$ mM, k_{45} , $k_{46} = 1$ mM, $k_{47} = 0.083$ mM.

In addition to the enzyme processes (expressions V1, V2, V3, V4), the model describes the constitutive synthesis of the active forms of pathway enzymes (parameters k_1, k_2, k_3, k_5), constitutively describes predecessor 3PG uptaking (parameter k_4), linearly describes enzyme degradation (constants k_6, k_7, k_8, k_9), the utilization of the substrate 3PG (constant k_{10}) and the utilization of the end products SER and GLY in other pathways in *E. coli* (constants k_{13}, k_{14}). Enzyme activity is described (expressions V1, V2, V3, V4) with due regard to regulation exerted on the enzymes by low-molecular compounds (Table 2). For example, the effects of H⁺ concentrations on PGDH activity are taken into account as proposed by Michaelis (Cornish-Bowden, 1979). The concentrations of NAD, NADH, GLU, AKG, THF, MTHF, PI and pH, involved in the regulation of enzyme activity and being a source of groups and energy, are included in the equations as parameters.

The values of the synthesis, degradation and utilization parameters and the physiological values of substance concentrations included in the model as parameters were selected in such a way that they best fit in experimental data and that stationary state of the mathematical model fit in experimental data on substance concentrations and enzyme activity in the gene network for serine and glycine biosynthesis.

Equilibrium concentration values for serine and glycine are shown in Fig. 3A (dotted lines). Curves for the time it takes the system to reach equilibrium after assigning 0 mM to SER or GLY concentrations were calculated using the model. As the calculations demonstrate, SER (solid line) reaches equilibrium (dotted line) in 100 seconds, but does not go beyond, plateauing at about 0.9 mM for 4 to 10 minutes and then declines to equilibrium during 100 seconds. It takes GLY 7 – 8 minutes to reach equilibrium. The time course of SER can be explained from looking at SER and PHP interactions (Fig. 3C). As a predecessor, PHP activates SER synthesis. SER inhibits PHP synthesis, which suggests that SER down regulates its own synthesis (Fig. 2). Four segments can be recognized in the curve presented in Fig. 3C: 1) The straight line ending with the first turn (coordinates 4, PHP × 0.25, SER (mM)). PHP is growing more rapidly than SER, because SER in deficient amounts fails to inhibit PHP synthesis. 2) The segment ending with the turn (coordinates 3.8, PHP × 0.86, SER). SER is growing more rapidly than PHP, because down regulation of PHP synthesis by SER is enhanced. Shortly before reaching the turn, SER passes by the equilibrium point, because excess PHP (Fig. 3B) enhances SER synthesis. SER goes beyond the equilibrium point because SER self-controlled synthesis is temporarily unaffected by PHP synthesis. SER stays beyond until excess PHP has exhausted. 3) The segment ending with the turn (coordinates 0.5, PHP × 0.9, SER). Because excess PHP has exhausted, SER is plateauing beyond the equilibrium point, and PHP is declining. 4) From 0.5, PHP × 0.9, SER on, after excess PHP has entirely exhausted, PHP concentrations again become dependent on SER-controlled SER synthesis and both PHP and SER return to equilibrium. The calculated time it takes SER to return to the physiological level is about 100 seconds, because from that point on SER do not overstep the physiologic limits.

In the future, we are planning to expand the model by including the regulatory mechanisms of gene expression in the serine and glycine biosynthetic pathway.

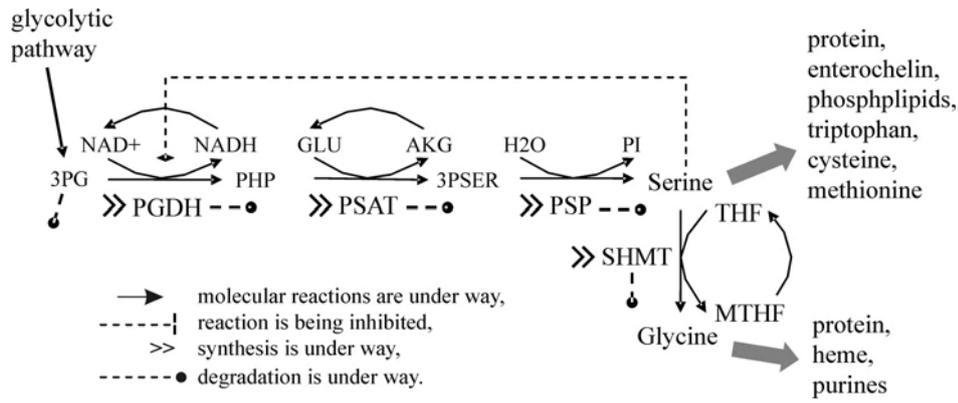


Figure 2. A general schematic of serine and glycine biosynthesis including synthesis and degradation.

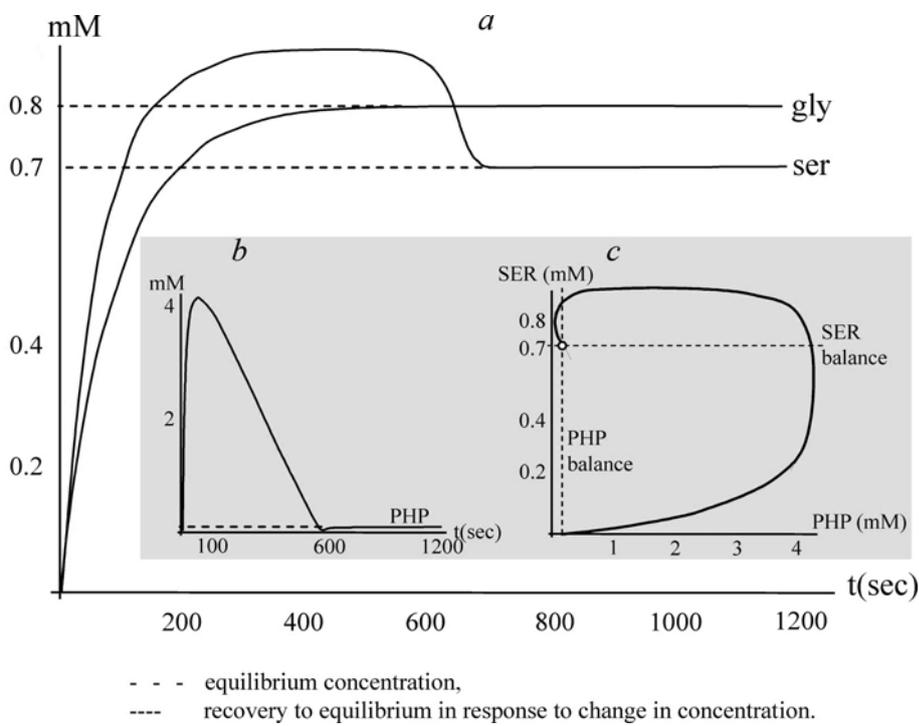


Figure 3. Dynamics calculations: *a* – serine, glycine and *b* – PHP. The X axis: time (in seconds); the Y axis: concentrations of serine, glycine and PHP (in mM). *c* – recovery to equilibrium on the interaction diagram for serine and PHP.

ACKNOWLEDGEMENTS

The authors are grateful to Vitaly Likhoshvai and Aleksandr Ratushny for valuable discussions and Irina Lokhova for bibliographical support. This work was supported in part by the Russian Government (Contract No. 02.467.11.1005), by Siberian Branch of the Russian Academy of Sciences (the project “Evolution of molecular-genetical systems:

computational analysis and simulation” and integration projects Nos 24 and 115), and by the Federal Agency of Science and Innovation (innovation project No. IT-CP.5/001).

REFERENCES

- Ananko E.A. *et al.* (2005) GeneNet in 2005. *Nucl. Acids Res.*, **33**, D425–D427.
- Cornish-Bowden A. (1979) The basis of enzyme kinetics, Mir, Moscow. (in Russ).
- Khlebodarova T.M. *et al.* (2006) *In silico* cell II. Information sources for modeling *Escherichia coli* metabolic pathways. *This issue*.
- Likhoshvai V.A., Ratushny A.V. (2006) *In silico* cell I. Hierarchical approach and generalized hill functions in modeling enzymatic reactions and gene expression regulation. *This issue*.
- Pizer L.I., Potochny M. L. (1964) Nutritional and regulatory aspects of serine metabolism in *Escherichia coli*. *J. Bacteriol.*, **88**, 611–619.
- Prub B.M., Matsumura P. (1996) A Regulator of the Flagellar Regulon of *Escherichia coli*, *flhD*, Also Affects Cell Division. *J. Bacteriol.*, **178**, 668–674.
- Stauffer G. (1996) Biosynthesis of serine, glycine and one-carbon units. *E.coli* and *Salmonella*: Cellular and Molecular Biology, Washington DC, (2nd ed Neidhart F.C.), **1**, 506–513.

REGULATION OF PYRUVATE BIOSYNTHESIS IN *ESCHERICHIA COLI*: GENE NETWORK RECONSTRUCTION AND MATHEMATICAL MODELING OF ENZYMATIC REACTIONS OF THE PATHWAY

Turnaev I.I.*, **Smirnova O.G.**

Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia

* Corresponding author: e-mail: turna@bionet.nsc.ru

Key words: mathematical modeling, gene network, regulation, pyruvate metabolism, D-Lactate dehydrogenase-2, *Escherichia coli*

SUMMARY

Motivation: Development of an *in silico* cell, a computer resource for modeling and analysis of physiological processes, is an urgent task of systems biology and computational biology. Mathematical modeling of the genetic regulation of the pyruvate metabolism is an important problem to be solved as part of this line of work.

Results: By using the GeneNet technology, we reproduced the gene network of pyruvate metabolism regulation in the *E. coli* cell. Mathematical models were constructed by the method of generalized Hill functions to describe the efficiency of enzyme systems.

Availability: Models are available on request. The gene network is available at <http://www.mgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/index.shtml>

INTRODUCTION

Under anaerobic conditions, competition for pyruvate between the branch point enzymes PFL⁸ and LDH determines the partition of carbon flux. Under anaerobic conditions, especially at low pH, *Escherichia coli* unidirectionally converts pyruvate to D-lactate by means of (LdhA). This LDH is present at substantial basal levels under all conditions but increases approximately 10-fold at low pH (Jiang *et al.*, 2001). The enzyme is allosterically activated by pyruvate (Tarmy, Kaplan, 1968). Pyruvate also caused a two- to fourfold increase in expression of *ldhA* (Jiang *et al.*, 2001).

The gene network of pyruvate metabolism regulation in the *E. coli* cell was reconstructed. Mathematical models of the efficiency of enzyme systems were constructed. A database storing experimental data on the behavior of components of this gene network was developed (Khlebodarova *et al.*, 2006). Parameters of the models were determined by numerical simulation. The results of calculation of steady-state properties and behavior of the components of the molecular system derived from the models are in agreement with experimental evidence.

⁸ The abbreviations used are: PFL, pyruvate formate lyase, LDH, fermentative lactate dehydrogenase.

METHODS AND ALGORITHMS

The gene network was reconstructed using the GeneNet system (Ananko *et al.*, 2005). The method of generalized Hill functions (Likhoshvai, Ratushny, 2006) was used to model the regulation of the operation of enzymatic systems.

RESULTS AND DISCUSSION

The gene network of pyruvate metabolism regulation was reconstructed (Fig. 1) with the use of the GeneNet technology (Ananko *et al.*, 2005) and available at <http://wwwmgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/index.shtml>.

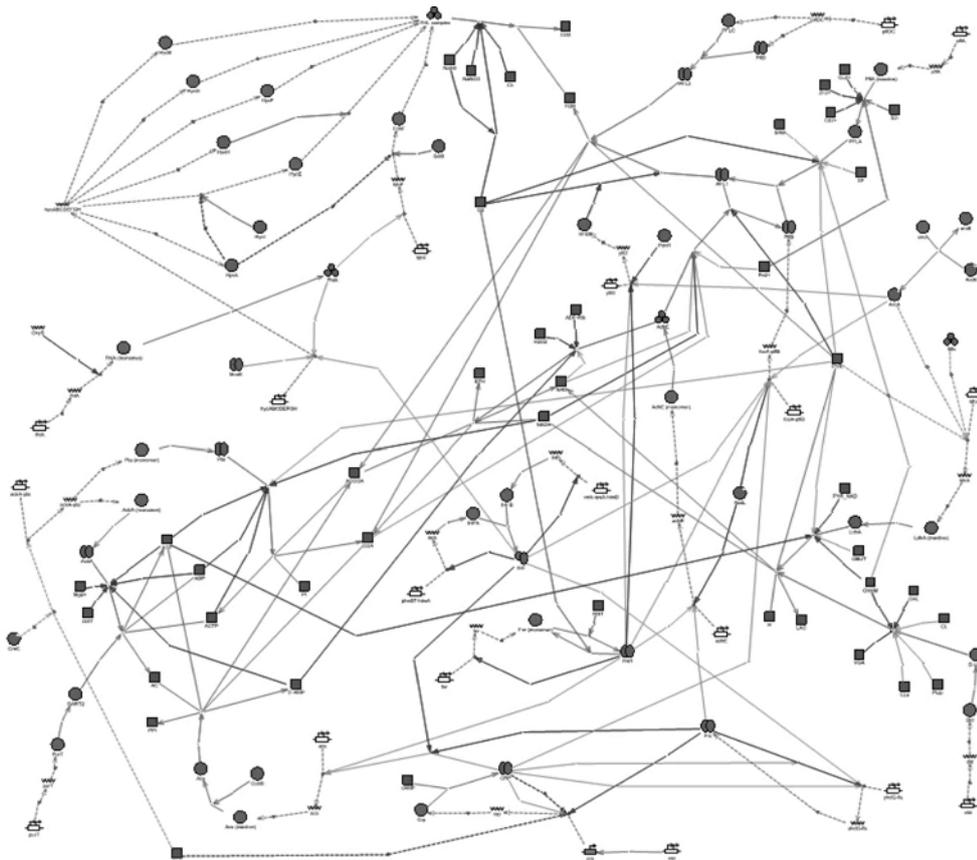


Figure 1. Pyruvate metabolism gene network reconstruction in the GeneNet.

Table 1 summarizes the components of the network. Table⁹ 2 shows the enzymatic reactions present in the pyruvate metabolism gene network, names of enzymes catalyzing corresponding reactions, and names of genes coding for the enzymes.

Table 1. Components of the pyruvate metabolism gene network

⁹ The abbreviations used are: NADH-dependent lactate dehydrogenase; PYR, pyruvate; LAC, lactate; COA, coenzyme A; ACCOA, acetyl COA; ETH, ethanol; FOR, formate; PI, phosphate; PPI, pyrophosphate; ACTP, acetylphosphate; AC, acetate; AKB, alpha-ketobutyrate; OX, oxamate; SAM, S-adenosylmethionine.

Operon	RNA	Protein	Reaction	Inorganic substance	Repressor	Transcription factor	Reference
18	18	52	178	43	12	7	255

Table 2. Enzymatic reactions constituting the pyruvate metabolism gene network

Enzyme	Gene	Reaction	Low M.w. regulators
D-Lactate dehydrogenase 2 (LDHA)	<i>ldhA</i>	PYR + NADH(NADPH) \leftrightarrow NAD(NADP) + LAC	AKB, ATP
Acetaldehyde dehydrogenase (ADHE)	<i>adhE</i>	ACCOA + 2 NADH \leftrightarrow ETH + 2 NAD + COA	
Pyruvate formate lyase 1 (PFL)	<i>pflAB</i>	PYR + COA \rightarrow ACCOA + FOR	Fe ²⁺ , SAM
Formate hydrogen lyase (<i>fdhF</i> , <i>hycBEFG</i>)	<i>fdhF</i>	FOR \rightarrow CO ₂	Nitrate, Azide
Phosphotransacetylase (PTA)	<i>pta</i>	ACCOA + PI \leftrightarrow ACTP + COA	ADP, NADH, PYR
Acetate kinase A (ACKA)	<i>ackA</i>	ACTP + ADP \leftrightarrow ATP + AC	Mg ²⁺
GAR transformylase T (PURT)	<i>purT</i>	ACTP + ADP \leftrightarrow ATP + AC	
Acetyl-CoA synthetase (ACS)	<i>acs</i>	ATP + AC + COA \rightarrow AMP + PPI + ACCOA	

Reactions of the PYR biosynthesis pathway were modeled. Application of the method of generalized Hill functions to modeling the molecular processes of the pyruvate metabolism can be exemplified by regulation of the activity of the enzyme D-Lactate dehydrogenase 2 (LDHA) in the *E. coli* cell (Table 2):

$$V = \frac{V_{\max} \cdot \left(\frac{[\text{NADPH}]}{K_{m_{\text{nadph}}} \cdot f(\text{PYR,ATP})} \right) \left(\frac{[\text{PYR}]}{K_{m_{\text{pyr}}} \cdot f_2(\text{AKB})} \right)^{n_p \cdot f_1(\text{AKB})}}{\left(1 + \left(\frac{[\text{PYR}]}{K_{m_{\text{pyr}}} \cdot f_2(\text{AKB})} \right)^{n_p \cdot f_1(\text{AKB})} \right) \left(1 + \frac{[\text{NADPH}]}{K_{m_{\text{nadph}}} \cdot f(\text{PYR,ATP})} \right)}$$

where

$$f_1(\text{AKB}) = \frac{1 + [\text{AKB}]/K_{\text{akb1}}}{1 + [\text{AKB}]/K_{\text{akb2}}}; \quad f_2(\text{AKB}) = \left(\frac{1 + [\text{AKB}]/K_{\text{akb3}}}{1 + [\text{AKB}]/K_{\text{akb4}}} \right);$$

$$f(\text{PYR,ATP}) = \frac{\left(1 + \frac{[\text{PYR}]}{K_{\text{pyr1}}} \right) \left(1 + \frac{[\text{ATP}]}{K_{\text{atp1}}} \right)}{\left(1 + \frac{[\text{PYR}]}{K_{\text{pyr2}}} \right) \left(1 + \frac{[\text{ATP}]}{K_{\text{atp2}}} \right)};$$

$$V_{\max} = e_0 \cdot k_f \cdot f(\text{NADPH,OX}) \quad \text{here } f(\text{NADPH,OX}) = \frac{\left(1 + \frac{[\text{NADPH}]}{K_{\text{nadph1}}} \right) \left(1 + \frac{[\text{OX}]}{K_{\text{ox1}}} \right)}{\left(1 + \frac{[\text{NADPH}]}{K_{\text{nadph2}}} \right) \left(1 + \frac{[\text{OX}]}{K_{\text{ox2}}} \right)}.$$

Values of the constants in the model were adapted to experimental curves: $k_f = 566 \text{ s}^{-1}$; $K_{m_{\text{pyr}}} = 4.2 \text{ mM}$; $K_{\text{md}} = 40 \text{ mM}$; $K_{\text{pyr1}} = 8 \text{ mM}$; $K_{\text{pyr2}} = 4 \text{ mM}$; $K_{\text{nadph1}} = 1 \text{ mM}$; $K_{\text{nadph2}} = 0.083 \text{ mM}$; $K_{\text{ox1}} = 100 \text{ mM}$; $K_{\text{ox2}} = 17.65 \text{ mM}$; $K_{\text{atp1}} = 1 \text{ mM}$; $K_{\text{atp2}} = 0.513 \text{ mM}$; $K_{\text{akb1}} = 1 \text{ mM}$; $K_{\text{akb2}} = 0.667 \text{ mM}$; $K_{\text{akb3}} = 1 \text{ mM}$; $K_{\text{akb4}} = 0.667 \text{ mM}$; $n_p = 2$.

The equation describes the dependence of the enzymatic reaction rate on concentrations of regulators (AKB, OX, and ATP), substrates (PYR and NADH), and products (LAC and NADP). The regulatory effects are complex and nonlinear. Addition of AKB reduces the Hill nonlinearity degree (term $n_p \cdot f_1(\text{AKB})$) of the substrate effect of PYR on LDHA activity from 2 to 1, simultaneously reducing the level of the Michaelis constant K_m for pyruvate (term $K_{m_{\text{PYR}}} \cdot f_2(\text{AKB})$). In turn, PYR and ATP alter the K_m value for NADPH ($K_{m_{\text{NADPH}}} \cdot f(\text{PYR,ATP})$), and OX alters V_{\max} ($e_0 \cdot k_f \cdot f(\text{NADPH,OX})$).

Comparison of the results of the prediction according to the model and experimental data is shown in Fig. 2.

In our further work, we will study the integral pyruvate metabolism model with regard to genetic regulation.

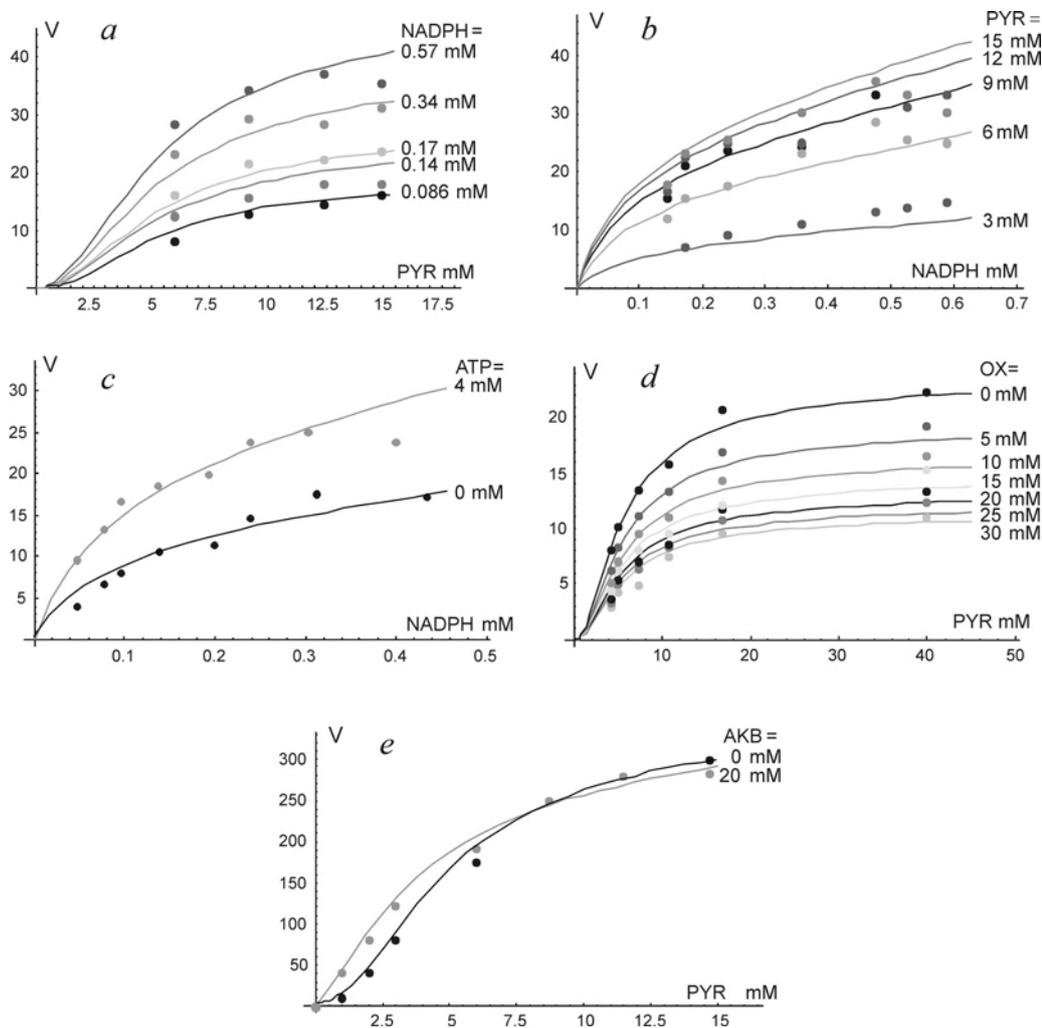


Figure 2. Y-axis: LDHA activity, μM NADPH oxidation per min at pH 7.5 (V); X-axis: regulator concentrations, mM. Panel a. Variation of V with PYR concentration at various NADPH concentrations. Panel b. Variation of V with NADP concentration at various PYR concentrations. Panel c. Variation of V with NADPH concentration with and without 4 mM ATP, PYR concentration equaling 4 mM. Panel d. Variation of V with PYR concentration at various OX concentrations, NADPH concentration equaling 0.33 mM. Panel e. Variation of V with PYR concentration with and without 1 mM AKB, NADPH concentration equaling 0.33 mM. Experimental data obtained by Tarmy and Kaplan (1968) were used for testing the model of LDHA activity regulation.

ACKNOWLEDGEMENTS

The authors are grateful to Vitaly Likhoshvai and Alexander Ratushny for valuable discussions and to Irina Lokhova for bibliographical support. Work was supported in part

by the Russian Foundation for Basic Research (Grant No. 04-01-00458), 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

- Ananko E.A., Podkolodny N.L., Stepanenko I.L., Podkolodnaya O.A., Rasskazov D.A., Miginsky D.S., Likhoshvai V.A., Ratushny A.V., Podkolodnaya N.N., Kolchanov N.A. (2005) GeneNet in 2005. *Nucl. Acids Res.*, **33**, D425–D427.
- Jiang G.R., Nikolova S., Clark D.P. (2001) Regulation of the *ldhA* gene, encoding the fermentative lactate dehydrogenase of *Escherichia coli*. *Microbiology*, **147**, 2437–2446.
- Khlebodarova T.M. *et al.* (2006) *In silico* cell II. Information sources for modeling *Escherichia coli* metabolic pathways. *This issue*.
- Likhoshvai V.A., Ratushny A.V. (2006) *In silico* cell I. Hierarchical approach and generalized hill functions in modeling enzymatic reactions and gene expression regulation. *This issue*.
- Tarmy E.M., Kaplan N.O. (1968) Kinetics of *Escherichia coli* B D-lactate dehydrogenase and evidence for pyruvate-controlled change in conformation. *J. Biol. Chem.*, **243**, 2587–2596.

BiotechPro: A DATABASE FOR MICROBIOLOGICALLY SYNTHESIZED PRODUCTS OF BIOTECHNOLOGICAL VALUE

*Ibragimova S.S.**, *Smirnova O.G.*, *Grigorovich D.A.*, *Khlebodarova T.M.*

Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia

* Corresponding author: isola@bionet.nsc.ru

Key words: database, microbial organism, industrially important products, microbial biotechnology, strain improvement

SUMMARY

Motivation: Development of bacterial strains producing biotechnologically important substances is a complex and many-staged process. For a better progress in the area, a knowledge base of biotechnologically valuable molecular-genetic bacterial systems and processes is required. This knowledge base should contain information on the organisms, which are the sources of a particular target product; on its biosynthesis pathways; on the genes whose products are involved in the metabolic pathways; and on the constants and dynamic variables of a particular process to enable a computer simulation of the process under various environmental conditions, given a desired yield of the target product. Eventually, on the basis of these data, the knowledge base should help scientists develop a gene engineering strategy for modification of a particular strain.

Results: We have developed a database of industrially important products (BiotechPro) produced by microbes. The database includes information about species producing a particular product, the metabolic pathway or functional components of the operon regulating this pathway, industrial strains, and associated substrates for each strain. BiotechPro is aimed to provide information for applied biotechnology and strain optimization projects.

Availability: the database is available at <http://wwwmgs.bionet.nsc.ru/mgs/dbases/biotech/>

INTRODUCTION

Many products of bacterial cellular metabolism possess physiological activity and are of practical value for various industries. Advancements in molecular biology and recombinant DNA technology allow the number of genes for a particular bacterial product and the level of its expression in that particular strain to be enhanced as desired (Stephanopoulos, 2002). To handle such tasks, informational support is required in the form of computer analysis of experimental data and computer simulation of the processes involved. Information on the metabolites synthesized by various bacterial species, on the strains in which a particular species can appear and on the molecular-genetic systems controlling its biosynthesis is scattered across multiple publications, each addressing but few aspects of the synthesis of the target product.

A specialized database compiling the information about biotechnologically important products, synthesized by microorganisms, their biosynthesis pathways, and source strains used in biotechnologies may be used for solving this problem.

The primary navigator in the field is “The Prokaryotes”, an electronic resource (<http://141.150.157.117:8080/prokPUB/index.htm>), which can be used as an external source of data on the most biotechnologically important products produced by microorganisms. The database contains information largely on industrial source strains that have extensively been used in biotechnology. BSD databases: the Biodegradative Strain Database and the Biocatalysis/Biodegradation Database are narrowly focused databases containing information on strains producing enzymes that utilize hazardous chemicals (Urbance *et al.*, 2003, Ellis *et al.*, 2006).

We have developed a database of industrially important products (BiotechPro) produced by microbes: we have since been collecting information about the species that are sources of a particular product, the metabolic pathway or functional component of the operon that regulates the corresponding pathway, industrial strains and associated substrates for each strain. Information of industrial products, strains and substrates is collected from scientific literature.

BiotechPro consists of two databases BiotechProduct and BiotechStrain cross-linked with one another. This provides a possibility to select a potential bacterial producer of a biotechnologically important product and information on how much of the product the strain can synthesize and on the associated substrate for that strain, with regard to data on its biosynthetic pathway.

BiotechPro will contain consolidated data on microorganisms whose products are or could be used in different biotechnological processes.

IMPLEMENTATION AND RESULTS

BiotechPro database is implemented on the SRS platform and consists of 2 subdatabases: database on product and database on strain. Both databases have cross-reference fields.

Structure and description of product database. This database compiles information of industrially important products produced by different microbial species, which are or could be used in biotechnology (Fig. 1a).

The PRODUCT field contains the full and short names of the product. The SPECIES field contains the Latin names of the organisms that synthesize this product. The PATHWAY field contains information on the metabolic pathway for the target product with the different references to the KEGG, MetaCyc databases (if the pathway is known) or the functional components involved in the synthesis of the target product (Caspi *et al.*, 2006; Kanehisa *et al.*, 2006). The field COMMENT details the biosynthetic pathways of the target product. The APPLICATION field contains information about biotechnological processes in which the product could be used. BiotechProduct subdatabases contain two cross-reference fields – PRODUCT_ID and STRAIN_ID to subdatabases BiotechStrain.

BiotechProduct database has links to databases containing additional information on product:

- GeneNet, a database that contains descriptions of the local gene networks and biosynthetic signal transduction pathways generating the target product (Ananko *et al.*, 2005);
- ProkaTEX, a database that is developed on the background of EcoTRRD, contains structural descriptions of the promoters of prokaryotic genes (including *E. coli*) whose products are involved in biotechnological processes (Khlebodarova *et al.*, this issue);
- KiNet, a database that contains the descriptions of the constants and dynamic variables required for modeling the process as a whole and in part (Khlebodarova *et al.*, this issue);
- ModelER, a database that contains computerized models for the functioning of the molecular-genetic and metabolic processes being investigated in different environments for different source organisms (unpublished data).

(A) PRODUCT

PRODUCT_ID glutamate
 PRODUCT L-glutamate, L-glutamic acid, glutamate
 DATE 18.04.2006
 AUTHOR Ibragimova S.
 SPECIES *Brevibacterium lactofermentum*
 PATHWAY Glutamate metabolism (KEGG), Glutamate biosynthesis (MetaCyc)
 COMMENT Production of large quantities of glutamate is observed when the bacteria are cultured in media containing a) limiting amounts ordinarily 2-5 mkg/l) of biotin, b) sub-lethal concentrations of penicillin or c) sub-optimal amounts of surface-active agents (surfactant).
 APPLICATION aminoacid production, L-glutamate production, food industry, pharmaceutical industry
 KEYWORD aminoacid, glutamate, monosodium glutamate (MSG)
 STRAIN_ID *B.lactofermentum*_glutamate
 Link_GeneNet: Glutamate and glutamine biosynthesis
 Link_ProkaTEX: PK01762 PK03212 PK03213 PK03214
 Link_KiNet: Glutamate and glutamine biosynthesis
 END

(B) STRAIN

STRAIN_ID *B.lactofermentum*_glutamate
 STRAIN *Brevibacterium lactofermentum* ATCC13869
 PRODUCTIVITY Glutamate production, approximately 230 mM/l, was observed in cultures of 3 mkg/l of biotin. In the PESP-added culture, total glutamate production is 105 mM/l.
 PRODUCT_ID glutamate
 REFERENCE Kawahara Y., Takahashi -Fuke K., Shimizu E., Nakamatsu T., Nakamori S. Relationship between the glutamate production and the activity of 2-oxoglutarate dehydrogenase in *Brevibacterium lactofermentum*. *Biosci. Biotechnol. Biochem.*, 1997, 61, 1109-1112.
 PMID 9255973
 END

Figure 1. Example of entries of BiotechPro database: (a) – product description, (b) – strain description.

Structure and description of strain database. This database accumulates information about strains that produce industrially important products. The strain description contains 3 fields. Each strain has an individual strain data page (Fig. 1b).

The STRAIN field contains the Latin name of the strain, the strain designation and a brief description of the origin of the strain. The PRODUCTIVITY field contains information on how much of the product the strain can synthesize and on the associated substrate for that strain. The reference stating that the strain produced this product, or degrades or transformed this substrate is included in the REFERENCE field. References are also provided as links to PubMed abstracts when available. BiotechStrain subdatabases contain cross-reference field to subdatabases BiotechProduct: – PRODUCT_ID.

The BiotechPro database is to be populated with data on the strains that synthesize products, which are or could be used in biotechnology. With BiotechPro, the biotechnologists can pick out source strains either to address particular biotechnological tasks or to modify the strain as deemed desirable.

SRS tools allow database fields to be indexed and searched using a query system.

The main field for the search of the subdatabase BiotechProduct is the PRODUCT field. The search by the field SPECIES provides information on the organisms that synthesize the product of your interest. The search by the field APPLICATION allows you to find the strains for particular tasks.

The fields STRAIN and PRODUCTIVITY are the main fields for search in BiotechStrain subdatabases. The search by the field PRODUCTIVITY allows one to find the references to the microorganisms that utilize a particular substrate.

At present, BiotechPro contains information on 50 industrially important products produced by microbes. It also contains information on more than 60 associated strains that are sources of amino acids, vitamins, food additives, industrially important thickeners and enzymes.

We plan to expand the content of BiotechPro in future.

ACKNOWLEDGEMENTS

The work was supported by the Siberian Branch of the Russian Academy of Sciences (Integration Project on Fundamental Research of SB RAS No. 24 “Role of microorganisms in live system functioning: fundamental aspects and bioengineering applications”) and by Innovation project of Federal Agency of Science and Innovation IT-CP.5/001 “Development of software for computer modeling and design in postgenomic system biology (system biology *in silico*)”. The authors are grateful to I.V. Likhova for bibliographical support.

REFERENCES

- Ananko E.A. *et al.* (2005) GeneNet in 2005. *Nucl. Acids Res.*, **33**, D425–D427.
- Caspi R. *et al.* (2006) MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucl. Acids Res.*, **34**, D511–D516.
- Ellis L.B.M. *et al.* (2006) The university of minnesota biocatalysis/biodegradation database: the first decade. *Nucl. Acids Res.*, **34**, D517–D521.
- Kanehisa M. *et al.* (2006) From genomics to chemical genomics: new developments in KEGG. *Nucl. Acids Res.*, **34**, D354–D357.
- Khlebodarova T.M. *et al.* (2006) *In silico* cell II. Information sources for modeling *Escherichia coli* metabolic pathways. *This issue*.
- Stephanopoulos G. (2002) Metabolic engineering by genome shuffling. *Nat. Biotechnol.*, **20**, 666–668.
- Urbance J.W. *et al.* (2003) BSD: the Biodegradative Strain Database. *Nucl. Acids Res.*, **31**, 152–155.

DATABASE GenSensor AS INFORMATIONAL SOURCE FOR DESIGN OF BIOSENSORS. EXPERIMENTAL DEVELOPMENT OF BIOSENSOR BASED ON *yfiA* GENE

Khlebodarova T.M.**, *Kachko A.V.*, *Stepanenko I.L.*, *Tikunova N.V.

Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia

* Corresponding author: e-mail: tamara@bionet.nsc.ru

Key words: database, biosensor, transcription regulation, environmental stress

SUMMARY

Motivation: Search for the genes-candidates aimed at designing testers of metabolic cell fitness.

Results: The database GenSensor has been developed for accumulation of information about sensitivity of prokaryotic genes to external stimuli. On the basis of information collected in the database GenSensor, we have developed and designed a biosensor construction serving for registration in the test environment of the agents disrupting not only DNA, RNA, and protein structure, but also physiological homeostasis of a cell as a whole. This construction is based on *yfiA* gene promoter of *Escherichia coli*. The sensitivity of the construction designed has been tested in the model experiments with the presence of hydrogen peroxide.

Availability: the database is available in a form of a flat-file that could be accessible on the basis of collaboration via kol@bionet.nsc.ru.

INTRODUCTION

Purity of water, air, and food substances becomes more and more important for the human beings. For ecological purity testing, we need to develop novel effective and productive methods that would be able to detect the substances toxic for the human organism in biological mediums, even though the information about their nature is absent. Designing of bacterial biosensors and microfluid systems for the registration of their signals has produced such a possibility (D'Souza, 2001; Groisman *et al.*, 2005). Usually, biosensors are based on an ability of cells to respond to the toxic impact by activation of its damaged repair systems. Under the presence of the agents the cascades of genes are activated in a cell disrupting the cell structure and function (Zheng *et al.*, 2001; Au *et al.*, 2005, etc). In particular, promoters of such genes may serve as a sensible unit of biosensor construction. Although efficacy of a biosensor as the testing system is determined by the whole integrity of the system's components (Hakkila *et al.*, 2002; Harkins *et al.*, 2004), the level of gene-sensor induction and its sensitivity to damaging stimuli compose two major factors of biosensor design. The higher are the values of these two parameters under other equal conditions, the simpler is to register the SOS response of a cell. So, descriptions of these parameters compose the main body of the novel database GenSensor. The necessity to develop this database has appeared in the process of searching for the genes-candidates for highly sensitive biosensors of metabolic destruction of a cell.

METHODS AND ALGORITHMS

Plasmid construction. Intermediate plasmid p-GFP containing a reporter gene encoding an intermediate stability variant of green fluorescent protein (*gfp-vaa*) was constructed at the first stage. DNA fragment encoding *gfp*-gene was obtained by PCR using pZE21 plasmid (Elowitz, Leibler, 2000) as a template. The primers, used for PCR included additional nucleotides containing *KpnI* and *HindIII* restriction sites. Following digestion with *KpnI* and *HindIII*, the PCR fragment was ligated to *KpnI*- and *HindIII*-digested pRS2 plasmid (Frolov *et al.*, 1992). p-GFP plasmid contains a promoterless *gfp*-gene flanked with *SphI* restriction site at the very 5'-end. The *yfiA* promoter was obtained by PCR amplification of a 288-bp fragment, using chromosomal DNA purified from *Escherichia coli* JM109 cells as a template. The primers used were 5'-CCTCGGGCCCCAGAAACCTGAAACACAAAACGG-3' and 5'-AAGGGCATGCATAAATTTACCTCTTGCTTCCCGTC-3'; these primers included additional nucleotides containing *ApaI* and *SphI* restriction sites, respectively. After digestion with *ApaI* and *SphI*, a 270 bp DNA fragment was ligated to *ApaI*- and *SphI*-digested p-GFP plasmid resulting in pYfi-GFP* plasmid. To restore first ATG, pYfi-GFP* plasmid was digested with *SphI* and Klenow fragment of polymerase I. The resulting plasmid pYfi-GFP contains a reporter gene *gfp* under transcriptional regulation of the *yfiA* promoter. The sequence of the insertion was verified by sequencing of the promoter region.

Strains and grown conditions. To assay for fluorescence, *E. coli* MC4100 cells, transformed with pYfi-GFP, grown overnight in LB medium with 50 µg/ml ampicillin were diluted in fresh LB medium and grown to mid-log phase. Then cells were washed with minimal media, and a 50-µl portion of culture was added to each well of a microtiter plate containing a series of twofold dilution of hydrogen peroxide in 50 µl of minimal media. Fluorescence was monitored by Perkin Elmer VICTOR³ fluorometer (0.1s, 485nm/535nm).

RESULTS AND DISCUSSION

Structure and the content of the GenSensor database

The database GenSensor is developed for accumulation of information that is necessary for designing of biosensor constructions on the basis of bacterial genes. It contains the data about (i) the structure of bacterial promoters that are activated and expressed in response to external stimuli, (ii) mechanisms of induction; and (iii) conditions that provoke the maximal response to the given type of stimulus. Importantly, the database accumulates only the experimentally confirmed data. The source of information for the database is annotation of scientific literature data. The entry of the database is a gene. Its description includes: (1) general information (name of a gene; its synonyms; name of an operon, to which the gene belongs; its localization in the chromosome; unique gene number in bacterial genome; links to the database GenBank/EMBL); (2) description of the gene promoters (boundaries of promoter, transcription starts; promoter sequences; names of transcription factors (TFs) interacting with promoter and their influence on the gene expression; (3) inducers activating promoter, as well as conditions in which promoter is being activated; (4) TFs, promoters, and sites responsible for activation or de-repression of a gene by means of this or that inducer; and (5) description of TF binding with promoter (names, synonyms, species origin, and active form).

The structure of the database entry reflects organization of regulatory regions of bacterial genes. It contains 61 different fields that are united into 6 informational blocks: general gene description contains 17 fields; description of promoters – 9 fields; description of inducers of gene expression – 11 fields; description of transcription factors – 8 fields; description of transcription factor binding sites – 8 fields; and the block of fields containing information about literary source – 8 fields.

Currently, the database GenSensor contains description of 86 promoters of *Escherichia coli*, *Helicobacter pylori*, and *Bacillus subtilis*, expression of which is essentially dependable upon different stress impacts causing the damage of DNA, RNA, proteins, and cell membranes. The influence of these factors on the level of induction of the genes described in the database is reflected by 187 expression patterns.

Although a set of other databases containing the information about promoter structure in bacteria are available (BioCyc, Regulon, etc.), the database GenSensor is a unique resource, because it additionally contains the information about sensitivity of gene promoters to inducers and about the levels of their induction in dependence upon the strength of the impact.

Development of the biosensor construction on the basis of the *yfiA* (*RaiA*) gene promoter

A considerable interest for the development of novel, highly sensitive biosensors have the genes with complex regulation, which respond to multiple stimuli. Usage of such genes in biosensor constructions will enable to design universal testers that could be activated in response to different aspects of unfavorable cell metabolism. An illustrative example of such sensors, from our viewpoint, could be the *yfiA* gene promoter. The exact mechanisms of *yfiA* gene activation are still not known. However, existing data about the function of this gene in a cell (Agafonov *et al.*, 2001), as well as its reaction to external conditions enable us to consider this gene as perspective one for designing of a sensor, which would register different agents damaging DNA, RNA, and proteins in the environment (Zheng *et al.*, 2001; Pomposiello *et al.*, 2001; Khil, Camerini-Otero, 2002). Besides, it could be sensible for violation of optimal conditions for the cell growth, thus, serving as a sensor of disruption of the general metabolic cell fitness (Agafonov *et al.*, 2001).

To develop the biosensor, pYfi-GFPvaa plasmid containing a reporter gene *gfp* under transcriptional regulation of the *yfiA* promoter was constructed. The response of *E. coli* MC4100 cells, transformed with pYfi-GFP plasmid, to different hydrogen peroxide concentrations at 25 °C is shown in Fig. 1.

A lag phase around 20 min long was followed by an increase in fluorescence, which peaked after approximately 70 min and then declined. The response was dose dependent, with concentration of more than 4 mM resulting in inhibition of fluorescence (data not shown). Approximately 4-fold induction of fluorescence level was observed at 25 °C, while 30-fold induction was detected at 37 °C (data not shown). This level is compatible with the induction level of biosensor designed on the *cda* gene promoter (Norman *et al.*, 2005), which is highly sensitive for oxidative stress. Further experiments are necessary to study the possibilities of the biosensor based on *yfiA* promoter, and some approaches for sensitivity increasing could be used. So, a novel biosensor, sensitive for oxidative stress, was designed as a result of first implementation of the GenSensor database.

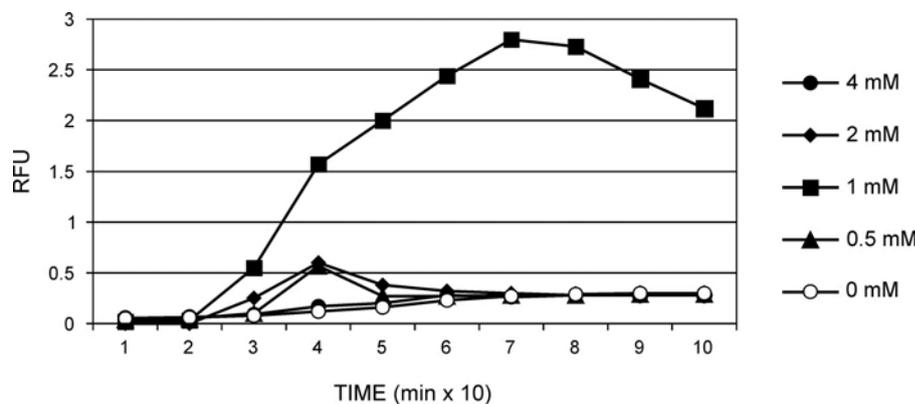


Figure 1. Effect of H₂O₂ on fluorescence of *E. coli* MC4100 cells, transformed with pYfi-GFP.

ACKNOWLEDGEMENTS

The authors are grateful to I.V. Likhova for bibliographical support and to G.V. Orlova for translation of the paper into English. The work was supported by the Russian Federal Agency for Science and Innovations (State contract No. 02.467.11.1005) and by Siberian Branch of the Russian Academy of Sciences (Integrational project No. 24).

REFERENCES

- Agafonov D.E. *et al.* (2001) Ribosome-associated protein that inhibits translation at the aminoacyl-tRNA binding stage. *EMBO Rep.*, **2**, 399–402.
- Au N. *et al.* (2005) Genetic composition of the *Bacillus subtilis* SOS system. *J. Bacteriol.*, **187**, 7655–7666.
- Elowitz M., Leibler S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature*, **403**, 335–338.
- Frolov I.V. *et al.* (1992) Study of the prospects for creating vector systems based on the alpha-virus genome. *Dokl. Akad. Nauk*, **326**, 738–741.
- Groisman A. *et al.* (2005) A microfluidic chemostat for experiments with bacterial and yeast cells. *Nat. Methods*, **2**, 685–689.
- D'Souza S.F. (2001) Microbial biosensors. *Biosens. Bioelectron.*, **16**, 337–353.
- Hakkila K. *et al.* (2002) Reporter genes lucFF, luxCDABE, gfp, and dsred have different characteristics in whole-cell bacterial sensors. *Anal. Biochem.*, **301**, 235–242.
- Harkins M. *et al.* (2004) The role of host organism, transcriptional switches and reporter mechanisms in the performance of Hg-induced biosensors. *J. Appl. Microbiol.*, **97**, 1192–1200.
- Khil P.P., Camerini-Otero R.D. (2002) Over 1000 genes are involved in the DNA damage response of *Escherichia coli*. *Mol. Microbiol.*, **44**, 89–105.
- Norman A. *et al.* (2005) Construction of a ColD cda promoter-based SOS-green fluorescent protein whole-cell biosensor with higher sensitivity toward genotoxic compounds than constructs based on recA, umuDC, or sulA promoters. *Appl. Environ. Microbiol.*, **71**, 2338–2346
- Pomposiello P.J. *et al.* (2001) Genome-wide transcriptional profiling of the *Escherichia coli* responses to superoxide stress and sodium salicylate. *J. Bacteriol.*, **183**, 3890–3902.
- Zheng M. *et al.* (2001) DNA microarray-mediated transcriptional profiling of the *Escherichia coli* response to hydrogen peroxide. *J. Bacteriol.*, **183**, 4562–4570.

ABOUT RECONSTRUCTION OF REGULATORY MECHANISM OF GENE ELEMENTS' EXPRESSION

Peshkov I.M.¹, Fadeev S.I.²

¹Novosibirsk State University, Novosibirsk, 630090, Russia; ²Sobolev Institute of Mathematics, SB RAS, Novosibirsk, 630090, Russia

*Corresponding author: e-mail: likho@bionet.nsc.ru

Key words: gene network, inverse problem, autonomous system

SUMMARY

Motivation: The core of gene networks is regulatory circuits – a set of genetic elements whose expression is subjected to mutual regulation. Discovering the laws defining this regulation is an actual problem of mathematical biology.

Results: The method of solving the inverse problem for systems of differential autonomous equations is proposed. These systems emerge as mathematical models of artificial gene networks that aimed for study of mechanism of promoters' activity regulation. In the work the case of periodic dynamics of changes of measured in experiment products concentration is taken up. The minimum of residual functional is searched by the method of quickest descent. The method takes into account the period T , which in autonomous case as the solutions depends on the parameters of problem.

INTRODUCTION

In the work the gene construction consisting of two operons A and B is taken up (Fig. 1).

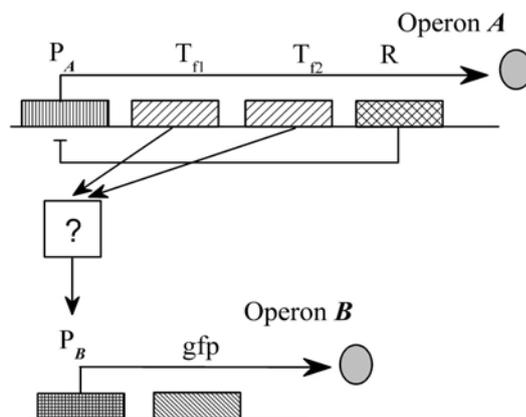


Figure 1. Fundamental scheme of the two-operon system for analysis of promoters functioning mechanism.

The self-repressing operon A contains gene R which encodes the repressor of activity of its own promoter P_A and l genes. These genes encode transcriptional factors (on the Fig. 1 $l = 2$, genes T_{f1} and T_{f2}) that regulate the promoter P_B activity of the operon B . The mechanism of action of transcription factors is unknown and must be determined. The

operon B contains reporter gene gfp which expression effectiveness can be measured in experiment. The mathematical model of this gene network is presented by the next system of differential autonomous equations

$$\left(\begin{array}{l} \frac{dz_1}{dt} = g(z_{k_z}) - \frac{k_e}{n_{z,1}} z_1, \\ \frac{dz_i}{dt} = \frac{k_e}{n_{z,1}} (z_{i-1} - z_i), \quad i = \overline{2, k_z - 1}, \\ \frac{dz_{k_z}}{dt} = \frac{k_e}{n_{z,1}} z_{k_z - 1} - \delta_z z_{k_z} \end{array} \right) \cdot \Delta_z, \\ \left(\begin{array}{l} \frac{dy_{j,1}}{dt} = g(z_{k_z}) - \frac{k_e}{n_{z,1}} y_{j,1}, \\ \frac{dy_{j,i}}{dt} = \frac{k_e}{n_{z,1}} (y_{j,i-1} - y_{j,i}), \quad i = \overline{2, k_{yj} - 1}, \\ \frac{dy_{j,k_{yj}}}{dt} = \frac{k_e}{n_{z,1}} y_{j,k_{yj} - 1} - \delta_{yj} y_{j,k_{yj}} \end{array} \right) \cdot \Delta_j, \quad j = \overline{1, l}, \\ \left(\begin{array}{l} \frac{dx_1}{dt} = f(y_{1,k_{y1}}, \dots, y_{l,k_{yl}}, p) - \frac{k_e}{n_{x,1}} x_1, \\ \frac{dx_i}{dt} = \frac{k_e}{n_{x,1}} (x_{i-1} - x_i), \quad i = \overline{2, k_x - 1}, \\ \frac{dx_{k_x}}{dt} = \frac{k_e}{n_{x,1}} x_{k_x - 1} - \delta_x x_{k_x}. \end{array} \right) \cdot \Delta_x \quad (1)$$

Here, k_e is the constant of average velocity of generalized process of elongation; $n_{z,1}, n_{x,1}$ is the length of elementary stages of elongation on which operons A and B are divided; k_z, k_{yj}, k_x is the quantity of elementary stages in model which is necessary for the corresponding protein synthesis after the initiation of synthesis; $\delta_z, \delta_x, \delta_{yj}, j = 1, \dots, l$ is the constants of utilization velocities of final products; z_{k_z} is the concentration of free active inhibitor's form of promoter's transcription P_A ; x_{k_x} is the concentration of reporter protein; $y_{j,k_{yj}}, j = \overline{1, l}$ is the concentration of active form of transcription factors that regulate the activity of promoter P_B ; $x_1, \dots, x_{k_x - 1}, z_1, \dots, z_{k_z - 1}, y_{j,1}, \dots, y_{k_{yj} - 1}, j = 1, \dots, l$ is the concentrations of the proteins on elementary stages of synthesis. The $\Delta_x, \Delta_z, \Delta_j, j = \overline{1, l}$ is the logical parameters that possess the value 0 or 1. The subsystem will absent in system (1) if the corresponding parameter is equal to 0. The functions $g(z_{k_z}): \mathbb{R}^1 \rightarrow \mathbb{R}^1$ and

$f(y_{1,k_{y_1}}, \dots, y_{l,k_{y_l}}, p) : (\mathbb{R}^l \times \mathbb{R}^m) \rightarrow \mathbb{R}^1$ initialize the regulation mechanism of promoters

P_A and P_B activity, correspondingly; $p \in \mathbb{R}^m$ – the vector of the parameters.

The problem of reconstruction of the functions f in the class of generalized Hill's function (Likhoshvai *et al.*, 2006) under the “experimental” data is set. The data correspond to dynamics of reporter gene expression and can be got by the series of models (1) for some set of the parameters $\Delta_z, \Delta_j, j = \overline{1, l}$ and k_z, k_{y_j}, k_x .

In this work the special case of periodic functioning of operon A is considered. The operon contains the monocistron that encodes self-repressor. The problem of searching of the parameters p of the function $f(y_{1,k_{y_1}}, \dots, y_{l,k_{y_l}}, p)$ is solved. The shape of f is known.

For solving the problem the results of calculation of the model with known value of the parameters are took as “experimental” data.

METHODS AND ALGORITHMS

For calculus the parameters the method of quickest descent is used. The residual

functional is taken in the shape $\Phi(p) = \sqrt{\sum_{j=1}^m (x_{k_x}(t_j, p) - M_j)^2 + (T(p) - T_e)^2}$. Here

T_e is the period of “experimental” data, $t_j \in [0, T_e]$. The period $T(p)$ of model solution is determined by the solving of boundary value problem. The boundary value problem is solved by the multiply shooting method.

RESULTS

The special case of model (1)

$$\frac{dz_1}{dt} = \frac{\alpha}{1 + (z_n/\theta)^\gamma} - \frac{k_e}{n_z} z_1; \quad \frac{dz_i}{dt} = \frac{k_e}{n_z} (z_{i-1} - z_i); \quad \frac{dz_n}{dt} = \frac{k_e}{n_z} z_{n-1} - \delta z_n; \quad i = \overline{2, n-1}, \quad (2)$$

describes the dynamics of self-repressing cistron functioning. The values of the constants k_e, n_z, n, δ are initialized. It is required to determine the values of the parameters α, γ, θ using changes of concentration z_n . The solution of the initial value problem of the system (2) with initial value on periodic solution and known value of the parameters $\alpha, \gamma, \theta, k_e, n_z, n, \delta$ (Fig. 2 and 3) is taken as the “experimental” data (experiment:

$\alpha = 3.1; \quad \gamma = 3.5; \quad \theta = 1.0; \quad \frac{k_e}{n_z} = 1.0; \quad \delta = 1.0; \quad n = 10;$). Even if the initial

approximation is “good”, the using of the method of quickest descent on the interval $[0, T_e]$ for solving the inverse problem doesn't allow to fit the solution to the “experimental” data so that theoretical curve and “experimental” data will not diverge. With all this going on the values of the parameters essentially differ from the parameters according to the “experiment”. The example of typical solution is shown on Fig. 2. The search of the parameters is realized on the interval $[0, T_e]$, however on the next interval

$[T_e, 2T_e]$ the period and amplitude of the theoretical curve and “experimental” data differ. The using of the boundary value problem for search periodic solutions and determine their period allows to raise the reliability of the method of quickest descent (Fig. 3).

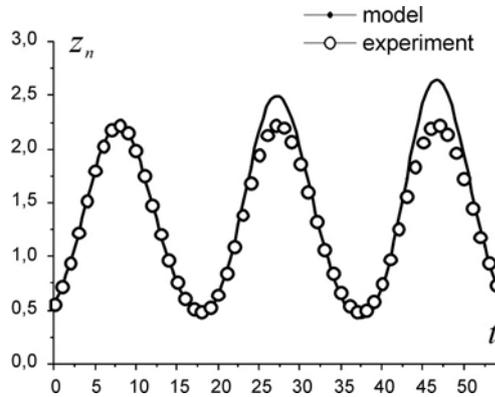


Figure 2. Founded parameters: $\alpha = 2.73$; $\gamma = 4.01$; $\theta = 0.86$.

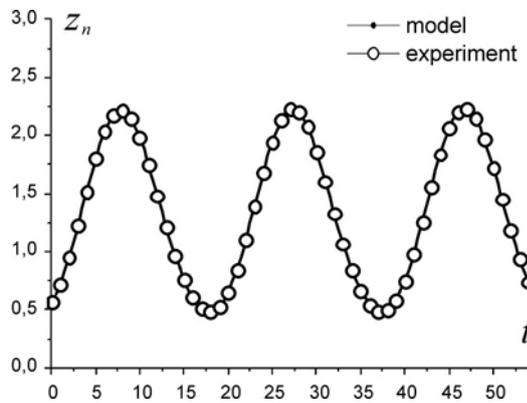


Figure 3. Founded parameters: $\alpha = 3.09$; $\gamma = 3.50$; $\theta = 0.99$.

DISCUSSION

This work is the part of actual theoretical problem of reconstruction of genetic elements regulation mechanism. One of the perspective ways of solving this problem is working out the universal genetic system whose construction will allow to get enough experimental information for reconstruction of regulatory mechanism. It is appropriate offer different type of such genetic construction. We choose the construction consisting of two operons (Fig. 1). The first operon is a self-repressing operon. Its structure and parameters are chosen so that periodic dynamics of concentration of protein which are contained in this promoter take place. The second operon consists of a target promoter whose regulatory mechanism is subject to determination and also reporter gene whose expression level can be easily measured. It is supposed that the construction of the first operon allows to place in it the genes coding the transcription regulator of the target promoter. As the first operon has a periodic regime of functioning, the concentrations of the transcription factor have periodic dynamics too. Therefore the second operon has oscillating dynamics of expression of the reporter gene. The problem of reconstruction of

target promoter regulator mechanism can be solved by studying this dynamics. In present work we suggest the model (1) of bicistron genetic system and give the general formulation of the problem. We studied the simplest case (2) of the inverse problem and suggest the modification of the quickest descent method. In future we suppose to construct two-operon genetic systems suggested above and to check up this model in experiment *in vivo*.

ACKNOWLEDGEMENTS

Work was supported in part by the Russian Foundation for Basic Research (No. 06-04-49556). The authors are grateful to Vitaly Likhoshvai for valuable discussions.

REFERENCES

Likhoshvai V.A., Ratushny A.V. (2006) *In silico* cell I. Hierarchical approach and generalized hill functions in modeling enzymatic reactions and gene expression regulation. *This issue*.



3.2. MODELLING OF MOLECULAR GENETIC SYSTEMS AND PROCESSES IN MULTICELLULAR ORGANISMS

A MATHEMATICAL MODEL FOR THE INFLUENZA VIRUS LIFE CYCLE

Akberdin I.R.^{1}, Kashevarova N.A.², Khlebodarova T.M.¹, Bazhan S.I.³,
Likhoshvai V.A.¹*

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ² Novosibirsk State University, Novosibirsk, 630090, Russia; ³ State Research Center of Virology and Biotechnology “Vector”, Koltsovo, 633159, Russia

* Corresponding author: e-mail: akberdin@bionet.nsc.ru

Key words: mathematical modeling, gene network, Influenza virus, intracellular ontogeny

SUMMARY

Motivation: The Influenza virus (Vinf) is a widespread infectious agent causing contagious disease in humans and animals. Its prominent feature is high and functional genetic variability. The mechanisms underlying virus give rise to strains possessing novel unpredictable properties. Their analysis and reliable prediction are the tasks of basic and applied biogenetics and virology. The pressing question is: What may be the replication pattern of the Vinf within the infected host cell? There is, as yet, no clarity. This is a stumbling block for advancement in this research area.

Results: Here, we have reconstructed the gene network and develop a mathematical model for the Vinf. They incorporate the processes of replication, transcription, translation, regulation them and assembly within the Vinf infected cell. Based on the mathematical model, we followed the dynamics of the developed gene network and analyzed the key regulatory stages of the Vinf replication.

Availability: The schematic presentation of the gene network “Influenza (life cycle)” is available through the GeneNet viewer at <http://www.mgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/index.shtml>.

INTRODUCTION

The Vinf is a member of the orthomyxovirus family. Its various strains induce infection involving the respiratory tract, in human and animals, provoking epidemics or pandemics. Because of its pernicious variability, the Vinf passes through the defensive barrier, and there appear to be no natural means to combat the Vinf infection. The mortality tolls of the periodical surges of Vinf epidemics are alarmingly high. Recently, in association with the emergence of the avian influenza, the H5N1 viral strain, and its hazardous transmission to human populations, (Claas *et al.*, 1998), overcoming of the stumbling block became an emergency. In fact, the H5N1 viral strain, most likely arisen in south-eastern Asia, has conquered the interspecific barrier and spread widely through 1996–2003 among avian species in Hong Kong and China. As a result, the more aggressive hazardous Vinf genotype attacks humans. The first cases of infection with H5N1 were reported in 2003. The number of humans infected with the highly pathogenic H5N1 keeps rising. Mortality among the H5N1 infected humans was recorded to be as high as 50 % (WHO site // http://www.who.int/csr/disease/avian_influenza/en/). The continual threat of the rise of novel, unpredictably pernicious Vinf strains require untraditional treatment approaches. The problem is that the viruses are able to reproduce

with genetic continuity and possess versatile possibilities of mutations and the joined attacks of the Vinf and H5N1 virus within host cells is another possible hazard. It is hoped that modeling of gene networks for the Vinf life history would serve as a reasonable basis for clarifying certain aspects of Vinf replication within the host cell. This would prompt how to search beyond and above the traditional approaches.

METHODS AND ALGORITHMS

The GeneNet technology was used to reconstruct the gene network development cycle of the Vinf within the host cell (Ananko *et al.*, 2005). The methods, we applied to a single Vinf replication cycle were the generalized chemical kinetic (Likhoshvai *et al.*, 2000) and the one in terms of generalized Hill functions (Likhoshvai *et al.*, 2006, this issue).

RESULTS AND DISCUSSION

Based on GeneNet (Ananko *et al.*, 2005), we reconstructed the sequence of molecular genetic events that unfold during Vinf development within the infected human host cell: from Vinf penetration– to assembly of novel viral particles and their release from the cell. In all, the gene network for the Vinf life cycle contains a description of 42 RNAs, 27 proteins, and 229 interrelations between components. The information has been extracted from 55 scientific papers. The schematic presentation of the gene network is available through the GeneNet viewer (see *Availability*).

Proceeding on the reconstructed sequence of molecular events, we developed a mathematical model for a single Vinf cycle in an Vinf susceptible cell. The model consists of blocks that describe vRNA replication (Mikulasova *et al.*, 2000), primary and secondary transcription, mRNA splicing, viral protein and RNA to and fro trafficking from the nucleus to the cytoplasm (Cros, Palese, 2003). The model describes the formation of defective viral particles, also how the infectious and defective viral particles are formed (Bancroft, Parslow, 2002; McCown, Pekosz, 2005). Moreover, the model contains the variables that control the cellular components that limit Vinf replication, also the ribosomes, the nuclear fraction of total cell mRNA, the surface membrane that are all synthesized and expended during viral RNA transcription and Vinf assembly (budding). The model for the time being does not track how the Vinf penetrates through the cell and spreads itself thereafter. It has been suggested that, at a starting point, the nucleus of the Vinf infected cell contains vRNP fragments and virion associated RNA polymerase consisting of the PB1, PB2 and PA proteins (Deng *et al.*, 2005). In all, the model contains 157 variables, each denotes the concentrations of the viral proteins, RNA and their complexes that form within the Vinf infected cell, the nucleus and cytoplasm at the various stepwise development of the Vinf infection. The implementation mechanism of the distinct developmental processes within the cell is not clear enough. Hence the model required the introduction of additional assumptions. The choice of the appropriate parameter values enabled us to describe the main dynamic characteristics of the Vinf replication within the cell as well as to reasonably explain the major replication mechanisms of the viral genome. There may be alternative amplification mechanisms of certain fragments of viral mRNA for the selective regulation of secondary transcription. Consequently, to reasonably explain how the model might work, a consideration of auxiliary specific amplification mechanisms of mRNA synthesis of distinct fragments is required. To justify the mechanism, we involved data according to which mutations affect Vinf late development. These mutations occur in fragment 8 of genomic RNA that encodes the nonstructural NS1 and NS2 proteins. We assumed that these proteins directly guide the passage from the primary transcription to the early and late secondary transcription. We were aware of experimental data that suggested that selective

gene amplification is specifically regulation at the vRNA level, whereas the synthesis of the respective mRNA and proteins is amplified nonspecifically, i.e. it results from vRNA amplification. It appeared reasonable to assume that the early amplification of fragments 5 and 8 is due to the NS1 action, while the late amplification of fragments 4, 6 and 7 is due to the NS2 effect, because NS2 level becomes detectable only during late the infection. The model reasonably well describes the main system's characteristics only after the introduction of the above mechanisms into the model (Fig. 1).

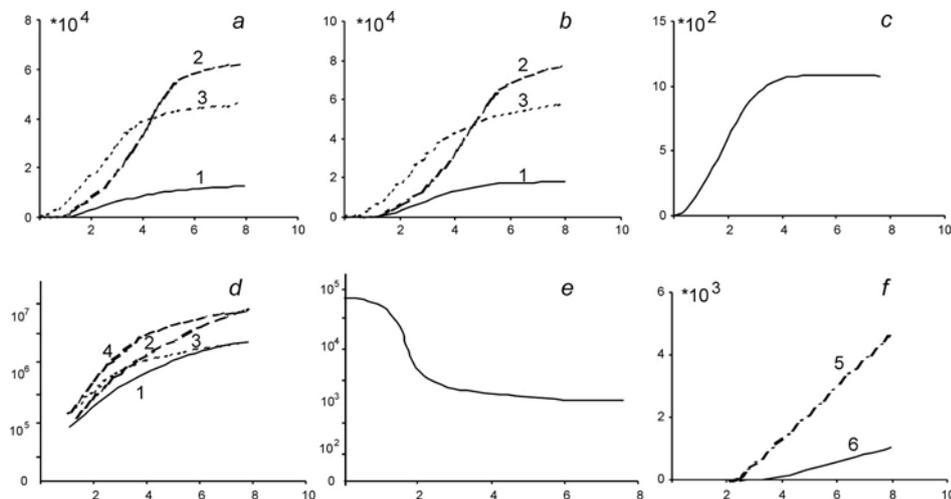


Figure 1. The time of course of changes in the major components of the model during the development of the Vinf within the host cell. a – vRNA; b – mRNA; c – cRNA; d – proteins; e – cell mRNA; f – virus release; 1 – PB1, PB2, PA; 2 – HA (hemagglutynine); 3 – NP (nucleoprotein); 4 – NS1; 5 – infectious virus; 6 – interfering defective viral particles.

Here we analyzed the model's susceptibility to changes in parameters. The model proved to be most susceptible to the constants that characterize the initiation and elongation rates during transcription, replication and translation, also to the constants that define the vRNP formation efficiency. As a rule, the effect of these parameters lies in low value range, and it is monotonous depending on the model's behavior. However, when the coefficients for the RNA synthesis initiation and RNA formation vary, the dependence is not monotonous any longer, this means that there are optimum parameters producing maximum effects. The model was used to study the effect of multiple of infections (MOI) on the release of infectious and defective virus particles, also to analyze the interference features upon combined infection with the infectious and defective viruses (InfV+DV). At decreasing MOI (i), the levels of viral and protein RNAs reduce considerably, at increasing MOIs (ii), the lag period decreases slightly. Analysis of the model demonstrated that the limiting stages are cRNA synthesis at (i) and cellular mRNA at (ii). According to the model for the all MOIs, the fraction of the defective particles remains unaltered. Calculations demonstrated that upon combined infection (InfV+DV), the total release and fraction of the defective particles of the daughter virions increases with increasing fraction of MI infected cells. This is consistent with the data in the literature indicating that the defective interfering particles inhibit the formation of the infectious Vinf.

CONCLUSIONS

Future prospects: Attention will be focused on the identification of key stages limiting viral replication. These might be the potential targets for the development of improved drugs.

ACKNOWLEDGEMENTS

The authors are grateful to I.V. Lokhova for bibliographical support. The work was supported in part by the Russian Foundation for Basic Research No. 04-01-00458, and NSF: FIBR (Grant EF-0330786) and by the Russian Federal Agency for Science and Innovations (State contract No. 02.467.11.1005).

REFERENCES

- Ananko E.A., Podkolodny N.L., Stepanenko I.L., Podkolodnaya O.A., Rasskazov D.A., Miginsky D.S., Likhoshvai V.A., Ratushny A.V., Podkolodnaya N.N., Kolchanov N.A. (2005) GeneNet in 2005. *Nucl. Acids Res.*, **33**, D425–D427.
- Bancroft C.T., Parslow T.G. (2002) Evidence for segment-nonspecific packaging of the influenza A virus genome. *J. Virol.*, **76**, 7133–7139.
- Claas E.C., Osterhaus A.D., van Beek R. *et al.* (1998) Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *Lancet*, **351**, 472–477.
- Cros J.F., Palese P. (2003) Trafficking of viral genomic RNA into and out of the nucleus: influenza, Thogoto and Borna disease viruses. *Virus Res.*, **95**, 3–12.
- Deng T., Sharps J., Fodor E., Brownlee G.G. (2005) *In vitro* assembly of PB2 with a PB1-PA dimer supports a new model of assembly of influenza A virus polymerase subunits into a functional trimeric complex. *J. Virol.*, **79**, 8669–8674.
- Likhoshvai V.A., Matushkin Yu.G., Vatolin Yu.N., Bazhan S.I. (2000) A generalized chemical kinetic method for simulating complex biological systems. A computer model of λ phage ontogenesis. *Comput. Technol.*, **5**, 87–99.
- Likhoshvai V.A., Ratushny A.V., Nedosekina E.A., Lashin S.A., Turnaev I.I. (2006) *In silico* cell III. Method of reconstruction of enzymatic reaction and gene regulation mathematical models in terms of generalized Hill functions. *This issue*.
- McCown M.F., Pekosz A. (2005) The influenza A virus M2 cytoplasmic tail is required for infectious virus production and efficient genome packaging. *J. Virol.*, **79**, 3595–3605.
- Mikulasova A., Vareckova E., Fodor E. (2000) Transcription and replication of the influenza A virus genome. *Acta Virol.*, **44**, 273–282.

A MINIMUM MATHEMATICAL MODEL FOR SUPPRESSION HCV RNA REPLICATION IN CELL CULTURE

**Bezmaternikh K.D., Mishchenko E.L.* , Likhoshvai V.A., Khlebodarova T.M.,
Ivanisenko V.A.**

Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia

* Corresponding author: e-mail: elmish@bionet.nsc.ru

Key words: mathematical modeling, human hepatitis C, HCV replication, inhibitors of HCV NS3 protease

SUMMARY

Motivation: Search of efficient therapeutic ways and means for combating the hepatitis C virus has become a most timely problem. It has been recently shown that a peptidomimetic, by competitively inhibiting viral NS3 protease, efficiently suppresses the replication of the viral RNA replicon in Huh-7 cells. Computer simulation of the process makes it possible to calculate the dynamics of the model and of its elementary steps, to identify the key links, to follow how the model changes under the inhibitor.

Results: Based on review of the reported experimental data and application of the GenNet technology, we reconstructed the gene network for the replication of the viral RNA replicon in the Huh-7 cell. This was done in the presence of a highly specific and efficient inhibitor of viral NS3 protease. The result was a mathematical model simulating the function of the gene network. There was a good agreement between the computer calculated kinetics of the decrease in the viral RNA level upon inhibitor and the experimentally obtained results.

INTRODUCTION

According to the worldwide scattered information, from 2–3 % of the human population is hepatitis C virus (HCV) infected. The morbidity and mortality tolls keep rising alarmingly. HCV possesses a positive-sense, single-stranded RNA genome (approximately 9.6 kb) that encodes a single, large polyprotein, which is translated by host cell ribosomes in the cytoplasm. The polyprotein must then be proteolytically cleaved by cellular and viral proteases to yield individual functional viral proteins (Polyak, 2003). Earlier our team has reconstructed a gene network that stores the main life cycle of HCV: its entry into the cell, translation of the positive RNA chain, processing of the primary polyprotein with yield of structural and nonstructural proteins, RNA replication and virion assembly (<http://wwwmgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/index.shtml>).

HCV NS3 protease appears to be a promising target for the search of anti-HCV drugs. Its advantage is capacity to cleave the nonconstructional region of the polyprotein with attendant formation of proteins, the components of the replicase complex NS4A, NS4B, NS5A, NS5B. Search of the most potent and specific NS3 inhibitors has ended up with the finding of a peptidomimetic. Its inhibitory properties were tested *in vitro* ($K_i=74$ pM) and on Huh-7 cells stably transfected with self-replicating subgenomic HCV RNA constructs, replicon, carrying the HCV nonstructural coding region and express the

nonstructural proteins (Pause *et al.*, 2003). The major aim of this study was to build a mathematical model for the kinetics of the decrease in the level of the HCV RNA replicon in cells incubated in the peptidomimetic containing medium.

METHODS AND ALGORITHMS

The generalized chemical kinetic method was used for building the model (Likhoshvai *et al.*, 2001). To verify the initial values, the parameters were specified on the basis of the data in the relevant HCV literature. The optimum values for the parameters were then searched. The parameter values were selected to reproduce the observed experimental data both in qualitative and quantitative terms.

RESULTS AND DISCUSSION

Taken together, the assessment of the previously reported experimental data and the current GenNet technology, enabled us to reconstruct the gene network for the HCV replicon replication upon the NS3 protein inhibitor (Fig. 1).

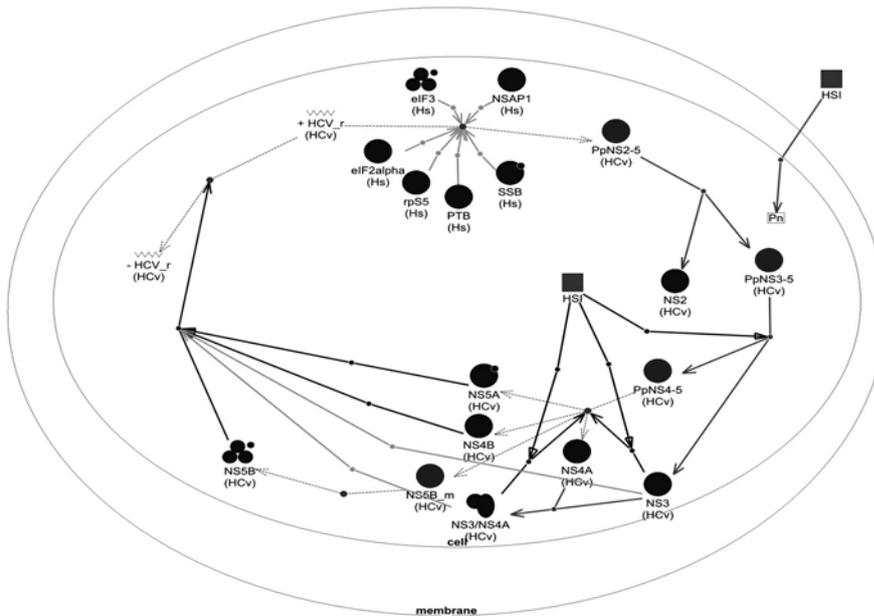


Figure 1. The gene network for the HCV replicon replication in the presence of the NS3 protease inhibitor.

The built gene network contains the following processes: a) translation of the HCV RNA replicon resulting in the polyprotein NS2-5; b) autocatalytic NS2-5 processing with the cleavage of the NS2, NS3 proteins and NS4-5 fragment; c) enzymatic processing NS4-5 through NS3 yielding NS5B and other non-structural HCV proteins; d) assembly of the HCV replicase; e) replication of the positive and negative HCV RNA chains (Polyak, 2003).

The gene network for the HCV replicon replication upon the NS3 protein inhibitor is a closed circuit of processes with a single negative feedback (inhibition of replication by NS5A protein, Shirota *et al.*, 2002). This provides the steady-state level of viral RNA and proteins two or three days after transfection). In the presence of NS3 protease inhibitor,

cleavage of the polyprotein is blocked, the concentration of inactive polyprotein rises in the cell, and the concentration of the viral proteins and RNA reduces (Pause *et al.*, 2003).

Based on the experimental data, the minimum mathematical model of HCV RNA propagation and its modification under NS3 inhibitor were developed:

$$\begin{aligned}
\dot{z} &= k_1 \cdot r^+ - k_2 \cdot z - k_3 \cdot L \cdot z + k_4 \cdot zL - k_{d1} \cdot z \\
NS5A &= k_2 \cdot z - k_5 \cdot rm \cdot NS5A + k_6 \cdot rm^* - k_5 \cdot rm^+ \cdot NS5A + k_6 \cdot rm^{+*} - k_5 \cdot rm^- \cdot NS5A + k_6 \cdot rm^{-*} - k_{d2} \cdot NS5A \\
rm &= k_2 \cdot z - k_5 \cdot rm \cdot NS5A + k_6 \cdot rm^* - k_i \cdot rm \cdot R^+ - k_i \cdot rm \cdot R^- + \bar{k}_e^+ \cdot rm^+ + \bar{k}_e^- \cdot rm^- - k_{d3} \cdot rm \\
rm^* &= k_5 \cdot rm \cdot NS5A - k_6 \cdot rm^* - k_i^* \cdot rm^* \cdot R^+ + k_e^* \cdot rm^{+*} - k_i^* \cdot rm^* \cdot R^- + k_e^* \cdot rm^{-*} - k_{d4} \cdot rm^* \\
r^+ &= \bar{k}_e^- \cdot rm^- + k_e^* \cdot rm^{-*} - k_{d5} \cdot r^+ \\
rm^+ &= k_i \cdot rm \cdot R^+ - \bar{k}_e^+ \cdot rm^+ - k_5 \cdot rm^+ \cdot NS5A + k_6 \cdot rm^{+*} - k_{d6} \cdot rm^+ \\
rm^{+*} &= k_i^* \cdot rm^* \cdot R^+ - k_e^* \cdot rm^{+*} + k_5 \cdot rm^+ \cdot NS5A - k_6 \cdot rm^{+*} - k_{d7} \cdot rm^{+*} \\
r^- &= \bar{k}_e^+ \cdot rm^+ + k_e^* \cdot rm^{+*} - k_{d8} \cdot r^- \\
rm^- &= k_i \cdot rm \cdot R^- - \bar{k}_e^- \cdot rm^- - k_5 \cdot rm^- \cdot NS5A + k_6 \cdot rm^{-*} - k_{d9} \cdot rm^- \\
rm^{-*} &= k_i^* \cdot rm^* \cdot R^- - k_e^* \cdot rm^{-*} + k_5 \cdot rm^- \cdot NS5A - k_6 \cdot rm^{-*} - k_{d10} \cdot rm^{-*} \\
\dot{L} &= k_7 - k_8 \cdot L - k_3 \cdot L \cdot z + k_4 \cdot zL \\
z\dot{L} &= k_3 \cdot L \cdot z - k_4 \cdot zL - k_{d11} \cdot zL
\end{aligned}$$

The designations for the concentrations are as follows: z – the primary polyprotein, $NS5A$ – NS5A, rm – replicase, rm^* – inactivated replicase, r^+ (r^-) – positive (negative) RNA strands, rm^+ (rm^-) – replicase on the positive (negative) RNA strands, rm^{+*} (rm^{-*}) – inactivated replicase on the positive (negative) RNA strands, L – peptidomimetic, zL – the peptidomimetic and the primary polyprotein complex.

$$R^+ = r^+ \cdot \left[1 - n \left(rm^+ + rm^{+*} \right) / \left(N \cdot r^+ \right) \right] - r^+ \text{ concentration with the free initiation site.}$$

$$R^- = r^- \cdot \left[1 - n \left(rm^- + rm^{-*} \right) / \left(N \cdot r^- \right) \right] - r^- \text{ concentration with the free initiation site.}$$

$$\bar{k}_e^+ = k_e - (k_e - k_e^*) \left[n \cdot rm^{+*} / \left(N \cdot r^+ \right) \right] - \text{the elongation constant for the replicase on } r^+,$$

$$\bar{k}_e^- = k_e - (k_e - k_e^*) \left[n \cdot rm^{-*} / \left(N \cdot r^- \right) \right] - \text{the elongation constant for the replicase on } r^-.$$

N denotes the length of the HCV RNA replicon in the nucleotides (~8000), n is the size of the replicase-RNA binding site in the nucleotides (~100). The description and values for the constants are tabulated (Table 1).

In the development of the mathematical model, we took it for granted that the NS5A protein is the sole factor that provides the steady-state HCV RNA level, however, the mechanism of its effects are not as yet identified. This prompted us to qualitatively analyse two possible application times of NS5A: (i) synthesis start of (–) and (+) RNA strands and (ii) elongation. Consideration of NS5A effect at replication initiation demonstrated

that the model has no steady-state. In contrast, when considered as a regulator of the activity of the replicase at initiation and elongation, NS5A allowed us to reach the system's steady-state and, moreover, to choose model parameters in such a way that they agree with the experimental data. Another achievement is estimation of the dynamics of the decrease in HCV RNA in the presence of the protease NS3 inhibitor. The achieved decrease is rather close to the experimentally observed (Fig. 2).

Table 1. The values for the mathematical model constants

Constant	Description	Value	Units
k_1	Translation r^+	10^{-3}	1/s
k_2	Polyprotein Z processing	5	1/s
k_3	zL Complex formation	10^6	1/[(molec/cell)*s]
k_4	zL Complex decay	1	1/s
k_5	Formation of the $rm - NS5A$ complex	10^6	1/[(molec/cell)*s]
k_6	$rm - NS5A$ Complex decay	$4 \cdot 10^3$	1/s
k_7	L Inflow	$6 \cdot 10^6$	(molec/cell)/s
k_8	L Outflow and decay	0.1	1/s
k_{d1}, k_{d2}, k_{d11}	$z, NS5A, zL$ decay	$6.8 \cdot 10^{-4}$	1/s
$k_{d3}, k_{d4}, k_{d5}, k_{d6}, k_{d7},$ k_{d8}, k_{d9}, k_{d10}	$rm, rm^*, r^+, rm^+,$ $rm^{+*}, r^-, rm^-, rm^{-*}$ decay	$1.375 \cdot 10^{-5}$	1/s
k_i	Replication initiation	0.01	1/[(molec/cell)*s]
k_i^*	Inactivated replicase	10^{-6}	1/[(molec/cell)*s]
k_e	Replication elongation	10^{-3}	1/s
k_e^*	Inactivated replicase	$3 \cdot 10^{-7}$	1/s

From the analysis of the minimum model incorporating NS5A as the sole factor limiting uncontrolled autocatalytic synthesis of (-) and (+) of the RNA replicon chains, it may be predicted that a NS5A function is a deceleration of replication elongation. We further intend to enhance the developed model. We will start with a description of the polyprotein translation processes and with its enzymic processing. Our other intention is to take into consideration the natural limitations imposed on the expended cell resources: energy, nucleotides, amino acids. With the enhanced model, we will continue to analyse in greater detail the patterns of replicon function in the presence of other known inhibitors of HCV development (other nonstructural protein inhibitors, translation process inhibitors, among others). The ultimate goal is the prediction of the most efficient targets for novel drugs.

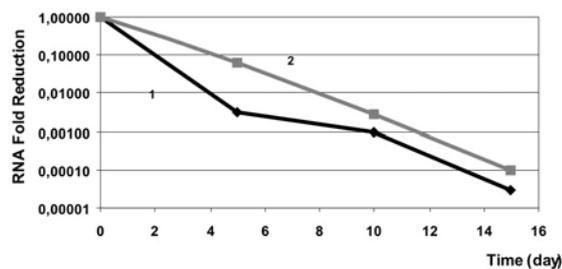


Figure 2. Kinetics of HCV RNA decay in the presence of inhibitor (100 mcM). Experimental data – 1, simulated data – 2.

ACKNOWLEDGEMENTS

Work was supported in part by Russian Foundation for Basic Research (Nos 04-01-00458, 05-04-49283, 06-04-49556), the State contract No. 02.434.11.3004 of 01.04.2005 and No. 02.467.11.1005 of 30.09.2005 with the Federal Agency for science and innovation “Identification of promising targets for the action of new drugs on the basis of gene network reconstruction” federal goal-oriented technical program “Study and development of priority directions in science and technique, 2002–2006”, Interdisciplinary integrative project for basic research of the SB RAS No. 115 “Development of intellectual and informational technologies for the generation and analysis of knowledge to support basic research in the area of natural science”, and NSF:FIBR (Grant EF-0330786). The authors are grateful to Aleksandr Ratushny for valuable discussions.

REFERENCES

- Likhoshvai V.A., Matushkin Y.G., Ratushny A.V., Anan'ko E.A., Ignat'eva E.V., Podkolodnaya O.A. (2001) Generalized chemokinetic method for gene network simulation. *Mol. Biol.*, **35**, 1072–1079.
- Pause A., Kukulj G., Bailey M., Brault M., Do F., Halmos T., Lagace L., Maurice R., Marquis M., McKercher G., Pellerin C., Pilote L., Thibeault D., Lamarre D. (2003) An NS3 serine protease inhibitor abrogates replication of subgenomic hepatitis C virus RNA. *J. Biol. Chem.*, **278**, 20374–20380.
- Polyak S.J. (2003) Hepatitis C virus–cell interactions and their role in pathogenesis. *Clin. Liver Dis.*, **7**, 67–88.
- Shirota Y., Luo H., Qin W., Kaneko S., Yamashita T., Kobayashi K., Murakami S. (2002) Hepatitis C virus (HCV) NS5A binds RNA-dependent RNA polymerase (RdRP) NS5B and modulates RNA-dependent RNA polymerase activity. *J. Biol. Chem.*, **277**, 11149–11155.

A MATHEMATICAL MODEL OF IMMUNE RESPONSE IN INFECTION INDUCED BY *MYCOBACTERIA TUBERCULOSIS*. PREDICTION OF THE DISEASE COURSE AND OUTCOMES AT DIFFERENT TREATMENT REGIMENS

*Bazhan S.I.*¹, *Schwartz Ya.Sh.*², *Gainova I.A.*³, *Ananko E.A.*⁴

¹ SRC VB "Vector", Novosibirsk, 633159, Russia; ² Institute of Internal Medicine, SB RAMS, Novosibirsk, Russia; ³ Institute of Mathematics, SB RAS, Novosibirsk, 630090, Russia;

⁴ Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia

* Corresponding author: e-mail: bazhan@vector.nsc.ru

Key words: mathematical model, mycobacteria tuberculosis, infection, immune response, treatment regimens

SUMMARY

Motivation: A mathematical model to describe the immune response and different chemotherapeutic regimens for the treatment of tuberculosis (TB) has been developed. The model can be used as a tool for predicting the course of the infectious process and outcomes in individuals infected with *Mycobacterium tuberculosis* (MBT). To identify what chemotherapeutic regimen is the best, one should clearly understand how different anti-TB drugs work. One of the most promising approaches to elaborate optimal anti-TB chemotherapeutic treatment/regimen is the gene-network technology.

Results: The model proposed herein consists of a set of ordinary differential equations for the dynamics of diverse populations of bacteria, macrophages, T lymphocytes, dendrite cells as well as for various concentrations of cytokines and antibacterial drugs. Different treatment regimens in individuals with different variants of the course of disease (latent and acute infection) were simulated, and so were different rates of drug inactivation (rapid and slow acetylators). Different regimes of the model, which yield different outcomes, correspond to different chemotherapeutic regimens, which either give recovery (adequate chemotherapy) or delay recovery and contribute to the emergence of drug-resistant strains (inadequate chemotherapy).

A further progression of the model will be connected with optimization of the treatment regimens for TB. To serve the purpose, we have reconstructed the gene network for the mechanisms of anti-TB drugs and for the mechanisms underlying the emergence of drug resistance developed by MBT due to mutation in separate target genes.

INTRODUCTION

There are at least three reasons which make TB one the most important public health problem worldwide: 1) the emergence of MTB strains with multiple drug resistance, 2) a high worldwide incidence of the disease and 3) its high mortality rate. About one-third of the world population is infected with MTB. What is unique about TB is that 90 % of those infected do not develop disease, 5 % have *de novo* acute infection and 5 % have acute infection after reactivation of latent infection.

It is not quite clear why the individuals infected with MTB develop different clinical forms of TB. Immune mechanisms of the host are believed to be critically important for the formation of individual course of the disease. On the other hand, the individual course of the disease depends on the pathogenicity of MTB strain, including its proliferative activity and characteristic features of its interaction with host defense mechanisms, and this could bring up the need for appropriate treatment regimen. We report a mathematical model of the immune response and different chemotherapeutic regimens for the treatment of TB infection designed to predict the course and the outcome of the process in MBT-infected individuals.

Multiple drug resistance is yet another point, which makes the fight against TB difficult. In order to gain a better understanding of the mechanisms of anti-TB drugs effects during infection, and multiple drug resistance formation, we have reconstructed a gene network, which provides insights into these mechanisms at the molecular-genetic level.

METHODS AND ALGORITHMS

An initial-value problem for a set of differential equations which describe the mathematical model of TB was addressed using the method proposed by Gear (1971) for stiff differential equations with a variable order of approximation. The reconstruction of the gene network for the mechanisms of anti-TB drugs effects and multiple drug resistance formation was performed using the GeneNet technology (Ananko *et al.*, 2005).

MODEL

The TB infection model presented herein is an extended modification of the mathematical model proposed by Marino S., Kirschner D.E. (2004) for the immune response induced by MBT in human lungs and lymph nodes. This extended model permits the simulation of TB treatment with one to five anti-TB drugs simultaneously and describes the emergence and further proliferation of genetically modified mycobacterial strains with resistance to more than one anti-TB drug used under different treatment regimens. The extended model consists of ordinary differential equations containing 25 variables, which describe four bacterial populations, four macrophage populations, five cytokines at different concentrations, five T-lymphocyte populations, two dendrite cell populations and five antibacterial drugs: isoniazid, rifampicin, pyrazinamide, streptomycin and ethambutol. The mechanisms of the drugs were described with regard to their bactericidal and bacteriostatic activity. Under the model, different treatment regimens were applied to individuals with rapid and slow drug inactivation (specifically, rapid and slow acetylators). Most coefficients and their value ranges were estimated from experimental literature data, the others, during the model verification.

RESULTS AND DISCUSSION

Antibacterial therapy is pivotal in the treatment of TB and should include co-administration of several drugs. Careful adherence to a treatment regimen is critical. Non-adherence to a prescribed chemotherapeutic regimen, namely, a reducing of the quantity of the drugs, reducing the doses or taking drugs at the wrong time, not only adversely affect the efficacy of treatment but also can result in the emergence of drug-resistant strains.

Using the model, different treatment regimens were simulated in individuals with different variants of the course of disease (latent and acute infection) and with different rates of drug inactivation (rapid and slow acetylators). The results of a simulation of an

adequate and a non-adequate treatment of acute TB infection in rapid acetylators infected with a drug-sensitive MBT strain are presented in Figs. 1 and 2.

Adequate treatment (Fig. 1) with the five drugs, namely isoniazid, rifampicin, pyrazinamide, streptomycin and ethambutol inhibits the exponential growth of MTB, and so the population of intra- and extracellular bacteria rapidly shrinks. The model suggests that recovery occurs in 180 days after the onset of treatment when the populations of intra- and extracellular bacteria and infected macrophages hit zero. Under these conditions the concentration of activated macrophages remains somewhat elevated for about two years, which suggests elevated resistance to infection.

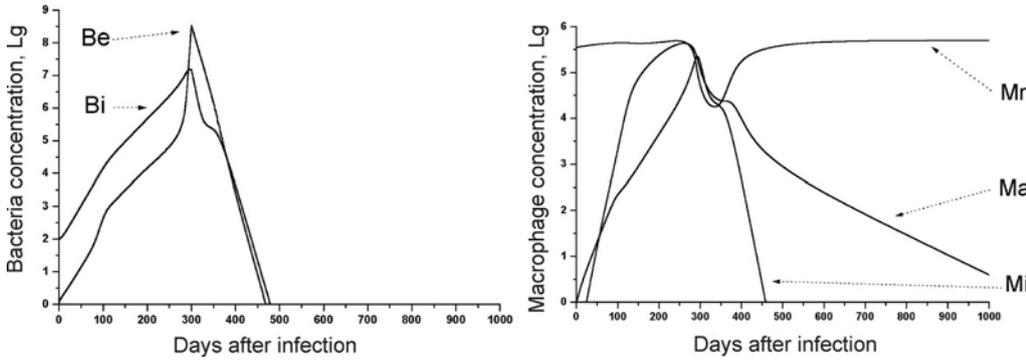


Figure 1. Simulation of an adequate treatment with five drugs applied to individuals infected with a drug-sensitive MBT strain (primary TB, acute infection, rapid acetylators). *t* from 0 to 300 days, untreated TB infection; *t* from 300 to 1000 days, TB infection responding to treatment. Chemotherapeutic regimen: Isoniazid, 30 mg/kg daily; Rifampicin, 10 mg/kg daily, Pyrazinamide, 25 mg/kg daily; Streptomycin, 15 mg/kg daily; Ethambutol, 30 mg/kg daily. Be, extracellular bacteria; Bi, intracellular bacteria; BRE, drug-resistant extracellular bacteria; BRi – drug-resistant intracellular bacteria; Mr, resting macrophages; Ma, activated macrophages; Mi, infected macrophages.

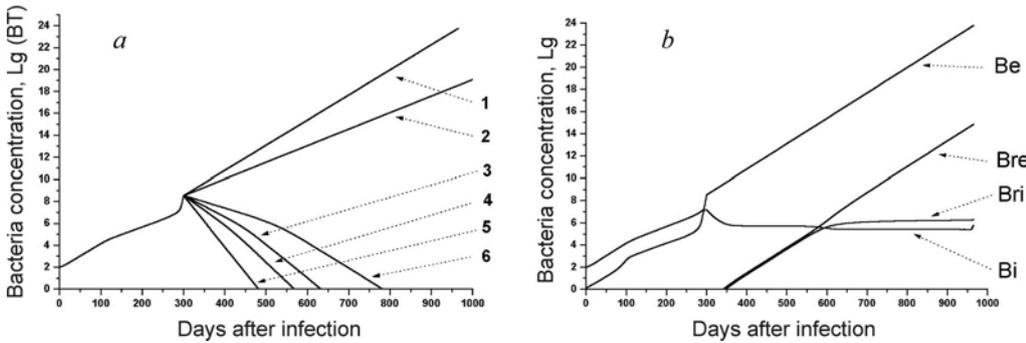


Figure 2. Simulation of non-adequate chemotherapeutic regimens applied to individuals infected with a drug-sensitive MBT strain (primary TB, acute infection, rapid acetylators). *t* from 0 to 300 days, untreated TB infection; *t* from 300 to 1000 days, TB infection responding to treatment. *a* – Bacterial populations (BT is the overall population of intra- and extracellular bacteria) responding to different chemotherapeutic regimens: 1, one drug (isoniazid, 30 mg/kg daily); 2, two drugs (isoniazid, 30 mg/kg daily; rifampicin, 10 mg/kg daily); 3, three drugs (isoniazid, 30 mg/kg daily; rifampicin, 10 mg/kg daily; pyrazinamide, 25 mg/kg daily); 4, four drugs (isoniazid, 30 mg/kg daily; rifampicin, 10 mg/kg daily; pyrazinamide, 25 mg/kg daily; streptomycin, 15 mg/kg daily); 5, five drugs (Isoniazid, 30 mg/kg daily; Rifampicin, 10 mg/kg daily; Pyrazinamide, 25 mg/kg daily; Streptomycin, 15 mg/kg daily; Ethambutol, 30 mg/kg daily); 6, five drugs at lower doses (Isoniazid, 5 mg/kg daily; Rifampicin, 10 mg/kg daily; Pyrazinamide, mg/kg daily; Streptomycin, 10 mg/kg daily; Ethambutol, 10 mg/kg daily). *b* – simulation of a treatment with one drug: isoniazid (30 mg/kg daily). Be, extracellular bacteria; Bi, intracellular bacteria; BRE, drug-resistant extracellular bacteria; BRi – drug-resistant intracellular bacteria.

Non-adequate treatment, which results from non-adherence to a standard chemotherapeutic regimen, yields different outcomes at the end of the disease “trajectories” (Fig. 2a). Taking three or four instead of five drugs delays recovery because periods of bacterial elimination are prolonged in these cases by 100 and 200 days, correspondingly. The model suggests that taking the five drugs at lower doses delays recovery even more (Fig. 2a, curve 6). Taking as few as two or one drug never brings recovery. Moreover, the model suggests that taking one (data not shown) or two drugs (Fig. 2b) brings up a drug-resistant MBT strain.

CONCLUSION

The mathematical model presented allows to predict possible variants of disease courses and outcomes in individuals infected with MBT. Using the model, treatments of TB infection under various different chemotherapeutic regimens were imitated. It has been demonstrated that different regimes of the model, which correspond to different chemotherapeutic regimens, yield different outcomes, which is either recovery (adequate chemotherapy) or delayed recovery and the emergence of drug-resistant strains (inadequate chemotherapy). Overall, the results obtained fit well with clinical observations.

The emergence of MBT drug resistance in response to inadequate anti-TB therapy hampers the further choice of adequate treatment. Studies of the molecular-genetic mechanisms of mycobacterial drug resistance provide insights into the mechanisms underlying the efficacy of anti-tuberculosis drugs and allow the timely choice of adequate therapy. For this purpose we have reconstructed the gene network, describing the mechanisms of anti-TB drugs activity and the mechanisms underlying the emergence of drug resistance developed due to mutations in the target genes (<http://www.mgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/index.shtml>). The sensitivity of mycobacteria with drug resistance mutations to anti-TB antibiotics was evaluated. The values obtained were transformed into the model’s parameters; test calculations were performed to simulate treatment regimens, which take into account the results of MBT genotyping (data not presented).

Further studies of the model behavior and the gene network functioning will focus on the optimization of possible treatment regimens for both drug-sensitive and drug-resistant tuberculosis.

ACKNOWLEDGEMENTS

The work was supported by the Russian Government (contract No. 02.467.11.1005).

REFERENCES

- Ananko E.A., Podkolodny N.L., Stepanenko I.L., Podkolodnaya O.A., Rasskazov D.A., Miginsky D.S., Likhoshvai V.A., Ratushny A.V., Podkolodnaya N.N., Kolchanov N.A. (2005) GeneNet in 2005. *Nucl. Acids Res.*, **33**, D425–D427.
- Gear C.W. (1971) The automatic integration of ordinary differential equations. *Communs. ACM*, **14**, 176–190.
- Marino S., Kirschner D.E. (2004) The human immune response to *Mycobacterium tuberculosis* in lung and lymph node. *J. Theor. Biology*, **227**, 463–486.

THE PACKAGE STEP+ FOR NUMERICAL STUDY OF AUTONOMOUS SYSTEMS ARISING WHEN MODELING DYNAMICS OF GENETIC-MOLECULAR SYSTEMS

Fadeev S.I.^{*1}, *Korolev V.K.*¹, *Gainova I.A.*¹, *Medvedev A.E.*²

¹ Sobolev Institute of Mathematics, SB RAS, Novosibirsk, 630090, Russia; ² Khristianovich Institute of Theoretical and Applied Mechanics, SB RAS, Novosibirsk, 630090, Russia

* Corresponding author: e-mail: fadeev@math.nsc.ru

Key words: mathematical modeling, numerical analysis, parameter continuation

SUMMARY

Motivation: Development and analysis of mathematical models of genetic-molecular systems is one of the actual problems of molecular biology. For this purpose it is necessary to develop specific algorithms and software.

Results: In this paper we present the package STEP+ for numerical study of autonomous systems of differential equations arising when modeling genetic-molecular systems.

Availability: This package is available by request.

INTRODUCTION

Intensive accumulation of data on structure functional organization and dynamics of gene networks put in the forefront the problem of development of tools for computer modeling and visualization of dynamic characteristics of gene networks. In this connection, one of the important problems as the problem of analysis of models under the development emerges. In turn, it requires elaboration of appropriate methods, approaches, and software. In this paper we present the package STEP+ realizing a complex of methods for numerical analysis of autonomous systems of differential equations modeling dynamics of gene networks (Fadeev *et al.*, 2002, 2004). Study of behavior of solutions of the given class of systems in dependence of parameters is one of the important parts of the general problem of constructing mathematical models adequate to experimental data. This package includes algorithms for numerical study of abstract autonomous systems consisting of N equations of the form

$$dy/dt = f(y,p), \quad (1)$$

where $p \in R^m$ is a vector of parameters. Among these algorithms there are integration of systems of the form (1) with given initial conditions (Cauchy problem), constructing stationary solutions in dependence of a parameter α , $\alpha \in p$, i.e., solutions of the system of nonlinear equations

$$f(y,\alpha) = 0, \quad (2)$$

and study of asymptotic stability of the obtained stationary solutions. The first version of the package was published in (Fadeev *et al.*, 2002).

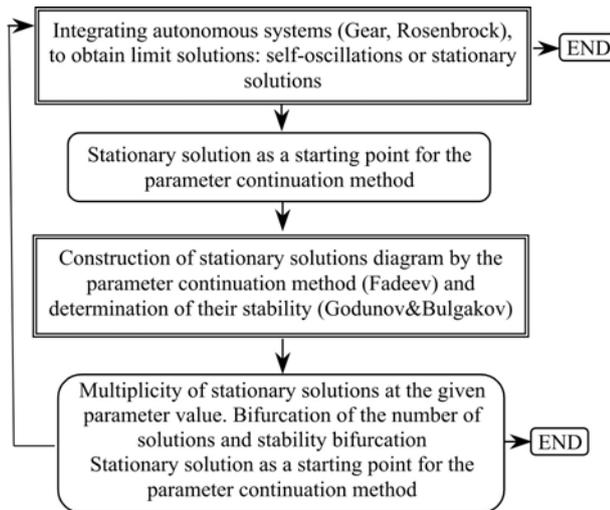


Figure 1. Operating opportunities of the package STEP+.

RESULTS

The package STEP+ is intended for numerical study of autonomous systems and systems of nonlinear transcendent equations. The present version of the package for Windows is a version of the package STEP (Fadeev *et al.*, 1998) and an update of the previous version (Fadeev *et al.*, 2002). The kernel of STEP+ is universal and allows the user to research numerically arbitrary systems consisting of N ordinary differential equations of the form (1). The package STEP+ uses original computational algorithms elaborated at the Sobolev Institute of Mathematics SB RAS for study of dynamical systems: a variant of the parameter continuation method, the Godunov-Bulgakov method (so called κ -criterion) for study of stability of solutions. These algorithms make it possible to research systems of arbitrary orders and take into account nonlinear effects (hysteresis, strong parametric sensitivity, self-oscillations and etc.) observed, as a rule, in mathematical models describing dynamics of genetic-molecular systems. In this package we realized also standard methods of numerical integration for stiff problems; in particular, multistep Gear's method of fractional accuracy order (Gear, 1971) and etc. Detailed descriptions of the methods can be found in the textbook (Fadeev *et al.*, 1998). The package STEP+ allows us:

- using the parameter continuation method, to establish numerically dependence of stationary solutions $y(\alpha)$ of the system (2) on parameter α and indicate domains of multiple solutions;
- to study stability of the stationary solutions and mark points, on the diagram of the stationary solutions, where an unstable stationary solution passes into a stable limit cycle (Hopf bifurcation);
- integrating autonomous systems, to obtain self-oscillations or start looking for stationary solutions which can be chosen as initial solutions for the continuation method;
- to point out domains of α , where all stationary solutions are unstable, i.e., self-excitation of self-oscillations for any initial data.

The structure of the package STEP+ is intended for organization of numerical experiments to study mathematical models presented by autonomous systems and systems of nonlinear equations. This package consists of three divisions: "Model development", "Cauchy problem", and "Nonlinear systems". The package STEP+ provides for automatic input right-hand sides of autonomous systems of differential equations by means of a

converter-program. To run any calculated program and input calculated parameters by means of user interface we produced a program for generation of DLL library. Owing to special input form of right-hand sides of systems of differential equations, this generator finds analytically the Jacobi matrix $df_i/dy_j, i=1, \dots, N, j=1, \dots, N$, and the matrix consisting of partial derivatives with respect to parameters $df_i/dp_j, i=1, \dots, N, j=1, \dots, m$. Admissible order of dynamical systems is equal to 1000. The package contains a series of models among which are mathematical models of gene networks, developed at the Institute of Cytology and Genetics SB RAS, and models from the literature (for example, the Lorentz model). The package STEP+ is realized in the visual programming environment Visual Basic.NET and intended for Windows 2000/XP.

Opportunities of the package STEP+ are demonstrated by a mathematical model describing dynamics of functioning of the genetic system controlling the cholesterol homeostasis in the cell (Ratushny *et al.*, 2003a, b; Ratushny, Likhoshvai, 2006). This model consists of 122 differential equations and contains 414 parameters. For given initial data and parameters we obtain a stationary solution, study its stability and sensibility on the parameters (Ratushny, Likhoshvai, 2006).

ACKNOWLEDGEMENTS

The authors are grateful to Vitaly Likhoshvai and Alexander Ratushny for valuable discussions. This work was supported in part by the Russian Foundation for Basic Research (No. 04-01-00458) and Siberian Branch of Russian Academy of Sciences (Interdisciplinary Integration Project No. 2006-46).

REFERENCES

- Fadeev S.I., Berezin A.Yu., Gainova I.A., Kogai V.V., Ratushny A.V., Likhoshvai V.A. (2002) Development of the program software for mathematic modelling of the gene network dynamics. In *Proc. III Intern. Conf. on Bioinformatics of Genome Regulation and Structure (BGRS'2002)*, **2**, 93–95.
- Fadeev S.I., Gainova I.A., Berezin A.Yu., Ratushny A.V., Matushkin Yu.G., Likhoshvai V.A. (2004) Study of stationary solutions in gene network models by the homotopy method. *Sib. Electronic Mathematical Reports*, **1**, 64–75.
- Fadeev S.I. Pokrovskaya S.A., Berezin A.Yu., Gainova I.A. (1998) *The Package STEP for Numerical Study of Systems of Nonlinear Equations and Autonomous System of General Form*. Description of work of the package STEP by examples of problems from the course “Engineering chemistry of catalytic processes”. Izdatel'stvo Novosibirskogo universiteta, Novosibirsk.
- Gear G.W. (1971) The automatic integration of ordinary differential equations. *Comm. ACM*, **14**, 76–179.
- Ratushny A.V., Likhoshvai V.A., Ignatieva E.V., Goryanin I.I., Kolchanov N.A. (2003a) Resilience of cholesterol concentration to a wide range of mutations in the cell. *Complexus*, **1**, 142–148.
- Ratushny A.V., Likhoshvai V.A., Ignatieva E.V., Matushkin Yu.G., Goryanin I.I., Kolchanov N.A. (2003b) A computer model of the gene network of the cholesterol biosynthesis regulation in the cell: analysis of the effect of mutations. *Dokl Biochem Biophys.*, **389**, 90–93.
- Ratushny A.V., Likhoshvai V.A. (2006) Mathematical modeling of the gene network controlling homeostasis of intracellular cholesterol. *This issue*.

MATRIX PROCESS MODELING: STUDY OF A MODEL OF SYNTHESIS OF LINEAR BIOMOLECULES WITH REGARD TO REVERSIBILITY OF PROCESSES

Fadeev S.I.^{*1}, *Shtokalo D.N.*², *Likhoshvai V.A.*³

¹ Sobolev Institute of Mathematics, SB RAS, Novosibirsk, 630090, Russia; ² Novosibirsk State University, Novosibirsk, 630090, Russia; ³ Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia

* Corresponding author: e-mail: fadeev@math.nsc.ru

Key words: genetic systems, mathematical models, matrix process, delay equations, limit theorem

SUMMARY

Motivation: Matrix processes of replication, transcription, and translation, including hundreds and thousands of elongation stages of the same type, is essential part of natural and artificial genetic systems. To reflect adequately these processes in models of gene networks it is necessary to elaborate effective theoretical and numerical mathematical methods.

Results: In this paper we present our research of a mathematical model of synthesis of linear biomolecules with regard to reversibility of processes of tailing of monomers to growing molecules and spontaneous termination of the synthesis process.

INTRODUCTION

Study of gene networks is one of the main problems of systemic biology. Matrix processes of replication, transcription, and translation, including hundreds and thousands of elongation stages of the same type, are essential part of natural and artificial genetic systems. Studying in detail elementary stages of synthesis, researchers confront with their fundamental property as micro-reversibility. Therefore we have the problem of correctness of models describing synthesis processes of linear molecules such as DNA, RNA, and proteins by means of equations in which intermediate stages are described by a delay. As was established, if microstages are irreversible then use of delay equations is correct (Likhoshvai *et al.*, 2004a). In the present paper we study a mathematical model with reversible microstages. We take into account also possible spontaneous termination of the synthesis process on any microstage.

In this paper, the synthesis process of linear molecules (protein, RNA, DNA) is described by the following autonomous system of differential equations:

$$\begin{aligned} \frac{dy_1}{dt} &= -(a + \omega)y_1 + by_2 + \alpha f(y_n), \quad a = \frac{n-1}{\tau_1}, \quad b = \frac{n-1}{\tau_2}, \quad S = a+b+\omega \\ \frac{dy_i}{dt} &= ay_{i-1} - Sy_i + by_{i+1}, \quad i=1, 2, \dots, n-2, \\ \frac{dy_{n-1}}{dt} &= ay_{n-2} - Sy_{n-1}, \quad \frac{dy_n}{dt} = ay_{n-1} - \theta y_n. \end{aligned} \quad (1)$$

Here y_n is a concentration of the synthesis product, τ_1, τ_2 are total transition times from the 1st stage to the n th stage, a is a rate constant of tailing of molecule length in the direct process, b is a rate constant of tailing of molecule length in the reverse process, ω is a rate constant of spontaneous termination of the process of molecule tailing (sink rate constant), α is a basal expression rate of genetic element, θ is a rate constant of degradation of the synthesis terminal product. All the constants are positive. The function $f(y_n)$ describes regulatory bond of negative type. In this paper we suppose that

$$f(y_n) = 1/(1 + \beta y_n^\gamma), \quad \gamma > 1, \quad \beta > 0. \quad (2)$$

We are interested in properties of the components $y_1(t), y_2(t) \dots, y_n(t)$ of solutions of autonomous systems of the form (1) for unboundedly increasing n . In the case $\tau_2 = \infty$, it was proved that the convergence holds $t > \tau_1, y_n(t) \rightarrow x(t)$ as $n \rightarrow \infty$;

Moreover,

$$dx/dt = \alpha f(x(t - \tau_1)) e^{-\omega \tau_1} - \theta x. \quad (3)$$

(Likhoshvai *et al.*, 2004a).

METHODS

We study the model in dependence on parameters by using both analytical and numerical methods. Proving numerically properties of the autonomous system (1) and the corresponding delay differential equation, we used a semi-implicit method from (Likhoshvai *et al.*, 2004a).

RESULTS

In this paper the result obtained for $\tau_2 = \infty$ is generalized to the case $\tau_1 < \tau_2$. Moreover, $y_n(t)$ tends to a solution of the delay differential equation as $n \rightarrow \infty$ (Fadeev *et al.*, 2005):

$$dx/dt = \alpha e^{-\omega \tau} f(x(t - \tau)) - \theta x, \quad \tau = \tau_1 \tau_2 / (\tau_2 - \tau_1), \quad (4)$$

For various parameters, solutions of (4) can give self-oscillations or stationary states of the autonomous system (1). We obtain that self-oscillations (Hopf bifurcation) arise for $\alpha > \alpha_0$,

$$\alpha_0 = e^{\omega \tau} \frac{\beta \gamma \theta}{[\beta(\gamma - 1)]^{\frac{\gamma+1}{\gamma}}}, \quad \tau = \tau_0, \quad (5)$$

where

$$\mu_0 = \theta \sqrt{[\gamma - \frac{1}{\beta} u^{\frac{\gamma}{\gamma+1}}]^2 - 1}, \quad u = \frac{\beta \gamma \theta}{\alpha_0} e^{\omega \tau}, \quad \tau_0 = \frac{1}{\mu_0} [1 + \arctg(\frac{\theta}{\mu_0})].$$

We establish numerically that, for the other relations between τ_1 and τ_2 , solutions arrive at steady-state regimes. Properties of stationary solutions of (1) as $n \rightarrow \infty$ are studied analytically (Fadeev *et al.*, 2005).

1) If $\tau_1 < \tau_2$, $\omega > 0$, $q < 1$, then the component y_n is finite, tends to a stationary solution of the delay differential equation with the delay τ , and is defined by the relation:

$$-e^{\omega\tau}y_n + \alpha f(y_n) = 0, \quad \tau = \tau_1\tau_2 / (\tau_2 - \tau_1). \quad (6)$$

Moreover, $y_1 \rightarrow 0$. 2) If $\tau = \tau_1 = \tau_2$, $\omega > 0$, and $q = 1$, all components of a stationary solution tend to zero. 3) If $\tau_1 > \tau_2$, $\omega > 0$, and $q > 1$, then $y_n \rightarrow 0$ and $y_1 \rightarrow \alpha\tau_2 / \omega\tau$, $\tau = \tau_1\tau_2 / (\tau_1 - \tau_2)$. Thus, for $\tau_1 \geq \tau_2$, $\omega > 0$, there exist finite limit values for components of a stationary solution of the system (1) as $n \rightarrow \infty$. 4) If $\tau_1 < \tau_2$, $\omega = 0$, and $q < 1$, then for any n the component y_n is defined by the relation (6) with $\omega = 0$. In particular, the relation (6) gives a stationary solution of the delay differential equation (4). 5) In the case $\tau_1 = \tau_2$, $\omega = 0$, and $q = 1$, a stationary solution is bounded for any $n > 1$. 6) If $\tau_1 > \tau_2$, $\omega = 0$, and $q > 1$, then the norm of a stationary solution grows unboundedly as $n \rightarrow \infty$. We mention in passing that the stationary solutions are uniquely defined for any $\alpha > 0$.

We study stability of stationary solutions of (1). If ω increases then the range of q , where stationary solutions are unstable, decreases. Consequently self-oscillations arise. One can assert that, for every n , there exists a ‘‘threshold’’ value of $\omega > 0$ such that, for all ω larger than the ‘‘threshold’’ value, stationary solutions of the system (1) are asymptotically stable for any τ_1 and τ_2 .

As a result, we have got a sufficiently complete view on properties of the model given by the autonomous system (1).

DISCUSSION

In the theory of gene networks, models of the form (1), considered in the present paper, arise as elements of more general systems of ordinary differential equations (Likhoshvai *et al.*, 2004b). In the paper (Likhoshvai *et al.*, 2004a) the case of irreversibility of synthesis processes ($\tau_2 = \infty$) and absence of sinks ($\omega = 0$) was studied. In this paper we raised the question on necessity of linearity and reversibility of intermediate stages for the proved limit theorems to be true. This question is rightful because in real situation we have nonlinearity and reversibility of elementary biochemical reactions forming more large blocks of genetic-molecular processes including stages of tailing of monomers to growing chains of proteins (RNA, DNA). In (Fadeev *et al.*, 2005) we provided an answer to the question on possibility of the limit passage from the system (1) to a delay differential equation when reversibility of processes of tailing of linear molecules and spontaneous termination of processes of tailing of bimolecular were taken into account. As it happens, the limit passage to the delay differential equation is possible if the total rate of the direct process is larger than the total rate of the reverse process. Presence of spontaneous termination is not a critical factor; it is evident as the factor $e^{-\omega\tau} < 1$ at α in the equation (4). We call it defect, and it describes decrease of the end product.

ACKNOWLEDGEMENTS

The work was supported in part by the Russian Foundation for Basic Research (No. 04-01-00458), NSF:FIBR (Grant EF-0330786) and by the Russian Federal Agency for Science and Innovations (State contract No. 02.467.11.1005). The authors are grateful to G.V. Demidenko for helpful discussion.

REFERENCES

- Fadeev S.I., Likhoshvai V.A., Shtokalo D.N. (2005) Study of a model of synthesis of linear biomolecules with regard to reversibility of processes. *Sib. Zh. Ind. Mat.*, **8**, 149–162. (In Russ.).
- Likhoshvai V.A., Fadeev S.I., Demidenko G.V., Matushkin Yu.G. (2004a) Modelling of multi-stage synthesis without branching by an equation with delay. *Sib. Zh. Ind. Mat.*, **7**, 73–94. (In Russ.).
- Likhoshvai V.A., Demidenko G.V., Kolchanov N.A., Matushkin Yu.G., Fadeev S.I. (2004b) Mathematical modeling of regulatory circuits of gene networks. *Zh. Vychisl. Mat. Mat. Fiz.*, **44**, 2276–2295; English transl. in *Comput. Math. Math. Phys.*, **44**, 2166–2183.

APPROXIMATE STATIONARY ATTRACTORS IN DROSOPHILA GAP GENE CIRCUITS IN THE LIMIT OF STEEP-SIGMOID INTERACTIONS

Gursky V.V.^{*1}, Samsonov A.M.¹, Reinitz J.²

¹ Ioffe Physico-Technical Institute, RAS, St. Petersburg, Russia; ² Stony Brook University, Stony Brook, New York, USA

* Corresponding author: e-mail: gursky@math.ioffe.ru

Key words: attractors, gene circuits, Drosophila, sigmoid interaction

SUMMARY

Motivation: Attractors in models of gene circuits govern the dynamics of gene expression. We aim at deriving explicit formulas for attractors as functions of model parameters, in order to approach the problem of parametric robustness of segmentation gene circuits in early Drosophila development.

Results: Plahte's method (Plahte *et al.*, 1998) is extended for derivation of approximate attractors in a model of segmentation gene circuits in early Drosophila development. Approximate regular and singular stationary points of model equations are calculated in terms of model parameters in the case of steep sigmoid function of gene regulation. The response of gene expression dynamics to variation of parameters is analysed by using these formulas.

INTRODUCTION

The regulation of expression of one gene by another in models of gene regulatory networks is usually represented by a sigmoid function which smoothly connects the zero expression level and a maximum expression level. The sigmoid depends on a fixed threshold, which is a characteristic quantity for the two genes. A method has been developed in Plahte *et al.* (1998) for gene circuit models with such sigmoid interactions to construct approximate stationary points in the case when the sigmoids are steep. We extend this method and apply in a circuit model of segmentation gene expression in the early Drosophila embryo (Mjolsness *et al.*, 1991; Jaeger *et al.*, 2004). This model possesses two features distinguishing it from previously considered: dynamical instead of fixed threshold in sigmoids and specific constraints on sigmoid functions. We have shown that these obstacles can be overcome in Plahte's approach and obtained approximate attractors in the steep-sigmoid version of a model incorporating basic features of the Drosophila gap gene system.

MODEL

We consider a modified model of Drosophila segmentation gene circuit (Mjolsness *et al.*, 1991; Jaeger *et al.*, 2004):

$$\frac{dv_i^a(t)}{dt} = R^a g_q \left(\sum_{b=1}^N T^{ab} v_i^b(t) + m^a v_i^{\text{Bcd}} + h^a \right) - \lambda^a v_i^a(t) + D^a \left(v_{i+1}^a(t) - 2v_i^a(t) + v_{i-1}^a(t) \right), \quad (1)$$

where N is a number of zygotic genes in the system, i marks the position of a cell nucleus along the A-P axis of the embryo, v_i^a is a concentration of a th gene's protein in nucleus i , v_i^{Bcd} is a concentration of protein from maternal gene *bicoid* in nucleus i , and vector $P = \{R^a, T^{ab}, m^a, h^a, \lambda^a, D^a\}$ includes the phenomenological parameters of the model.

The function g_q incorporates the sigmoid interactions between genes. We use the following specific sigmoid:

$$g_q(\eta) = \frac{1}{2} \left(\frac{\eta/q}{\sqrt{1 + \eta^2/q^2}} + 1 \right), \quad (2)$$

in which parameter q controls the steepness of the sigmoid curve: g_q tends to the step function limit as q tends to zero.

It is seen from (1) that the threshold of sigmoid interactions in the equations is not fixed, but has a dynamical structure. The expression level of a th gene in nucleus i is controlled not by many fixed thresholds from other genes, but by the only threshold represented via the time dependent variable

$$u_i^a(t) = \sum_{b=1}^N T^{ab} v_i^b(t) + m^a v_i^{\text{Bcd}} + h^a. \quad (3)$$

This variable incorporates inputs from all regulators of gene a in nucleus i .

RESULTS AND DISCUSSION

There are stationary points of two types in the circuits with continuous sigmoids, regular and singular, depending on their behavior in the step function limit for sigmoids (Plahte *et al.*, 1998). The regular ones lie in the domains of sigmoid saturation, while the singular stationary points occur exactly at the threshold as sigmoid tends to the step function. We have investigated the behavior of solutions in the gap gene circuits obtained by fitting to data in Jaeger *et al.*, 2004 in the step function limit $q \rightarrow 0$. The stationary attractors in these circuits have components approaching the threshold very close at $q \ll 1$. This allows to propose a hypothesis that the threshold dominated regulation is a characteristic feature of the pattern formation mechanism in the gap gene system.

In order to apply Plahte's approach to equations (1) with the dynamical threshold, we have first changed variables v_i^a to new variables u_i^a by using (3). In terms of the new variables, the thresholds of sigmoids in the model equations are fixed to zero value for all genes. We have also shown that results of the method can be extended to a wider class of sigmoid functions, including the chosen g_q from (2) which does not obey all constraints from Plahte *et al.*, 1998.

The extended Plahte's method was applied in a simplified gene circuit specifically designed to capture basic properties of the gap gene network in the steep-sigmoid case ($q \ll 1$). Namely, approximate stationary attractors in the circuit were derived in the general form:

$$w(P) = w_0(P) + qw_1(P) + O(q^2), \quad 0 < q \ll 1, \quad (4)$$

where w is a stationary point of equations (1) at a given parameter set P , which is a vector in the phase space of dimension $G \times N$ with G and N being numbers of genes and cell nuclei, respectively.

By using (4) we analysed the parametric dependence of dynamics in the circuit. Thus, the input of each component of the vector P in the parametric robustness of the circuit was quantified at the level of stationary attractors in the steep-sigmoid case.

ACKNOWLEDGEMENTS

The support of this work by NIH Grants RUB1-1578 and CRDF GAP award RBO1286, and by RFBR-NWO grant No. 047.011.2004.013 is gratefully acknowledged.

REFERENCES

- Jaeger J., Surkova S., Blagov M., Janssens H., Kosman D., Kozlov K.N., Manu, Myasnikova E., Vanario-Alonso C.E., Samsonova M., Sharp D.H., Reinitz J. (2004) Dynamic control of positional information in the early *Drosophila* embryo. *Nature*, **430**, 368–371.
- Mjolsness E., Sharp D.H., Reinitz J. (1991) A connectionist model of development. *J. Theor. Biol.*, **152**, 429–453.
- Plahte E., Mestl T., Omholt S.W. (1998) A methodological basis for description and analysis of systems with complex switch-like interactions. *J. Math. Biol.*, **36**, 321–348.

SIGNAL TRANSDUCTION PATHWAYS INVOLVED IN TRANSCRIPTIONAL REGULATION OF TYROSINE HYDROXYLASE

Kalashnikova E.V.^{*1}, *Dymshits G.M.*¹, *Kolpakov F.A.*^{2,3}

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ² Design Technological Institute of Digital Techniques, SB RAS, Novosibirsk, Russia; ³ Institute of Systems Biology OOO, Novosibirsk, Russia

* Corresponding author: e-mail: kalashnikova@ucdavis.edu

Key words: signal transduction, tyrosine hydroxylase, CREB, AP-1, Ca²⁺ signaling

SUMMARY

Motivation: Tyrosine hydroxylase (TH) has been shown to be very intensively regulated on different levels, not only the levels of transcription and translation, but posttranslational modification as well. Theoretical analysis of the pathways involved in regulation of TH expression would allow to understand the mechanisms of development of diseases associated with alterations in TH function such as Parkinson, Alzheimer disease and arterial hypertension.

Results: Experimental evidence obtained from experiments of activation of six TH promoter motifs in response to a number of various stimuli in *Rattus norvegicus* has been analyzed. Based on the facts reviewed in literature a series of diagrams have been created that describe the mechanisms and pathways involved in this regulation.

Availability: <http://biopath.biouml.org>.

INTRODUCTION

Primary interest in TH gene expression comes from its being a cornerstone enzyme in the biosynthesis of dopamine, adrenaline and noradrenalin, – the signaling molecules known to have dual functions as both neurotransmitters and hormones. TH dysfunction is thought to be associated with human neurodegenerative Parkinson and Alzheimer diseases. Adrenaline and noradrenalin synthesized in adrenal medulla also act as hormones causing vasodilatation or vasoconstriction of local blood vessels in different organs. Hypersensitivity of this system together with any deregulation in production or secretion of hormones may lead to such diseases as arterial hypertension. Sympathoadrenal system is activated in response to emotional stressors – the most common type of stimuli in human society known to be a cause of this disease; this results in increased levels of catecholamines and elevated blood pressure. TH is known to be regulated very intensively on different levels. Studying the signal transduction pathways involved in regulation of TH activity seems promising in terms of understanding the mechanisms of the diseases associated with TH dysfunction.

METHODS

BioUML technology (Kolpakov, 2004; <http://www.biouml.org>) was used for formal description of mechanisms and pathways involved in TH regulation in rats. All diagrams and description diagram elements (genes, proteins, substances, concepts, reactions and semantic relationships) were stored in Biopath database. Corresponding information was obtained from literature annotation and different biological databases (TRANSPATH, TRANSFAC and KEGG). Web interface developed using BeanExplorer Enterprise Edition (<http://www.beanexplorer.com>) provides public access to Biopath database.

RESULTS AND DISCUSSION

Table 1. The effect of different stimuli on the transcription rate of TH in *Rattus norvegicus*

Stimulus	Pathways, molecules involved	Cis-elements activated	Response	Diagram
Protein stimuli				
Ang II	ERK1/2, p90RSK; cAMP	AP-1, CRE	+	DGR_TH4
EGF	MAPK, JNK, ERK1/2	<i>Egr1/SP1, AP-1, CRE, Egr1</i>	+	DGR_TH7
IGF	MAPK, JNK, ERK1/2; HIF-1,2- α	<i>Egr1/SP1, AP-1, CRE, Egr1</i>	+	DGR_TH2
NGF	Ras/Raf/MEK/ERK1/2; HIF-1,2- α	<i>Egr1/SP1, AP-1, CRE, Egr1</i>	+	DGR_TH1
Insulin	HIF-1,2- α	HIF-1	+	DGR_TH7
PrRp	PKC, p90RSK, CaMK II, PKA	<i>AP-1, CRE</i>	+	DGR_TH2
TNF	JNK, p38- α , - β , - γ	AP-1, CRE	+	DGR_TH2
Urocortin	cAMP	CRE	+	DGR_TH12
VIP	adenylyl cyclase	AP-1, CRE	+	DGR_TH3
Other stimuli				
Glucocorticoids	GR Ras/Raf/MEK/ERK1/2	GRE <i>AP-1, CRE, Egr1</i>	+ +	DGR_TH1
Cell density	-	-	+	DGR_TH12
Cold stress	-	-	+	DGR_TH9
Depolarization	c-Fos	AP-1	+	DGR_TH5
Immobilization stress	JNK	AP-1, CRE,	+	DGR_TH9
Hypoxia	ERK1/2, CaMK II, PKA	HIF-1, AP-1, <i>CRE</i>	+	DGR_TH7
Cycloheximide	pCREB	AP-1, CRE	+	DGR_TH3
TPA	PKA	AP-1, <i>CRE</i>	+	DGR_TH5
	Raf/ MEK/ ERK 1/2	AP-1, <i>CRE, Egr1</i>	+	
	Raf/ MEK/ MAPKAPK2	<i>CRE</i>	+	
Forskolin	Raf/ MEK/ p90RSK	AP-1, <i>CRE</i>	+	DGR_TH3
	ICER	CRE	-	
	Reserpine	CRE	-	
Retinoic acid	RXR	-	-	DGR_TH3

“+” (in the “Response” column) indicates induction, “-” – decrease in the level of TH transcription.

Table 1 summarizes all the different stimuli that have been analyzed in this study and pathways involved (diagrams DGR_TH1 through 12). There is still little known about how the response to such stressful stimuli as cell density, immobilization and cold stress, depolarization and hypoxia is formed (data obtained from experiments on cultured *Rattus norvegicus* cell lines and also *in vivo* studies). In general most of these factors lead to

induction of stress-reactive sympathoadrenal system by inducing transcription of TH. A series of diagrams have been created that describe the mechanisms and pathways involved in biosynthesis of catecholamines (DGR_TH10), activation of sympathoadrenal system (DGR_TH11), levels of regulation of TH expression (DGR_TH12), regulation via phosphorylation/dephosphorylation (DGR_TH8), regulation of TH transcription (DGR_TH5), regulation of transcription via CRE/CaRE element (DGR_TH6), AP-1/CRE crosstalk (DGR_TH3), AP-1/Egr1/SP-1 crosstalk (DGR_TH8), effects of angiotensin II on TH activity (DGR_TH4), AP-1/ GRE crosstalk (DGR_TH1), AP-1/HIF-1 crosstalk (DGR_TH7), effects of stress on TH activity (DGR_TH9) and AP-1 complexes/dimers (DGR_TH13). The next diagram demonstrates all the pathways that have been shown to be involved in signaling in response to the stimuli described above (Fig. 1). Six different cis-elements in the promoter of TH have been studied: AP-1, CRE/CaRE, CRE-2, GRE, Egr1/Sp-1 and HIF-1. More detailed analysis of the pathways that lead to activation of one separate cis-element revealed that each of these elements required activation of AP-1 site (Sun, 2003; Sun, Tank, 2003; Suzuki *et al.*, 2004).

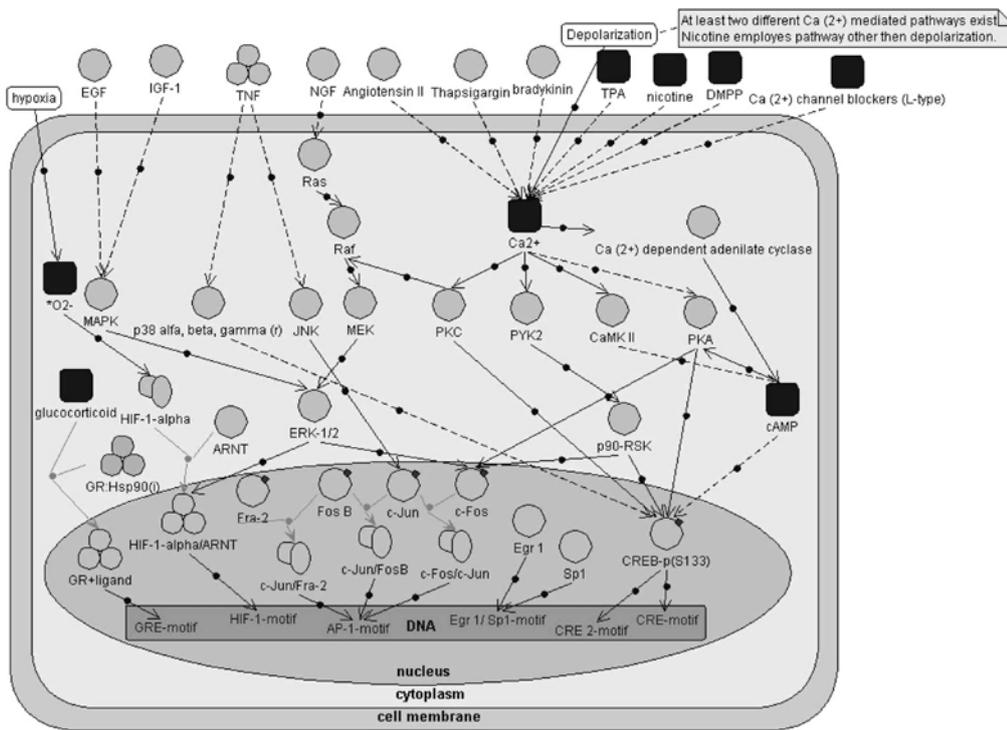


Figure 1. (DGR_TH5). Summary of all the pathways analyzed that are involved in regulation of TH transcription in *Rattus norvegicus*.

One of such examples is activation of both AP-1 and CRE elements in TH promoter by the same stimulus. Several important protein kinases are positioned under control of Ca²⁺ concentration, among these are PKC, PKA, CaMKII, p90RSK (Chae, Kim, 1995; Lim *et al.*, 2000; Lewis *et al.*, 2003). They are potent to directly activate CREB via phosphorylation, but also they may initiate other kinase pathways such as Raf/ MEK/ ERK1/2 by means of PKC. ERK1/2 is a central figure in several cascades, including cell response to growth factors (IGF, EGF, NGF), prolactine releasing peptide (PrRp). At this level a cross-talk with the family of AP-1 factors occurs (ERK1/2 directly transactivates c-Fos), which leads to activation of another cis-element in TH promoter (Nagamoto-Combs *et al.*, 1997). Another important stimulus for CREB-mediated induction is cAMP. This second messenger in turn activates PKA and stimulates phosphorylation of CREB.

Activation by TNF involves p38 alpha, beta, gamma protein kinases and JNK (leading to activation of yet another AP-1 family member – c-Jun). These kinases phosphorylate other transcription factors such as STAT and Elk-1, in turn activating genes encoding various factors of AP-1 family and other genes. This way Ca^{2+} and cAMP signaling lead to induction of both CREB and AP-1 family members and induction of one of the sites leads to induction of the other. The same mechanism has been observed between AP-1 family members and some of the members of the analyzed pathways involved in regulation of TH transcription. The fact that AP-1 activation is required for functioning of all the transcription sites analyzed suggests an idea that binding of AP-1 factors to the corresponding cis-element is needed to achieve a favorable conformation. c-Jun/c-Fos heterodimers are known to bend DNA strands upon binding to the consensus site (Kerppola, Curran, 1993); this bending might facilitate recognition of other cis-elements and provide yet another level of regulation. To support this model further experimental evidence is required.

ACKNOWLEDGEMENTS

This work was supported by and Siberian Branch of Russian Academy of Sciences (interdisciplinary project No. 46).

REFERENCES

- Chae H.D., Kim K.T. (1995) Cytosolic calcium is essential in the basal expression of tyrosine hydroxylase gene. *Biochem. Biophys. Res. Commun.*, **17**, 659–666.
- Kerppola T., Curran T. (1993) Selective DNA Bending by a Variety of bZIP Proteins. *Mol. Cell. Biol.*, **13**, 5479–5489.
- Kolpakov F.A. (2004) BioUML – open source extensible workbench for systems biology. *Proceedings of BGRS'2004*, **2**, 77–80.
- Lewis J.B., Randol T.M., Lockwood P.E., Wataha J.C. (2003) Effect of subtoxic concentrations of metal ions on NFkappaB activation in THP-1 human monocytes. *J. Biomed. Mater. Res.*, **1**, 217–224.
- Lim J., Yang C., Hong S.J., Kim K.S. (2000) Regulation of tyrosine hydroxylase gene transcription by the cAMP-signaling pathway: involvement of multiple transcription factors. *Mol. Cell. Biochem.*, **212**, 51–60.
- Nagamoto-Combs K., Piech K.M., Best J.A., Sun B., Tank A.W. (1997) Tyrosine hydroxylase gene promoter activity is regulated by both cyclic AMP-responsive element and AP1 sites following calcium influx. Evidence for cyclic amp-responsive element binding protein-independent regulation. *J. Biol. Chem.*, **28**, 6051–6058.
- Suzuki T., Kurahashi H., Ichinose H. (2004) Ras/MEK pathway is required for NGF-induced expression of tyrosine hydroxylase gene. *Biochem. Biophys. Res. Commun.*, **5**, 389–396.
- Sun B. (2003) Chronic nicotine treatment leads to sustained stimulation of tyrosine hydroxylase gene transcription rate in rat adrenal medulla. *J. Pharmacol. Exp. Ther.*, **304**, 575–588.
- Sun B., Tank A.W. (2003) c-Fos is essential for the response of the tyrosine hydroxylase gene to depolarization or phorbol ester. *J. Neurochem.*, **85**, 1421–1430.

Cyclonet – AN INTEGRATED DATABASE ON CELL CYCLE REGULATION AND CARCINOGENESIS

Kolpakov F.^{*1,2}, **Poroikov V.**³, **Sharipov R.**^{1,2,4}, **Milanesi L.**⁵, **Kel A.**⁶

¹Design Technological Institute of Digital Techniques, SB RAS, Novosibirsk, 630090, Russia; ²Institute of Systems Biology OOO, Novosibirsk, 630090, Russia; ³Institute of Biomedical Chemistry, RAMS, Moscow, Russia; ⁴Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ⁵Institute of Biomedical Technologies, CNR, Segrate (MI), Italy; ⁶BIOBASE GmbH, Wolfenbuettel, Germany

* Corresponding author: e-mail: fedor@biouml.org

Key words: database, cell cycle regulation, cancer, drug design, BioUML

SUMMARY

Motivation: Success of the biologically reasonable modeling of cellular systems depends on the completeness of our knowledge and integration of all fundamental molecular processes, such as signal transduction, regulation of gene expression and metabolism. Mammalian cell cycle is a good example of the system where such natural integration of all the molecular processes plays important role in their regulation. Despite of the massive development of various biological databases, no specialized repository was created so far that would integrate all knowledge on cell cycle.

Results: We have developed the Cyclonet – an integrated database on cell cycle regulation and carcinogenesis. It contains information about known specific genes, proteins and their complexes involved in regulation of cell cycle and carcinogenesis; diagrams of cell cycle regulation and related processes; models of cell cycle and results of their simulation; links on microarray data on cell cycle and on various types of cancer, information on anticancer drug targets as well as their ligands, broad literature references and other related resources.

Availability: <http://cyclonet.biouml.org>.

INTRODUCTION

The main goal of the Cyclonet database is to integrate information from genomics, proteomics, chemoinformatics and systems biology and provide specialized and comprehensive resource in order to enable molecular biologists working in the field of anticancer drug development to analyze systematically all this data and generate experimentally testable hypothesis (Fig. 1).

METHODS

Novel software technologies were used for development of Cyclonet database:

- BioUML technology (Kolpakov, 2004; <http://www.biouml.org>) was used for formalized description, visual modeling of eukaryotic cell cycle and for query and editing of the database content. BioUML workbench also allows simulate the described systems behavior using Java or MATLAB simulation engines;

- Biopath database (<http://biopath.biouml.org>) stores diagrams, description diagram elements (genes, proteins, substances, concepts, reactions and semantic relationships) as well as for storage mathematical models and results of simulation;
- using BeanExplorer Enterprise Edition (<http://www.beanexplorer.com>) web interface for Biopath database was developed.

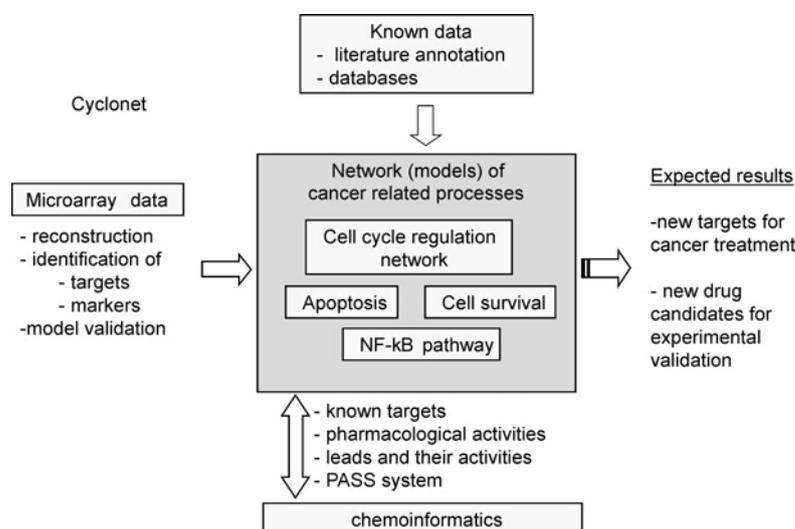


Figure 1. Diagrams (models) of carcinogenesis and related processes as the basis of data integration in Cyclonet database.

The data in Cyclonet are compiled mainly from manual literature annotation. Links to the public databases, GeneOntology, RefSeq and Ensembl, are provided from genes, proteins and other corresponding entries. Known cell-cycle models are imported from SBML (Hucka *et al.*, 2003; <http://www.sbml.org>) and CellML (<http://www.cellml.org>) model repositories and annotated manually based on literature. Cyclonet also contains a vast body of literature references that are arranged by categories.

RESULTS

Cyclonet database integrates data of genomics, proteomics, chemoinformatics, and systems biology for their use in drug design:

- genomics – we have collected and categorized (arranged) links to available microarray experiments (Table 1). During next stage of work (BioUML team is now working on support of microarray data) we will merge microarray data on breast cancer with corresponding BioUML diagrams and analyses tools. Cyclonet includes information about 196 genes related with cell cycle and cancer development.
- proteomics – Cyclonet contains information about 2465 proteins, their complexes and interactions (protein, its modified form – for example, phosphorylation, protein complexes and protein families are considered as a separate entries in protein table).
- chemoinformatics – Cyclonet contains information about 40 key targets for anticancer treatments, 33 related pharmacological activities. 422 ligands known to be related with these activities are placed in database. Chemical and structural formulae are available for these ligands (Fig. 2). This information is used by computer program PASS (Porokov, Filimonov, 2005; <http://www.ibmcm.sk.ru/PASS>) to predict new ligands with anticancer activities.

- systems biology – Cyclonet contains 351 diagrams describing cell cycle regulation and related systems. Cyclonet also includes 32 mathematical model of cell cycle regulation annotated from literature (Fig. 3). The number of collected diagrams related to cell cycle exceeds the content of any other related databases and resources.

DISCUSSION

New high-throughput methods allow generation of massive amounts of molecular biological data. These, mainly phenomenological, data are often difficult to relate with the activation/inhibition of particular signal transduction pathways and/or transcriptional regulators. A way to facilitate data interpretation is to construct gene regulatory networks that include signal transduction mediators, transcriptional regulators and target genes. This is a complex task, not only because of the huge number of molecules involved, but also because of variations across tissues, developmental stages and physiological conditions. However, these networks hold the key to the understanding of the regulatory processes within a cell.

The aim of the Cyclonet database is, therefore, to develop an integrative approach that will help researchers to understand the cell cycle process through modeling and simulation of gene regulatory networks. It combines and puts results of different high-throughput experimental methods together under the roof of bio- and chemoinformatics, to exploit the full potential of the included methods as well as the generated data.

Table 1. Microarray web resources and microarray series related with cell cycle and cancer

Category	Entries	Comment
Resources → microarray		
data → processed	24	This category includes two groups of resources related to cancer and cell cycle experiments: 1) Resources containing raw or processed published data. 2) Resources containing published papers or description of the experiments and in some cases a link to the site with raw data. Resources in this category do not contain information about published papers concerning experiment and raw data but allow exploring expression of a particular gene in the different conditions.
data → web interface	10	
data	7	Unclassified microarray data.
Total microarray resources	41	
Series (microarray experiments)		
Cell cycle	24	Microarray data related with cell cycle
For modeling	36	Data are believed to be useful for modeling
Treatment	50	Data related with different treatment (chemotherapy, hormones, growth factors, etc.) of normal and cancer tissues or cell lines
Cancer → targets	11	Experiments where some genes were identified as targets for cancer treatment
Cancer → markers	4	Experiments where some genes were identified as cancer markers
Cancer total	310	All cancer data.
Total microarray series	354	
Publications		
Publications, microarray	287	354 microarray series refer to 287 articles

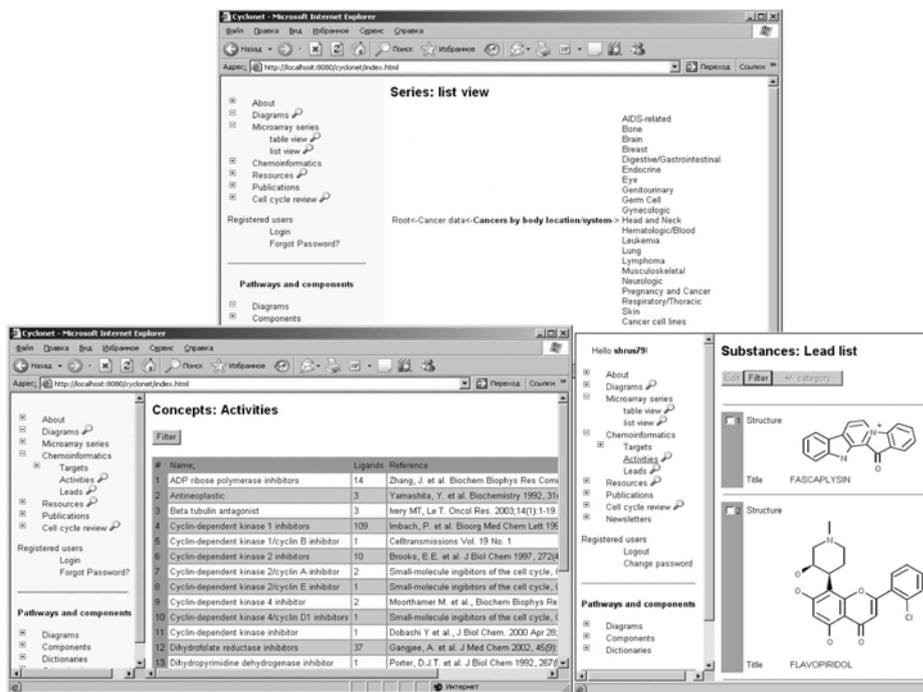


Figure 2. Web interface of Cyclonet database generated with BeanExplorer technology. Top screen displays fragment of microarray series classification in Cyclonet database, bottom left screen demonstrates fragment of list of pharmacological activities for anticancer therapy, bottom right – examples of chemical structure of two cyclin-dependent kinase 4 inhibitors.

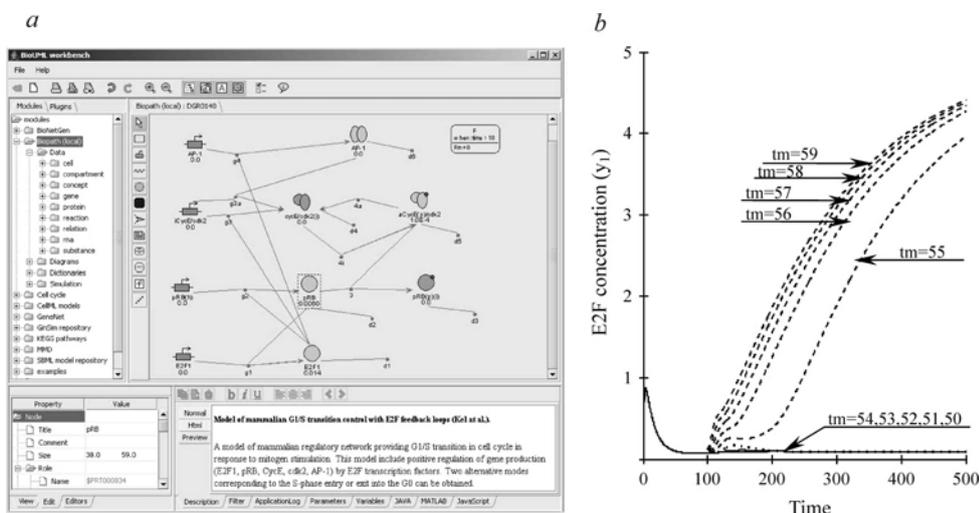


Figure 3. On the left – G1/S entry model (DGR0140, Kel *et al.*, 2000) simulating two different modes of the system: quiescence or cycle progression – described using BioUML technology. On the right – dynamics of E2F-1 concentration – simulation results of the G1/S entry model. tm – duration of system stimulation by mitogen (in seconds).

ACKNOWLEDGEMENTS

This work was supported by INTAS grant No. 03-51-5218, MIUR-FIRB grant No. RBLA0332RH Laboratory for Interdisciplinary Technologies in Bioinformatics and BIOINFOGRID No. 026808. Authors are grateful to V. Komashko and V. Valuev for microarray data annotation and E. Cheremushkina for annotation of some diagrams.

REFERENCES

- Hucka M. *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.
- Kel A. *et al.* (2000) Modeling of Gene Regulatory Network of Cell Cycle Control. Role of E2F Feedback Loops. *Proceedings of the German Conference on Bioinformatics (GCB 2000)*, pp. 107–114.
- Kolpakov F.A. (2004) <http://www.bioiml.org>.
- Poroikov V., Filimonov D. (2005) PASS: Prediction of Biological Activity Spectra for Substances. In Helma C. (ed), *Predictive Toxicology*. Taylor & Francis, pp. 459–478.

A GENE NETWORK OF ADIPOCYTE LIPID METABOLISM: COORDINATED INTERACTIONS BETWEEN THE TRANSCRIPTION FACTORS SREBP, LXR α AND PPAR γ

Kouznetsova T.N.^{*1}, *Ignatieva E.V.*^{1,2}, *Oshchepkov D.Yu.*¹, *Kondrakhin Y.V.*¹,
Katokhin A.V.^{1,2}, *Shamanina M.Y.*¹, *Mordvinov V.A.*¹

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ² Novosibirsk State University, Novosibirsk, 630090, Russia

* Corresponding author: e-mail: verzun@bionet.nsc.ru

Key words: adipocyte, transcription regulation, lipid metabolism, SREBP family, LXR α , PPAR γ

SUMMARY

Motivation: It is of great importance to keep studying lipid metabolism regulation in the adipose tissue, because the tissue excess is a risk factor for serious human diseases (dyslipidemia, diabetes, cardiovascular ones). This is why the accumulation and analysis of data on the regulation of the expression of adipocyte genes are expected to help elaborate possible scenarios for development of pathological processes and propose possible ways for their correction.

Results: Based on analysis of experimental data on the mechanisms of gene expression regulation in adipocytes and data on metabolic reactions, we have developed the second version of the gene network for the adipocyte within the GeneNet database. We herein report a logical analysis of a gene network fragment for lipid metabolism regulation in adipocytes, including cholesterol and fatty acid level control. Analysis of the gene network revealed that the coordinated interaction of transcription factors Sterol Regulatory Element Binding Proteins (SREBPs), Liver-X-Receptor alpha (LXR α) and Peroxisome Proliferator-Activated Receptor gamma (PPAR γ) plays a key role in gene expression regulation of this system. Using the SITECON method, seven new genes, which have been identified as being potential targets for SREBP factors, have been revealed among human genes for lipolysis and lipid droplet formation.

Availability: <http://www.mgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/index.shtml>
(object: `_Adipocyte_pathways`).

INTRODUCTION

Adipose tissue is specialized in the storage of energy in the form of triacylglycerol (TAG), a neutral lipid that is synthesized from glycerol-3-phosphate and CoA-derivatives of fatty acids. TAG is located in special structures known as lipid drops, including also phospholipids, cholesterol, structure proteins and lipid metabolism enzymes (Brasaemle *et al.*, 2004). Some data provide a link between the presence of cholesterol on the surface of the lipid drop, its biogenesis and TAG accumulation. Adipocytes (fat cells) are poorly active in the *de novo* cholesterol synthesis. The majority of adipocyte sterols originate from circulating lipoproteins. The fat cells also are very active in the exogenous fatty acid uptake. Little is known about the mechanism of adipocyte lipid drop formation and its structural reorganization in response to the lipolysis stimulation.

Systematization and logical analysis of GeneNet data allowed us to reveal a scheme of regulatory interactions between the transcription factors of the SREBP family and PPAR γ and LXR α . This scheme underlies the molecular mechanism of cholesterol and fatty acid metabolism regulation in adipocytes. The presence of putative binding sites for SREBPs that we have recognized using SITECON in the regulatory regions of 7 genes controlling lipolysis and lipid drop formation suggests that these genes are likely regulated by the family SREBPs and are promising candidates for further experimental trials.

METHODS

The adipocyte gene network was reconstructed using the GeneNet system (Ananko *et al.*, 2005). In its current version, the gene network for the adipocyte includes the subschema (*_Adipocyte_lipolysis*) and totally contains information on 51 genes, 173 proteins, 40 mRNAs and 352 relationships between the objects, which were included in the database on the basis of analysis of 314 experimental works. Recognition of potential SRE-type sites in human genes promoters was performed using the SITECON method, which efficacy to recognize conservative physicochemical and conformational properties of SREBPs binding sites was described earlier (Oshchepkov *et al.*, 2004). As a recognition threshold, the SITECON method employs the level of necessary conformational similarity which was 0.73 for SRE. The nucleotide sequences of the promoters of the genes that are expressed in adipocytes were retrieved from RefSeq database genomic contig entries, using Feature Table data on position of first exon of respective RefSeq mRNA.

RESULTS AND DISCUSSION

Regulation of cholesterol and fatty acid metabolism by the transcription factors SREBPs, LXR α and PPAR γ . The level of intracellular cholesterol in adipocytes is controlled by two processes: biosynthesis and transport into the cell from circulating lipoproteins complexes such as low- and high-density lipoprotein particles (LDL and HDL). Genes controlling cholesterol biosynthesis (HMG-CoA synthase (HMGCS), HMG-CoA reductase (HMGCR)) and for LDL-receptor (LDLR), containing classic sterol response elements (SREs), are efficiently activated by SREBP-2 and SREBP-1a, but not SREBP-1c (Amemiya-Kudo *et al.*, 2002) (Fig. 1). However the formation rate of transcriptionally active SREBPs from membrane-bound precursors inversely depends upon cellular cholesterol concentration. Thus, cholesterol biosynthesis is feedback-regulated: the higher cholesterol concentration, the lower the activity of SREBP factors. Additionally, SREBP1a and SREBP2 activate the expression of the gene INSIG-1, which suppresses the processing of SREBP proteins (Fig. 1). It is of importance, that INSIG-1 expression is also under positive control of PPAR γ . Note that the SREBPs are capable of stimulating PPAR γ gene transcription (Fajas *et al.*, 1999) and regulating their activity through the production of endogenous ligand such as free fatty acids (FFA) (not shown in the Fig.). The genes that are regulated by PPAR γ in the adipocytes include those for fatty acids transport proteins FATP and CD36, and for the adipocyte-specific fatty acids binding transport protein aP2, which transports fatty acids to the cell nucleus (Fig. 1). Thus, activation of PPAR γ by fatty acid enhances their uptake by the cell from plasma. Another important implication of the transcriptional activity of PPAR γ is a considerably enhanced supply of plasma cholesterol from lipoprotein particles of different density. PPAR γ positively regulates the expression of the oxidized low-density lipoprotein receptor 1 (OLR1), which is a receptor for oxidized LDL (oxLDL). The LDL receptor-

related protein (LRP) is another target gene for PPAR γ . LRP controls cholesterol access to the cell through specific uptake of cholesterol ethers (CEs) coming from HDL. This process is responsible for at least one-third of selective uptake of HDL-supplied CE by fat cells (Fazio, Linton, 2004). CEs come from HDL and go to an adipocyte via the CE-transfer protein (CETP) too. CETP is highly expressed in fat cells. CETP activity is associated with the adipocyte plasma membrane. In adipocytes, CETP expression is up-regulated by LXR α being under positive control of PPAR γ .

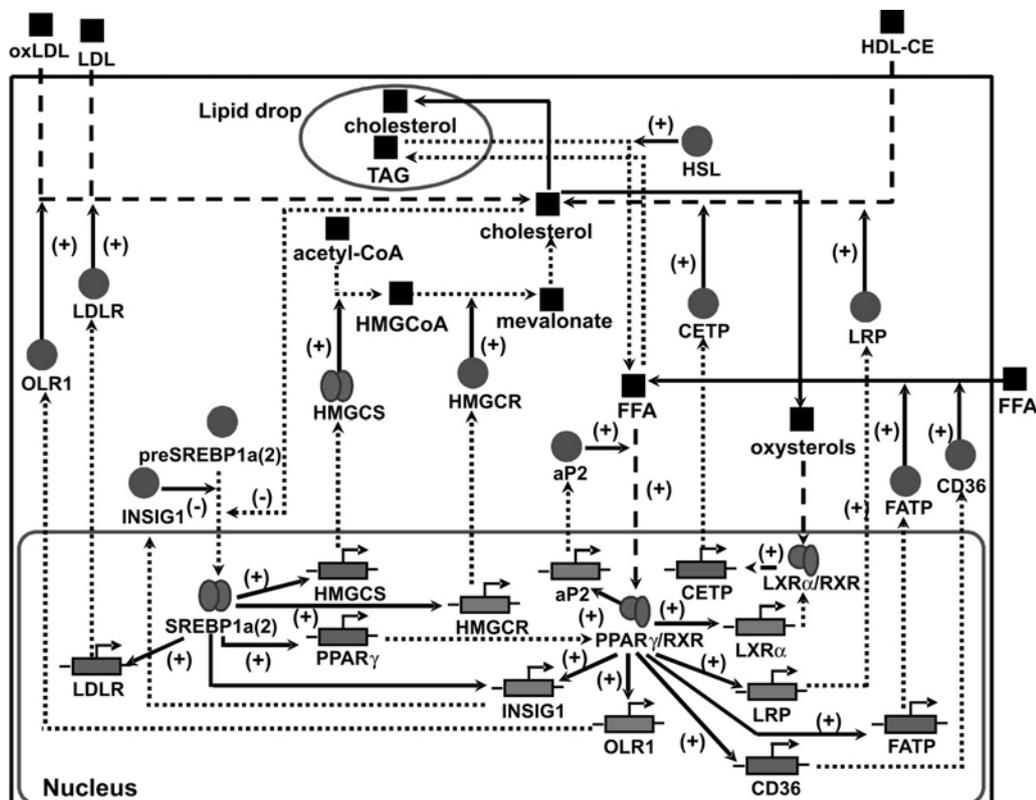


Figure 1. A fragment of adipocyte gene network: regulation of cholesterol metabolism by transcription factors SREBP1a(2), LXR α and PPAR γ .

In summary, we conclude: in adipocytes there is a sequence of regulatory events that coordinated interactions between PPAR γ , LXR α and SREBP unfolds, that supports fatty acid and cholesterol metabolism regulation.

Recognition of putative binding sites for SREBPs in the regulatory regions of a range of human genes for lipolysis and lipid drop formation. The role of SREBPs in regulation of FFA and cholesterol synthesis is established, however this does not for process of lipolytic stimulation and lipid drop formation. One mean to reveal this role is to apply bioinformatic algorithm for recognition putative binding sites for SREBPs (SRE-type) in the regulatory regions of genes involved in the control of lipolysis and lipid drop formation. Using the SITECON method, which allows SRE-type sites to be recognized, we analyzed the promoter regions (–600 to –1 relative to the transcription start site) of 10 human genes involved in regulation of mentioned processes. We have identified 9 SRE-like motives in the promoter regions of 7 genes. The results are presented in Table 1. If the ability of these sites to interact with SREBP proteins can find further experimental support, then a much better understanding of the role SREBPs play in the regulation of metabolic processes has been gained.

Table 1. Putative binding sites for the transcription factors SREBPs. Using the SITECON method, these sites were revealed in the promoter regions (600 bp) of human genes expressed in adipocytes

Gene	N _{SREs}	Putative SRE-site				GenBank acc. number
		Chain	Position*	P**	Sequence	
Control of lipolysis						
GPR109b	1	+	-585/-576	0.771	agcactccat	NT_009755; NM_006018
GPR109a	1	+	-575/-566	0.771	agcactccat	NT_009755; NM_177551
HSL	1	+	-314/-305	0.758	tgactccag	NT_011109; NM_005357
ADRA2A	1	-	-557/-566	0.769	agcagcccag	NT_030059; NM_000681
PRKACB	2	+	-329/-320	0.773	tgactccag	NT_032977;
		-	-576/-585	0.777	ctcagcccac	NM_002731
Proteins associated with lipid droplets						
CAVI	1	+	-13/-4	0.803	agcaccag	NT_007933; NM_001753
CGI-58 (ABHD5)	2	-	-144/-153	0.759	gtcagcccag	NT_022517;
		-	-223/-232	0.748	atcaccag	NM_016006

* Positions are specified concerning the beginning of NM-sequence in the contig (NT); ** Level of conformational similarity with known SRE-type sites.

ACKNOWLEDGEMENTS

The work was supported by Government Contract with the Federal Agency for Science and Technology “Identification of potential targets for novel medicinal drugs based on reconstructed gene networks” 02.434.11.3004 01.04.2005 and Innovation Project of Federal Agency of Science and innovation IT-CP.5/001 “Development of software for computer modeling and design in postgenomic system biology (system biology *in silico*)” 02.467.11.1005 30.09.2005. The authors are grateful to I.V. Likhova for bibliographical support, A.L. Proskura for taking part in populating the gene network, and to Vladimir Filonenko for translating this paper from Russian into English.

REFERENCES

- Amemiya-Kudo M. *et al.* (2002) Transcriptional activities of nuclear SREBP-1a, -1c, and -2 to different target promoters of lipogenic and cholesterologenic genes. *J. Lipid Res.*, **43**, 1220–1235.
- Ananko E.A. *et al.* (2005) GeneNet in 2005. *Nucl. Acids Res.*, **33**, D425–D427.
- Brasaemle D.L. *et al.* (2004) Proteomic analysis of proteins associated with lipid droplets of basal and lipolytically stimulated 3T3-L1 adipocytes. *J. Biol. Chem.*, **279**, 46835–46842.
- Fajas L. *et al.* (1999) Regulation of peroxisome proliferator-activated receptor gamma expression by adipocyte differentiation and determination factor 1/sterol regulatory element binding protein 1: implications for adipocyte differentiation and metabolism. *Mol. Cell. Biol.*, **19**, 5495–5503.
- Fazio S., Linton M.F. (2004) Unique pathway for cholesterol uptake in fat cells. *Arterioscler. Thromb. Vasc. Biol.*, **24**, 1538–1539.
- Oshchepkov D.Y. *et al.* (2004) SITECON: a tool for detecting conservative conformational and physicochemical properties in transcription factor binding site alignments and for site recognition. *Nucl. Acids Res.*, **32**, W208–W212.

NEW MIGRATION SCHEME FOR PARALLEL DIFFERENTIAL EVOLUTION

Kozlov K.N.^{*1}, *Samsonov A.M.*²

¹Dept. of computational biology, State Polytechnical University, St.Petersburg, 195251, Russia;

²A.F. Ioffe Physico-technical Institute, RAS, St.Petersburg, 194021, Russia

* Corresponding author: e-mail: kozlov@spbcas.ru

Key words: differential evolution, optimization, regulatory gene networks

SUMMARY

Motivation: Modern molecular biology has massive amounts of quantitative data already at its disposal. The crucially important problem for getting closer insights into mechanisms of development is to reduce the complexity of finding the parameters of mathematical models by fitting to experimental data. Parallel Differential Evolution is the modern optimization technique that posses natural parallelization.

Results: The new migration scheme for Parallel Differential Evolution has been developed that showed high speed of convergence on the test problem. The dependency of the accuracy of the final result on the period of communication between branches was analyzed.

Availability: available on request from the authors.

INTRODUCTION

Modern molecular biology has massive amounts of quantitative data already at its disposal, and robust and reliable algorithms' development to treat them becomes a foreground job. Mathematical modeling is essential for systematic treatment of experimental results and for getting insights into the structure of underlying natural objects. The crucially important problem is to reduce the complexity of finding the parameters of mathematical models by fitting to experimental data.

We are developing the Combined Optimization Technique (COT) for biological data fitting and perform the experiments in the context of one biological system namely the segment determination gene network of a fruit fly *Drosophila* embryo (Jaeger *et al.*, 2004). In previous studies we used Simulated Annealing (SA) method with a weak quality criterion to obtain the rough approximation of the parameter set, which is refined afterwards by the Optimal Steepest Descent Algorithm (OSDA), see (Kozlov, Samsonov, 2003; Gursky, 2004).

In this work we introduce a new migration scheme for the Parallel Differential Evolution (DE) that is going to be used as an alternative for the SA at the first stage of COT. In contrast to traditional approach to parallelization of DE in which the best member of one branch is used to substitute the randomly chosen member in another branch we suggest to substitute the oldest member of the target branch. In our numerical experiments we used a ring topology for the network of computational nodes.

We present numerical results on optimization using the developed algorithm for the test problem of finding parameters in a network of two genes and the analysis of the dependency of the accuracy of the final result on the period of communication between branches.

METHODS AND ALGORITHMS

Differential Evolution. DE is a stochastic optimization technique that was invented at the end of the previous century by Storn and Price (Storn, Price, 1995). It is an iterative optimization technique that starts from the set of the randomly generated parameter vectors. The set is called population, and the vectors are called individuals. The population on each iteration is referred to as generation. The size of population is fixed. The algorithm applies the following formula on each iteration g for each individual i :

$$v_{g+1}^i = w_g^{r1} + F(w_g^{r2} - w_g^{r3}),$$

where w_g^a is the member of the current generation g with index a , v_{g+1}^i is new individual, F is a predefined scaling constant and $r1, r2, r3$ are random indices of the members of population. The new individual v_{g+1}^i replaces its parent w_g^i if the value of the quality functional for its set of parameters is less than for the latter one.

The original algorithm was highly dependent on internal parameters as reported by different authors, for example (Gaemperle *et al.*, 2002). An efficient adaptive scheme for selection of internal parameters based on the control of the population diversity was developed by Zaharie (Zaharie, 2002). As any evolutionary algorithm, DE can be easily parallelized due to the fact that each member of the population is evaluated individually. The whole population is divided into subpopulations that are sometimes called islands or branches one per each computational node. The individual members of branches are then allowed to migrate, i.e. move, from one branch to another according to predefined topology (Tasoulis *et al.*, 2004). The number of iterations between migrations is called communication period.

We have developed a new migration scheme for the Parallel Differential Evolution in which the best member of one branch is used to substitute the oldest member of the target branch. The computational nodes are organized in a ring and individuals migrate from node k to node $k + 1$ if it exists and from the last one to the first one.

Gene Network Model. The dynamics of the model is described by the system of coupled differential-difference NRD equations formulated in (Reinitz, Sharp, 1995)

$$\frac{\partial v_n^a}{\partial t} = R^a g \left(\sum_{b=0}^{G-1} T^{ab} v_n^b + m^a v_n^{bcd} + h^a \right) + D^a (v_{n-1}^a - 2v_n^a + v_{n+1}^a) - \lambda^a v_n^a. \quad (1)$$

The equation is written for *each* gene product a and *each* nucleus n , and G is the number of genes. A matrix element T^{ab} , one for each pair of proteins, and coefficients $m^a, h^a, R^a, D^a, \lambda^a$ for each protein are unknown parameters which should be determined by means of minimization of a functional equal to the sum of squared differences between the concentrations of the gene products (say, proteins), observed experimentally and calculated independently, e.g., by means of gene network approach.

IMPLEMENTATION AND RESULTS

The algorithms were implemented in ANSI C programming language and MPI was used for parallelization. Runs were performed on the cluster of Ioffe Physical-Technical Institute equipped with Pentium III Xeon working at 2.0 GHz.

To study the convergence of a new method in a lab conditions we produced the artificial gene expression data for the network of two genes in eight nuclei by integration

the model equations, using the set of parameters that represents already known solution. We took the model output for 9 time moments to calculate the functional value. We used the following measure for the accuracy of the optimal approximation of parameter set q^{opt} :

$$\kappa = \max_i \frac{|q_i^{true} - q_i^{opt}|}{|q_i^{true}|} \times 100\%, \quad (2)$$

where q^{true} is the known solution.

To evaluate the new communication strategy we performed 100 runs using 4 mpi nodes with communication periods 5, 10, 15, 20, 25, 30, 35, 40 and 45 keeping all other settings the same. The size of population was set to 7 and equals the number of unknown parameters in the model. Numerical results are given in the Table below. Because of stochastic nature of the DE method the results provided are averaged over all performed experiments.

Table 1. The accuracy of parameter optimization

Communication period	K	Number of functional evaluations
5	417.03	223067
10	7.04	141772
15	64.08	103023
20	22.45	92446
25	5.32	91431
30	5.84	87628
35	4.91	75550
40	5.62	85400
45	5.15	85165

Parameters were recovered with 4.91 % accuracy. The average number of functional evaluations used by parallel DE equals 75,550.

As can be seen from the Table the accuracy of the final result of optimization depends on the communication period. To analyze this dependency we calculated the maximal age A of the member that is substituted at the migration step for each run. Then we calculated the number of runs $N(A)$ that have the same value of A . Fig. 1 shows the plots of $N(A)$ versus A for communication periods 15 (F15), 35 (F35) and 45 (F45).

The curves for periods 35 and 45 that produces more accurate results take the maximal value at 35 and 45 respectively, that is, the quotient of division of $\text{argmax}(N(A))$ by corresponding period is an odd number, 1 in these cases. In contrast, for F15 that produces the result of poor accuracy this quotient equals 2, that is even. This rule holds for other periods also (data not shown).

DISCUSSION

Parallel Differential Evolution is a rather new and promising optimization technique that is able to converge to the global minimum. The number of functional evaluations determines the time necessary to obtain the solution of the data fitting problem because of a huge number of species and, therefore, the differential equations that are to be integrated. In real gene networks this number can exceed three hundreds, and to include more genes in the network under consideration is crucially important for getting closer insights into mechanisms of development.

The results of the analysis of the new migration scheme performed in this work will allow to determine the communication period automatically and, consequently, to use parallel DE as an alternative to SA at the first stage of COT.

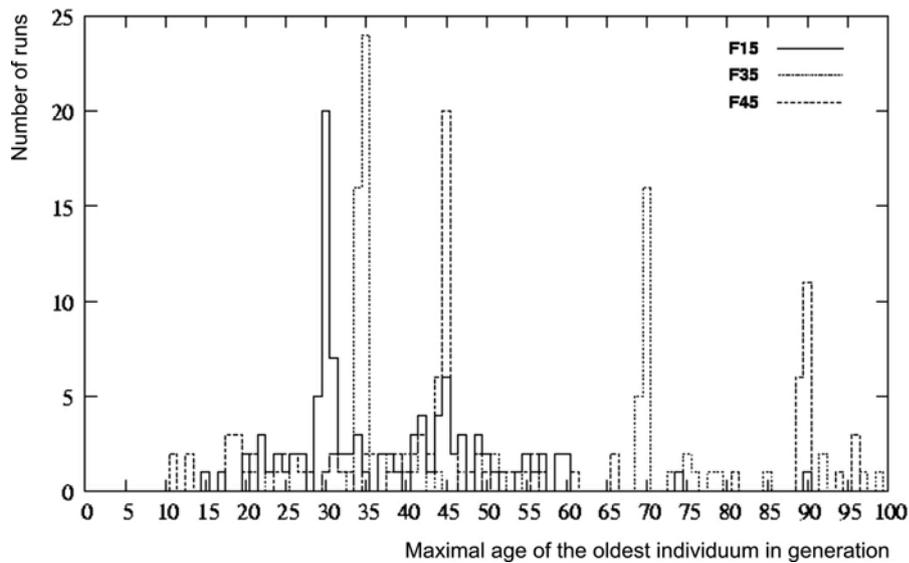


Figure 1. Plots of $N(A)$ vs A , where A is the maximal age of the member that is substituted at the migration and $N(A)$ is the number of runs with the same value of A .

ACKNOWLEDGEMENTS

Our sincere thanks to Dr J. Reinitz and Manu for many valuable discussions. The support of the study by the NIH Grants RR07801, TW01147, the CRDF GAP Awards RBO685, RBO1286 and NWO No. 047.011.2004.013 is gratefully acknowledged.

REFERENCES

- Gaemperle R., Mueller S. D., Koumoutsakos P. (2002) A Parameter Study for Differential Evolution. In Gmela A., Mastorakis N.E. (eds), *Advances in Intelligent Systems, Fuzzy Systems, Evolutionary Computation*, WSEAS Press, pp. 293–298.
- Gursky V.V., Jaeger J., Kozlov K.N., Reinitz J., Samsonov A.M. (2004) Pattern formation and nuclear divisions are uncoupled in *Drosophila* segmentation: Comparison of spatially discrete and continuous models. *Physica D.*, **197**, 286–302.
- Jaeger J., Surkova S., Blagov M., Janssens H., Kosman D., Kozlov K.N., Manu, Myasnikova E., Vanario-Alonso C.E., Samsonova M., Sharp D.H., Reinitz J. (2004) Dynamic control of positional information in the early *Drosophila* embryo. *Nature*, **430**(6997), 368–371.
- Kozlov K.N., Samsonov A.M. (2003) New data processing technique based on the optimal control theory. *Techn. Physics*, **48**, 11, 6–14.
- Reinitz J., Sharp D. (1995) Mechanism of Formation of Eve Stripes. *Mechanisms of Development*, **49**, 133–158.
- Storn R., Price K. (1995) Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces, *Technical Report TR-95-012, ICSI*.
- Tasoulis D.K., Pavlidis N.G., Plagianakos V.P., Vrahatis M.N. (2004) Parallel differential evolution. In *Congress on Evolutionary Computation (CEC 2004)*, Portland, Oregon.
- Zaharie D. (2002) Parameter Adaptation in Differential Evolution by Controlling the Population Diversity. In D. Petcu et al. (eds), *Proc. of 4th International Workshop on Symbolic and Numeric Algorithms for Scientific Computing*, Timisoara, Romania, 385–397.

HCCDB: DATA MINING SYSTEM FOR HUMAN CELL CYCLE GENE

*Milanesi L. *, Alfieri R., Merelli I.*

CNR Institute of Biomedical Technologies, Milan, Italy

* Corresponding author: e-mail: luciano.milanesi@itb.cnr.it

Key words: cell cycle, systems biology, database, bioinformatics, data mining

SUMMARY

Motivation: The cell cycle is a complex biological process that implies the interaction of a large number of genes. Disease studies on tumour proliferation and de-regulation of human cell cycle have to face the problem of finding as quickly as possible the information related to all the genes that are involved in this cellular process. This work aims to implement a new database containing information about the human cell cycle to support studies on genetic diseases related to this crucial biological process.

Results: HCCdb is a data mining system that integrates information related to genes and proteins involved in human cell cycle.

Availability: <http://cellcycle.itb.cnr.it/>

INTRODUCTION

Several resources that collect many biological pathways, such as the cell cycle, are available for different organisms, but in the state of the art there are no specific resources for human cell cycle data integration. The most important resources are Kegg Pathway Database (<http://www.genome.ad.jp/kegg/pathway.html>) and Reactome (<http://www.reactome.org/>).

Kegg acts in a larger range because it is a collection of pathway maps for metabolic processes, genetic and environmental data, such as signal transductions and human diseases. Reactome is a resource for human biological processes, which relies on information about single reactions grouped into pathways.

METHODS

For the models of the gene networks introduced in Likhoshvai *et al.* (2001), we consider here the corresponding 3-dimensional dynamical systems.

The HCCdb database “Human Cell Cycle Database” is a resource, which relies on the data taken from Kegg and Reactome. To integrate the data, we query many resources to collect the information related to each gene and protein previously selected. The database infrastructure is designed to make possible an automated data integration and to query a set of selected biological databases retrieving information about genes and proteins. Moreover, we create a database automatic updating through a pipeline that queries public databases to integrate new data in our resource. The database administrator can access a specific page where he can insert a gene name and perform the pipeline for data integration. As a result, an updating of all tables of the database occurs: in this way the

resource can be maintained up to date. For example, the genes involved in the complete cell cycle pathway, in apoptosis pathway and in MAP kinase signaling pathway are taken from Kegg, while the genes involved in mitotic and checkpoint pathways are taken from Reactome. The main goal of this work is the integration of the data related to each gene or protein. For this reason, users can query the database contents both inserting the gene/protein name and using the IDs of public databases. The query results page is a complete report, and users can browse the data using direct links to the different biological related database. Users can also query the database using keywords: the result is a list of genes related to the query. HCCdb data are stored in a relational database based on MySQL DBMS. HCCdb data mining system is made up using a “snowflakes” schema, which present the important information about genes and proteins and auxiliary data. In particular, for each gene we give the promoter information, validated experimental data about gene’s expression and quantitative PCR primers. The HCCdb database can be used to find useful information about each protein, such as complexes, protein–protein interactions, protein structures and surfaces, as show in Fig. 1 for BCL2 gene structure. The HCCdb database is accessible through a web interface made up of a set of HTML pages dynamically generated from PHP scripts.

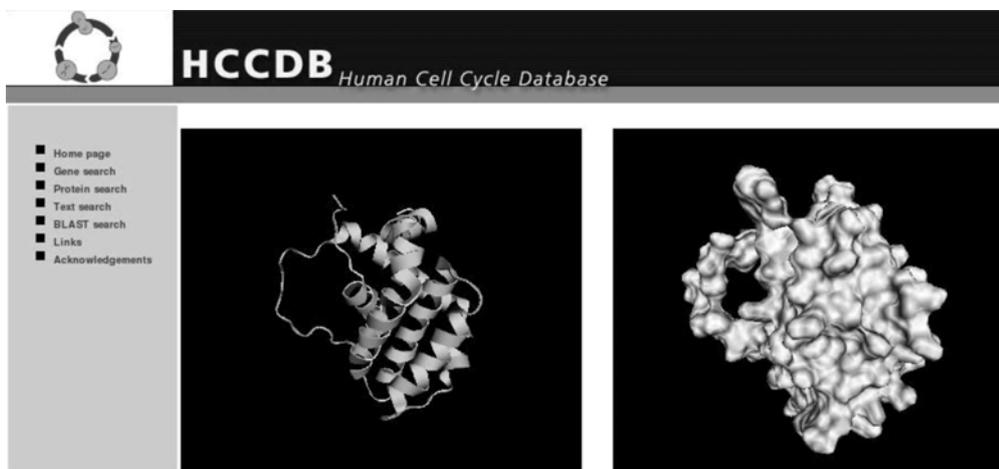


Figure 1. An example of an HCCdb entry for BCL2 protein structures and surfaces visualizations.

RESULTS AND DISCUSSION

We have developed a data integration system by integrating data from many biological resources: NCBI, Ensemble, Kegg, Reactome, dbSNP, MGC, DBTSS, Unigene, QPPD, TRANSFAC, UniProt, InterPro, PDB, TRANSPATH, BIND, MINT, and IntAct. HCCdb database use the data warehouse approach that allows for bringing all the related data in a single database and experimental data. This approach is used in order

- To develop a unified data model that can accommodate all the information stored in various source databases;
- To develop programs that will take data from the source databases, transform them to match the unified data model and load them into the warehouse;
- To increase the efficiency of retrieving the specific information related to a specific query;
- To perform data mining of different kinds of information through a single query; and
- To increase the information accuracy and better control over the information standard.

The HCCDB is a MySQL relational database with a “snowflake” schema aimed to collect principal information about genes and different kind of auxiliary data related to genes and proteins. An example of the “Gene report” from HCCdb by using BCL2 gene as input query is shown in Fig. 2.

HCCDB Human Cell Cycle Database

- Home page
- Gene search
- Protein search
- Text search
- BLAST search
- Links
- Acknowledgements

Protein report: BCL2_HUMAN

Apoptosis regulator Bcl-2

Alternative names:

Belongs to Bcl-2 family

Gene name: BCL2

Sequence information:

protein length	protein sequence
239 aa	view sequence

Protein links to other biological databases:

UNIPROT	ENTREZ PROTEIN	GO
P10415	NP_000624	GO:0005741

InterPro Domains:

domain	domain database	position
IPR000712	PS01080	137-155
IPR000712	PS01258	188-199
IPR000712	PS01259	93-107
IPR000712	SM00337	97-195
IPR002475	PS50062	97-197
IPR003093	PF02180	7-33
IPR003093	PS01260	10-30
IPR003093	PS50063	11-30
IPR003093	SM00265	7-33
IPR004725	PIRSF500115	1-239
IPR004725	TIGR00865	1-239
IPR012238	PIRSF001714	1-239
IPR013278	PRO1863	1-18 36-60 81-98 204-221

Protein structure and surface:

- 1G5M view
- 1GJH view

Protein interactions:

- Bind
- Mint
- IntAct

Protein complexes:

- Transpath molecule

Figure 2. An example of a “Gene report” from HCCdb by using BCL2 gene as an input query. The report collects the information integrating data from different resources.

ACKNOWLEDGEMENTS

The work was supported by the grant No.03-51-5218 INTAS, European SSH BioinfoGRID and MIUR FIRB LITBIO projects. The authors are grateful to J. Hatton and C. Bishop, for helpful discussions and Web interface.

HCV-KINET DATABASE: KINETIC PARAMETER REACTIONS AND REGULATORY PROCESSES OF THE LIFE CYCLE OF THE HEPATITIS C VIRUS

*Mishchenko E.L.¹, Korotkov R.O.², Ivanisenko V.A.^{*1,2}*

¹Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ²Novosibirsk State University, Novosibirsk, 630090, Russia

* Corresponding author: e-mail: salix@bionet.nsc.ru

Key words: hepatitis C virus, database, kinetic parameters

SUMMARY

Motivation: To build mathematical models that describe the dynamic behavior of the life cycle of the hepatitis C virus, the development of a database for the kinetic parameters of reactions and the regulatory processes of a given gene network is required.

Results: The created HCV-KINET contained the kinetic parameters for the enzymic reactions and the regulatory processes (K_m , V_{max} , k_{cat} , specific activity, K_i , IC_{50} , the direct and reverse reaction constants, the equilibrium constants of the dissociation (K_d) and association (K_a) of protein-protein and nucleoprotein complexes, half-life of proteins, enzymes, RNA, process duration, kinetic parameters for mutant proteins and enzymes). The total number of kinetic parameter stored is 363, the coordinates for the points of the kinetic plots are stored in the HCV-KINET, too.

INTRODUCTION

Infection with the hepatitis C virus (HCV) is a major cause of chronic liver diseases leading to liver cirrhosis and hepatocellular carcinoma (Alter, 1997; Okuda, 2000). According to current estimates about 170 million people are worldwide infected with HCV, and vaccine has not been as yet produced to combat HCV. Earlier a gene network that stores the main life cycle of HCV have been reconstructed: its entry into the cell, translation of the positive RNA chain, processing of the primary polyprotein with yield of structural and nonstructural proteins, RNA replication and virion assembly (Bezmaternikh *et al.*, 2006). The gene network contains 63 proteins, 16 genes, 19 RNAs, and 141 reaction and regulatory processes; it incorporates the main functions of both viral and human objects and their complex interplay. The idea was to develop a mathematical model that would adequately express gene network function as differential equations of reaction rates (the formal dynamic variables in the descriptions are genes, RNA, protein-protein and nucleoprotein complexes, low molecular weight compounds (Likhoshvai *et al.*, 2001). To achieve this information about the kinetic parameters of the reactions and the regulatory processes could be gathered. Mathematical modeling would allow us to calculate the dynamics of the processes of the gene network, to identify the limiting stages, to predict the behavior of the mathematical model in the presence of the inhibitors of enzymes that catalyse the key reactions and may be potentially useful as drugs. This study is concerned with the development of the HCV-KINET database for the kinetic parameters and the regulatory processes of the HCV life cycle.

METHODS AND ALGORITHMS

HCV-KINET is a relational database. Its main requirement during implementation was the feasibility of enhancement. Access to HCV-KINET is available through the web-site (the development of the web-site is underway). A manager providing the user with ways and means to navigate at various levels of data organization and also with processing methods are supplied. The HCV-KINET manager is an appendix that enables to manipulate, view, and administer the data stored in this database. The appendix is implemented on the NET platform, in the C# language for the main reason that. NET provides good possibilities for enhancing the appendices. An objective model of data is implemented in HCV-KINET. The Composite design pattern is modified. It allows to uniformly process the individual and composite objects and also to easily add new ones. Moreover, new data and methods for their processing are provided through the use of the pattern: Observer and Visitor.

RESULTS AND DISCUSSION

The HCV-KINET database for the kinetic parameters of reactions and regulatory processes of the gene network for the HCV life cycle was derived from about 100 relevant scientific reports. HCV KINET contains the kinetic parameters for enzymic reactions (K_m , V_{max} , k_{cat} , specific activity, K_i , IC_{50} , the direct and reverse reaction constants, the equilibrium constants of the dissociation (K_d) and association (K_a) of protein-protein and nucleoprotein complexes, half-life of proteins, enzymes, RNA, process duration, kinetic parameters for mutant proteins and enzymes. The total number of kinetic parameter stored is 363, the coordinates for the points of the kinetic plots are stored in the HCV-KINET as well. HCV-KINET consists of blocks for protein description, description of methods for defining kinetic parameters and experimental conditions, kinetic parameters, literature references. The HCV-KINET format amply reflects the input experimental data. Fig. 1 presents the HCV-Kinet database tables to illustrate its structure. The Investigation Table describes a process, a reaction, and the applied methods; the Experiment Table contains data on the name of the substrate or inhibitor, also references to the other tables that define the particular experimental conditions, for example, the Main Conditions Table contains data on the temperature, pH and other relevant data; the Buffer Table gives the name and concentration of the buffer; the Constants Table gives the types and values of the kinetic parameters, with each parameter listed in this Table. Furthermore, the HCV-Kinet database is provided with Tables with references to the literature (the Article Table), structural data on protein (the Molecule Table), data on mutations (the Mutant Table), organism is also tabulated (the Source Table), the HCV-Kinet contains other tables not mentioned here.

There currently exist numerous useful HCV databases. The InterPro database contains general information about the viral protein genome (<http://www.ebi.ac.uk/interpro/IEntry?ac=IPR000745>), most others contain extensive information about HCV nucleotide sequences and genotypes (Kuiken *et al.*, 2005, <http://hcv.lanl.gov/content/hcv-db/HTML/otherweb.html>, <http://pbil.univ-lyon1.fr/events/jobim2005/proceedings/E08Charavay.pdf>), others are designed for immunology, immune epitopes (Yusim *et al.*, 2005), antibodies (http://hcv.lanl.gov/content/immuno/ab_search), antiviral therapy, chronic disease carriers, risk factors, HCV-affected populations (<http://www.healthyhepper.com/db/dbsearch.htm>, http://www.cdc.gov/ncidod/diseases/hepatitis/c/plan/HCV_infection.htm), diagnostic tests, (http://www.researchandmarkets.com/reportinfo.asp?cat_id=0&report_id=53207&q=global%20market%20for%20hcv%20diagnostic%20testing&p=1) and others serve many other purposes. Regrettably, all the HCV databases do not contain quantitative data about the kinetic parameters of the reactions and regulatory processes ongoing in the gene network of

the HCV life cycle. This lack of knowledge must be bridged. HCV-KINET is unique, requisite for building mathematical models for quantitative and, hence, more accurate estimation of the gene network of the HCV life cycle.

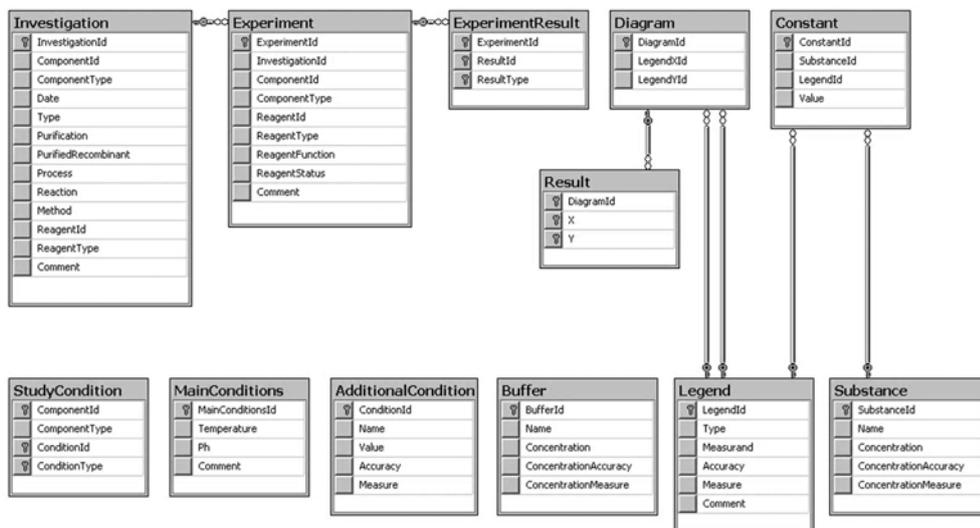


Figure 1. Fragment of database structure for the Gene network HCV life cycle reactions and regulatory processes kinetic parameters.

ACKNOWLEDGEMENTS

Work was supported in part by Russian Foundation for Basic Research No. 05-04-49283, the State contract No. 02.434.11.3004 of 01.04.2005 and No. 02.467.11.1005 of 30.09.2005 with the Federal Agency for science and innovation "Identification of promising targets for the action of new drugs on the basis of gene network reconstruction" federal goal-oriented technical program "Study and development of priority directions in science and technique, 2002–2006, Interdisciplinary integrative project for basic research of the SB RAS No. 115 "Development of intellectual and informational technologies for the generation and analysis of knowledge to support basic research in the area of natural science".

REFERENCES

- Alter M.J.H. (1997) Epidemiology of hepatitis C. *Hepatology*, **26**, 62–65.
- Bezmaternikh K.D., Mishchenko E.L., Ratushny A.V., Likhoshvai V.A., Khlebodarova T.M., Ivanisenko V.A. (2006) A minimum mathematical model for suppression HCV RNA replication in cell culture. *This issue*.
- Kuiken C., Yusim K., Boykin L., Richardson R. (2005) The Los Alamos hepatitis C sequence database. *Bioinformatics*, **21**, 379–384.
- Likhoshvai V.A., Matushkin Y.G., Ratushny A.V., Anan'ko E.A., Ignat'eva E.V., Podkolodnaya O.A. (2001) Generalized chemokinetic method for gene network simulation. *Mol. Biol.*, **35**, 1072–1079.
- Okuda K. (2000) Hepatocellular carcinoma. *J. Hepatology*, **32**, 225–237.
- Yusim K., Richardson R., Tao N., Dalwani A., Agrawal A., Szinger J., Funkhouser R., Korber B., Kuiken C. (2005) Los alamos hepatitis C immunology database. *Appl. Bioinformatics*, **4**, 217–225.

MATHEMATICAL MODELING OF RECEPTOR MEDIATED ENDOCYTOSIS OF LOW-DENSITY LIPOPROTEINS AND THEIR DEGRADATION IN LYSOSOMES

Ratushny A.V. *

Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia

* Corresponding author: e-mail: ratushny@bionet.nsc.ru

Key words: mathematical modeling, endocytosis, LDL, receptor

SUMMARY

Motivation: Receptor-mediated endocytosis of low-density lipoproteins (LDL), their transport within endosomes, and subsequent degradation in lysosomes are essential components of the molecular system for cholesterol homeostasis in vertebrate cells. Construction of a detailed mathematical model of these processes would allow comprehensive study of their molecular mechanisms and evaluation of the effect of various mutations and disorders of this system on the gene net controlling intracellular cholesterol homeostasis.

Results: The receptor-mediated endocytosis of LDL particles and their subsequent degradation in the cell has been modeled. The network of mono- and bimolecular reactions best describing the system has been constructed. The results of calculation of kinetic parameters of the molecular system obtained with the use of the model are in agreement with experimental evidence.

INTRODUCTION

Receptor-mediated endocytosis of LDL particles and their subsequent degradation in lysosomes involve at least seven stages: (1) exposure of synthesized receptor molecules on the cell surface; (2) binding of the receptor to the ligand (LDL particle); (3) internalization of the LDL-receptor complex; (4) dissociation of the receptor from the ligand; (5) return of the receptor on the cell surface; (6) fusion between endosomes and lysosomes; (7) degradation of LDL in lysosomes to amino acids, cholesterol, and other lipids (Brown, Goldstein, 1979).

We developed a mathematical model of receptor-mediated endocytosis of LDL particles and their subsequent degradation in the cell. A sequence of mono- and bimolecular reactions was developed to provide an adequate description of the system under study in terms of chemical kinetics. The results of calculation of kinetic parameters of the components of the molecular system derived from the model are in agreement with experimental evidence.

METHODS AND ALGORITHMS

The generalized chemical-kinetic method (Likhoshvai *et al.*, 2001) was used for simulation of receptor-mediated endocytosis of LDL particles, their transport within endosomes, and subsequent degradation in lysosomes.

RESULTS

A minimal mathematical model of receptor-mediated endocytosis of LDL particles and their subsequent degradation in the cell was constructed in terms of mono- and bimolecular reactions as shown in Fig. 1a. In the model, all receptors on the cell surface are internalized at a certain rate, including those not associated with the ligand. All internalized LDL within endosomes reach lysosomes at a certain speed to be degraded there.

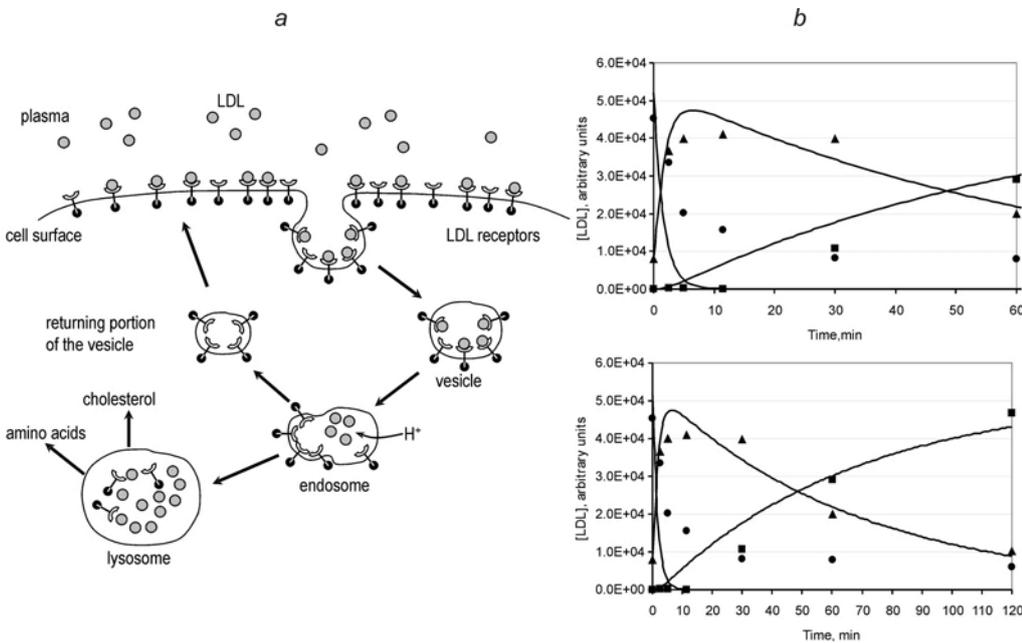


Figure 1. (a) Scheme I. Receptor-mediated endocytosis of LDL particles followed by LDL degradation in lysosomes (for details see text). (b) Comparison of the results of simulation according to scheme I (panel a) with experimental data on the kinetics of receptor internalization and LDL degradation in the cell (Brown, Goldstein, 1979). Experimental data are designated as follows: ● – amount of receptor-bound LDL particles on the cell surface; ▲ – internalized LDL concentration; ■ – degraded LDL concentration. Simulation results are shown with solid curves.

Fig. 1b presents the results of simulation according to the model under discussion compared with experimental data on the kinetics of receptor internalization and LDL degradation in the cell. The Fig. shows that the simulation does not illustrate the experimental data (Brown, Goldstein, 1979) quite perfectly. The simulation assumes exponential decrease in the concentration of ligand-bound receptors on the cell surface, so that the kinetic curve of LDL degradation does not flatten out within the time span of measurements from 5 to 30 min.

Analysis of the kinetic curves obtained by Brown and Goldstein (1979) and shown in Fig. 1b suggests that not all receptors on the cell surface can be internalized during endocytosis. The experiment shows that endocytosis is confined to certain sites on the plasma membrane. These sites are known as coated pits (Roth, Porter, 1964), and their

formation involves the protein clathrin (Pearse, 1975). Coated pits cover about 2 % of the surface of human fibroblasts and contain 50 to 80 % of LDL receptors (Orci *et al.*, 1978). This fact provides an explanation for the kinetic curves (Figs. 1*b* and 2*b*) for internalization of LDL-bound receptors. Rapid uptake of 70–80 % of the receptors occurs within the first 5–10 min. In the simulation, the mean time of the first stage of endocytosis (invagination of the coated pit together with receptors and its uptake into the cells) is estimated to be 2 min. The subsequent flattening-out of the kinetic curve is determined by buffering of the system owing to receptors not reaching coated pits during diffusion. To correct for this fact in the model, receptors on the cell surface are divided into two exchanging pools. Receptors of the first pool are located outside coated pits; thus, they are not involved in endocytosis, and vice versa (Fig. 2*a*).

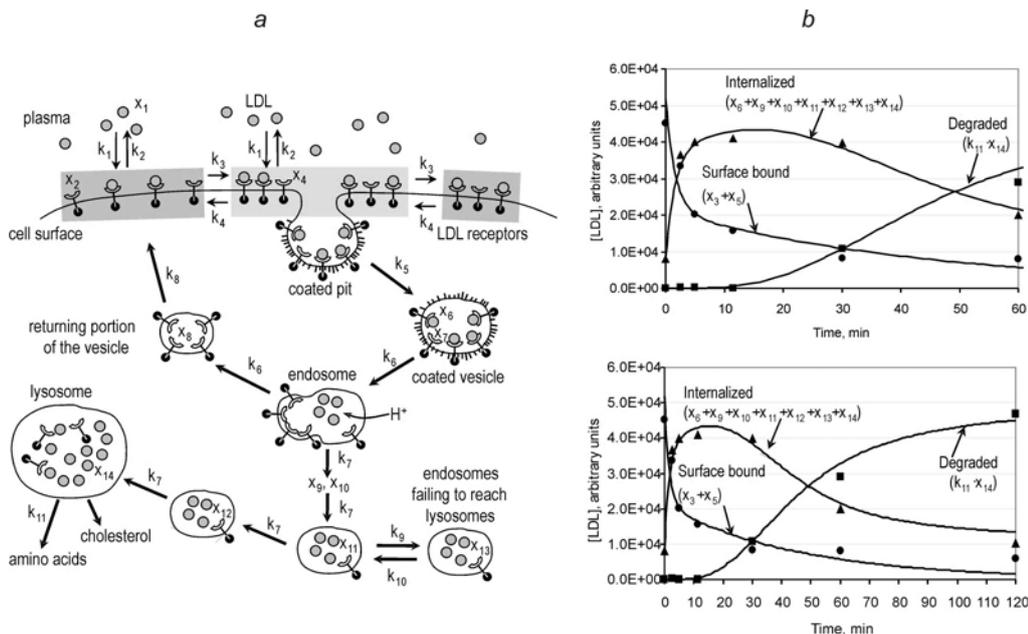


Figure 2. (a) Scheme II. Receptor-mediated endocytosis of LDL particles followed by LDL degradation in lysosomes (for details see text). (b) Comparison of predictions according to scheme II (panel a) with experimental data on the kinetics of receptor internalization and LDL degradation in the cell (Brown, Goldstein, 1979). Designations follow Fig. 1. The initial distribution of LDL-bound receptors was taken to be: receptors in the area of coated pits, 61.5 %; outside coated pits, 38.5 %.

It is known that after internalization bordered vesicles interact with small cellular organelles, smooth vesicles, to form endosomes (Klimov, Nikul'cheva, 1999). We explain the flattened kinetics of the concentration of uptaken LDL within the time span of measurement, 5–30 min, by the fact that LDL particles within endosomes cannot immediately reach lysosomes, which contain LDL-degrading enzymes. Actually, a lag occurs that is related to the time required for LDL transport by endosomes to lysosomes. To take into account the LDL delivery lag, the model includes four endosome pools passing into one another, and a "deadlock" pool, which mimics a certain amount of undegraded LDL in the cell. As shown in Fig. 2*b*, LDL degradation slows down after the second hour of measurements and virtually flattens out. The mean time of pass from one endosome pool to the next one in the simulation is taken to be ca. 8 min. The release of LDL receptors from ligands in the simulation occurs at the first stage owing to the acidic pH (~5) in endosomes (Klimov, Nikul'cheva, 1999). The time of receptor return on the cell surface is taken to be ca. 10 min. The overall time of receptor recycling is 20 min (2+8+10), which is in agreement with experimental evidence (Klimov, Nikul'cheva, 1999). The mean time of LDL degradation in the lysosome and release of about

475 cholesterol molecules and 1310 cholesterol ester molecules was estimated from the kinetic curve of LDL degradation in the cell (Fig. 2b, ■) and taken to be ca. 1 min.

Thus the new model constructed according to the processes schematically presented in Fig. 2a takes the form:

$$\left\{ \begin{array}{l} \frac{dx_1}{dt} = -k_1 \cdot x_1 \cdot x_2 + k_2 \cdot x_3 - k_1 \cdot x_1 \cdot x_4 + k_2 \cdot x_5 \\ \frac{dx_2}{dt} = -k_1 \cdot x_1 \cdot x_2 + k_2 \cdot x_3 - k_3 \cdot x_2 + k_4 \cdot x_4 + k_8 \cdot x_8 \\ \frac{dx_3}{dt} = k_1 \cdot x_1 \cdot x_2 - k_2 \cdot x_3 - k_3 \cdot x_3 + k_4 \cdot x_5 \\ \frac{dx_4}{dt} = -k_1 \cdot x_1 \cdot x_4 - k_2 \cdot x_5 - k_3 \cdot x_2 - (k_4 + k_5) \cdot x_4 \\ \frac{dx_5}{dt} = -k_1 \cdot x_1 \cdot x_4 + k_2 \cdot x_5 + k_3 \cdot x_2 - (k_4 + k_5) \cdot x_5 \\ \frac{dx_6}{dt} = k_5 \cdot x_5 - k_7 \cdot x_6 \\ \frac{dx_7}{dt} = k_5 \cdot x_4 - k_7 \cdot x_7 \end{array} \right. \quad \left\{ \begin{array}{l} \frac{dx_8}{dt} = k_6 \cdot (x_6 + x_7) - k_8 \cdot x_8 \\ \frac{dx_9}{dt} = k_7 \cdot (x_6 - x_9) \\ \frac{dx_{10}}{dt} = k_7 \cdot (x_9 - x_{10}) \\ \frac{dx_{11}}{dt} = k_7 \cdot (x_{10} - x_{11}) - k_9 \cdot x_{11} + k_{10} \cdot x_{13} \\ \frac{dx_{12}}{dt} = k_7 \cdot (x_{11} - x_{12}) \\ \frac{dx_{13}}{dt} = k_9 \cdot x_{11} - k_{10} \cdot x_{13} \\ \frac{dx_{14}}{dt} = k_7 \cdot x_{12} - k_{11} \cdot x_{14} \end{array} \right. \quad (1)$$

where x_i are concentrations of LDL particles and their receptors in corresponding states (Fig. 2a) and k_i are constants of the processes presented in Fig. 2a.

Model (1) describes the experimental evidence on the kinetics of LDL internalization and degradation more precisely (Fig. 2b). The kinetic curves were simulated for the following parameters: $k_1 = 1.11 \cdot 10^{-5}$ (molecules/cell) $^{-1} \cdot s^{-1}$, $k_2 = 10^{-4} s^{-1}$, $k_3 = 10^{-4} s^{-1}$, $k_4 = 3 \cdot 10^{-4} s^{-1}$, $k_5 = 10^{-2} s^{-1}$, $k_6 = 2 \cdot 10^{-3} s^{-1}$, $k_7 = 2 \cdot 10^{-3} s^{-1}$, $k_8 = 1.67 \cdot 10^{-3} s^{-1}$, $k_9 = 5 \cdot 10^{-4} s^{-1}$, $k_{10} = 10^{-5} s^{-1}$, and $k_{11} = 2.78 \cdot 10^{-2} s^{-1}$. Nevertheless, the uptake of receptors on the cell surface in this simulation from the first to second hour of measurement slightly exceeds the experimental values (Fig. 2b, ●). To correct this discrepancy, we accept an addition pool of receptors on the cell surface. This corrected model allows virtually perfect agreement with experimental data (data not shown).

Although this model provides a more precise description of the actual situation, it raises several questions lacking unambiguous answers, for example: how to describe receptor exposure on the cell surface, to what pool they come, and how they are distributed. To meet these questions, we accept the previous model, which provides simple answers. As the area of coated pits constitutes about 2 % of the cell surface, it is safe to assume that receptor exposure occurs mainly outside them.

DISCUSSION

A mathematical model of receptor-mediated endocytosis of LDL proteins, endosome transport, LDL degradation in lysosomes, and regeneration of LDL receptors on the cell surface has been constructed and tested. The example of this model illustrates the importance of consideration of buffering and delay, occurring in each living organism and reflecting the actual kinetics (response time etc.) of cell operation. Neglect of these features of living systems highly probably results in errors in the description of the behavior of a system under study.

Receptor-mediated endocytosis is essential for many cell processes. An example is transport of iron ions in the protein transferrin, which is bound by corresponding receptors on the cell surface and absorbed by the cell according to the above-described mechanisms. Thus, the mathematical simulation can also be used for describing other molecular mechanisms of interaction between cells and environment by means of vesicular transport.

ACKNOWLEDGEMENTS

Work was supported in part by the Russian Foundation for Basic Research No. 04-01-00458, and NSF:FIBR (Grant EF-0330786) and by Innovation project of Federal Agency of Science and innovation IT-CP.5/001 "Development of software for computer modeling and design in postgenomic system biology (system biology *in silico*)". The author is grateful to I.V. Lokhova for bibliographical support and to V.V. Gulevich for translating the manuscript from Russian into English.

REFERENCES

- Brown M.S., Goldstein J.L. (1979) Receptor-mediated endocytosis: insights from the lipoprotein receptor system. *Proc. Natl. Acad. Sci. USA*, **76**(7), 3330–3337.
- Klimov A.N., Nikul'cheva N.G. (1999) *Lipid and Lipoprotein Metabolism and Its Disturbances*. St. Petersburg: Piter Kom.
- Likhoshvai V.A., Matushkin Iu.G., Ratushnyi A.V., Anan'ko E.A., Ignat'eva E.V., Podkolodnaia O.A. (2001) A generalized chemical-kinetic method for modeling gene networks. *Mol. Biol. (Mosk)*, **35**(6), 1072–1079.
- Orci L., Carpentier J.L., Perrelet A., Anderson R.G., Goldstein J.L., Brown M.S. (1978) Occurrence of low density lipoprotein receptors within large pits on the surface of human fibroblasts as demonstrated by freeze-etching. *Exp. Cell. Res.*, **113**(1), 1–13.
- Pearse B.M. (1975) Coated vesicles from pig brain: purification and biochemical characterization. *J Mol Biol.*, **97**(1), 93–98.
- Roth T.F., Porter K.R. (1964) Yolk protein uptake in the oocyte of the mosquito *Aedes Aegypti*. *L. J. Cell. Biol.*, **20**, 313–332.

CONSERVED PROPERTIES OF ENZYMATIC SYSTEMS: PRENYLTRANSFERASE KINETICS

*Ratushny A.V.**, *Bezmaternikh K.D.*

Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia

* Corresponding author: e-mail: ratushny@bionet.nsc.ru

Key words: mathematical modeling, enzymatic reaction, isoprenoid and cholesterol biosynthesis, prenyltransferase

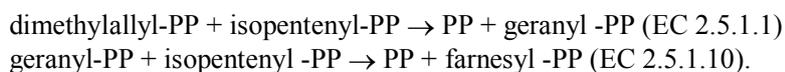
SUMMARY

Motivation: Prenyltransferase is a key enzyme of the biosynthesis of isoprenoids. Closer examination of the kinetic properties of prenyltransferase, development of mathematical models of its regulatory activity would broaden our understanding of the properties of the molecular systems that control the synthesis of sterols, sesquiterpenes, dolichols, ubiquinones and carotenoids, and also farnesylation and geranylgeranylation of proteins.

Results: A computer-aided mathematical model that describes the steady-state rate of the enzymatic reaction of farnesyl diphosphate synthesis was developed. Optimum parameter values were chosen using data that describes the function of human prenyltransferase. It was demonstrated that the function of chicken prenyltransferase can be described using the same proposed model.

INTRODUCTION

Prenyltransferase is a key enzyme along the biosynthesis pathway of isoprenoids that act as the molecular precursors of various classes. These classes contain sterols, sesquiterpenes, dolichols, ubiquinones and carotenoids, and also serve as substrates of the farnesylation and geranylgeranylation of proteins (Szkopinska, Plochocka, 2005). Prenyltransferase catalyzes the sequential irreversible 1'-4 condensation of dimethylallyl diphosphate (DPP¹⁰) and geranyl diphosphate (GPP) with isopentenyl diphosphate (IPP) (Barnard, Popjak, 1981):



In this work, a mathematical model that describes the steady-state rate of the enzymatic reaction of farnesyl diphosphate (FPP) synthesis was developed and verified using the data on human hepatocytes. As a result, the model using the same parameter set describes the function of chicken prenyltransferase.

¹⁰ The abbreviations used are: DPP, Dimethylallyl diphosphate; FPP, Farnesyl diphosphate; GPP, Geranyl diphosphate; IPP, Isopentenyl diphosphate.

MODEL

From the available experimental data on the kinetics of FPP synthesis (Barnard, Popjak, 1981) catalyzed by human prenyltransferase (Fig. 1, 2) it may be inferred that the effects of the substrates GPP and IPP on the rate of the enzymatic reaction are complex and nonlinear. Thus, Fig. 2 shows that IPP at high concentrations acts as a reaction inhibitor. Barnard and Popjak (1981) have described the ordered-sequential bi-bi mechanism of this reaction. Cleland (1963) has proposed a mathematical model that describes the steady state of the reaction involving this mechanism. Barnard and Popjak have also reported that, at high concentrations, IPP is inhibitory. To take fuller consideration of this observation and of the inhibitory effect of the FPP product on the reaction rate, we offer the following mathematical model:

$$V = \frac{k_{cat} \cdot e_0 \cdot \frac{S_1}{K_{m,S_1}} \cdot \frac{S_2}{K_{m,S_2}}}{\left(1 + \frac{S_1}{K_{m,S_1}} + \frac{S_2}{K_{m,S_2}} \cdot \left(1 + \frac{S_2}{K_{i,S_2}}\right) + \frac{S_1}{K_{m,S_1}} \cdot \frac{S_2}{K_{m,S_2}}\right) + \frac{P}{K_{i,P}}}, \quad (1)$$

where V is steady-state reaction rate; e_0 is the total prenyltransferase concentration; S_1 , S_2 and P are the GPP, IPP and FPP concentrations, respectively; k_{cat} is the catalytic constant; K_{m,S_1} and K_{m,S_2} are the Michaelis-Menten constants for GPP and IPP, respectively; K_{i,S_2} denotes the IPP inhibitory constant; $K_{i,P}$ is the substrate inhibitory constant for FPP. The effect of diphosphate on the enzymatic reaction rate is disregarded.

RESULTS AND DISCUSSION

To implement model (1) we chose an optimum parameter set such that provides reasonable well agreement with the experimental data reported by Barnard and Popjak, (1981) (Figs. 1–2). Fig. 1 compares the results calculated using model (1) with the experimental data on the effect of various IPP and GPP concentrations on the human hepatocyte prenyltransferase catalyzed reaction rate. The experimental data are represented as dots. The continuous lines are the estimates calculated using model (1). The calculated curves were drawn using the following parameter values: $k_{cat} = 1.5$ 1/s; $K_{m,S_1} = 0.23$ mM; $K_{m,S_2} = 0.65$ mM; $K_{i,S_2} = 1.11$ mM.

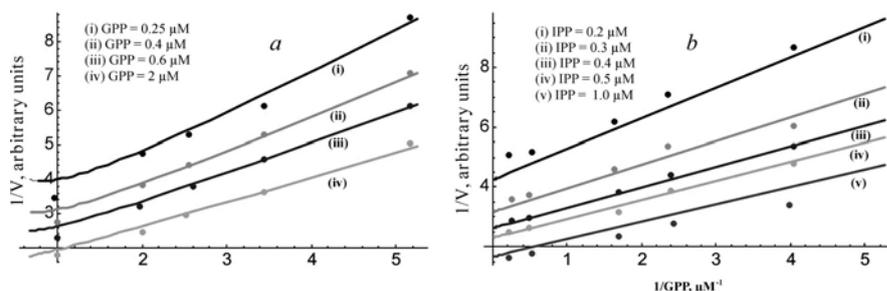


Figure 1. The kinetics of human prenyltransferase at low substrate concentrations. Fig (a) shows the dependency of the inverse reaction rate on the inverse value of IPP concentrations at various GPP concentrations: (i) 0.25 μ M; (ii) 0.4 μ M; (iii) 0.6 μ M; (iv) 2 μ M. Fig (b) presents the dependence of the inverse reaction rate on the fixed IPP concentrations: (i) 0.2 μ M; (ii) 0.3 μ M; (iii) 0.4 μ M; (iv) 0.5 μ M; (v) 1.0 μ M. Dots indicate experimental data according to (Barnard and Popjak, 1981); the curves are the result of simulation using model (1); parameter values: $k_{cat} = 1.5$ s⁻¹; $K_{m,S_1} = 0.23$ mM; $K_{m,S_2} = 0.65$ mM; $K_{i,S_2} = 1.11$ mM.

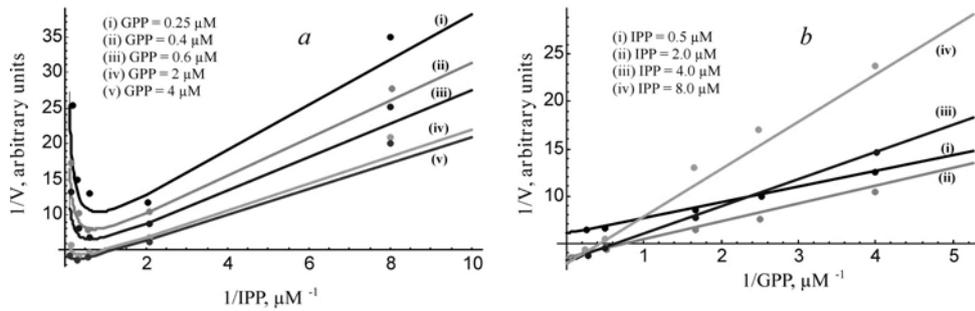


Figure 2. The kinetics of human prenyltransferase at low substrate concentrations. Fig (a) shows the dependency of the inverse reaction rate on the inverse value of IPP concentrations at various fixed GPP concentrations: (i) 0.25 μM ; (ii) 0.4 μM ; (iii) 0.6 μM ; (iv) 2 μM ; (v) 4 μM . Fig (b) presents the dependence of the inverse reaction rate on the various fixed IPP concentrations: (i) 0.5 μM ; (ii) 2.0 μM ; (iii) 4.0 μM ; (iv) 8.0 μM . Dots indicate experimental data according to (Barnard and Popjak, 1981); the curves are the result of simulation using model (1); parameter values: $k_{\text{cat}} = 1.5 \text{ s}^{-1}$; $K_{\text{m},\text{S}1} = 0.23 \text{ mM}$; $K_{\text{m},\text{S}2} = 0.65 \text{ mM}$; $K_{\text{i},\text{S}2} = 1.11 \text{ mM}$.

The kinetics curves in Fig. 1a for the low substrate concentrations show that there is no substrate inhibitory effect of IPP activity by human prenyltransferase. This feature becomes more prominent when the reaction rate is measured in a wider range of changes in IPP concentration (Fig. 2a). The proposed model (1) describes these inhibition by IPP ($K_{\text{i},\text{S}2} = 1.11 \text{ mM}$).

The model of the function of human hepatocyte prenyltransferase was supported by experimental kinetic data for chicken hepatocyte prenyltransferase (Fig. 3). It is noteworthy that model (1) described very well kinetic data for avian hepatocytes using the same kinetic parameters as for human prenyltransferase. The dots on the graphs represent the experimental data from Reed and Rilling (1975). The estimates calculated using model (1) are represented by continuous lines.

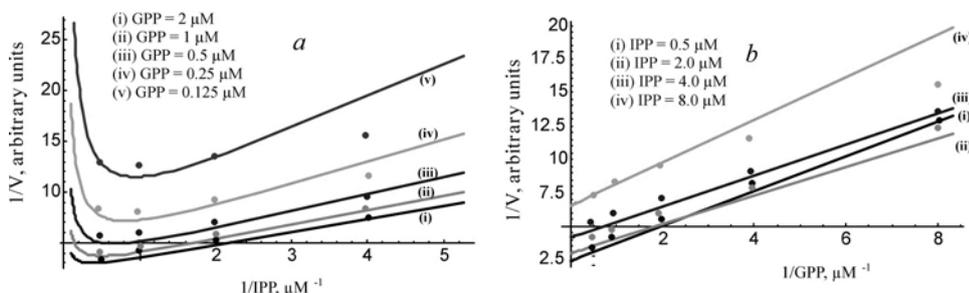


Figure 3. The kinetics of chicken prenyltransferase. The dependence of the inverse reaction rate on the inverse concentration at various fixed GPP concentrations is depicted in (a): (i) 2 μM ; (ii) 1 μM ; (iii) 0.5 μM ; (iv) 0.25 μM ; (v) 0.125 μM . The dependence of the inverse reaction rate on the inverse GPP concentration at various fixed IPP concentrations is obvious in (b): (i) 0.5 μM ; (ii) 2.0 μM ; (iii) 4.0 μM ; (iv) 8.0 μM . Dots indicate experimental data according to (Reed and Rilling, 1975); the curves are the result of simulation using model (1); parameter values: $k_{\text{cat}} = 1.5 \text{ s}^{-1}$; $K_{\text{m},\text{S}1} = 0.23 \text{ mM}$; $K_{\text{m},\text{S}2} = 0.65 \text{ mM}$; $K_{\text{i},\text{S}2} = 1.11 \text{ mM}$.

The results are suggestive. There may exist a universal molecular mechanism and similarities between the kinetic characteristics of prenyltransferase catalyzed reaction at least in human and chicken hepatocytes. The results are also gratifying because, in many cases, the molecular biologists can combine the experimental data obtained for distinct species in refining models for other species.

ACKNOWLEDGEMENTS

This work was supported by Innovation project of Federal Agency of Science and innovation IT-CP.5/001 “Development of software for computer modeling and design in postgenomic system biology (system biology *in silico*)”. The authors are grateful to Vitaly Likhoshvai for valuable discussions and to I.V. Lokhova for bibliographical support.

REFERENCES

- Barnard G.F., Popjak G. (1981) Human liver prenyltransferase and its characterization. *Biochim. Biophys. Acta.*, **661**(1), 87–99.
- Cleland W.W. (1963) The kinetics of enzyme-catalyzed reactions with two or more substrates or products. I. Nomenclature and rate equations. *Biochim. Biophys. Acta.*, **67**, 104–137.
- Reed B.C., Rilling H.C. (1975) Crystallization and partial characterization of prenyltransferase from avian liver. *Biochemistry*, **14**(1), 50–54.
- Szkopinska A., Plochocka D. (2005) Farnesyl diphosphate synthase; regulation of product specificity. *Acta Biochim Pol.*, **52**(1), 45–55.

MATHEMATICAL MODELING OF THE GENE NETWORK CONTROLLING HOMEOSTASIS OF INTRACELLULAR CHOLESTEROL

Ratushny A.V. ^{*1,2}

¹Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ²Novosibirsk State University, Novosibirsk, 630090, Russia

* Corresponding author: e-mail: ratushny@bionet.nsc.ru

Key words: mathematical modeling, cholesterol homeostasis, gene network

SUMMARY

Motivation: Study of the possible modes of the gene network controlling homeostasis of intracellular cholesterol requires the development of appropriate mathematical models.

Results: Here we present mathematical models of 11 subsystems of the gene network controlling intracellular cholesterol homeostasis. A database was created that contains quantitative data about components of this gene network.

Availability: Models are available on request.

INTRODUCTION

Cholesterol is an essential component of cells and a precursor of numerous other sterols that sustain the cell and the entire organism. However, cholesterol excess may become pathogenic. Hence, cholesterol level normally is under stringent control.

Cholesterol homeostasis in vertebrate cells is mainly regulated by two transcription factors (TF), sterol regulatory element-binding proteins (SREBP)-1a и -2. SREBPs activate the expression of numerous genes whose products are involved in cholesterol biosynthesis and its transport from blood plasma into the cell as a part of lipoproteins (Horton *et al.*, 2003). The formation of the active SREBP is directly related to cholesterol content in the cell. Transport of the cleavage activating protein (SCAP)–SREBP complex (SSC) from the endoplasmic reticulum (ER) to the Golgi complex (GC) is the link sensitive to intracellular cholesterol content. In the GC, there is a two-step proteolysis of SSC catalyzed by the S1 and S2 proteases resulting in a release of the active SREBP form which migrates into the nucleus and activates transcription of target genes. The key step of SSC transport from the ER to GC is the conformational change in the SCAP. At high cholesterol content in the cell, SCAP conformation allows SSC to bind to the insulin-induced gene (Insig-1, 2) retention protein in the ER (Adams *et al.*, 2003), thereby preventing its transport to the GC. At low cholesterol content in the cell, SCAP assumes a conformation that dissociates from Insig to be transported to the GC (Gimpl *et al.*, 2002).

In addition to the regulation of gene transcription by SREBPs there exists control of intracellular cholesterol homeostasis at the posttranscriptional level, at least in the case of HMG-CoA reductase (HMGCR), which catalyzes the limiting stage of cholesterol synthesis in the cell (Goldstein, Brown, 1990). The number of HMGCR in the cell is regulated at various levels (Ness, Chambers, 2000): mRNA stabilization, translation efficiency, activity modulation due to reversible phosphorylation, enzyme degradation (Sever *et al.*, 2003). It is also known that free cholesterol activates the microsomal

enzyme acyl-coenzyme A:cholesterol transferase (ACAT) esterifying cholesterol and thereby promoting the pooling of cholesterol esters in the cytoplasm (Klimov, Nikul'cheva, 1999). The major components of the gene network controlling intracellular cholesterol homeostasis are comprised in the GeneNet database (<http://wwwmgs.bionet.nsc.ru/mgs/gnw/genenet/>). Here we present mathematical models of 11 subsystems of the gene network under investigation, which support and extend our previously published model (Ratushny *et al.*, 2003).

METHODS AND ALGORITHMS

Models were developed by using the generalized chemical-kinetic approach (Likhoshvai *et al.*, 2001) and the generalized Hill function method (Likhoshvai *et al.*, 2006, this issue).

RESULTS AND DISCUSSION

The created database contains quantitative characteristics of the gene network components controlling cholesterol homeostasis in vertebrate cells. The database contains a description of 95 experimental dependencies of the gene network characteristics on concentrations of the molecular components of this system and quantitative characteristics of 477 steady-states of the gene network components.

Mathematical models were developed for 11 subsystems of the gene network.

Subsystem 1. Influx and utilization of low molecular weight substances in the intracellular cholesterol synthesis. Fluxes of CoA, acetyl-CoA, acetoacetyl-CoA, acetoacetate, isopentenyl diphosphate, geranylgeranyl and cholesterol were introduced into the model. The constants of the processes were derived from experimental data for their steady-state concentrations in vertebrate cells in normal conditions.

Subsystem 2. Modeling of the metabolic pathway for cholesterol biosynthesis in the cell. The model describes 32 enzymatic reactions along the pathway. For the results of the simulation of the enzymatic reaction catalyzed by prenyltransferase see (Ratushny *et al.*, 2006, this issue).

Subsystem 3. Regulation of SREBP transport from ER to the nucleus via GC. Fig. 1a is a schematic representation of regulation of the SREBP ER→GC→nucleus transport.

Transport rate is calculated as $V = k \cdot d = k \cdot k_0 \cdot S_{n,ER} / (1 + (C/k_i)^h)$, (1)

where $S_{n,ER}$ is the SSC concentration in the ER; C is the free cholesterol concentration; d denotes the fraction of SSC unbound to Insig; k is the SSC ER→GC transport rate constant; k_0 is the constant defining the maximum fraction of the transportable SSC at zero cholesterol concentration in the medium; k_i is the cholesterol inhibition constant of the SSC ER→GC transport; h is Hill coefficient defining the degree of nonlinearity of the cholesterol inhibition for the SSC ER→GC transport. The estimates yielded by model (1) were supported by experimental data (Adams *et al.*, 2003) (Fig. 1b).

Subsystem 4. Gene transcription regulation by SREBP factors. The following equation expresses the regulation of the transcription rate of the target gene by SREBPs:

$$V_{R_x} = k \cdot \frac{k_0 + k_{1,S_{1a}} \cdot (S_{1a}/k_{2,S_{1a}})^2 + k_{1,S_2} \cdot (S_2/k_{2,S_2})^2 + k_{1,S_{1a2}} \cdot S_{1a} \cdot S_2 / (k_{2,S_{1a2}})^2}{1 + (S_{1a}/k_{2,S_{1a}})^2 + (S_2/k_{2,S_2})^2 + S_{1a} \cdot S_2 / (k_{2,S_{1a2}})^2} \quad (2)$$

where V_{Rx} is mRNA synthesis rate encoded by the x gene; S_{1a} , S_2 denote the concentrations of the SREBP-1a and SREBP-2 monomers; k is the transcription rate constant for the x gene; k_0 is the basal transcription level; $k_{1,S1a}$, $k_{1,S2}$, $k_{1,S1a2}$ are the maximum transcription rates for the x gene attainable under SREBP-1a, SREBP-2 and their combined action, respectively; $k_{2,S1a}$, $k_{2,S2}$, $k_{2,S1a2}$ are the efficiency constants of the SREBP-1a, SREBP-2 and their combined effects on x gene transcription, respectively.

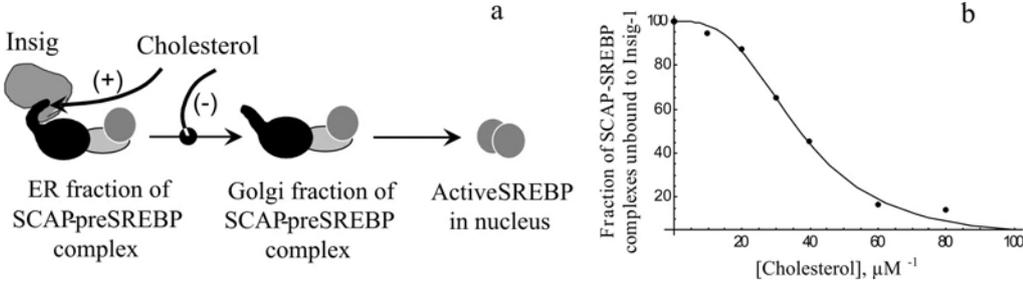


Figure 1. Schematic representation of regulation of the SREBP transport from the ER to the nucleus via GC (a). The effect of the various cholesterol contents in the medium on the SCAP conformation and its ability to form a complex with Insig-1 (b). Dots indicate experimental data according to (Adams *et al.*, 2003); the curve is the result of simulation using model (1); parameter values: $k_0 = 0.94$; $k_i = 35 \mu\text{M}$; $h = 3$.

The estimates yielded by model (2) were supported by experimental data (Datta and Osborne, 2005) (Fig. 2).

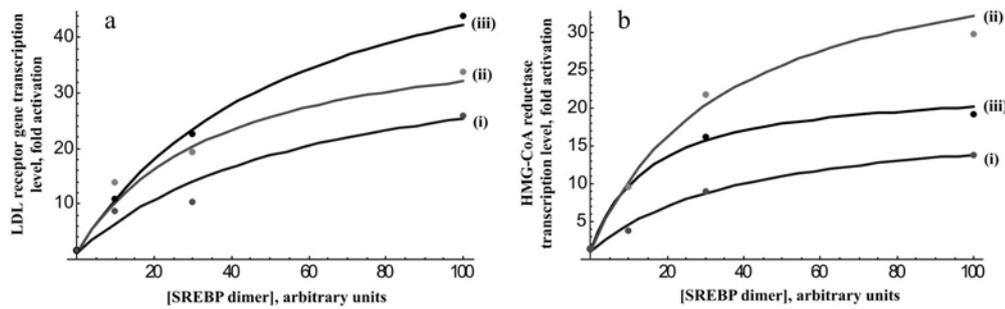


Figure 2. Transcription activation of genes encoding the LDL receptor (a) and HMG-CoA reductase (b) by the homodimer SREBP-1a/ SREBP-1a (i), homodimer SREBP-2/SREBP-2 (ii) и heterodimer SREBP-1a/ SREBP-2 (iii). Dots indicate experimental data using (Datta, Osborne, 2005); the curves are the result of simulation using model (2).

Subsystem 5. Maturation and translation of mRNAs. mRNAs transport from the nucleus and their maturation are described as monomolecular reactions. The total reaction time was estimated as 17 min ($k = 10^{-3} \text{ sec}^{-1}$). Protein synthesis is also described in terms of monomolecular reactions ($k = 10^{-1} \text{ sec}^{-1}$).

Subsystem 6. Influx and utilization of LDL in blood plasma. The model describes (I) LDL influx into plasma at a constant rate; (II) receptor mediated LDL internalization into cells and (III) LDL utilization.

Subsystem 7. Receptor mediated LDL endocytosis and their degradation in lysosomes. The model is described in more detail in (Ratushny, 2006, this issue).

Subsystem 8. Cholesterol pooling in the cell. The steady-state rate equation for the ACAT catalyzed reaction is as follows:

$$V = k_{cat} \cdot e_0 \cdot (C/K_C)^2 / (1 + (C/K_C)^2 + C_e/K_{i,C_e}), \quad (3)$$

where e_0 is the total ACAT concentration; C_e denotes the cholesterol ester concentration; k_{cat} is the catalytic constant; K_C is the efficiency constant of the cholesterol effect on the reaction rate; K_{i,C_e} is the inhibition constant of the reaction rate by cholesterol ester. Support for model (3) comes from (Zhang *et al.*, 2003) (Fig. 3).

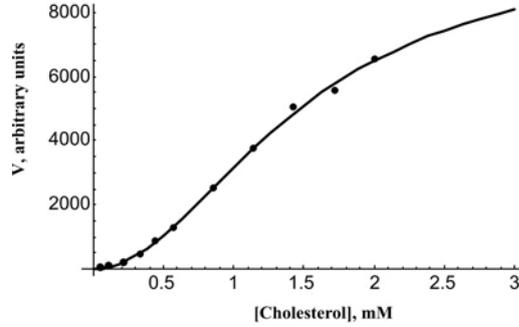


Figure 3. Effect of the free cholesterol concentration on the rate of the reaction of cholesteryl oleate formation catalyzed by ACAT. Dots indicate experimental data according to (Zhang *et al.*, 2003); the curve is the result of simulation according to model (3) ($K_C = 1.47$ mM).

Subsystem 9. Modulation of HMG-CoA reductase activity. A cholesterol-dependent regulation of HMGCR activity is described as follows:

$$-\frac{dR}{dt} = \frac{dR_p}{dt} = \frac{k_f \cdot e_{o,rk} \cdot d \cdot R / K_{m,R}}{1 + R/K_{m,R} + R_p/K_{i,R_p}} - \frac{k_r \cdot e_{o,rp} \cdot R_p / K_{m,R_p}}{1 + R_p/K_{m,R_p} + R/K_{i,R}}, \quad d = \frac{C^h}{k_a^h + C^h}, \quad (4)$$

where $e_{o,rk}$, $e_{o,rp}$ denote the total concentrations of HMGCR kinase and phosphatase; R , R_p are the concentrations of active and inactive HMGCR; k_f , k_r are the catalytic constants of HMGCR kinase and phosphatase; $K_{m,R}$, K_{i,R_p} are the Michaelis-Menten constants for active and inactive HMGCRs ($K_{m,R} = 0.85$ μ M, $K_{m,R_p} = 0.01$ μ M); $K_{i,R}$, K_{i,R_p} are the inhibition constants of HMGCR dephosphorylation and phosphorylation by the active and inactive HMGCRs, respectively; d denotes the fraction of the active HMGCR kinases; k_a is the efficiency constant of the cholesterol effect on the activation of HMGCR kinases; h is Hill coefficient defining the degree of nonlinearity of the activation of HMGCR kinases by cholesterol.

Subsystem 10. A cholesterol-dependent regulation of HMG-CoA reductase degradation. This process is described as follows:

$$V = k_d \cdot R \cdot (C/k_a)^h / (1 + (C/k_a)^h), \quad (5)$$

where V is the degradation rate of HMGCR bound to Insig; k_d is the degradation rate constant; k_a is the efficiency constant of the cholesterol effect on the HMGCR binding to Insig; h is Hill coefficient defining the degree of nonlinearity of the cholesterol effect on the HMGCR binding to Insig. The constants of model (5) were estimated by using experimental data (Adams *et al.*, 2003) ($k_a = 37$ μ M; $h = 3$).

Subsystem 11. Utilization and degradation of high-molecular substances. The constants of the nonspecific degradation of RNAs and proteins were estimated as $k_{d,mRNA} = 10^{-3}$ sec^{-1} ($\tau_{1/2} = 17$ min) and $k_{d,prot} = 5 \cdot 10^{-5}$ sec^{-1} ($\tau_{1/2} = 5.5$ hours), respectively.

Thus, we succeeded in modeling 11 subsystems of the gene network controlling intracellular cholesterol homeostasis. We have defined parameter values and demonstrated agreement between the calculated estimates and reported experimental data. Models of the subsystems were assembled to form a single model that describes 176 processes and contains 122 dynamic variables. We have also demonstrated agreement between the steady-state characteristics yielded by the unified model and the experimental data. The parametric stability of the steady-state characteristics of the gene network components was analyzed. It was demonstrated that the stationary concentration of free cholesterol is robust to changes in the rates of most gene network processes. The results of this analysis can be helpful in resolving problems aimed at the detection of targets for pharmacological regulation. We intend to study in more detail the phase portrait of the model.

ACKNOWLEDGEMENTS

This work was supported by Innovation project of Federal Agency of Science and innovation IT-CP.5/001 “Development of software for computer modeling and design in postgenomic system biology (system biology *in silico*)”. The author is grateful to Vitaly Likhoshvai for valuable discussions and to I.V. Likhova for bibliographical support.

REFERENCES

- Adams C.M. *et al.* (2003) Cholesterol-induced conformational change in SCAP enhanced by Insig proteins and mimicked by cationic amphiphiles. *Proc. Natl. Acad. Sci. USA*, **100**(19), 10647–10652.
- Datta S., Osborne T.F. (2005) Activation domains from both monomers contribute to transcriptional stimulation by sterol regulatory element-binding protein dimmers. *J. Biol. Chem.*, **280**(5), 3338–3345.
- Gimpl G. *et al.* (2002). A closer look at the cholesterol sensor. *Trends Biochem. Sci.*, **27**(12), 596–599.
- Goldstein J.L., Brown M.S. (1990) Regulation of the mevalonate pathway. *Nature*, **343**, 425–430.
- Horton J.D. *et al.* (2003) Combined analysis of oligonucleotide microarray data from transgenic and knockout mice identifies direct SREBP target genes. *Proc. Natl. Acad. Sci. USA*, **100**(21), 12027–12032.
- Klimov A.N., Nikul’cheva N.G. (1999) *Lipid and Lipoprotein Metabolism and Its Disturbances*. St. Petersburg: Piter Kom.
- Likhoshvai V.A. *et al.* (2001) A generalized chemical-kinetic method for modeling gene networks. *Mol. Biol. (Mosk)*, **35**(6), 1072–9.
- Likhoshvai V.A. *et al.* (2006) In silico cell III. Method of reconstruction of enzymatic reaction and gene regulation mathematical models in terms of generalized Hill functions. *This issue*.
- Ness G.C., Chambers C.M. (2000) Feedback and hormonal regulation of hepatic 3-hydroxy-3-methylglutaryl coenzyme A reductase: the concept of cholesterol buffering capacity. *Proc. Soc. Exp. Biol. Med.*, **224**(1), 8–19.
- Ratushny A.V. (2006) Mathematical modeling of receptor mediated endocytosis of low-density lipoproteins and their degradation in lysosomes. *This issue*.
- Ratushny A.V. *et al.* (2003) Resilience of Cholesterol Concentration to a Wide Range of Mutations in the Cell. *Complexus*, **1**, 142–148.
- Ratushny A.V. *et al.* (2006) On some conservative properties of enzymatic systems: kinetic properties of prenyltransferase. *This issue*.
- Sever N. *et al.* (2003) Accelerated degradation of HMG CoA reductase mediated by binding of insig-1 to its sterol-sensing domain. *Mol. Cell*, **11**(1), 25–33.
- Zhang Y. *et al.* (2003) Cholesterol is superior to 7-ketocholesterol or 7 alpha-hydroxycholesterol as an allosteric activator for acyl-coenzyme A:cholesterol acyltransferase 1. *J. Biol. Chem.*, **278**(13), 11642–11647.

CONSTRUCTION OF THE HCV-HEPATOCYTE SYSTEM MODEL

Rudenko V.M.*, **Korotkov E.V.**

Center of Bioengineering, RAS, Moscow, Russia

* Corresponding author: e-mail: u102095@dialup.podolsk.ru

Key words: hepatitis C virus (HCV), mathematical model, computer analysis, genetic net

SUMMARY

Motivation: the hepatitis C is one of the most dangerous diseases of the human and there are no effective medical products. Overwhelming majority of cases of disease pass in a chronic stage (Racanelli, Rehermann, 2003). For creation of new vaccines and medical products detailed studying life cycle of a virus and revealing of the factors influencing its reproduction and protein synthesis is necessary. As research of behaviour HCV *in vitro* is complicated, creation of computer model of a genetic network infected HCV hepatocyte and its use for the experimental analysis is expedient.

Results: the mathematical model of the hepatitis C virus evolution in the hepatocyte has been developed. Model allows estimate concentration of various components of the system virus-cell: DNA, RNA, proteins, virus particles and other components eventually. The model assumes consideration discrete time intervals, within number of various chemical and biological reactions, possible at presence, is maximized. The problem of creation of adequate model consists that many parameters of model are unknown. Therefore the primary goal of the given work was the selection of parameters of model and demonstration of its correctness for simple cases of behaviour of system a virus-hepatocyte.

INTRODUCTION

Now medicine and pharmacology have problems of development of highly effective drugs for treatment of the various virus and new and existing infections. The general problem in struggle against virus diseases is the high degree of mutations of virus genomes. This problem concerns as concerning easy infections, such various forms of a flu, and fatally dangerous viruses: AIDS, H5N1 and HCV. The given work is devoted to studying of a virus of a hepatitis C, as one of representing the greatest danger to the human population. Unpleasant feature of the HCV virus is the resistance to the host immune answer and therefore a high degree of chronic diseases. According to statistical data only 20 % diseases of hepatitis C finish with recover. In other cases disease gets chronic character, and within the next 15–20 years after acute stage the probability of development of a cirrhosis and a cancer of a liver is great. Now in the world about 180 million people suffer a chronic hepatitis C (Ivanov, 2005). Experimental studying of a virus of a hepatitis C is complicated because it replicates badly in cellular cultures *in vitro*. The well suitable animal models are absent too, except for a chimpanzee. However, recent researches have shown an opportunity of allocation and studying of a virus in cultures of cells of a brain of newborn mice and in HUH7 cell lines, but their application is expensive enough. Therefore an actual problem is development of the mathematical model describing life cycle of a virus in the hepatocyte, and also processes influencing on distribution of an infection and its persist.

Uses of computer modeling allows to predict in dynamics the stage of disease depending on presence (absence) of different substances in a cell, speed of immune reaction of an organism on a virus infection, without animal models.

MODEL

The virus of a hepatitis C does not possess ability to replication outside of cells of a liver, behind very rare exception. Therefore creation the model of the cell inside of which the virus functions is expedient. We represent a cell in the form of the closed system consisting of components or substances. The term “substance” we shall designate various biological molecules, such as virus RNA, virus and cellular proteins, ions of various metals, complexes of molecules and even separate virions of HCV. Between substances of the system chemical and “biological” reactions pass with formation of other substances. For example, reaction of formation polypeptide chain (HCVpp) on virus RNA (HCV_rna) which simultaneously serves as a matrix for translation of proteins HCV can be described by the equation:



The given equation shows, that from one RNA chain is copied λ virus polypeptides at presence of λ cellular ribosomes. By means of the similar equations it is possible to describe all stages of life cycle of a virus in the cell, and also various metabolic pathways in infected hepatocyte in particular synthesis of proteins of the immune answer. The purpose of modeling is to estimate quantities of molecules of each substance (or concentration of substances) in a cell eventually. We shall designate substances through (X_1, X_2, \dots, X_N) , and through $(C_1^0, C_2^0, \dots, C_N^0)$ – concentration of these substances in the zero moment of time. Concentration of substances in time can vary owing to processes of their natural degradation and during the reactions existent in a cell. All possible reactions we shall present in the form of system of the equations:

$$\sum_{i=1}^N a_{ij} X_i \rightarrow \sum_{i=1}^N b_{ij} X_i \quad 1 = 1 \dots m \quad (2)$$

Where m – number of different type of reactions which can potentially proceed in a cell under condition of presence of the necessary substances. Parameters a_{ij} and b_{ij} – show quantity of molecules of each substance participating in reaction and products of reaction accordingly. If the substance is not entrance or target product of the reaction, corresponding coefficients are equal 0. Let's consider the discrete moments of time equal to average time duration of reactions of system (2). It is supposed, that all reactions have necessary catalysts (as entrance and target products) and time of performance of reactions is small and approximately is the same for various reactions. If there are reactions which speeds considerably exceed of the speeds of other reactions of the system, they should be united in one equation with adjacent to reactions. Adjacent with reaction Y we shall name reactions of two types: 1) reactions, target products from which are used as entrance in reaction Y and 2) reactions that use the result product of Y as the entrance product.

The model is constructed in the assumption, that during each moment of time all reactions for which there are initial products can proceed, and the system always aspires to maximize the general number of proceeding reactions. Thus it is necessary to consider, that quantities of substances of each kind which is taking part reactions, do not exceed their concentration at present time. The algorithm describing behavior of

system a virus-hepatocyte can be formulated as follows. We make steps 1 and 2 consistently in a cycle for each moment of time. Step 1 permits to solve a problem of maximization of total quantity of reactions (3), at performance of restrictions on concentration of substances (4).

$$MAX(F^i) = \sum_{j=1}^m Y_j^i \quad (3)$$

Where Y_j – number of reactions j type.

$$\sum_{j=1}^m a_{jk} Y_j^i \leq C_k^i, \quad k = 1.. N. \quad (4)$$

The problem (3), (4) is a task of linear dynamic programming which can be solved, using a simplex-method (Glebov *et al.*, 2000).

Step 2. To calculate new concentration of substances under the formula:

$$C_k^{i+1} = (1 - \delta_k) \cdot \left(C_k^i - \sum_{j=1}^m a_{jk} Y_j^i + \sum_{j=1}^m b_{jk} Y_j^i \right), \quad k = 1.. N. \quad (5)$$

The first multiplier of the formula considers natural degradation of biological molecules eventually δ_k – a share of molecules k substances which degrade for a time unit, a_{jk} – the factor showing quantity of units k of substance participating in reaction j of type, b_{jk} – shows how many units of substance k is formed as a result of reaction j type.

RESULTS AND DISCUSSION

The mathematical model describing behaviour infected HCV cell, has been realized as a program complex in Visual C 6.0. Entrance parameters of model which include

- The list of components of the system,
- Initial components concentrations,
- Degrees of degradation of everyone for a time unit,
- Descriptions of reactions

are stored in the Microsoft Access database. These parameters can change by the user. Also there is an opportunity of addition/removal of substances and reactions from system. The user sets quantity of time intervals during which the behavior of system is modeled. During performance of the program files of concentration of substances of system during each moment of time are formed. From the received data for the components of system chosen by the user, for example virions, virus RNA, proteins, etc. are under construction, allowing visually to estimate speed and a degree of infection of the host cell. There are two basic problems with the creation and application described model of the virus-hepatocyte system. One of them consists in presence of huge quantity of unknown parameters, in particular ratios of speeds of reactions among themselves. Second is the absence of full data about HCV infection processes and host immune answer mechanisms. Thereof we can vary parameters of model proceeding from statistical data about a likelihood parity of outcomes of diseases HCV which are available in the literature.

Despite of the specified discrepancies, the created model evidently reflects dynamics of behavior of system a cell-virus. We modeled some simple cases of evolution of a virus in a cell at various initial concentration of substances at presence and absence of the

immune answer. The results show that the behavior of system coincides with expected. Diagrams of dependences of quantity of substances during time are shown in Fig. 1, 2.

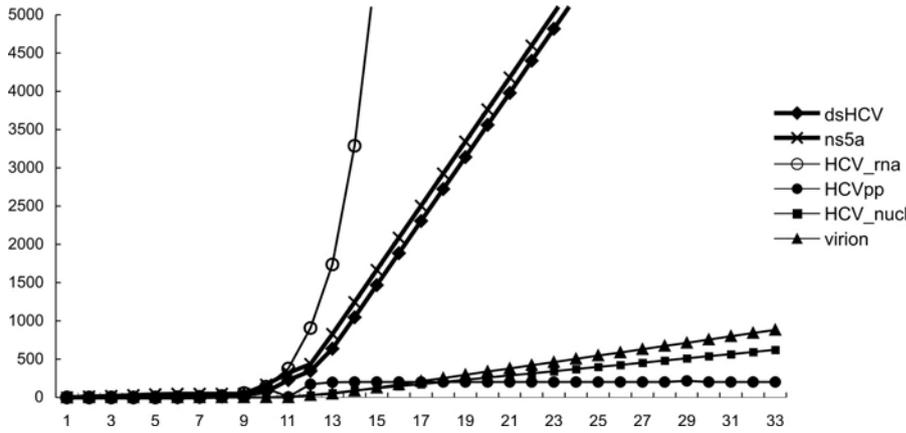


Figure 1. Development HCV in a cell at absence of the immune answer. The fast gain of virus proteins (it is shown ns5a) and dsRNA is observed, the growth of the virus RNA (HCV_rna) is exponential. For polyproteins (HCVpp), nucleocapsids (HCV_nucl) and HCV virions (virion) slower increase in concentration is observed.

Results show that the proposed approach to modeling complex biological systems with using linear dynamic programming is perspective. The approach can adequately describe the evolution of the system virus-hepatocyte at various initial conditions. We plan to add the database of system components and interactions between them, using available information on HCV-induced genes from the site: <http://palladin.bionet.nsc.ru/mgs/papers/stepanenko/hcv-trrd>. We hope that developed computer model of system virus-hepatocyte will allow to test various combinations of biochemical reactions and to define conditions at which probably to rich full virus elimination, or considerably to reduce its negative influence on hepatocyte.

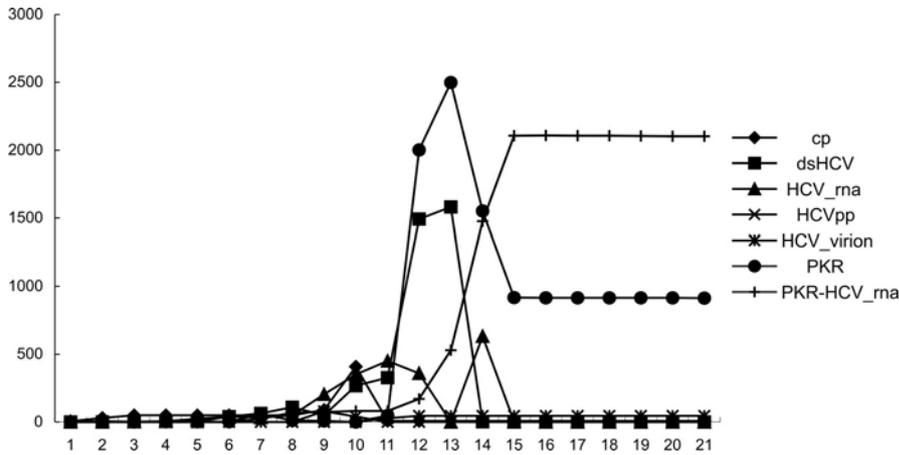


Figure 2. Development HCV in a cell with occurrence of PKR (interferon-induced protein kinase). PKR catalyse phosphorylation of the α -subunit of eIF2 therefore translation of virus proteins is blocked. From figure we can see that the increase the quantity of dsRNA of HCV (dsHCV) lead to growth of PKR. Full elimination of HCV follows by high concentration of PKR. PKR-HCV_rna line show number of virus RNA that were neutralized by PKR.

ACKNOWLEDGEMENTS

The work was supported by grant of Russian Ministry of Education and Sciences No. 102205/03-1 at 03.10.2005.

REFERENCES

- Glebov A.V., Kochetov Yu.A., Plyasunov A.V. (2000) *Optimization methods*. Novosibirsk State University.
- Ivanov A.V., Kuzyakin A.O., Kochetkov S.N. (2005) Molecular biology of the hepatitis C virus. *Uspekhi biologicheskoi khimii*, **45**, 37–86. (in Russ.).
- Racanelli V., Rehmann B. (2003) Hepatitis C virus infection: when silence is deception. *Trends Immunolog.*, **24**, 8, 456–464.

FORMAL DESCRIPTION OF NF- κ B PATHWAY, ITS ROLE IN INFLAMMATION, INHIBITION OF APOPTOSIS, CARCINOGENESIS, AND WAYS OF INACTIVATION FOR PREDICTION OF NEW TARGETS FOR ANTI-INFLAMMATORY AND ANTI-CANCER TREATMENT

Sharipov R.N.^{*1, 2, 3}, *Kolpakova A.F.*⁴

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ² Design Technological Institute of Digital Techniques, SB RAS, Novosibirsk, Russia; ³ Institute of Systems Biology OOO, Novosibirsk, Russia; ⁴ State Research Institute for Medical Problems of the North, SB RAMS, Krasnoyarsk, Russia

Key words: NF- κ B, transcription factor, apoptosis, cell survival, inflammation, cancer, Biopath

SUMMARY

Motivation: Hyperactivation of transcription factor NF- κ B plays important part in development of number of human pathologies: autoimmune diseases, many types of inflammation, and various viral infections. Progress in cancer research allowed to reveal importance of NF- κ B in induction of anticancer drug resistance and cell proliferation. Formal description of NF- κ B pathways and related issues will help to predict new targets for NF- κ B inactivation and design better drugs and therapeutic strategies.

Results: NF- κ B pathway, mechanisms of its activation, regulation of dependent genes, crosstalks with other regulatory pathways, participation in development of human diseases, as well as known mechanisms of drug actions were described formally using literature annotation and BioUML workbench. All data are stored and classified in Biopath database.

Availability: <http://biopath.biouml.org>.

INTRODUCTION

Since discovery of NF- κ B in 1986 it has been attracting attention because of its unusual regulation, diversity of activating stimuli, regulated genes and biological effects, striking evolutionary conservatism of structure and functions among its family members (Ghosh *et al.*, 1998). NF- κ B hyperactivation was observed associated with block of apoptosis and increased cell proliferation in many cases of cancer in breast, thymus, colon, non-Hodgkin lymphomas, T- and B-cell leukemia, and various types of melanoma (Wu, Kral, 2005). Hyperactivation of NF- κ B also plays important role in inflammation, autoimmune diseases and various viral infections (Baldwin, 2001).

A number of drugs and substances were developed to inhibit NF- κ B pathway. However in many cases their specificity is not sufficient that causes side effects (Baldwin, 2001). Complete inhibition of NF- κ B pathway could result in unexpected side effects due to its importance for cell survival. Formal description of NF- κ B pathways and related issues will help to predict new targets for NF- κ B inactivation and design better drugs and therapeutic strategies for treatment of cancer, inflammation and autoimmune diseases.

METHODS

Data about NF- κ B pathway and related issues were obtained by literature annotation and from following biological databases: TRANSPATH, TRANSFAC, InterPro, KEGG, OMIM and Gene Ontology. BioUML technology (Kolpakov, 2004; <http://www.biouml.org>) was used for formal description of NF- κ B pathway and related issues. Three BioUML diagram types were used for formal description: semantic networks, gene networks, and pathways. All diagrams and description diagram elements (genes, proteins, substances, concepts, reactions and semantic relationships) were stored in Biopath database. BeanExplorer Enterprise Edition (<http://www.beanexplorer.com>) was used for web interface developing for Internet access to Biopath database (Fig. 1).

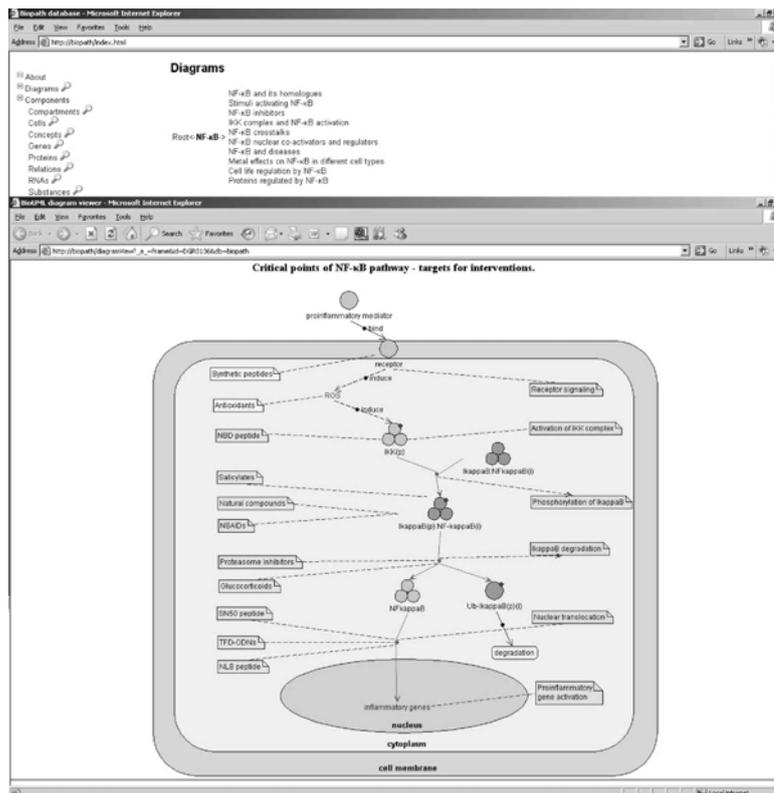


Figure 1. Points of NF- κ B pathway – targets for NF- κ B inhibitors (DGR0136).
Web interface of Biopath database.

RESULTS

Data obtained from more than 300 articles were analysed and represented in the form of 66 diagrams with descriptions (the most important diagrams are listed in Table 1). These diagrams can be classified into following groups:

- structure of NF- κ B pathway (7 diagrams);
- stimuli and mechanisms of NF- κ B activation (35 DGRs);
- role of NF- κ B in inflammation, cancer and other human diseases (13 DGRs);
- regulation of apoptosis (7 DGRs);
- mechanisms and targets of NF- κ B inactivation by NF- κ B inhibitors (4 DGRs; Fig. 1).

Table 1. List of the most important diagrams about NF- κ B

Structure of NF-κB pathway	
Diagram ID	Name of diagram
DGR0010	Rel/NF- κ B family
DGR0003	I κ B family
DGR0027	Proteins, regulated by NF- κ B
DGR0012	NF- κ B interactions with transcription co-activators in nucleus
DGR0135	p100- and p105-processing
DGR0219	NF- κ B and cell cycle regulation
Stimuli and mechanisms of NF-κB activation	
DGR0026	Stimuli, that activate NF- κ B
DGR0051	NF- κ B activation pathways
DGR0118	TNF- α -induced NF- κ B activating pathway
DGR0119	IL-1 β -induced NF- κ B activating pathway
DGR0120	LPS(lipopolysaccharide)-TLR4(Toll-like receptor 4)-dependent NF- κ B activating pathway
DGR0134	dsRNA-TLR3(Toll-like receptor 3)-dependent NF- κ B-activating pathway
DGR0122	p38 mitogen-activated protein kinase pathway
DGR0109	Immune cell activation as response to bacterial infection
DGR0106	LPS-induced effects in mammal cells
DGR0110	TNF- α and IL-1 β effects in VSMCs(vascular smooth muscle cells)
DGR0033	Cell-type specific role of ROS in NF- κ B activation by IL-1
DGR0045	Activation of the I κ B kinase (IKK)
DGR0013	Activation of NF- κ B by non-IKK and other stimuli
DGR0102	Akt (PKB) regulation of NF- κ B
DGR0128	Metals and metals-induced ROS influence on NF- κ B signaling
Role of NF-κB in inflammation, cancer and other human diseases	
DGR0047	Diseases associated with NF- κ B activation
DGR0104	NF- κ B in the initiation of chronic inflammation
DGR0105	NF- κ B and the perpetuation of chronic inflammation
DGR0050	Acute lung injury and NF- κ B
DGR0052	Development of multiple organ dysfunction syndrome in human lung
DGR0049	NF- κ B involvement in asthma development
DGR0113	RSV(respiratory syncytial virus)- and TNF- α -induced differential activation of p65 and p50 in human airway epithelium
DGR0048	Role of NF- κ B in oncogenesis
DGR0132	Rearrangements of NF- κ B and I κ B- α genes loci in oncogenesis
DGR0220	NF- κ B: a possible link between obesity and breast cancer
DGR0221	Role of NF- κ B in chemoresistance
Regulation of apoptosis	
DGR0131	NF- κ B in regulation of apoptosis
DGR0234	Cell death pathways
DGR0235	The Extrinsic Cell Death Pathway: Fas
DGR0238	Multiple Roads to Cell Death from E2F-1
Known mechanisms and targets of NF-κB inactivation, and NF-κB inhibitors	
DGR0136	Critical points of NF- κ B pathway - targets for interventions
DGR_NF- κ B_inhibs	NF- κ B inhibitors
DGR0021	Effects of glucocorticoids on NF- κ B

DISCUSSION

Proinflammatory stimuli (e.g., cell damage, UV radiation, reactive oxygen species (ROS), cytokines) activate NF- κ B through signal transduction pathways and induce overexpression of a set of genes encoding cytokines (e.g., IL-1, IL-2, IL-8), interferons (IFN β , IFN γ) and enzymes producing mediators of inflammation (e.g., inducible NO synthase (iNOS), cyclooxygenase-2 (COX-2), and 5-lipoxygenase (5-LOX)). Highly increased levels of active NF- κ B were found in asthma, different types of arthritis and

lung inflammations in human (Ghosh *et al.*, 1998). A general scheme of NF- κ B activation by proinflammatory stimuli is represented on Fig. 1.

NF- κ B also regulates expression of genes encoding inhibitors of apoptosis cIAPs, Bfl/A1 and Bcl-XL. These proteins function through inhibition of activity of proapoptotic proteins (e.g., caspases, tBid), stabilization of mitochondrial membrane and promotion of cell survival. Expression of gene of p53 protein – important apoptosis inducer – is also regulated by NF- κ B. In certain conditions NF- κ B was shown to induce apoptosis mediated by p53. Complete detailed scheme of NF- κ B-mediated regulation of apoptosis is not known. Every year brings reports about new proteins that involved in regulation of NF- κ B pathway. Complexity of NF- κ B regulating pathway can be characterized by the fact that in certain conditions well-known apoptosis-inducing factor TNF- α is able to induce cell survival mediated by this transcription factor. It was suggested that induction of cell survival by TNF- α was mediated by TRAF2 protein interacting with death-inducing signal complex (DISC) and leading to NF- κ B activation. Expression of TRAF2 gene is also under control of NF- κ B (Wu, Kral, 2005).

NF- κ B participates in human carcinogenesis through blocking of apoptosis and induction of cell survival and cell proliferation. Elevated levels of NF- κ B were revealed in many types of cancer (e.g., breast, colon, prostate). This fact was explained partly by observed in cells amplifications of NF- κ B family genes or deletion of nucleotide sequences encoding sites for interactions with intracellular inhibitor of NF- κ B (I κ B). In addition, NF- κ B was shown to inhibit action of anti-cancer drugs – inducers of apoptosis – in tumor cells and cause chemotherapy resistance (Wu, Kral, 2005).

Different types of NF- κ B inhibitors are used extensively for treatment of inflammation and as supportive drugs against chemotherapy resistance and target different steps of NF- κ B pathway (Fig. 1). Many of them are non-specific (antioxidants) or possess significant side effects (glucocorticoids) (Baldwin, 2001). Comprehensive data about NF- κ B pathway united in one database will help in design of specific and effective NF- κ B inhibitors without side effects.

The NF- κ B data collected in Biopath is used by our colleagues from Institute of Biomedical Chemistry (Moscow) for prediction of chemical substances – effective inhibitors of NF- κ B pathways – by use of PASS (Prediction of Biological Activity Spectra for Substances) system (Poroikov, Filimonov, 2005).

ACKNOWLEDGEMENTS

This work was supported by INTAS grant No. 03-51-5218 and RFBR grant No. 04-04-49826-a.

REFERENCES

- Baldwin A.S.Jr. (2001) Series introduction: the transcription factor NF-kappaB and human diseases. *J. Clin. Invest.*, **107**, 3–6.
- Ghosh S. *et al.* (1998) NF-kappaB and Rel proteins: evolutionarily conserved mediators of immune responses. *Annu. Rev. Immunol.*, **16**, 225–260.
- Kolpakov F.A. (2004) BioUML – open source extensible workbench for systems biology. *Proceedings of BGRS'2004*, **2**, 77–80.
- Poroikov V., Filimonov D. (2005) PASS: Prediction of Biological Activity Spectra for Substances. In Helma C. (ed), *Predictive Toxicology*. Taylor & Francis, pp. 459–478.
- Wu J.T., Kral J.G. (2005) The NF-kappaB/IkappaB signaling system: a molecular target in breast cancer therapy. *J. Surg. Res.*, **123**, 158–169.

RECONSTRUCTION AND STRUCTURE ANALYSIS OF APOPTOSIS GENE NETWORK IN HEPATITIS C

Stepanenko I.L.^{*1}, Podkolodnaya N.N.¹, Paramonova N.V.¹, Podkolodny N.L.^{1,2}

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia;

² Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, 630090, Novosibirsk, Russia

* Corresponding author: e-mail: stepan@bionet.nsc.ru

Key words: gene network, feedback loop, cluster analysis, signal pathway, TNF α , NF- κ B

SUMMARY

Motivation: Apoptosis plays a significant role in the pathogenesis of hepatitis C. This process may be viewed as a host defense mechanism against viral infection and hepatocarcinogenesis.

Results: Reconstruction of TNF α induced apoptosis gene network allows identifying of cell protein-targets of hepatitis C virus and provides the framework for the application of mathematical methods. Cluster analysis of the feedback loops regulating NF- κ B detects the modules of gene network that are associated with signaling proteins such as I κ B, c-FLIP, TRAF2, IEX-S, A20, c-IAP1, and c-IAP2. This analysis brings to light the hierarchical structure of the gene network.

Availability: <http://www.mgs.bionet.nsc.ru/mgs/papers/stepanenko/hcv-trrd/http://www.mgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/index.shtml> Hepatitis C (apoptosis).

INTRODUCTION

The interactions between viral proteins and host factors, such as cellular proteins of signal transduction and transcription regulation, make a significant influence on replication, persistence, and pathogenesis of hepatitis C virus (HCV), which infects 3 % of the world's human population and is a significant cause of liver disease. HCV-host interactions also affect the outcome of interferon antiviral therapy, which is effective only in certain patients.

Apoptosis plays a significant role in the pathogenesis of hepatitis C and is a host defense mechanism against viral infection and hepatocarcinogenesis (Kountouras *et al.*, 2003). Tumor necrosis factor (TNF α) and its receptors are the major components of the immune response involved in the control of viral infection and induction of apoptosis in infected cells. TNF α is a proinflammatory cytokine that plays a key role in both inflammatory and infectious diseases, especially in viral infection. TNF α is produced by a wide variety of cells including T cells and macrophages. The balance between viral life cycle and cytokine network, especially TNF α -signal pathway, is a key component that influences pathogenesis of HCV disease. The aim of this work is to summarize recent experimental data on induction and repression of TNF- α induced apoptosis by HCV products by using GeneNet technology (Ananko *et al.*, 2005). The reconstruction of apoptosis gene network in hepatitis C and analysis of its properties enables a systematic understanding of network functions and offer the potential of new therapeutic applications.

METHODS AND ALGORITHMS

HCV-regulated gene network in apoptosis was reconstructed on the basis of the data extracted from 273 experimental papers. It includes 36 genes, 121 proteins, and 280 molecular interactions. The network is graphically represented in GeneNet database (Ananko *et al.*, 2005). The information about hepatitis C virus-induced genes, transcription regulation, regulatory regions, and transcription factor binding sites is collected in TRRD database (Kolchanov *et al.*, 2002), section HCV-TRRD.

For searching for all cycles in the network graph, we have designed special program software using the method of breadth-first search (Cormen *et al.*, 2001). This software tools enables to reveal regulatory contours of different types including regulatory contours with the positive feedback (switch on, increase) and regulatory contours with the negative feedback (switch out, decrease). The contour is considered positive in case the number of links with the negative feedback is even or equals to 0.

Representation of the gene network in a form of bipartite graph was realized by applying the editor of gene networks, Gened (Ananko *et al.*, 2005).

By definition, a cluster is a set of cycles that are alike, whereas the cycles from different clusters are not alike. A cycle is characterized by a set of nodes and links included in a cycle. Let x be a vector of values (0,1), where $x_i = 1$, if node i is a cycle component ($i \in [1, N]$, где N – number of nodes of gene network), otherwise $x_i = 0$.

In order to search for the clusters, we have used the statistics package STATISTICA 5.0 for Windows (Statsoft Inc., USA). Ward's method was used for clustering of cycles (Ward, 1963). In accordance with the Ward's method, we suppose that cluster membership is assessed by calculating the total sum of squared deviations from the mean of a cluster. The criterion for fusion is that it should produce the smallest possible increase in the error sum of squares. We use the Euclidean distance measure between the clusters x and y : $\text{distance}(x, y) = \sum_i (x_i - y_i)^2$.

RESULTS

The apoptosis gene network in hepatitis C is transcription regulatory network including the reactions from the binding of $\text{TNF}\alpha$ to its receptor through subsequent activation of $\text{NF-}\kappa\text{B}$ transcription factor that induces expression of the target genes. Tumor necrosis factor receptor 1 (TNFR1) may trigger distinct signaling pathways leading either to activation of the $\text{NF-}\kappa\text{B}$ transcription factors or to apoptosis. $\text{TNF}\alpha$ -induced trimerization of TNFR1 receptors recruits TNF-receptor death domain protein, TRADD, and several signaling molecules to activate TNFR1. The adaptor proteins, TRAF2 and RIP, stimulate the pathway leading to activation of $\text{NF-}\kappa\text{B}$ and survival of a cell, whereas TNFR1-TRADD-FADD complex mediates activation of caspase cascade and apoptosis. Activated $\text{NF-}\kappa\text{B}$ inducing kinase (NIK) and MEKK1 phosphorylate and activate IKK. Activated IKK phosphorylates $\text{NF-}\kappa\text{B}$ inhibitor $\text{I}\kappa\text{B}$, which associates with $\text{NF-}\kappa\text{B}$ in the cytoplasm. $\text{I}\kappa\text{B}$ becomes ubiquitinated and degraded, and then $\text{NF-}\kappa\text{B}$ translocates into the nucleus and initiates transcription of various survival genes. The discriminative feature of the $\text{NF-}\kappa\text{B}$ -regulated gene network is a closed feedback loop system. Caspase 8 inhibitor FLIP is an important mediator of $\text{NF-}\kappa\text{B}$ -controlled antiapoptotic signals. Several antiapoptotic genes, such as caspase inhibitors c-IAP1, c-IAP2, adaptor proteins TFAF1, TRAF2, Bcl family protein genes A1/Bfl1, A20, Bcl-x, are regulated by the transcription factor $\text{NF-}\kappa\text{B}$ and are involved in feedback loops controlling $\text{NF-}\kappa\text{B}$ transcription activity and gene network behaviour. A divaricated structure of the signal pathways and autoregulatory loops form a set of cycles with positive or negative relationships regulating $\text{NF-}\kappa\text{B}$ activity. We have used cluster

analysis to identify hierarchical structure of the gene network. The results of clustering are demonstrated as the tree diagram (Fig. 1).

The resulting clusters are interpretable in terms of the network. The 157 negative and positive autoregulatory circuits are clustered due to the coalescing nodes NF- κ B/I κ B, c-FLIP, TRAF2, and caspase 3. The network is likely to be partitioned into a set of circuits including NF- κ B-regulated genes and a set of loops containing NF- κ B/I κ B complex. The last module is associated with activation and cytoplasm-nuclear shuttling of transcription factor NF- κ B. NF- κ B/I κ B cluster includes I κ B and NIK subclusters, NF- κ B cluster joins c-FLIP, TRAF2, c-IAP1, c-IAP2, A20, IEX submodules. The separate module of gene network unites the proteins that regulate their own transcription or activity, e.g., A20, Sp1, caspase 3 and 8.

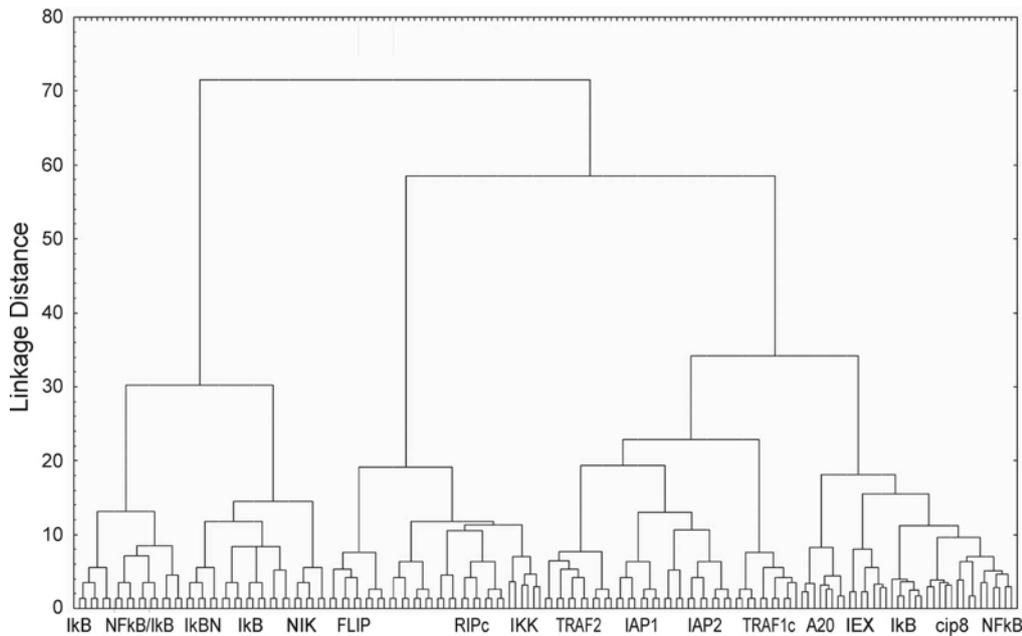


Figure 1. A hierarchical tree clustering feedback 157 circuits of the NF- κ B regulated network.

The apoptosis gene network in hepatitis C has a scale-free topology. A few nodes with a very large number of links, which are called hubs, are represented by the central transcription factors, NF- κ B and p53. HCV proteins modulate various signal transduction pathways binding directly with the key signal molecules and transcription factors. As an important target for antiviral drug development may be viewed the adaptor protein TRAF2. HCV core protein significantly activates NF- κ B pathway through TRAF2 and inhibits TNF α -induced apoptosis. In addition to hepatitis C virus, the nonstructural protein NS5A directly interacts with TRAF2 and inhibits NF- κ B activation. Also, NS5A is a positive regulator of TRAF2-mediated JNK signal pathway, which increases AP1 transcription activity and may play the significant role in pathogenesis of hepatitis C. The activation of transcription factor NF- κ B by TNF α involves phosphatidylinositol 3-kinase (PI3K) signal pathway. IEX-1S that is positively regulated by NF- κ B, promotes TNF α -induced hepatocyte apoptosis through the blockage of the PI3K/Akt survival signaling. NS5A is able to interact directly and specifically with the p85 subunit of PI3K. The association of NS5A with p85 PI3K enhances activity of the PI3K-Akt pathway. In its turn, this fact causes inhibition of apoptosis, therefore, facilitates to HCV persistence in host cells and development of liver cancer. NS5A physically interacts with the tumor suppressor p53 and promotes the cell growth. The role of HCV proteins with regard to p53-induced apoptosis is controversial. HCV core protein also binds directly to the master

transcription factor p53 and modulates apoptosis. While the low level of the core protein enhances transcription activity of p53, the high level of HCV core protein inhibits this activity. Thus, the virus is capable to use the cellular signaling pathways and transcription factors via controlling them for its own advantage.

DISCUSSION

A new paradigm for antiviral therapy focuses on targeting cellular genes that are critical for viral replication and pathogenesis. Several new technologies have contributed to the identification of such genes for hepatitis C virus. These technologies include proteomic approaches that identify the cellular proteins that physically interact with the viral proteins, genomic approaches that identify essential cellular genes, and gene networks approaches that facilitate the studies of virus-host interactions. High-throughput technologies have resulted in more complete description of protein-protein interactions and signaling pathways that have generated reconstruction of the gene networks. By microarray analysis, several potential gene markers of HCV-associated liver disease and more than 100 genes with the altered expression levels were identified. Novel data allow to create the subsequent level of detailization in the reconstruction of HCV-regulated gene network. The application of mathematical methods that quantitatively describe the properties of large-scale gene network is crucial for studying complex diseases such as hepatitis C and cancer.

ACKNOWLEDGEMENTS

The work was supported by the Russian Government Contract No. 02.434.11.3004 and Siberian Branch of the Russian Academy of Sciences (integration project No. 115).

REFERENCES

- Ananko E.A., Podkolodny N.L., Stepanenko I.L. *et al.* (2005) GeneNet in 2005. *Nucl. Acids Res.*, **33**, D425–427.
- Cormen T.H., Leiserson C.E., Rivest R.L., Stein C. (2001) Introduction to Algorithms. MIT Press and McGraw-Hill, 531–539.
- Kolchanov N.A., Ignatieva E.V., Ananko E.A., Podkolodnaya O.A., Stepanenko I.L. *et al.* (2002) Transcription regulatory regions database (TRRD): its status in 2002. *Nucl. Acids Res.*, **30**, 312–317.
- Kountouras J., Zavos C., Chatzopoulos D. (2003) Apoptosis in hepatitis C. *J. Viral. Hepat.*, **10**, 335–342.
- Ward J.H. (1963). Hierarchical grouping to optimize an objective function. *J. Ame. Statistical Association*, **58**, 236–244.

DYNAMIC FILTRATION OF VARIABILITY WITHIN EXPRESSION PATTERNS OF ZYGOTIC SEGMENTATION GENES IN *DROSOPHILA*

Surkova S.Yu.^{*}, *Samsonova M.G.*

St. Petersburg State Polytechnical University, 195251, Russia

^{*} Corresponding author: e-mail: surkova@spbcas.ru

Key words: *Drosophila*, segmentation genes, gene expression, variability in expression

SUMMARY

Motivation: Cycle 14A is a crucial period in the development of *Drosophila*. At this time positions of the segment boundaries are determined by the set of segmentation genes acting as a cascade. During this short period the early variable zygotic expression is canalized into the highly uniform patterns. Analysis of this process using the quantitative data on segmentation gene expression enables to validate the model of segment determination in *Drosophila* (Jaeger *et al.*, 2004).

Results: Variability of patterns at early times takes the form of: variation in the level of gene expression; uncertainty in the time and sequence of formation of individual domains; uncertainty in the way by which a domain can be formed; variation in the position of expression domains. All these sources of variability are decreased by gastrulation. The positional error within the expression domains of zygotic genes is filtered dynamically with time.

Availability: FlyEx database: <http://urchin.spbcas.ru/flyex>

INTRODUCTION

During *Drosophila* embryogenesis the segmented body plan is established through a cascade of maternally and zygotically expressed segmentation genes. Maternal genes set the anteroposterior (A-P) polarity of the egg and form the anterior and posterior protein gradients that act at different points along the egg axis in a concentration-dependent manner. The zygotic genes have been classified according to their mutant phenotypes and expression patterns. "Gap" genes are expressed in one to three broad domains, "pair-rule" genes form seven transverse stripes and "segment polarity" genes manifest in patterns of fourteen stripes about one-cell wide. Of these genes, maternal, gap, and pair-rule genes act during the blastoderm stage, giving rise to initial expression of segment polarity genes at the onset of gastrulation, by which time the segmental pattern is determined (Ingham, 1988).

The morphogenetic field is a fundamental object in developmental biology. In such a field the groups of cells undergo collective determination events in which the developmental fate is stably assigned to individual cells with exquisite spatial precision.

A major biological question motivating this work is that of the regulation, or error-correction, properties of this morphogenetic field. In this work we show that there is substantial variation between individual expression pattern early, but this variation is canalized into highly uniform patterns by the onset of gastrulation. Using the high-throughput methods of acquisition and processing of the quantitative data we characterize

in detail all sources of variability inherent to the segmentation morphogenetic field as well as the dynamic reduction of this variability.

As we perform the analysis of the *Drosophila* segmentation gene expression in parallel with the mathematical modeling of this process, our findings are applied for the validation of the model of segment determination (Reinitz, Sharp, 1995; Jaeger *et al.*, 2004).

MATERIALS AND METHODS

Wild-type (OregonR) *Drosophila* blastoderm embryos were collected, fixed and immunostained for three segmentation gene products as described (Kosman *et al.*, 1998). Embryo confocal images were taken using a laser confocal scanning microscope.

Image processing procedures resulted in the reduction of image information to a quantitative data on gene expression (Janssens *et al.*, 2005). At present our dataset contains confocal scans of 1600 wild type embryos belonging to cleavage cycles 10-14A. Embryos were scanned for the expression of 14 segmentation genes. The images of embryos from cycle 14A were distributed by visual inspection of pair-rule gene expression patterns into 8 temporal equivalence classes (Myasnikova *et al.*, 2001). In this study we used 1D data in the 10 % central strip along the A-P axis of an embryo. As the characteristic features of segmentation gene expression domains we considered the A-P positions of peaks and points, where the level of gene expression reached the predefined threshold. The positions and fluorescence intensity of expression maxima for the late pair-rule stripes were extracted by means of the wavelet decomposition of the signal (Kozlov *et al.*, 2000), while patterns of gap genes and early pair-rule domains were approximated by quadratic splines (Myasnikova *et al.*, 2001). The positional variability of expression patterns and the variability in the level of gene expression were estimated by computing the standard deviations of the positions and fluorescence intensity of the characteristic features using StatSoft Statistica package, version 6.0.

RESULTS

The results of our analysis demonstrate that in the early-middle cycle 14A there is a high variation in the level of expression, shape, time and sequence of formation of the segmentation gene expression domains of the zygotic origin. However all these sources of variability are significantly decreased by gastrulation. Moreover the positional error can be filtered dynamically at a level of each zygotic gene.

Variability of the early gene expression patterns takes the form of:

Variation in the level of gene expression

There is a dramatic variation in the level of early expression of *Kr* and *eve* (Fig. 1a, c), however it is strongly filtered by the late cycle 14A (Fig. 1b, d). For the central *Kr* domain the standard deviation of the fluorescence intensity levels in the individual embryos almost halves by temporal class 6.

Uncertainty in the time and sequence of formation of individual domains

We demonstrate that in the early patterns of the pair-rule genes stripes can be formed at different time and in different sequence. The temporal delay in formation of one stripe may last up to 24 minutes. Fig. 2 shows the early variability in the time of formation of *eve* stripes. Moreover the stripes within the pair-rule expression patterns are formed in different sequence (Fig. 1C) and in one temporal class there can be two quite dissimilar patterns of the same gene. Nevertheless by gastrulation all pair-rule patterns exhibit seven stripes of equal intensity.

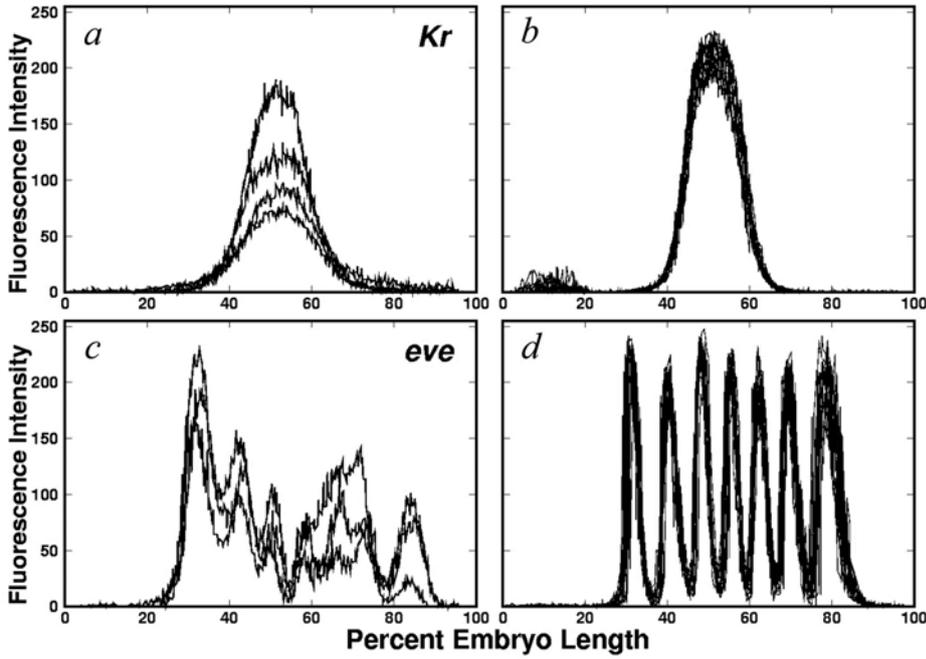


Figure 1. Variability in the intensity level of the early *Kr* patterns (a) is decreased by time class 6 (b). *Eve* expression pattern in time class 3 during the formation of stripes (c) show high variation in shape and intensity. Before gastrulation (d) all *eve* patterns have the identical shape and are much less variable in the expression level.

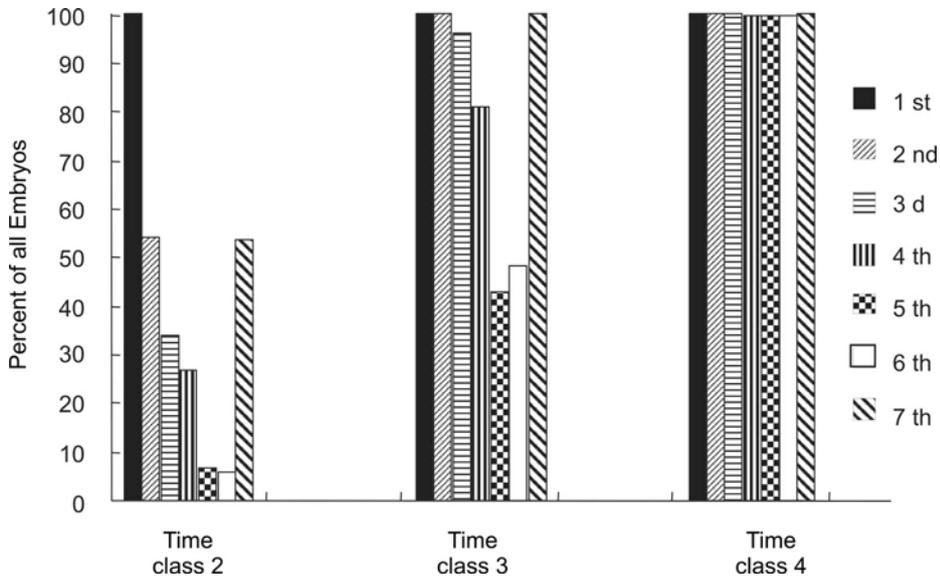


Figure 2. Variability in time of formation of *eve* stripes in the early time classes. For each time point the degree of stripe formation is considered as a percent of the total number of embryos.

Uncertainty in the way by which a domain can be formed

According to our observations of quantitative expression data, the stripes within the pair-rule patterns may be formed in three different ways: splitting, budding and *de novo* formation (not shown). For some domains these ways of formation may even vary from

embryo to embryo. This uncertainty is also abolished by the end of cycle 14A when the shape of all pair-rule domains becomes identical.

Variation in the position of expression domains

We have recently demonstrated that the level of positional error of the eve domain in cycle 13 is as high as that of the Bcd protein gradient (Surkova, Samsonova, 2004). After the formation of stripes the variability within the eve expression pattern becomes lower. The latest results showed that the positional variability is dynamically filtered in all examined zygotic genes (Fig. 3). The positional error of the eve pattern in the middle of cycle 14A is three times lower than that in cycle 13 and at time class 1. In *Kr* and *hb* the early variability also decreases dynamically before time class 2 and doesn't change thereafter (Fig. 3). We conclude that the positional error is dynamically filtered during the first 20 minutes of cycle 14A.

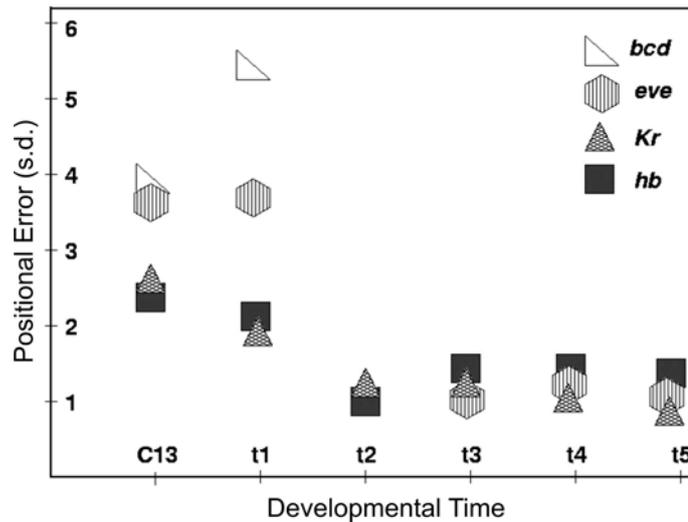


Figure 3. Dynamic filtration of the positional error at a level of zygotic gene expression. We consider the standard deviations (s.d.) in the positions of the posterior border of the anterior *hb* domain, maximum of the central *Kr* domain, and the posterior border of the early *eve* domain, that later corresponds to the position of the *eve* 3rd eve stripe. Changes in the positional error of these domains are compared to the level of variability in position of the 12 % concentration threshold of the Bcd protein gradient. All these features have approximately the same positions along the A-P axis of an embryo (45–55 % embryo length). In cycle 14A we show the Bcd variability only for time class 1, as it remains at the same level thereafter.

DISCUSSION

In this paper we show that the precision of the segment determination is established dynamically. During about 50 minutes the extremely variable patterns develop into the highly uniform expression domains with the stable shape and the precise positions along the main body axis. Such filtration occurs by the time of initial positioning of the parasegment boundaries, one of the main patterning decisions during the *Drosophila* development. The dynamic filtration of the variability within the genetic network controlling *Drosophila* segmentation may be interpreted in terms of the classical regulatory and error-correcting properties of the morphogenetic fields and used for the validation of the mathematical model of segment determination (Jaeger *et al.*, 2004).

ACKNOWLEDGEMENTS

This work is supported by NIH grant RR07801.

REFERENCES

- Ingham P. (1988) The molecular genetics of embryonic pattern formation in *Drosophila*. *Nature*, **335**, 25–34.
- Jaeger J., Surkova S., Blagov M., Janssens H., Kosman D., Kozlov K., Manu, Myasnikova E., Vanario-Alonso C.-E., Samsonova M., Sharp D., Reinitz J. (2004) Dynamic control of positional information in the early *Drosophila* embryo. *Nature*, **430**, 368–371.
- Janssens H., Kosman D., Vanario-Alonso C.-E., Jaeger J., Samsonova M., Reinitz J. (2005) A high-throughput method for quantifying gene expression data from early *Drosophila* embryo. *Development, Genes and Evolution*, **225**(7), 374–381.
- Kosman D., Small S., Reinitz J. (1998) Rapid preparation of a panel of polyclonal antibodies to *Drosophila* segmentation proteins. *Development, Genes and Evolution*, **208**, 290–294.
- Kozlov K., Myasnikova E., Samsonova M., Reinitz J., Kosman D. (2000) Method for spatial registration of the expression patterns of *Drosophila* segmentation genes using wavelets. *Computational Technologies*, **5**, 112–119.
- Myasnikova E., Samsonova A., Kozlov K., Samsonova M., Reinitz J. (2001) Registration of the expression patterns of *Drosophila* segmentation genes by two independent methods. *Bioinformatics*, **17**, 1, 3–12.
- Reinitz J., Sharp D. (1995) Mechanism of *eve* stripe formation. *Mechanisms of Development*, **49**, 133–158.
- Surkova S.Yu., Samsonova M.G. (2004) Temporal and spatial precision in formation of segmentation gene expression domains in *Drosophila*. In *Proceedings of the 4th International Conference on Bioinformatics of Genome Regulation and Structure (BGRS'2004)*, 25-30 July, Novosibirsk, Russia, **2**, 142–145.

3.3. MODELLING OF MORPHOGENESIS

A CELLULAR AUTOMATON TO MODEL THE DEVELOPMENT OF SHOOT MERISTEMS OF *ARABIDOPSIS THALIANA*

Akberdin I.R.^{*2}, Ozonov E.A.¹, Mironova V.V.², Komarov A.V.¹, Omelianchuk N.A.²

¹ Novosibirsk State University, Novosibirsk, 630090, Russia; ² Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia

* Corresponding author: e-mail: akberdin@bionet.nsc.ru

Key words: cellular automaton, mathematical model, development of shoot meristems, parameters of model, period of division

SUMMARY

Motivation: Development of organisms is a very complex process in that a lot of gene networks of different cell types are to be integrated. Development of a cell automaton (Ermentrout, Edelshtein-Keshet, 1993) that models the morphodynamics of different cell types is the first step in the understanding and analysis of the regulatory mechanisms that underlie the functioning of the developmental gene networks.

Results: A model of a cell automaton has been developed, which simulates the embryonic development of the shoot meristems of *Arabidopsis thaliana*. The model adequately describes the basic stages in the development of this organ in wild and mutant types.

INTRODUCTION

Postembryonic development of the above-ground part of higher plants depends on the expression of apical shoot meristems, a dynamic structure which forms leafage, flowers and scape. The formation of the apical shoot meristems occurs at the earliest stages of embryogenesis. Furthermore, the functioning of the promeristem triggers the development of the germ layers, and at that time a complex meristem structure is being formed (Fig. 1) (Sharma, Fletcher, 2003). The cells of the apical and basal parts have different expression of different genes in a 16-cell-like embryo. No further development of embryos is possible without shoot meristems, or at least their apical parts (Kaplan *et al.*, 1997).

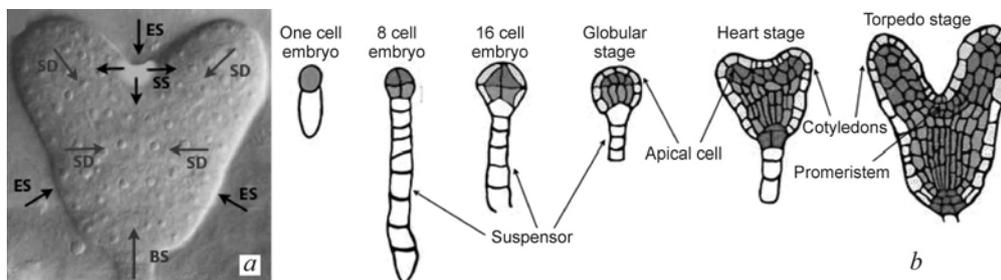


Figure 1. Directions of the model's hypothetical signal distributions in heart-stage embryo tissues (a): ES is External Signal, SS is Stem Signal, SD is Signal of Differentiation, and BS is Basal Signal; developmental stages of a plant embryo with the indication of organs and tissues significant to the model (b).

The cell automaton was developed to model the development of shoot meristems of the *Arabidopsis thaliana* in embryogenesis. Modeling covered the initiation of meristems, the formation of their complex structure and functioning. Here the embryo is described as a two-dimensional array of cells, the rates of division of which depend on the cellular environment. The cells in the model can receive and, depending on the cell type, produce signals that can be received by other cells in the model. The biological meaning of a signal is the concentration of certain diffusing substances, morphogenes, which have a specific influence on the cell.

METHODS AND ALGORITHMS

The model assumes that the stage and rates of division of an individual cell depend on the influence of signals that are coming from other cells of the embryo. Under the model, four biologically meaningful signals, which unambiguously simulate the morphodynamic of the cell tissues at the generally accepted level of abstraction, were selected (Fig. 1).

All the cells in the model of the embryo plants can be classified according to the type of the signal they produce:

1. *Null*. These cells mean vacancy neither produce signals nor divide.
2. *NullEx*. Cells of the epidermis layer. They produce *ES* (*External Signal*) and are represented around the entire perimeter of the embryo. They do not divide, but we had them surround cell embryos in the model.
3. *NullSus*. Cells of the suspensor. They produce *BS* (*Basal Signal*) and are confined to the lower part of the embryo. There are two *NullSus* cells in the model.
4. *Lateral*. These cells imitate “differentiated” cells, which produce *SD* (*Signal of Differentiation*).
5. *Promeristem*. Cells of the embryo meristems. These cells produce *SS* (*Stem Signal*) and are confined to the upper part of the embryo. During development, they change into *L2meristem* and *L3meristem* type cells.
6. *Transit*. Cells near the meristem. They also produce *SD*, but have the highest rates of division.
7. *L2meristem*. Cells of the meristem. They that situated in second layer from the epidermic layer of the upper part of the embryo. These cells produce *SS*.
8. *L3meristem*. Cells one layer down from *L2meristem* type cells. These cells produce *SS*.

Each cell has a set of internal parameters to characterize its state:

1. *Type*. Cell type.
2. *ES0*, *BS0*, *SS0*, *SD0* are the values of the signals produced by the cell.
3. *ES*, *BS*, *SS*, *SD* are the values of the signals accepted by the cell.
4. K_{ij} is the characteristic of cell state at position (i, j) calculated as the ratio of *SS* to *SD*. At the current point of time, the cell is influenced in the state characterized by the

$$\text{parameter } K_{ij} = \frac{SS_{ij}}{SD_{ij}} .$$

5. T_{ij} is the period of cell division, which depends on the current value of the characteristic K_{ij} .
6. Tp_{ij} is the number of iterations after the last division of the cell at position (i, j).

We calculate the overall influence of all the cells on the cell at position (i,j) by the following formulas:

$$\begin{aligned}
 ES_{ij} &= \alpha_{ij}^E \sum ESO_{km} e^{-\frac{n}{R_E}}, \\
 BS_{ij} &= \alpha_{ij}^B \sum BSO_{km} e^{-\frac{n}{R_B}}, \\
 SS_{ij} &= \alpha_{ij}^S \sum SSO_{km} e^{-\frac{n}{R_S}}, \\
 SD_{ij} &= \alpha_{ij}^D \sum SDO_{km} e^{-\frac{n}{R_D}}.
 \end{aligned}$$

The summation is performed over all the cells of a particular tissue, including the epidermis layer and the suspensor; (k, m) is the position of the cell, the influence of which was taken into account; $n = |k - i| + |m - j|$ where i, j are the considered cell coordinates, k, m are the affected cell coordinates; R_B, R_S, R_D are the constants which characterize penetrance for ES, BS, SS and SD respectively; $\alpha_{ij}^B, \alpha_{ij}^S, \alpha_{ij}^D$ are the constants which characterize the sensitivity of the cell to a certain type of signal. This constant depends only on the type of the cell.

The qualitative behavior of the function T_{ij} is shown in the Fig. 2 according to the biological position:

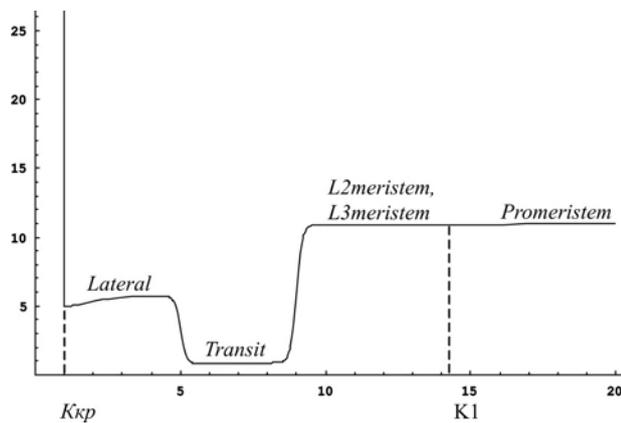


Figure 2. Dependence of division periods on the parameter K for different cell types.

As can be seen from Fig. 2, the embryo meristem cells (*Promeristem, L2meristem, L3meristem*) in the tissue divide slower than others; the *Transit* type cells divide very quickly; the rates of division of the *Lateral* type cells are medium. When the parameter K takes on the value K_{xp} , the period of division becomes infinity and the cells will not divide anymore. When the parameter K takes on the threshold value K_1 , the *Promeristem* type cells divide into *L2meristem* and *L3meristem* cells. This dependence was revealed

using the function $f(x) = \frac{1}{1 + e^{\frac{x-\mu}{\sigma}}}$, where x is value of K_{ij} , μ equals critic values of K_{ij}

and σ equals 0.1. If T_{ij} is not an integer, the function T_{ij} is rounded in accordance with the standard rules of adjustment. To ascertain whether or not a cell is undergoing division, Tp_{ij} and T_{ij} are compared. At the next iteration, the value Tp_{ij} is increased by one for each non-dividing cell. If $Tp_{ij} > T_{ij}$, division into two daughter cells is ascertained. For each daughter cell, Tp_{ij} is equal to zero. Noteworthy, if T_{ij} increases more rapidly than Tp_{ij} , the cell will not divide.

RESULTS

The cell automaton adequately describes the developmental morphodynamics of shoot meristems of *Arabidopsis thaliana* in embryogenesis (Fig. 3).

Varying the parameters of the cell automaton, we have successfully modeled the following mutations described in the literature:

a) the meristem forms but cannot cope with auxin flowing in from the germ layers, and so it differentiates. This leads to the formation of joined germ layers and plant development stops (Aida *et al.*, 1999);

б) the meristem forms, and so do germ layers, but no regulation is exerted on the meristem cells population. As a result, the meristem cells differentiate and the meristem zone shrinks (Clark *et al.*, 1995; Sharma, Fletcher, 2003).

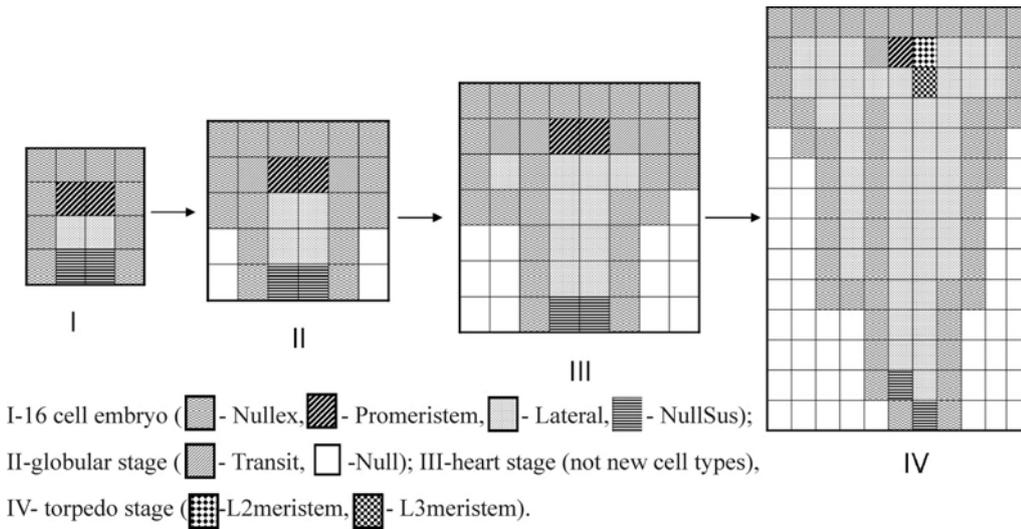


Figure 3. Developmental stages of a plant embryo modeled by the cell automaton.

DISCUSSION

The creation of a cell automaton, which imitates the morphodynamics of embryo development by regulation of signals produced by different embryonic cells, is the first step in modeling the process of development in general and in modeling the gene network for morphogenesis in particular. The formation of plant meristems in embryogenesis is characterized by a combination of a violent development of differentiating tissue and a stable development of its stem cells. Both processes were modeled in the cell automaton being reported. Not only is this automaton a tool for predicting the dynamics of the division process and the cell differentiation process which underway in the systems being considered, but also for the examination of how real mutations influence the system. As a progression of this work, we plan to develop a cell automaton, which makes for the modeling of various experimentally induced mutations. Sophistication of characteristic cell definitions will be added for a more adequate description. We plan to enable graph visualization of this cell automaton.

ACKNOWLEDGEMENTS

This work was supported in part by Innovation project IT-CP.5/001 “Development of software for computer modeling and design in post genomic system biology (system biology *in silico*)”. The authors are grateful to Vitaly Likhoshvai for valuable discussions.

REFERENCES

- Aida M., Ishida T., Tasaka M. (1999) Shoot apical meristem and cotyledon formation during Arabidopsis embryogenesis: Interaction among the CUP-SHAPED COTYLEDON and SHOOT MERISTEMLESS genes. *Development*, **119**, 823–831.
- Clark S.E., Running M.P., Meyerowitz E.M. (1995) CLAVATA3 is a specific regulator of shoot and floral meristem development affecting the same processes as CLAVATA1. *Development*, **121**, 2057–2067.
- Ermentrout G. B., Eldenshtein-Keshet L. (1993) Cellular automata approaches to biological modeling. *J. Theor. Biol.*, **160**, 97–133.
- Kaplan D.R., Cooke T.J. (1997) Fundamental concepts in the embryogenesis of Dicotyledons: a morphological interpretation of embryo mutants. *The Plant Cell*, **9**, 1903–1919.
- Sharma V.K., Fletcher J.C. (2003) Maintenance of Shoot and Floral Meristem Cell Proliferation and Fate. *PNAS*, **100**, 11823–11829.

THE GENE NETWORK DETERMINING DEVELOPMENT OF *DROSOPHILA MELANOGASTER* MECHANORECEPTORS

Bukharina T.A.^{*1}, Katokhin A.V.^{1,2}, Furman D.P.^{1,2}

¹Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ²Novosibirsk State University, Novosibirsk, 630090, Russia

* Corresponding author: e-mail: bukharina@bionet.nsc.ru

Key words: gene network, proneural cluster, *achaete-scute complex (AS-C)*, signaling pathways

SUMMARY*

Motivation: The integration and comprehension of the experimental data on the genetic control of the development of *Drosophila melanogaster* mechanoreceptors accumulated so far require a formalized representation of the data in question in the context of a gene network for their further analysis and modeling of morphogenesis molecular mechanisms.

Results: This work reports reconstruction of the gene network Neurogenesis: Determination, describing determination of the precursor cells for drosophila mechanoreceptors and analysis of the information contained herein. The main functional elements with this gene network are detected, and the general principles underlying its function are considered.

Availability: <http://www.mgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/>

INTRODUCTION

Mechanoreceptors (external sensory organs) are elements of the *Drosophila* peripheral nervous system. By now, a considerable volume of the information related to the regulation of mechanoreceptor development has been accumulated (see, for example, reviews by Ghysen, Thomas, 2003; Lai, Orgogozo, 2004; and references therein). Certain key moments of this process are known, and individual stages of the mechanoreceptor development are modeled (Marnellos, Mjolsness, 1998; Meir *et al.*, 2002); however, understanding of the integral pattern of their development requires systematization of the available data and their representation in a unified formalized form. The GeneNet technology allows this demand to be met via a reconstruction of the gene network of mechanoreceptor morphogenesis. This forms the background for a future construction of the model reproducing adequately not only the development of various sensory organs, but also the overall peripheral nervous system. This work reports the first stage in the development of mechanoreceptors—determination of the cells of proneural clusters.

METHODS AND ALGORITHMS

The gene network was reconstructed using the GeneNet technology, developed at the Laboratory of Theoretical Genetics with the Institute of Cytology and Genetics SB RAS (Ananko *et al.*, 2005). The gene network Neurogenesis: Determination was constructed based on annotation of 132 scientific publications and at the moment comprises the

information about 187 objects, including 31 genes, 54 proteins and protein complexes, 85 interrelations between the network components and 3 processes.

RESULTS

Each mechanoreceptor consists of four cells, which originate from one precursor cell (SOP, Sensory Organ Precursor cell)¹¹ via two sequential divisions. Individualization of these SOP from the cells of imaginal discs is the determinative moment in the morphogenesis of mechanoreceptors; this proceeds in two stages, each having its individual genetic support.

The first stage involves the separation of proneural cluster—a group of cells, each displaying a potential of developing into SOP. The leading part here is assigned to the so-called proneural genes, constituents of the *achaete-scute* (*AS-C*) complex. The topology of proneural clusters is determined by interaction between the regulatory elements located within the complex with transcription factors, the proteins of the EGFR (Epidermal Growth Factor Receptor) signaling pathway, and the products of *u-shaped*, *iroquois*, and *pannier* genes (Culi *et al.*, 2001).

The binding of EGFR to the ligand SPITZ initiates a cascade of events leading to activation of *pointed* gene transcription. In turn, two POINTED isoforms further play the role of transcription activators for the proneural genes (Fig. 1a).

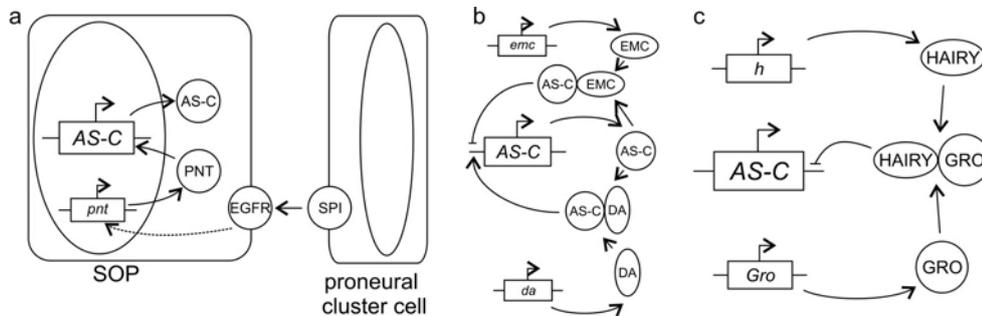


Figure 1. A scheme of the processes involved in regulation of the first stage in mechanoreceptor development: *a* – EGFR signal pathway; *b*, *c* – expression regulation of proneural genes. (*b*) The heterodimers of AS-C and DA proteins enhance the expression of proneural genes, while the heterodimers of AS-C and EMC are involved in a negative expression regulation of proneural genes. (*c*) The negative regulation of AS-C gene expressions by the repressor complex HAIRY/GRO.

Another positive regulator of *AS-C* gene expressions is the product of gene *daughterless* (*da*), a protein of bHLH type. It forms heterodimeric complexes with AS-C proteins; these heterodimers are capable of binding to E boxes (CANNTG) in the regulatory regions of the corresponding genes, thereby activating their expression (Fig. 1b). On the other hand, the heterodimers of AS-C proteins and the product of *extramacrochaete* (EMC) gene are involved in a negative regulation of *AS-C* gene expression, since EMC, an HLH-type protein, lacks the DNA-binding basic domain. Competing with DA for binding AS-C proteins, EMC decreases the transcription level of the genes of this complex. The protein product of *hairy* gene, HAIRY, is the repressor for

¹¹ **The following abbreviations are used:** *AS-C* – *achaete-scute* complex; *SOP* – *Sensory Organ Precursor cell*; *EGFR* – *Epidermal Growth Factor Receptor*; *PNT* – *pointed*; *DA* – *daughterless*; *EMC* – *extramacrochaete*; *GRO* – *groucho*; *h* – *hairy*; *E(spl)-C* – *Enhancer of split complex*; *D* – *Delta*; *N* – *Notch*; *N_{id}* – *NOTCH intracellular domain*.

the transcription activity of *AS-C* genes. It needs cofactors, in particular, GRO, the product of gene *groucho*, to fulfill its regulatory activity (Fig. 1c).

The second stage is connected with a more precise SOP location within the proneural cluster, which occurs during lateral inhibition, controlled by the products of genes from Notch signaling pathway (Ghysen, Thomas, 2003). The rest cells of proneural cluster develop into epidermal cells. The mandatory condition for SOP separation within the cluster is the threshold level of AS-C proteins in this cell.

Insignificant distinctions in the expression levels of *AS-C* genes result in the different concentrations of NOTCH and DELTA proteins, thereby triggering the molecular mechanism of intercellular interactions—the lateral inhibition. Consequently, the necessary level of AS-C proteins is reached in only one cell of the proneural cluster; this is the particular cell that subsequently becomes the SOP of mechanoreceptor. In the neighbor cells, genes *AS-C* are repressed, and these cells develop into ectodermal cells.

The central part of the Notch signaling pathway is represented by NOTCH receptor, its ligand DELTA, and transcription factor SUPPRESSOR OF HAIRLESS, whose target is the genes of *Enhancer of split (E(spl)-C)* complex. The Notch signaling pathway functions in the following manner. The ligand DELTA, localized to the SOP surface, binds to the NOTCH receptor, localized to the membrane of a cell adjacent to SOP. The action of a number of proteins causes the cleavage of NOTCH intracellular domain (N_{id}), which, once it enters the cell, forms the complex with SUPPRESSOR OF HAIRLESS. The resulting complex activates the transcription of *E(spl)-C* genes, whose products repress the transcription of their own targets, *AS-C* genes (Fig. 2).

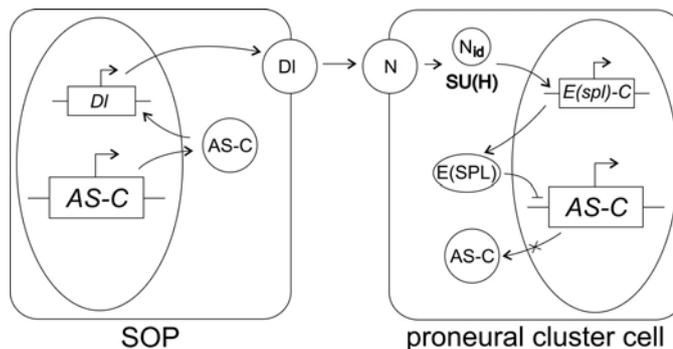


Figure 2. A scheme of expression regulation of the AS-C complex genes via the NOTCH signaling pathway.

Concurrently, the Notch signaling pathway ceases stimulating the proneural determination of cells of the cluster via the EGFR signaling pathway.

Then each SOP divides twice, giving rise to socket, shaft, and sheath cells and to bipolar neuron. Differentiation of these cells is controlled by the Notch signaling pathway and an ensemble of genes, where *numb* and *Suppressor of Deltex* are the main components (Rhyu *et al.*, 1994).

CONCLUSIONS

Analysis of the reconstructed gene network Neurogenesis: Determination allowed the authors to detect the key components and molecular genetic processes: 1) genes *AS-C*, whose expression is necessary for determining the position of mechanoreceptor on drosophila body; 2) the mechanism enhancing the transcription of *AS-C* genes, which allows the proteins AS-C to reach the threshold concentrations via a self-regulation by DA/AS-C heterodimers; 3) the mechanisms repressing the transcription of *AS-C* genes with

involvement EMC/AS-C heterodimers and transcription factor HAIRY; and 4) the signaling pathways EGFR and Notch, providing transducing of signaling information between the cells of proneural cluster. The direction of proneural cluster cell development depends on the interactions of all described molecular mechanisms of regulation of key component – AS-C genes –transcription both inside cell and between the cells of proneural cluster.

ACKNOWLEDGEMENTS

The work was supported by the Program “The Origin and Evolution of the Biosphere” (state contract No. 10104-34/P-18/155-270/1105-06-001/28/2006-1), Program of the Basic Research of the RAS Presidium “Molecular and Cellular Biology” (10.4), the interdisciplinary integration project on the basic research of the SB RAS no. 119, and the Program of the Federal Agency for Science and Innovations (state contract No. 02.467.11.1005, project no. IT/KP No. 5/001). The authors are grateful to G.B. Chirikova for translating this paper into English.

REFERENCES

- Ananko E.A., Podkolodny N.L. *et al.* (2005) GeneNet in 2005. *Nucl. Acids Res.*, **33**, D425–D427.
- Culi J., Martin-Blanco E., Modolell J. (2001) The EGF receptor and N signaling pathways act antagonistically in *Drosophila* mesothorax bristle patterning. *Development*, **128**, 299–308.
- Ghysen A., Thomas R. (2003) The formation of sense organs in *Drosophila*: a logical approach. *Bioassays*, **25**, 802–807.
- Lai E.C., Orgogozo V. (2004) A hidden program in *Drosophila* peripheral neurogenesis revealed: fundamental principles underlying sensory organ diversity. *Dev. Biol.*, **269**, 1–17.
- Marnellos G., Mjolsness E. (1998) A gene network approach to modeling early neurogenesis in *Drosophila* Pac. *Symp. Biocomput.*, 30–41.
- Meir E., von Dassow G. *et al.* (2002) Robustness, flexibility, and the role of lateral inhibition in the neurogenic network. *Curr. Biol.*, **12**, 247–349.
- Rhyu M.S., Jan L.Y., Jan Y.N. (1994) Asymmetric distribution of numb protein during division of the sensory organ precursor cell confers distinct fates to daughter cells. *Cell*, **76**, 477–491.

A MODEL FOR SENSING THE MORPHOGEN HEDGEHOG CONCENTRATION GRADIENT I: A NUMERICAL STUDY

Kogai V.V.^{*2}, *Gunbin K.V.*¹, *Omelyanchuk L.V.*¹, *Fadeev S.I.*²

¹Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ²Sobolev Institute of Mathematics, SB RAS, Novosibirsk, 630090, Russia

* Corresponding author: e-mail: kogai@math.nsc.ru

Key words: pattern formation, positional information, mathematical model

SUMMARY

Motivation: In the present study, our aim was to reveal the molecular machinery responsible for interpretation of the Hedgehog morphogene gradient.

Results: The model for sensing the concentration gradient of the Hedgehog morphogen has been developed. Using the proposed numerical method, stationary solutions that experimental observations have been found.

INTRODUCTION

Development of a specific type of cells that form the anterior-posterior compartment boundaries of the wing imaginal disc is controlled by a specific molecular module. The transcription factor Ci in either form has identical binding sites in the enhancers of the genes, transcription of which it regulates (Muller, Basler, 2000). In the absence of the morphogen Hedgehog (Hh), the receptor of the morphogen Ptc inhibits the activity of the protein Smo, which lead to site-specific cleavage of the transcription factor Ci. The N-terminal cleavage product of Ci is transported into the nucleus and suppresses the gene cassette. When Hh interacts with its Ptc receptor, Ptc does not have any inhibitory effect on Smo signaling activity. Thus activated, Smo initiates a series of reactions, which lead to stabilization of full-length Ci and its transport into the nucleus and. Full-length Ci activates transcription of the gene cassette (Lum, Beachy, 2004).

MODEL

A chemical kinetic schema for the block, which receives Hh concentration gradient information, is presented in (Fig. 1). For simplicity we introduced the protein complex CYT which includes four proteins (Cos2, Fu, Su(Fu), PKA). In the $CYT_{(p)}$ form, the active proteins are Cos2 and Fu; in the $CYT_{(c)}$ form, Su(Fu) and PKA (Fig. 1). Since morphogenesis is a much slower process than transcription, morphogen diffusion or protein complex assembly, we may analyze its stationary phase.

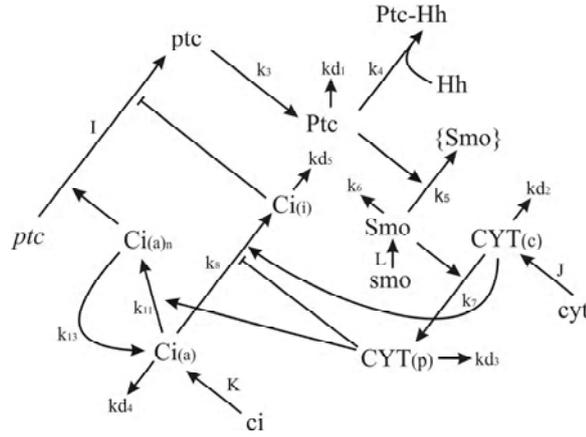


Figure 1. A chemical kinetics schema of the block, which receives Hh concentration gradient. Non-active Smo is given in curly brackets. Pointed arrows symbolize reactions or flows; T-shaped arrows, suppression. kd_i , where $i = 1, 2, 3 \dots$, describes degradation reactions; k_i , the chemical kinetic constants of other reactions. Letters I, J and K stand for constant inflow rates.

The dynamics of pattern formation of the anterior-posterior compartment boundary of the *Drosophila* wing is described by the following system of differential equations:

$$\begin{cases}
 \frac{\partial[\text{ptc}]}{\partial t} = I + \frac{k_0 \cdot [\text{Ci}_{(a)_n}]}{k_1 + k_2 \cdot [\text{Ci}_{(i)}] + [\text{Ci}_{(a)_n}]} - k_3 \cdot [\text{ptc}] \\
 \frac{\partial[\text{Ptc}]}{\partial t} = k_{3a} \cdot [\text{ptc}] - \frac{[\text{Hh}] \cdot [\text{Ptc}]}{k_4 + [\text{Hh}]} \cdot M - kd_1 \cdot [\text{Ptc}] \\
 \frac{\partial[\text{Smo}]}{\partial t} = \frac{1 + k_4^{-1} \cdot [\text{Hh}]}{1 + k_4^{-1} \cdot [\text{Hh}] + k_5^{-1} \cdot [\text{Ptc}]} \cdot L - k_6 \cdot [\text{Smo}] \\
 \frac{\partial[\text{CYT}_{(c)}]}{\partial t} = J - k_7 \cdot [\text{Smo}] \cdot [\text{CYT}_{(c)}] - kd_2 \cdot [\text{CYT}_{(c)}] \\
 \frac{\partial[\text{CYT}_{(p)}]}{\partial t} = k_7 \cdot [\text{Smo}] \cdot [\text{CYT}_{(c)}] - kd_3 \cdot [\text{CYT}_{(p)}] \\
 \frac{\partial[\text{Ci}_{(a)}]}{\partial t} = K - \frac{k_8 \cdot [\text{Ci}_{(a)}] \cdot [\text{CYT}_{(c)}]}{k_9 + k_{10} \cdot [\text{CYT}_{(p)}] + [\text{Ci}_{(a)}]} - kd_4 \cdot [\text{Ci}_{(a)}] + k_{13} \cdot [\text{Ci}_{(a)_n}] - \frac{k_{11} \cdot [\text{Ci}_{(a)}] \cdot [\text{CYT}_{(p)}]}{k_{12} + [\text{Ci}_{(a)}]} \\
 \frac{\partial[\text{Ci}_{(a)_n}]}{\partial t} = \frac{k_{11} \cdot [\text{Ci}_{(a)}] \cdot [\text{CYT}_{(p)}]}{k_{12} + [\text{Ci}_{(a)}]} - k_{13} \cdot [\text{Ci}_{(a)_n}] \\
 \frac{\partial[\text{Ci}_{(i)}]}{\partial t} = \frac{k_8 \cdot [\text{Ci}_{(a)}] \cdot [\text{CYT}_{(c)}]}{k_9 + k_{10} \cdot [\text{CYT}_{(p)}] + [\text{Ci}_{(a)}]} - kd_5 \cdot [\text{Ci}_{(i)}],
 \end{cases} \quad (1)$$

where $k_0 = 0.5 \text{ nmol/s}$, $k_1 = 1.0 \text{ nmol}$, $k_2 = 0.1$, $k_3 = k_{3a} = 0.1 \text{ s}^{-1}$, $k_4 = 0.58 \text{ nmol}$, $k_5 = 0.083 \text{ nmol}$, $k_6 = 0.5 \text{ s}^{-1}$, $k_7 = 0.1 \text{ nmol}^{-1} \times \text{s}^{-1}$, $k_8 = 0.1 \text{ s}^{-1}$, $k_9 = 0.001 \text{ nmol}$, $k_{10} = 0.001$, $k_{11} = 1.0 \text{ s}^{-1}$, $k_{12} = 0.01 \text{ nmol}$, $k_{13} = 0.3 \text{ s}^{-1}$, $kd_i = 0.1 \text{ s}^{-1}$ ($i = 1, 2, \dots, 5$), $M = 1 \text{ s}^{-1}$, $L = 0.00001 \text{ nmol/s}$, $J = K = 0.1 \text{ nmol/s}$, $L = 1 \text{ nmol/s}$, $D_H = 0.1 \text{ nm}^2/\text{s}$ (von Dassow *et al.*, 2000; Lai *et al.*, 2004).

The effects of the different forms of Ci on *ptc* gene transcription were described using the Michaelis-Menten equation, assuming that the intracellular fraction of the repressor form of Ci ($\text{Ci}_{(i)_n}$) is a competitive repressor of the intracellular fraction of the activator form of Ci ($\text{Ci}_{(a)_n}$). In Equation 2 we considered only active membrane form of Ptc protein¹². By analogy with the work of Lai and the associates, we introduced in Equation 3 the Scatchard equations describing “effective interactions” between Ptc and Smo (Lai *et*

¹² There are at least 3 Ptc protein forms: cytoplasmic inactive form, active (unbound with Hh) and inactive (bound with Hh) membrane forms.

al., 2004). Equations 4 and 5 describe the dynamics of the high-molecular-weight complex $CYT_{(c)}$ and $CYT_{(p)}$ correspondingly. The transformation process of Ci from activator form to inhibitor form is described by Equations 6 and 8. The translocation of the activator form of Ci into the nucleus and back into the cytoplasm is described by Equation 7 of System (1).

The Hh morphogen is secreted specifically by posterior compartment cells and diffuse into anterior compartment cells, thus forming a stationary concentration gradient in the anterior compartment. Under the model being considered, these processes are represented by a diffusion equation which describes a flow.

$$0 \leq r \leq R, \frac{\partial[Hh]}{\partial t} = D_H \frac{\partial^2[Hh]}{\partial r^2} - \frac{[Hh] \cdot [Ptc]}{k_4 + [Hh]} \cdot M, \quad Hh(t, 0) = 1, Hh(t, R) = 0$$

(2)

Here, this diffusion is one-dimensional, r is the distance between the anterior-posterior compartment boundary and a location in the anterior compartment.

METHODS AND ALGORITHMS

Consider stationary solutions of (1), (2). Let $u = [Hh]$, y is vector of solution of system (1), where $y_2 = [Ptc]$. The stationary solutions are defined from boundary value problem and a nonlinear system:

$$0 \leq r \leq R, \frac{d^2 u}{dr^2} - M \frac{y_2 u}{k_4 + u} = 0, \quad u(0) = 1, u(R) = 0 \quad (3)$$

$$f(y; u, p) = 0, \quad 0 \leq u \leq 1. \quad (4)$$

where p is the vector of parameters.

Step 1 of the method is the following. Let

$$\frac{dy}{dt} = f(y; u, p) \quad (5)$$

be the equations of chemical kinetics, where u , $0 \leq u \leq 1$, is one of the system's parameters. By integration (5) with respect to t over a sufficiently large interval, we find an approximate stationary solution. This solution is then used as the initial approximation in Newton's method, with which we seek to find a better approximation of the system's solution (5) at $u=0$ for finding other stationary solutions of (5) depending on u using the continuation method.

Step 2. Applying the continuation method (Fadeev, Kogai, 2004) to u , we define, on the interval $[0, 1]$, the vector function $y(u; p)$, which is the solution to system (5). In particular, this is how we find the dependence of y_2 on u .

Step 3. We introduce a mesh for u and obtain the interpolation of the function $y_2(u; p)$ using the interpolation cubic spline $g(u)$ of the class $C_{[0,1]}^2$. With this, we can define the boundary value problem to get the values of the function $u(r)$:

$$0 \leq r \leq R, D_H \frac{\partial^2 u}{\partial r^2} - \frac{u \cdot g(u) \cdot M}{k_4 + u} = 0, \quad u(0) = 1, u(R) = 0. \quad (6)$$

Step 4. We seek the solution to the system of nonlinear equations, which is the discrete case of boundary value problem (5),

$$D_H \frac{u^{j-1} - 2u^j + u^{j+1}}{h^2} - \frac{u^j \cdot g(u^j) \cdot M}{k_4 + u^j} = 0, \quad j = 1, 2, \dots, n-1, \quad u^0 = 1, u^n = 0,$$

which follows from (6). The efficient, quickly diverging method was the one of simple iteration with the initial approximation as follows: $u^j = 1 - r_j / R$, $j=0,1,\dots,n$.

Note that the next approximation of the mesh values of $u(r)$ is found by solving a system of linear algebraic equations with a tridiagonal matrix of coefficients. The all parameters in the model only take on positive values, thus the matrix is diagonally dominant and therefore the system of equations can be solved using the efficient sweep method.

Step 5. We now defined the mesh values of the other components of the vector $y^j = y(u^j)$, $u^j = u(r_j)$, using interpolation cubic spline of the class $C_{[0,1]}^2$ for each of them.

RESULTS

In our model, the concentrations of all the items that constitute the molecular mechanism of the Hh cascade behave exactly as in the experiment (Fig. 2). The total concentration of all the forms (membrane active and inactive, and cytoplasmic) of the protein Ptc on the cell surface is calculated empirically¹³ as $[Ptc_T] = [Ptc] + \alpha \cdot [Ptc] \cdot [Hh]$ and is in good agreement with experimental data (Omelyanchuk, Gunbin, 2004).

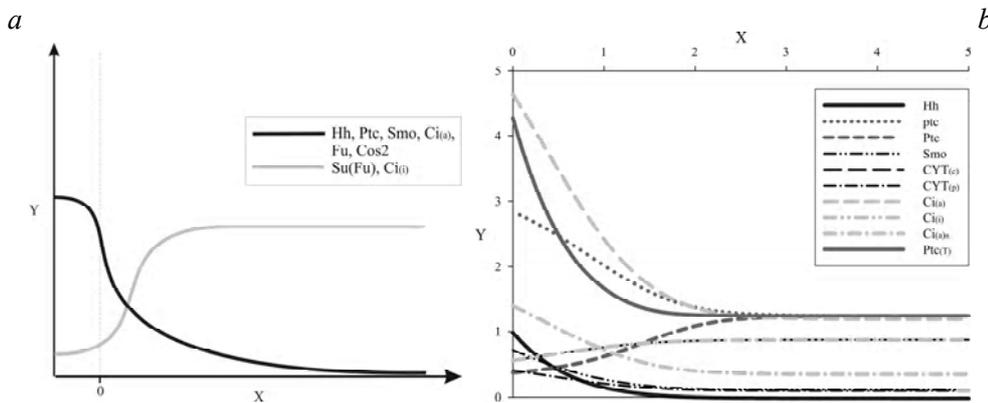


Figure 2. Comparison of experimental and predicted concentration dynamics for the system's components in the anterior compartment of the wing imaginal disc. *a* – qualitative experimental data reflecting the distribution of concentrations of various components of the system being modeled depending on the distance between the anterior-posterior compartment boundary and a location in the anterior compartment (**Hh** – after Tabata and Takei, 2004; **Ptc** and protein complex **Ptc-Hh** – after Torroja *et al.*, 2004; activated **Smo** – after Zhu *et al.*, 2003; **Ci_(a)** and **Ci_(i)** – after Muller and Basler, 2000; activated **Fu**, **Su(Fu)** and **Cos2** – interpretation of qualitative results in work of Zhu *et al.*, 2003). *b* – modeling results. On the *x* axis, the relative distance between the anterior-posterior compartment boundary and a location in the anterior compartment; on the *y* axis, the relative concentration of the system's components. Point 0 on the *x* axis indicates the anterior compartment boundary.

¹³ α is assumed to be equal to 10.0. Empirical formula was used because the intensity of FITC (antibody for Ptc) fluorescence is proportional to concentration of all Ptc forms (Omelyanchuk, Gunbin, 2004).

ACKNOWLEDGEMENTS

This work was supported by grant No. 03-04-48506-a from RFBR “*Studies of Structural and Functional Organization and Evolution of Gene Networks: Computer Analysis and Modeling*”. The authors are grateful to V.V. Filonenko for translating this paper into English.

REFERENCES

- Fadeev S.I., Kogai V.V. (2004) Using parameter continuation based on multiple shooting method for numerical research of nonlinear boundary value problems *Intern. J. of Pure and Appl. Math.*, **14**, 467–498.
- Lai K. *et al.* (2004) The sonic hedgehog signaling system as a bistable genetic switch. *Biophys. J.*, **86**, 2748–2757.
- Lum L., Beachy P.A. (2004) The Hedgehog response network: sensors, switches, and routers. *Science*, **304**, 1755–1759.
- Muller B., Basler K. (2000) The repressor and activator forms of *Cubitus interruptus* control Hedgehog target genes through common generic gli-binding sites. *Development*, **127**, 2999–3007.
- Omelyanchuk L.V., Gunbin K.V. (2004) An elementary module recognizing morphogenetic gradients in tissues. *Proceedings of the Fourth Intern. Conf. on Bioinformatics of Genome Regulation and Structure*, **2**, 117–120.
- Tabata T., Takei Y. (2004) Morphogens, their identification and regulation. *Development*, **131**, 703–712.
- Torroja C. *et al.* (2004) Patched controls the Hedgehog gradient by endocytosis in a dynamin-dependent manner, but this internalization does not play a major role in signal transduction. *Development*, **131**, 2395–2408.
- von Dassow G. *et al.* (2000) The segment polarity network is a robust developmental module. *Nature*, **406**, 188–192.
- Zhu A.J. *et al.* (2003) Altered localization of *Drosophila* Smoothed protein activates Hedgehog signal transduction. *Genes Dev.*, **17**, 1240–1252.

A MODEL FOR SENSING THE HEDGEHOG CONCENTRATION GRADIENT II: A CHECK FOR ADEQUACY

Gunbin K.V.^{*1}, Kogai V.V.², Fadeev S.I.², Omelyanchuk L.V.¹

¹Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ²Sobolev Institute of Mathematics, SB RAS, Novosibirsk, 630090, Russia

* Corresponding author: e-mail: genkvg@bionet.nsc.ru

Key words: pattern formation, positional information, mathematical model, virtual mutagenesis, hypothesis testing

SUMMARY

Motivation: In the present study, our aim was to reveal the molecular machinery responsible for interpretation of the Hedgehog morphogene gradient.

Results: The model for sensing the gradient of the morphogen Hedgehog concentration is checked for adequacy using “virtual” gene mutation (Kogai *et al.*, 2006). A few hypotheses for how the transmembrane proteins Patched, Smoothed and the morphogen Hedgehog can interact have been tested.

INTRODUCTION

The initial stage of modeling any developmental process is the development of gene networks (GNs) and their logical analysis, which should identify which blocks are responsible for the dynamics of the GN functioning. On the basis of the GN blocks thus identified, mathematical models describing the GN functioning are developed. In this work, the mathematical model for sensing the concentration gradient of the morphogen Hedgehog developed on the basis of analysis of the gene network responsible for the development of the anterior-posterior compartment boundary of the wing imaginal disc in *Drosophila melanogaster* is tested for adequacy (Gunbin *et al.* 2004; Omelyanchuk, Gunbin, 2004; Kogai *et al.*, 2006).

MODELS

As Lai *et al.* (2004) put it, “the amount of “freed” Smo can be modeled as the fraction that is “effectively” not interacting with Ptc.” This fraction is modeled by the Scatchard relationship in the original system (Model 0) (Lai *et al.*, 2004; Kogai *et al.*, 2006):

$$\frac{\partial[\text{Smo}]}{\partial t} = \frac{1 + k_4^{-1} \cdot [\text{Hh}]}{1 + k_4^{-1} \cdot [\text{Hh}] + k_5^{-1} \cdot [\text{Ptc}]} \cdot L - k_6 \cdot [\text{Smo}] \cdot$$

Three more hypotheses for how the proteins Ptc, Smo and the morphogen Hh can interact (Casali and Struhl, 2004) have been considered:

1) “unliganded Ptc function catalytically to inhibit Smo, and the presence of Hh-liganded Ptc would titrate the catalytic activity of unliganded Ptc”:

$$\frac{\partial[\text{Smo}]}{\partial t} = \frac{k_{5a} \cdot [\text{Smo}] \cdot [\text{Hh}] \cdot [\text{Ptc}]}{k_{5b} \cdot k_4 \cdot [\text{Ptc}] + (k_4 + [\text{Hh}] \cdot ([\text{Smo}] + k_{5c}))} - k_6 \cdot [\text{Smo}], \quad \text{where } k_{5a}=5.0 \text{ s}^{-1},$$

$k_{5b}=0.001, k_{5c}=0.1 \text{ mol}$ (Model 1),

2) “liganded and unliganded Ptc might function stoichiometrically by competing for access to Smo and exerting opposing effects on its activity”:

$$\frac{\partial[\text{Smo}]}{\partial t} = L \cdot \frac{k_{5b} \cdot [\text{Hh}] \cdot (k_5 \cdot [\text{Hh}] + k_4 \cdot (k_5 + [\text{Ptc}]))}{(k_{5a} \cdot k_4 + [\text{Hh}] \cdot (k_{5a} + [\text{Ptc}])) \cdot k_4} - k_6 \cdot [\text{Smo}], \quad \text{where } k_{5a}=3.0 \text{ mol},$$

$k_{5b}=0.1 \text{ mol}$ (Model 2),

3) “binding of Hh to one monomer in given Ptc multimer might block the ability to inhibit Smo”:

$$\frac{\partial[\text{Ptc}]}{\partial t} = k_3 \cdot [\text{ptc}] - \frac{k_{5a} \cdot [\text{Hh}] \cdot [\text{Ptc}]}{k_4 + [\text{Hh}]} \cdot M - kd_1 \cdot [\text{Ptc}],$$

$$\frac{\partial[\text{Smo}]}{\partial t} = L \cdot \frac{1 + [\text{Hh}] \cdot k_4^{-1}}{1 + [\text{Hh}] \cdot k_4^{-1} + [\text{Ptc}] \cdot k_5^{-1} - (k_{5a} - 1) \cdot [\text{Ptc}] \cdot [\text{Hh}] \cdot k_5^{-1} \cdot k_4^{-1}} - k_6 \cdot [\text{Smo}],$$

where $k_{5a}=1.5$ (Model 3).

RESULTS AND DISCUSSION

In model 0, the concentrations of all the items that constitute the molecular mechanism of the Hh cascade behave exactly as in the experiment (Kogai *et al.*, 2006, *see references herein*). The numerical analysis of the model revealed that the number of active receptors of Ptc should be lowest on the anterior-posterior compartment boundary, which is in agreement with the most recent experimental data (Casali, Struhl, 2004; Torroja *et al.*, 2004).

Testing the other hypotheses for interaction between Ptc, Hh and Smo revealed the following: when tested with the same parameters that give a good description of the dynamics of system 0, system 1, which proposes that Ptc exerts a catalytic effect on Smo and gives it a description in terms of an equation of catalytic inhibition, does not stand up to reality (Fig. 1b); system 2, which proposes a stoichiometrical binding for Ptc and Smo (Fig. 1c); system 3, which proposes the presence of Ptc multimers, exerts regulation on the activator form of the protein Ci much more efficiently than the original system (Fig. 1d).

Thus, the processes in the Hh signaling cascade are best described by the original system and the one which proposes the presence of Ptc multimers.

To verify model 0, the system was brought to virtual mutational analysis: we compared the behavior of the system while its certain components were mutant (switched off) and the behavior components (expression of genes or proteins) in the cells within the clones that were mutant for each particular component (gene) (Fig. 2; Kogai *et al.*, 2006).

For example, experimental observations suggest that when the gene *ptc* is switched-off ($k_0 = 0$), the cells are not responding to Hh signaling, and the activator form of Ci is everywhere in the nuclei, which leads to overexpression of the genes that are regulated through the activator form of Ci (Fig. 2a; Chen and Struhl, 1996). In contrast, when the gene *ptc* is overexpressed ($k_0 = 5.0 \text{ nM/sec}$), cells actively absorb Hh (for convenience, shown on a logarithmic scale) and no overexpression of the genes regulated by the protein Ci is observed (Fig. 2b; Casali, Struhl, 2004; Torroja *et al.*, 2004).

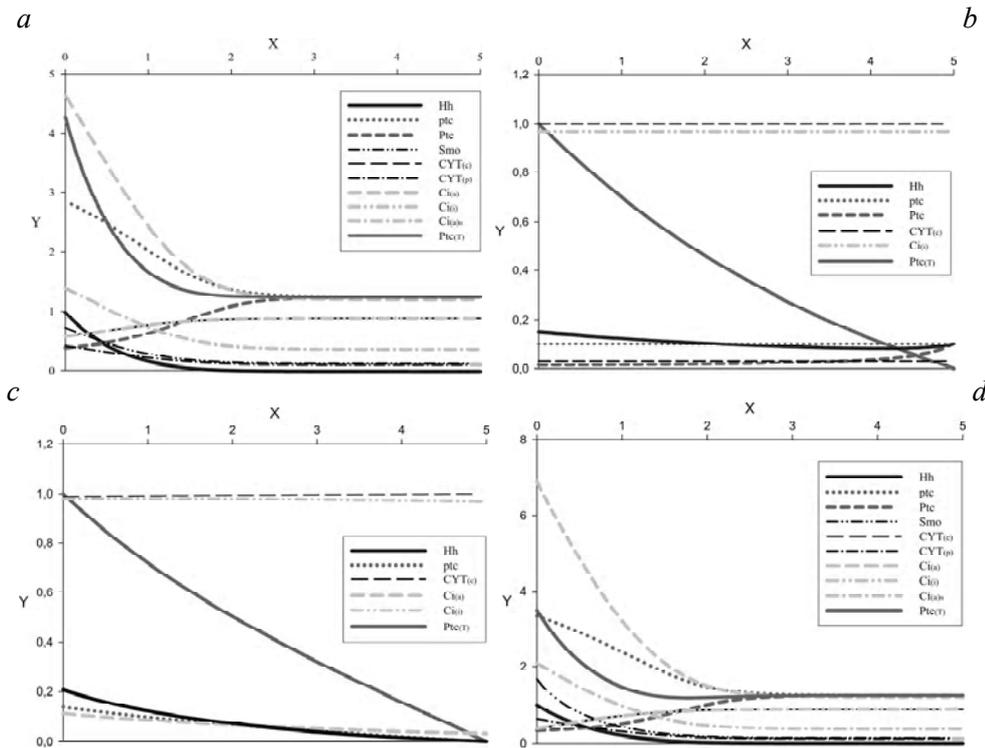


Figure 1. Comparison of various hypotheses describing interactions between the proteins Ptc, Smo and Hh¹⁴. Modeling results using *a* – system 0; *b* – system 1; *c* – system 2 and *d* – system 3. On the *x* axis, the relative distance between the anterior-posterior compartment boundary and a location in the anterior compartment; on the *y* axis, the relative concentration of the system’s components.

When the gene *smo* is mutated, elevated concentrations of the repressor form of Ci in the nucleus are observed both in the experiment and the model ($k_5 \rightarrow 0$) (Fig. 2c; Chen, Struhl, 1996).

When the high-molecular-weight complex CYT is mutated ($J = 0, k_7 = 0$), the activator form of Ci is accumulated only in the cytoplasm, no repressor form is ever arised (Fig. 2d; Zhu *et al.*, 2003). When the protein Ci cannot switch from activator to repressor form ($k_8 = 0$), the activator form of Ci is accumulated both in the cytoplasm and the nucleus, hence a continuous cell response to Hh achieved (Fig. 2e; Chen *et al.*, 1999). When the gene *ci* itself is mutated ($K = 0, k_8 = 0, k_{11} = 0$), no response to Hh signaling is possible, which is exactly what the model shows (Fig. 2f; Dahmann, Basler, 2000).

¹⁴ At the pictures: Hh – Hh protein; ptc – *ptc* gene mRNA; Ptc - Ptc protein, active form; CYT_(c) - high-molecular-weight complex CYT, form (c); CYT_(p) - high-molecular-weight complex, CYT form (p); Ci_(i) – protein Ci inhibitor form; Ci_(a) – protein Ci activator form; Ci_{(a)n} – protein Ci activator form, nuclear fraction; Ptc_(T) - Ptc protein, total forms (Kogai *et al.*, 2006).

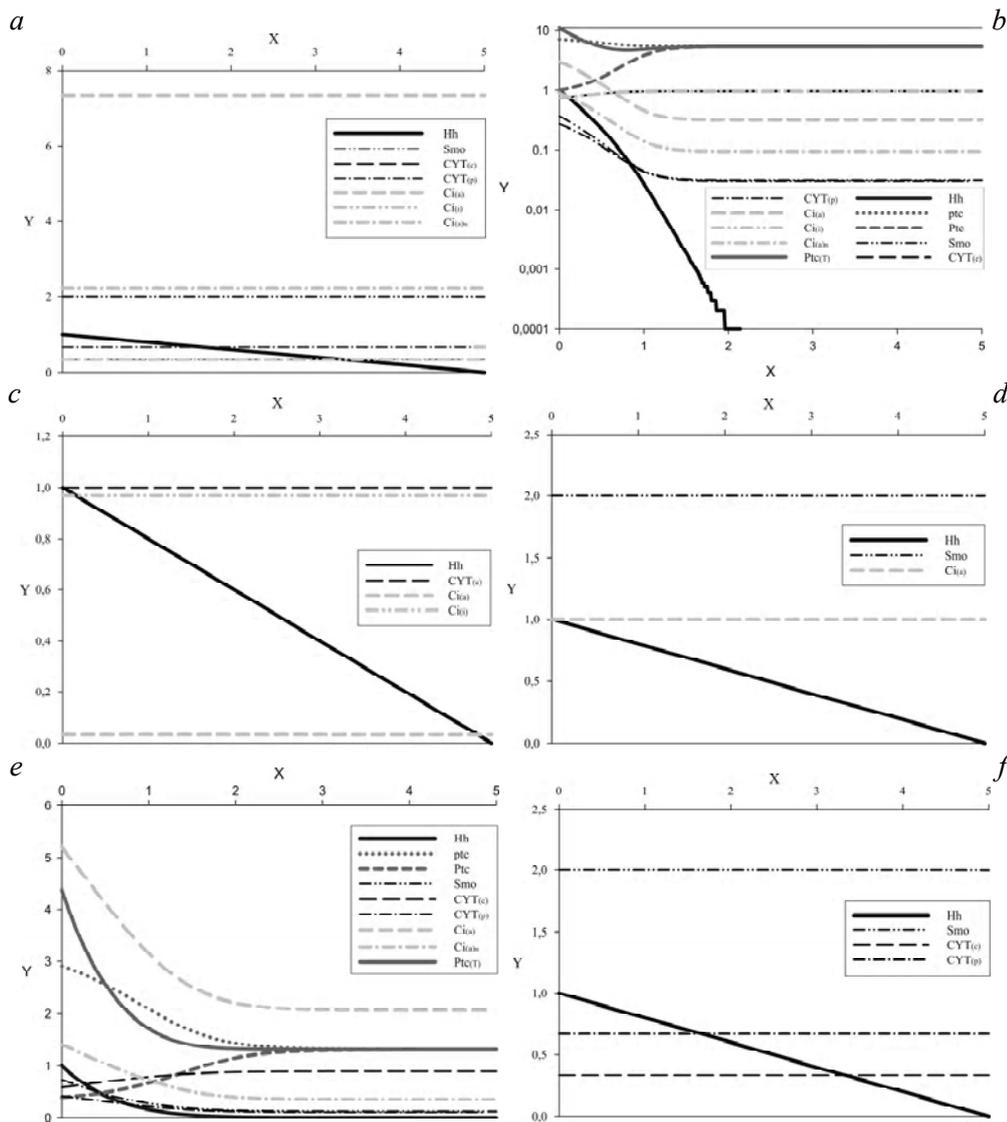


Figure 2. The behavior of the system in the anterior compartment of the wing imaginal disc in response to mutation: *a* – *ptc* switched off; *b* – *ptc* is expressed ectopically; *c* – *smo* switched off; *d* – components missing from the high-molecular-weight protein complex CYT; *e* – the protein Ci does not switch from activator to repressor form; *f* – *ci* switched off.

On the x axis, the relative distance between the anterior-posterior compartment boundary and a location in the anterior compartment; on the y axis, the relative concentration of the system's components.

ACKNOWLEDGEMENTS

This work was supported by grant No. 03-04-48506-a from RFBR “*Studies of Structural and Functional Organization and Evolution of Gene Networks: Computer Analysis and Modeling*”. The authors are grateful to V.V. Filonenko for translating this paper into English.

REFERENCES

- Casali A., Struhl G. (2004) Reading the Hedgehog morphogen gradient by measuring the ratio of bound to unbound Patched protein. *Nature*, **431**, 76–80.
- Chen Y. *et al.* (1999) Mutants of *cubitus interruptus* that are independent of PKA regulation are independent of hedgehog signaling. *Development*, **126**, 3607–3616.
- Chen Y., Struhl G. (1996) Dual roles for patched in sequestering and transducing Hedgehog. *Cell*, **87**, 553–563.
- Dahmann C., Basler K. (2000) Opposing transcriptional outputs of Hedgehog signaling and engrailed control compartmental cell sorting at the *Drosophila* A/P boundary. *Cell*, **100**, 411–422.
- Gunbin K.V. *et al.* (2004) Two gene networks underlying the formation of the anterior-posterior and dorso-ventral wing imaginal disc compartment boundaries in *Drosophila melanogaster*. *Proceedings of the Fourth Intern. Conf. on Bioinformatics of Genome Regulation and Structure BGRS' 2004*, **2**, 56–59.
- Kogai V.V. *et al.* (2006) A model for sensing the morphogen Hedgehog concentration gradient I: a numerical study. *This issue*.
- Lai K. *et al.* (2004) The sonic hedgehog signaling system as a bistable genetic switch. *Biophys. J.*, **86**, 2748–2757.
- Omelyanchuk L.V., Gunbin K.V. (2004) An elementary module recognizing morphogenetic gradients in tissues. *Proceedings of the Fourth Intern. Conf. on Bioinformatics of Genome Regulation and Structure BGRS' 2004*, **2**, 117–120.
- Torroja C. *et al.* (2004) Patched controls the Hedgehog gradient by endocytosis in a dynamin-dependent manner, but this internalization does not play a major role in signal transduction. *Development*, **131**, 2395–2408.
- Zhu A.J. *et al.* (2003) Altered localization of *Drosophila* smoothed protein activates Hedgehog signal transduction. *Genes Dev.*, **17**, 1240–1252.

METHODOLOGY FOR BUILDING OF COMPLEX WORKFLOWS WITH PROSTAK PACKAGE AND ISIMBIOS

*Matveeva A. *, Kozlov K., Samsonova M.*

Department of Computational Biology, Center for Advanced Studies, St. Petersburg State Polytechnical University, St. Petersburg, 195251, Russia

* Corresponding author: e-mail: mdespb@mail.ru

Key words: *Drosophila*, segmentation, image processing, quantitative information, workflow

SUMMARY

Motivation: The development of multicellular organisms involves the differential expression of many genes. Thus knowledge about spatial and temporal patterns of gene expression is crucial for understanding development. Typically researchers analyze spatial patterns of gene expression in multicellular organisms by visual inspection of photographic images obtained in the *in situ* hybridization experiments. A profound limitation of this approach is that it captures little of 2D and 3D information and provides only rough quantitative information. Thus the development of new algorithms and methods for processing of images to accurately quantify the level of gene expression at cellular resolution remains one of the important tasks for bioinformatics.

Results: Here we describe a method to quantify protein and mRNA levels per nucleus in the *Drosophila* blastoderm embryos without loss of spatial information.

Availability: All materials are available from authors upon request.

INTRODUCTION

One of the important problems of bioinformatics is the development of an integrated problem-solving environment that allows users to perform a variety of tasks related to scientific computing, including processing and analysis of digital images of biological objects acquired with light, electron or confocal microscope, data manipulation, and scientific visualization.

Efforts with different rate of success have been made in this direction for years by both corporations and communities. The most interesting examples are VisiQuest from AccuSoft Corporation and SCIRUN. These packages have their own advantages and disadvantages but unfortunately lack important problem specific features and are not easily extendable.

In this work, we present a new software package Prostak (**P**rocessing **S**tacks) integrated in information management system known as iSIMBioS (Integrated Service Infrastructure for Molecular Biology Systems).

Prostak is capable to process 2D and 3D digital images of biological objects. It includes all methods needed to extract quantitative information from experimental data, such as patterns of gene expression in early *Drosophila* embryo. Quantitative information on gene expression is essential to discover dynamical mechanisms underlying gene regulation by mathematical modeling (Jaeger *et al.*, 2004a, b).

We describe the method for building complex workflows to process the digital images of patterns of gene expression. This method is applied to images of segmentation gene expression obtained by confocal scanning microscopy (Kosman *et al.*, 1998). The segmentation genes are involved in *Drosophila* embryogenesis at the syncytial blastoderm stage. We also present preliminary results on development of automated method to control the quality of the whole processing procedure.

MATERIAL AND METHODS

Package description. Prostak implements more than 40 operations that include domain specific methods to correctly orient the embryo images and to extract quantitative information from nuclei and cytoplasm. Domain-independent procedures include rotation and cropping, plain and adaptive thresholding, edge detection, non-linear contrast enhancement, image reconstruction, noise reduction, interpolation, distance and watershed transformations and pixel sub-sampling.

The processing methods are available for a user through command line interface prostak and application program interface parus that can be linked statically or dynamically. All processing methods are exported to the iSIMBioS information management system as workflow modules (WMs) by using wrappers written in Perl programming language. In this system the workflow can be visually constructed by use of the graphical user interface Pegas (Fig. 1) that provides convenient environment for all groups of scientists ranging from the beginners to experts.

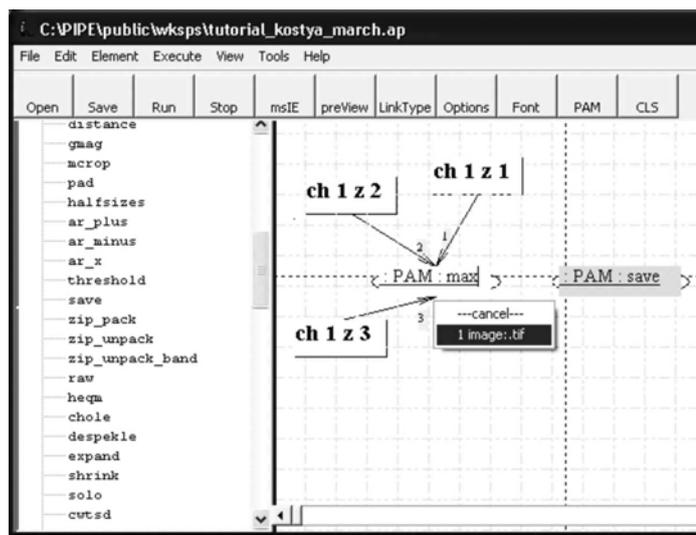


Figure 1. The screenshot of Pegas. The main window consists of two frames. The left one displays the list of available WMs on each server machine. The right bigger frame with two scrollbars shows how to construct the workflow. We illustrate how to select the output port for the WM called as max.

Experimental data. The images were obtained by confocal laser scanning microscopy using fluorescence tagged antibodies. Each embryo was stained for Even-skipped protein, *lacZ* mRNA of the *even-skipped* promoter-reporter construct and histones. For each stain three optical sections were scanned. As a result for each embryo nine 1024×1024 pixels grayscale images in TIFF format without compression were obtained. These experimental data were used to extract quantitative information on expression of the native *even-skipped* gene (by monitoring the level of Even-skipped protein), as well as on expression

of the reporter *lacZ* gene under control of the *even-skipped* promoter. The histone proteins were used to visualize nuclei.

Image segmentation. Previously, the extraction of quantitative information on gene expression was performed with Khoros, now available as VisiQuest from AccuSoft Corp. (Kosman *et al.*, 1998, Janssens *et al.*, 2005).

The implemented method can be subdivided into three major steps, which consist in (1) preprocessing of raw images to put them into standard orientation, (2) determination of the embryo boundaries and (3) extraction of quantitative information. At these steps different image processing operations are applied, e.g. averaging, thresholding, median filtering, histogram equalization, watershed transformation. To extract quantitative information from each nucleus of an embryo a nuclear mask is generated. This mask is a binary image in which all the nuclei are separated from each other. Usually visual inspection is used to estimate the mask quality. However, as the number of nuclei in an embryo is about 2,500, an automated error checking procedure is required.

We propose to check the quality of the nuclear mask by estimating the ratio of the variances in pixel values between classes and inside classes for each nucleus, where one class represents the nucleus itself and another class represents the nuclear surrounding (cytoplasm). For correctly segmented nucleus this measure should be larger than for the incorrectly segmented one and consequently can serve as an accuracy estimator.

RESULTS AND DISCUSSION

We have used Pegas and Prostack to construct three scenarios for the segmentation of images of gene expression patterns.

The first workflow called as *Import And Orientation* serves to combine three optical section obtained in each channel (i.e. for each stain) into one image and to put the resultant images into standard orientation.

Those optical sections, which were obtained by scanning for Even-skipped protein or *even-skipped* promoter-reporter construct are averaged, while pixel by pixel maximum image is calculated from three optical sections obtained for histones. The three resulting images are compared pixel by pixel to obtain the pixel maximum image. The pixel maximum image is thresholded to acquire a binary image in which only those pixels that cover the area occupied by the embryo are “on”. The mask is used to calculate the rotation angle and to remove spurious pixels at the embryo borders. At the next step the embryo images are rotated and cropped (Fig. 2).

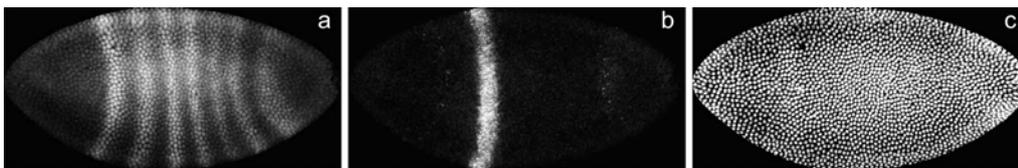


Figure 2. The images of *Drosophila* embryo stained for Even-skipped (Eve) (a), *lacZ* mRNA expressed under control of the *even-skipped* (*eve*) promoter (b), and histone proteins (c) after combining z-sections, rotation and cropping.

The second *Smooth Mask* workflow is constructed to precisely find the area occupied by the embryo in the image. The smooth mask is a binary image in which this area is marked by the foreground pixels (Fig. 3a). Again, three embryo images obtained after rotation and cropping are compared to construct the pixel maxim image and the histogram of this resulting image is equalized. Next several median filters with different structural

elements are applied. To smooth the border of the mask we apply distance transform followed by thresholding, edge detection and filling.

The third workflow called as *Segmentation* serves to extract quantitative information. The nuclear mask is constructed by smoothing the histones image with local histogram and median filters followed by watershed transformation and edge detection. The outlines of nuclei in the histones image overlaid with nuclear mask are presented in Fig. 3*b*. The coordinates of the nuclei centroids expressed as per cent of the embryo length and width are then calculated by averaging the pixel coordinates for each segmented nucleus in the mask. The average fluorescence intensities (relative gene expression levels) for Even-skipped protein and *lacZ* mRNA in segmented nuclei are obtained by averaging the pixel values within each nucleus of each channel.

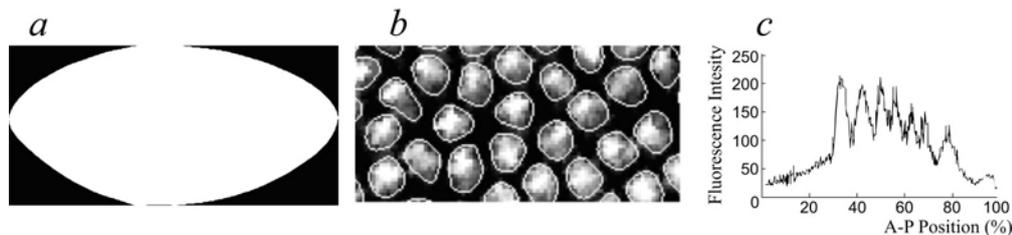


Figure 3. Summary of quantification of gene expression in a triple stained embryo. *Drosophila* blastoderm embryo at late cleavage cycle 14, immunostained for Even-skipped (Eve), *lacZ* mRNA of the *even-skipped* (*eve*) promoter–reporter construct and histones. *a* – smooth whole-embryo mask; *b* – detail of an overlay of the outline of the nuclear mask with the original nuclear channel (histones) image; *c* – quantified gene expression data for Eve. Vertical axis represents fluorescence intensity, horizontal axis, position along the A-P axis (where 0 % is the anterior pole). Only values from nuclei located in the middle 10 % (D-V) of the embryo are plotted.

To estimate the quality of nuclear mask we use nuclear channel (histones) image, watershed image and nuclear mask itself. The estimation is performed by considering two classes of pixels for each nucleus. The pixels of the first class come from the area occupied by this nucleus and outlined in the histones image with nuclear mask, while the second class encompasses all the pixels of the corresponding watershed domain with the exception of pixels that are “off” in the nuclear mask. To estimate the accuracy of segmentation we calculate the ratio of variances in pixel values between and inside these two classes.

To check the validity of the estimator we have compared its maximal values over all nuclei for the nuclear mask obtained with the algorithm proposed here and after one and two additional erosions applied to the mask. These erosions destroy the shape of nuclei. It appeared that the initial value of the estimator (4.46) decreases significantly after the first (2.24) and second erosion (0.76). For the mask obtained with the VisiQuest package the corresponding values of the estimator are 4.00, 3.73 and 1.68 correspondingly.

This result shows, that the estimator proposed here can be used to evaluate the mask quality, however further experiments are necessary to develop a robust method for estimation of the segmentation quality.

ACKNOWLEDGEMENTS

The support of this work by NIH Grant RR07801, NWO-RFFI No. 047.011.2004.013 grant and Federal Contract No. 02.467.11.1005 is gratefully acknowledged.

REFERENCES

- Jaeger J., Blagov M., Kosman D., Kozlov K.N., Manu, Myasnikova E., Surkova S., Vanario-Alonso C.E., Samsonova M., Sharp D.H., Reinitz J. (2004a) Dynamical analysis of regulatory interactions in the gap gene system of *Drosophila melanogaster*. *Genetics*, **167**, 1721–1737.
- Jaeger J., Surkova S., Blagov M., Janssens H., Kosman D., Kozlov K.N., Manu, Myasnikova E., Vanario-Alonso C.E., Samsonova M., Sharp D.H., Reinitz J. (2004b) Dynamic control of positional information in the early *Drosophila* embryo. *Nature*, **430**, 368–371.
- Janssens H., Kosman D., Vanario-Alonso C.E., Jaeger J., Samsonova M., Reinitz J. (2005) A high-throughput method for quantifying gene expression data from early *Drosophila* embryos. *Development Genes and Evolution*, **215**, 374–381.
- Kosman D., Small S., Reinitz J. (1998b) Automated assay of gene expression at cellular resolution. *Development Genes and Evolution*, **208**, 290–294.

ONTOLOGY OF ARABIDOPSIS GENE NET SUPPLEMENTARY DATABASE (AGNS), CROSS DATABASE REFERENCES TO TAIR ONTOLOGY

Mironova V.V.^{*1}, *Poplavsky A.S.*¹, *Ponomaryov D.K.*², *Omelianchuk N.A.*¹

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ² Institute of Informatics System, SB RAS, Novosibirsk, 630090, Russia

* Corresponding author: e-mail: kviki@bionet.nsc.ru

Key words: ontology, plant development, controlled vocabulary, Arabidopsis

SUMMARY

Motivation: Inter-database queries on the same subject domain require a coordinated vocabulary of terms. The existing plant ontology (PO) has been made mainly for description of large-scale experiments and contains gaps, which should be filled to provide terms for description of gene expression patterns revealed by *in situ* hybridization and phenotype abnormalities in mutant and transgenic plants.

Results: We have created an Arabidopsis PO based on controlled vocabularies of AGNS (Arabidopsis Gene Net Supplementary) database, which contains a detailed information on anatomical elements and development stages from annotated papers on gene expression and phenotype abnormalities in mutant and transgenic plants. The comparison of AGNS ontology with PO allowed us to develop a common ontology providing both inter-database queries based on general standards and an opportunity to describe the whole scale of transcriptomic and phenomic data in the course of Arabidopsis development.

Availability: <http://www.mgs2.bionet.nsc.ru/agns>

INTRODUCTION

To study development in Arabidopsis one needs to have an access to various and scattered information in publications and to different databases storing results of large-scale experiments. Description of this growing body of data using a common format would considerably facilitate analysis of the information and support inter-database queries. The Plant Ontology Consortium (POC) founded in 2002, has developed ontology for describing anatomical elements and developmental stages in Arabidopsis. This ontology consists of two controlled vocabularies: a controlled vocabulary describing morphological and anatomical structures representing organ, tissue and cell types and a controlled vocabulary on growth and developmental stages (Jaiswal *et al.*, 2005). The main purpose of these vocabularies is to facilitate the cross-database querying and to foster consistent use of these vocabularies in annotation of tissue and/or growth stage specific expression of genes, proteins and phenotypes. Relationships between terms of such PO within controlled vocabularies and between them are represented as Directed Acyclic Graphs (DAGs.) A DAG is similar to a hierarchical structure but is superior because terms (representing concepts) within a DAG structure have the ability to have one or more parents. For example, in the dictionary of anatomy and morphology of plants, the following initial

classes of terms are defined: “cell”, “tissue”, “sporophyte”, etc. In this case, the term “root epidermis” corresponds to both parental classes: “tissue” and “sporophyte”. There are three types of relationships between PO classes: “is a”, “part of” and “develops from”. At present, the PO of Arabidopsis contains information on structure and stages of germplasm development and information on organs and stages of Arabidopsis development, which is sufficient for description of microarray experiments and gene annotation.

Our aims comprise accumulating data on gene expression from various experiments and information on phenotypic anomalies of Arabidopsis. These tasks are addressed in the AGNS database (Arabidopsis GeneNet Supplementary DataBase) (Omelyanchuk *et al.*, 2005). Controlled vocabularies of AGNS contain detailed information on stages of development of separate organs and the whole plant, anatomy and morphology of the plant, as detailed as it could be found in numerous publications on Arabidopsis development. We have constructed the Arabidopsis PO on the basis of these controlled vocabularies. Besides the function of data systematization, this PO can be used for studying the plant development. We compare our data in PO AGNS with PO TAIR and discuss the results of this work and application of PO AGNS in this paper.

METHODS AND ALGORITHMS

There are two ways of representation of PO Arabidopsis in the AGNS: the navigation system of AGNS and the Protégé ontology editor 3.1.

In the AGNS database one can see informational content of controlled vocabularies on separate pages, which are marked by corresponding tabs: Development (the vocabulary for stages of organs development) and Morphology (the vocabulary of organs morphology). On the same pages a search in vocabularies for detailed description of the particular stage of development and morphology of a selected organ is provided (Fig. 1a).

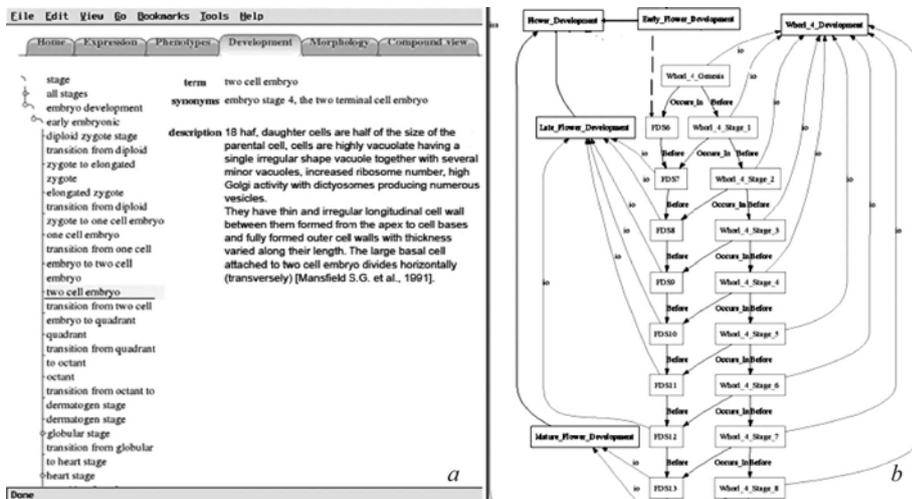


Figure 1. Representation of PO Arabidopsis in navigation system of AGNS by the example of hierarchy and description of early embryonic stages in the controlled vocabulary on development stages (a). Representation of PO Arabidopsis in Protégé system by the example of Whorl-4 development during flowering (b).

In the Protégé system we have constructed an alternative version of PO using the OWL language (Ponomaryov *et al.*, 2006). The proposed ontology (see Fig. 1b) describes plant anatomy in development and is filled in with data from Arabidopsis development studies that have been summarized by development of the AGNS controlled vocabulary.

IMPLEMENTATION AND RESULTS

The PO of the AGNS database consists of two hierarchically structured controlled vocabularies for stages of development and for anatomy and morphology of Arabidopsis. These vocabularies have been constructed around the contents of annotated data based on both the data provided by the authors of the annotated papers, and from specialized publications on plant developmental stages and morphology of the wild type. The most frequently used names of the stages and organs are highlighted and their synonyms are given, also description of stage and organ includes quantitative data and accompanied by detailed comments. The vocabularies are being supplemented with new research data as they become available. During construction of the controlled vocabularies we were consistent with the hierarchy requirement: organs and stages of development are organized with enclosure from the smaller to the greater, where it is possible. A characteristic feature of PO AGNS is that the controlled vocabulary for anatomy and morphology is presented in two parts: (1) as a structured terminological vocabulary like in TAIR, (2) as a complete vocabulary for descriptions of anatomy and morphology of Arabidopsis. In the terminological vocabulary anatomical elements are presented in sections “organs”, “tissues”, and “domains” with a brief description. This vocabulary is used in AGNS input system, and we have carried out its comparison and linking with PO TAIR. The complete controlled vocabulary is filled automatically with the entered combinations from sections of AGNS terminological vocabulary. The description of an anatomical element in this dictionary can be generalized as “organ, organ, ..., domain, ..., domain, tissue”, where description of organs is represented in the hierarchical form, and the name parts “domain” and “tissue” can be skipped or reiterated in various combinations. In the complete controlled vocabulary the information on description of an anatomical element, quantitative characteristics and other useful data with reference to publication are entered. This vocabulary is used for navigation within the database and for further analysis of the data. It may be also useful as a reference book for educational and research purposes.

In the controlled vocabulary for development stages we did not separate “plant growth stages” or “body part developmental stages” in PO AGNS. Both types are joined and “plant growth stages” are higher in the hierarchy. Such structure divides whole ontogenesis into four stages: an embryogenesis, a seedling, a transition to flowering, and a reproductive phase. A term “vegetative growth” is used for description of development of a vegetative part during the last three stages.

We have carried out a comparison of terms in the controlled vocabularies of PO Arabidopsis AGNS with terms from PO Arabidopsis TAIR. Terms concerning the description of the same plant structures and identical stages of development in PO AGNS have been replaced by corresponding terms from PO TAIR. New terms that have not described in AGNS have been added.

In the controlled vocabulary for stages of development about 50 % of the terms have been compared. For the rest of the terms in PO TAIR no analogues have been found. These are the terms for development stages of some organs, such as embryo, leaf primordium, midvein, endosperm, and others. During large-scale experiments such details are usually omitted, but they are necessary for studying gene expression using in situ hybridization and for description of phenotypic anomalies.

The comparison of controlled vocabularies for anatomy and morphology has been carried out for the terminological dictionary at first. We have found out a lot of differences in terms of POs. This has occurred because there is no detailed description of anatomical structure for many organs in TAIR, which are selected by researchers during analysis of plant development. For example, there is no description of domains in the embryo, no description of some organs of flowers and leaves, etc. On the other hand, in TAIR PO description for germplasm, tissue culture and root is much better developed.

The comparison and in some cases replacement of terms has been performed for AGNS PO to be according to international nomenclature. Therefore, AGNS and TAIR vocabularies are partly overlapping in common terms, and comprehensively describe different parts of the plant organism. Their integration may become an important breakthrough in ontology development.

DISCUSSION

Earlier the construction of PO has been an additional task during description of large-scale experiments on gene expression. This approach has resulted in a too general character of such information. Recently a tendency towards description and systematization of other data type has appeared. These types of data include phenotypic abnormalities of plants in mutants, insertion and transgene lines, patterns of gene expression in normal and mutant types, obtained not only by large-scale experiments. A high accuracy of determining the plant stage and organ is typical for these data (up to number of cells expressing the selected gene). Systematization of such data requires to construct a more comprehensive PO, which would allow to describe not only general data, but all necessary details as well. The proposed PO *Arabidopsis* in AGNS meets these demands: it is used for description of gene expression obtained in various experiments, from large-scale down to *in situ* hybridizations, and for description of phenotypic anomalies in mutant and transgenic lines. Furthermore, the Protégé-version of the AGNS ontology can be a useful tool for studying the development, since it will allow to analyze phenotypic abnormalities and to reveal the causes and time of their appearances. We believe that the developed controlled vocabularies and ontology will be a useful resource for the entire plant science community.

ACKNOWLEDGEMENTS

This work was supported in part by INNOVATION PROJECT of Federal Agency of Science and innovation IT-CP.5/001 “Development of software for computer modeling and design in postgenomic system biology (system biology *in silico*)”, Russian Foundation for Basic Research (grants Nos 05-07-98012 and 03-04-48506), Russian Academy of Sciences (grant No. 10.4), Siberian Branch of Russian Academy of Sciences (Integration Project No. 119) and the US National Science Foundation (FIBR EF-0330786 Development Modeling and Bioinformatics).

REFERENCES

- Jaiswal P. *et al.* (2005) Plant Ontology (PO): a controlled vocabulary of plant structures and growth stages. *Comparative and Functional Genomics.*, **6**, 388–397.
- Omelyanchuk N. *et al.* (2005) AGNS-A database on expression of *arabidopsis* genes. In Kolchanov N.A., Hofestadt (eds), *Bioinformatics of Genome Regulation and Structure II*. Springer Science+Business Media, Inc. 433–442.
- Ponomyov D.K. *et al.* (2006) Semantically Rich Ontology of Anatomical Structure and Development for *Arabidopsis thaliana*. *This issue*.

A ONE-DIMENSIONAL MODEL FOR THE REGULATION OF THE SIZE OF THE RENEWABLE ZONE IN BIOLOGICAL TISSUE

Nikolaev S.V.^{*1}, *Fadeev S.I.*², *Kogai V.V.*², *Mjolsness E.*³, *Kolchanov N.A.*¹

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ² Institute of Mathematics, SB RAS, Novosibirsk, 630090, Russia; ³ School of Information and Computer Science and Institute for Genomics and Bioinformatics, University of California, Irvine, CA, USA

* Corresponding author: e-mail: nikolaev@bionet.nsc.ru

Key words: shoot apical meristem, SAM, tissue, morphogenesis, mathematical model

SUMMARY

Motivation: Spatial structure of biological tissues develops during growth and morphogenesis of an organism. In certain cases such a special structure remains invariant during the all organism's life despite constant renewal of the cells in this tissue. The question of interest is: what scheme underlies this phenomenon?

Results: A simple mathematical model for the regulation of shoot apical meristem compartments sizes was developed. A mathematical analysis had revealed numerous of solutions, some of which were biologically plausible and interesting.

INTRODUCTION

Differentiated cells make up the bulk of the cells of the adult organism. Each differentiated cell does not divide, as a rule, being specialized to perform a particular function in a particular organ tissue. However, in the adult organism, there exist cells that are undifferentiated, although perhaps "predetermined". This means that their fate is, to a certain extent, predestined in the sense that they can become cells of a particular type or of a restricted set of types. There are stem cells that continue to divide at a definite rate. The stem cells are important for the life of the adult. In the animal tissue, they serve as sources of continuously renewable tissue (the skin, for example) and certain plants, the stem cells of the shoot meristem (the growing tip of the shoot) provides the growth of the plant through its entire life.

In this work, a "cell-oriented" model for the structural-functional organization of the renewable zone – the shoot apical meristem (SAM) is suggested. The model is cell-oriented because its purpose is to describe the observed cell behavior (in the given case these are the types of cells and the switching over from one cell type to another) within the framework of the minimal model (Merks *et al.*, 2005). The minimal model is not intended to describe real molecular-genetic systems that control cell behavior.

The group at the growing tip of the shoot, referred to as the shoot apical meristem (SAM), is of importance. The SAM contains stem cells that continuously divide, ultimately giving rise to all the cells of the plant. Although the cells of the SAM are undifferentiated, they are determined with respect to the expression of certain genes, and, on this basis, the SAM is divided into compartments that are specifically positioned relative to each other in space through the entire life of the plant. The cells that are located within the vertical axis of the meristem in the radius of 2–4 cells at the uppermost layers 3–4 (see

Fig. 1) express, i.e. switch on-off the corresponding genes and, as a result, synthesize a protein called CLV3, belong to the central zone (CZ). Cells that express the WUS gene are located at the lower layer of the CZ cells. These cells are referred to the organizational center (OC), that is about 2–3 cells thick in the vertical direction. It is thought that the constancy of the SAM structure is required for the maintenance of the pool of the stem cells (Groß-Hardt *et al.*, 2003). The mechanisms that provide the constancy of the SAM structure are the subjects of both applied and basic in-depth studies.

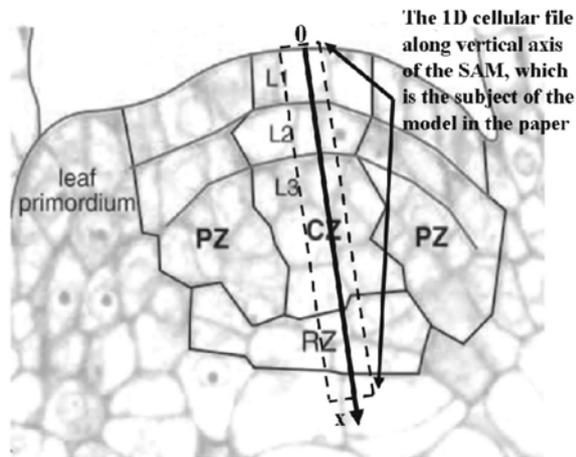


Figure 1. Cross-section of the apical meristem shoot in *Arabidopsis thaliana*.

Fig. 1 presents a cross-section of the shoot apical meristem of *Arabidopsis thaliana*. (adapted from Groß-Hardt *et al.*, 2003). The external layer is denoted by L1, the second by L2; L3 is arbitrarily called the third layer because it actually results from the cells dividing in all the planes, it is no longer a layer, rather an accumulation (a collection) of cells; CZ is the control zone, PZ is the peripheral zone, RZ is the rib-zone where cells start to differentiate into the cells of the vascular system. The X axis is pointed downwards from the shoot apex. Cells along the axis are considered as a one-dimensional array in the proposed model.

MODEL

There may be, in principle, two mechanisms that maintain the vertical compartmentalization of the meristem: first, symmetric division of the cells at the compartment boundaries with their determination in the morphogen fields and the second, asymmetric division of the cells at the boundaries. In fact, division proceeds in all the planes in the L3 layer. This makes more likely the mechanism of the vertical structure maintenance. Furthermore, in mutants whose division orientation pattern is impaired at the early stages, seedlings with a normal structural framework are formed (Berleth, Chatfield, 2002). For this reason, a possible mechanism for cell determination controlled by positional information will be considered.

Fields of the concentrations of substances that spread over from different sources (for example, by diffusion) are the physical carriers of positional information. Fig. 2 shows a possible mechanism zonal structure formation in the 1D domain. The distance is plotted along the axis X in arbitrary units which corresponds to the vertical axis that passes through the center of the apical shoot. The concentration of the morphogen Y that spreads from the apical shoot (from the point O) is plotted along the axis Y. As a result of diffusion of Y and of its continuous destruction (decay), a steady-state distribution (a

decreasing function from x) is established. At concentration above the thresholds, Y may activate gene expression in the C and W substances. It should be noted that the activation threshold for the C is higher than that for the W gene. The assumption is made that the C substance is a repressor of the expression of the W gene. It follows that where the C gene is expressed, the W gene is repressed, and the W gene is actually expressed in the zone that is remote from the shoot apex (the axis origin).

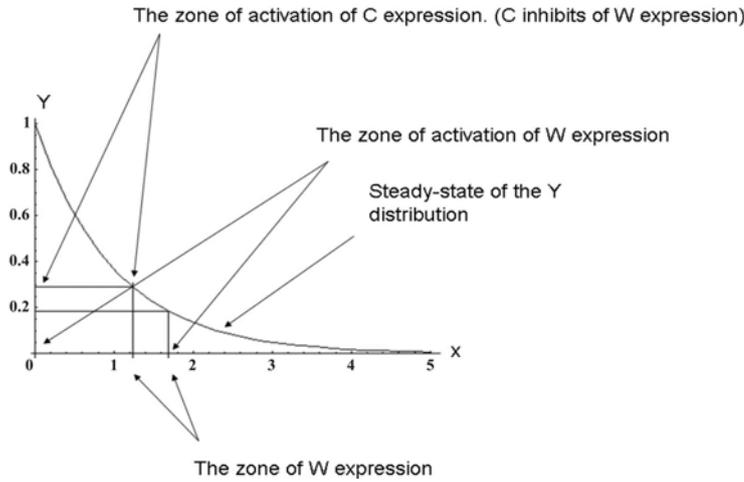


Figure 2. The simulated mechanism. The Y is distributed depending on the distance from the apical meristem and its threshold values at which the C and W expression is activated (enhanced).

$$\frac{dY(1)}{dt} = \frac{1}{\tau_Y} g(h_Y + T_{YW}W(1)) - d_Y Y(1) + D_Y (Y(2) - Y(1))$$

$$\frac{dY(i)}{dt} = -d_Y Y(i) + D_Y (Y(i-1) - 2 \cdot Y(i) + Y(i+1)), \quad 1 < i < n-1$$

$$\frac{dY(n)}{dt} = -d_Y Y(n) + D_Y (Y(n-1) - Y(n))$$

$$\frac{dC(i)}{dt} = \frac{1}{\tau_C} g(h_C + T_{CY}Y(i)) - d_C C(i), \quad 1 \leq i \leq n$$

$$\frac{dW(i)}{dt} = \frac{1}{\tau_W} g(h_W + T_{WY}Y(i) + T_{WC}C(i)) - d_W W(i) + D_W (W(i-1) - 2 \cdot W(i) + W(i+1)),$$

$$1 < i < n-1$$

$$\frac{dW(1)}{dt} = -d_W W(1) + D_W (W(2) - W(1))$$

$$\frac{dW(n)}{dt} = -d_W W(n) + D_W (W(n-1) - W(n)),$$

where D_S are the coefficients of S -substance diffusion, d_S are the coefficients of S -substance degradation, $\frac{1}{\tau_S}$ are the maximal rates of S substance expression. The sigmoid

function is in the form: $g(x) = \frac{1}{2} (1 + \frac{x}{\sqrt{1+x^2}})$. The arguments x of the sigmoid

functions are the linear combinations of h_S and $T_{SR} \cdot R$, where h_S – the threshold of regulation of S substance expression, and T_{SR} reflect influence of regulator R on S substance expression.

RESULTS AND DISCUSSION

The mathematical model was analyzed and its results are the subject of the paper (Nikolaev *et al.*, 2006). This simple model has numerous solutions.

An example of the model steady state solution for $Y(n)$, $W(n)$, and “synthesis” of $W(n)$ is given in Fig. 3. It was shown that the proposed mechanism can stabilize the position of the OC relative to the upper point of the meristem in the vertical direction while the resident cells of the meristem compartments are substituted by the other cells. With some values of model parameters the OC positioning can periodically displace in range of 1–2 cells about a “point of attraction” (Fig. 4), which is consistent with some experimental observations.

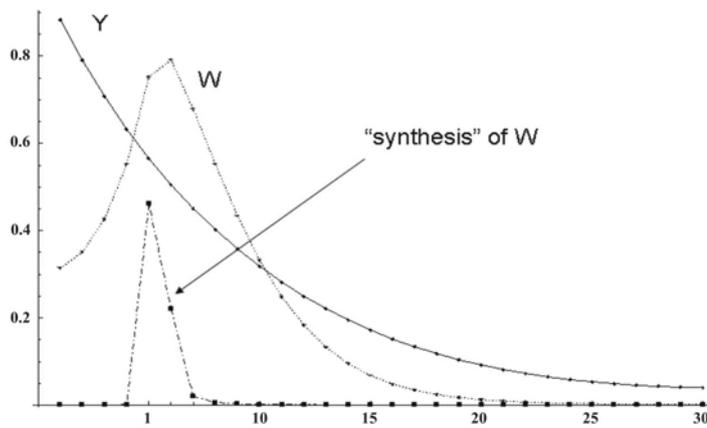


Figure 3. A steady state solution for $Y(n)$, $W(n)$, and “synthesis” of W .

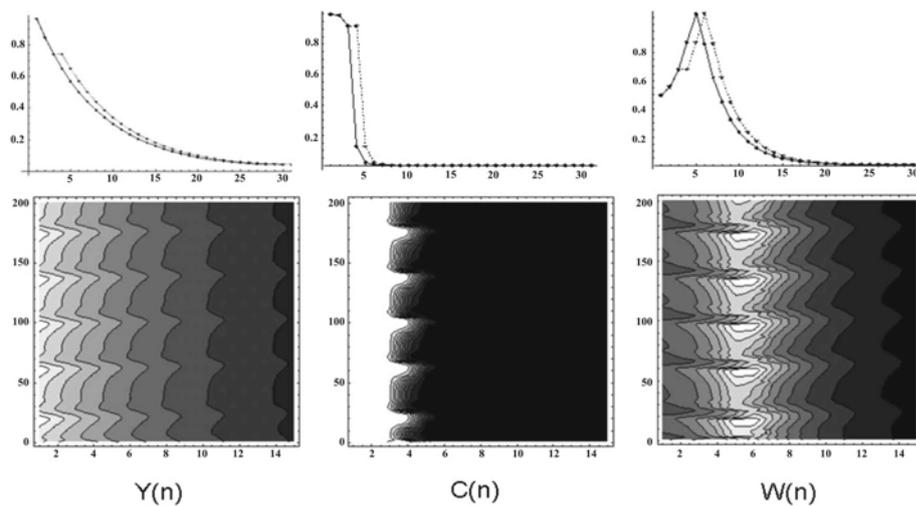


Figure 4. An example with 4-th cell division (upper). This perturbation of a stable solution can induce a periodical solution.

ACKNOWLEDGEMENTS

This work was supported by the NSF:FIBR EF-0330786 grant, Innovation project of Federal Agency of Science and innovation IT-CP.5/001 “Development of software for computer modeling and design in postgenomic system biology (system biology *in silico*)”.

REFERENCES

- Berleth Th., Chatfield S. (2002) Embryogenesis: Pattern Formation from a Single Cell. *The Arabidopsis Book. American Society of Plant Biologists*. (<http://www.bioone.org/pdfserv/i1543-8120-007-01-0001.pdf>).
- Groß-Hardt R., Laux Th. (2003) Stem cell regulation in the shoot meristem. *J. of Cell Sci.*, **116**, 1659–1666.
- Merks R.M.H., Glazier J.A. (2005) A cell-centered approach to developmental biology. *Physica A.*, **352**, 113–130.
- Nikolaev S.V., Kolchanov N.A., Fadeev S.I., Kogay V.V., Mjolsness E. (2006) Analysis of a one-dimensional model for the regulation of the size of the renewable zone in biological tissue. *Computational Technologies*, **11**, 65–79 (in Russ.).

A SYSTEM FOR SIMULATION OF 2D PLANT TISSUE GROWTH AND DEVELOPMENT

*Nikolaev S.V.^{*1}, Penenko A.V.², Belavskaya V.V.¹, Mjolsness E.³, Kolchanov N.A.¹*

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ² Institute of Mathematics, SB RAS, Novosibirsk, , 630090, Russia; ³ School of Information and Computer Science, and Institute for Genomics and Bioinformatics, University of California, Irvine, CA, USA

* Corresponding author: e-mail: nikolaev@bionet.nsc.ru

Key words: biomechanics, osmotic pressure, water potential, viscoelastic deformation, cell wall, cell ensemble, tissue, morphogenesis, mathematical model

SUMMARY

Motivation: At the tissue level growth and development are determined by cellular biomechanics, molecule transportation, and cellular responses to their microenvironments. Biologically motivated models of such processes, when incorporated into models of tissue growth and development, would help to study the roles of these processes in tissue growth and morphogenesis.

Results: A program system for simulation of “planar” plant tissue growth and development was developed. The system includes graphical user interface to prepare an initial configuration of system to be modeled and to perform simulations.

INTRODUCTION

Morphogenesis on the scales from cells - tissues - organs and up to the whole organism depends on differential cell growth and division. It is shown that deformation tensions that arise in different parts of growing biological tissue have dramatic effect on the all biological processes up to differential gene expression (Nelson et al., 2005). The gene expression controls the cell functions. Thus biomechanics and biological functions form a regulatory loop. While development processes including morphogenesis in animals involve cell movement, it dose not occurs in plants, so the plant tissue development is, in a sense, easier to model and simulate. The growth of plant cells and tissues is shown to be dependent on water potential in the cells. The growing plant cells are enveloped with primary cell wall, which can yield under intracellular pressure (Cosgrove, 1986).

We used models of the main biophysical processes in cells and tissue, which are listed above, to construct a model of plant tissue growth and a program system for its simulation.

METHODS AND ALGORITHMS

Biophysical considerations. (Nobel, 1973) Water flow between compartments with semipermeable boundaries is proportional to the difference between water potentials in the compartments:

$$\frac{dv_k}{dt} = \sum_{i(k)} (\Psi_{i(k)} - \Psi_k) \cdot L_{k,i(k)}, \quad (1)$$

where $L_{k,i(k)}$ is hydrodynamical conductance; and $i(k)$ are numbers of the cells in the neighborhood of the i -th cell.

Water potential is dependent on osmotic (π_k) and hydrostatic (turgor) (P_k) pressures:

$$\Psi_k = P_k - \pi_k \quad (2)$$

In turn, osmotic pressure is dependent on content of osmotically active particles s_k in a compartment, the pressure for diluted solutions is:

$$\pi_k = \frac{s_k \cdot R \cdot T}{V_k}, \quad (3)$$

where V_k, R, T are the compartment volume, the absolute gas constant, and the Kelvin temperature respectively.

Applied pressure and volume for condensed matter obey to the state equation (Prigogine, Defay, 1966):

$$v(T, p) = v(T, 0) \cdot (1 - \kappa \cdot p) \quad (4)$$

where κ is thermodynamical compressibility, the typical value for liquids is $\sim 10^{-4} \text{ Bar}^{-1}$; $v(T, 0), v(T, p)$ are molar volumes for reference (0) and an excess (p) pressures respectively.

The equation (4) is used to calculate intracellular hydrostatic pressures resulting from water flows (1). The pressures enforce deformation of the tissue, which consists of the cells.

There are cell wall deformations at the base of the deformation of the tissue. In a certain range of intracellular pressure, the cell wall deformation is elastic and above a threshold pressure this deformation is plastic –the cell wall behaves like viscous liquid.

Elastic deformation obeys the Hooke's law: $T = \frac{E \cdot S \cdot \Delta l}{l}$. Viscous behavior is described

as $\frac{dl}{dt} = \eta \cdot F$; appearing viscosity η reflects cellular capacity to remodel its walls, and F is the applied force.

The cell deformations arise as result of cell wall vertex movements caused by applied forces, as illustrated in Fig. 1 in the elastic case.

Features of calculation. The processes that are described in terms of different nature are involved in the model. For example, a process changes the system's topology (mainly, the cell divisions). Other processes are responsible for evolution of distributed system parameters such as the concentration of chemical substances (biomass growth, differential gene expression, and transport processes). It is convenient to use systems of ordinary differential equations to describe one kind of processes. The other processes are described as operations with algebraic structures like graphs. Operator formalism is used to unify these different approaches.

Each cell is represented as a polyhedron with straight elastic walls. Walls of neighbor cells are tightly bound to each other. Each wall separating two neighbor cells consists of two straight elastic rods (one for each cell) (Fig. 2).

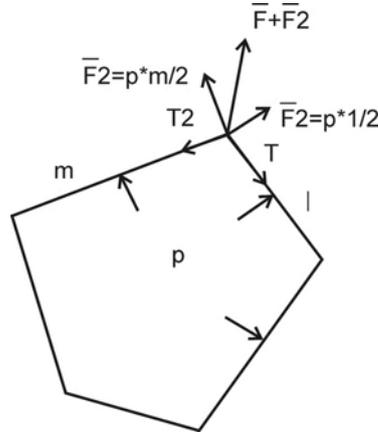


Figure 1. Vector diagram for forces on a cell wall vertex.

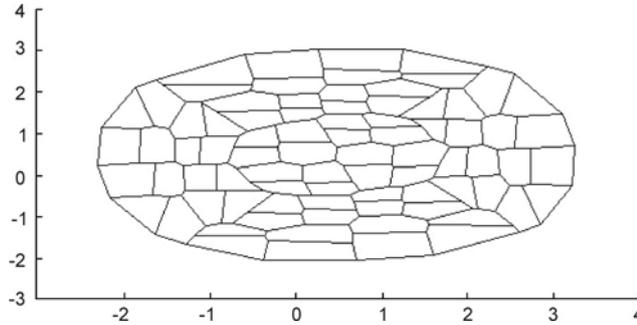


Figure 2. Spatial structure of a toy tissue.

The concentration of chemical substances can be different in different cells but it is considered to be uniform within a single cell. Some chemical substances can freely penetrate through the cell walls. The transport of the other substances is restricted.

Data structure of the model consists of the following groups of parameters: (a) global model parameters; (b) individual cell parameters such as topology, cell wall properties, and chemical cell contents; (c) vertex properties.

Let us demonstrate an example of an operator's inner structure. This operator describes the processes of mechanical deformation and turgor dynamics.

$$\bar{x}_i = \sum_{\substack{\text{rod } \bar{l} \\ \text{incident} \\ \text{to vertex } i}} k_{elastic} \cdot (|\bar{l}| - |l_n|) \cdot \frac{\bar{l}}{|\bar{l}|} \cdot \text{sgn}(\bar{l}) + \sum_{\substack{\text{rod } \bar{l} \text{ from cell} \\ \text{incident} \\ \text{to vertex } i}} \frac{1}{2} \cdot \frac{1}{k} \left(1 - \frac{1}{\nu p_0} \cdot \frac{V_{cell}}{\mu_{cell}}\right) \cdot |\bar{l}| \cdot h \cdot \bar{n} \quad (5a)$$

$$\dot{\mu}_k = \sum_{\substack{\text{nigbor } i \\ \text{of cell } k}} \left[\frac{h \cdot L_w}{k \cdot \nu p_0} \cdot l(i, k) \cdot \left[\frac{V_k}{\mu_k} - \frac{V_i}{\mu_i} \right] \right] + \sum_{\substack{\text{nigbor } i \\ \text{of cell } k}} \left[h \cdot L_w \cdot RT \cdot l(i, k) \cdot \left[\frac{\mu S_k}{V_k} - \frac{\mu S_i}{V_i} \right] \right] \quad (5b)$$

Here \bar{x}_i are the coordinates of the vertex i , μ_k is molar water content in cell k , V_k is the volume of the cell k and the other variables are model parameters.

Another group of operators implements external interaction with the model. These are so-called dialog operators. They are implemented as User Interface.

IMPLEMENTATION AND RESULTS

The graphical user interface is an organizing component. It is designed to provide convenient interaction with the program. Its purpose is to join the stages of creation and development of cellular ensemble, and also to build and execute various scripts of cellular ensemble dynamics.

Interface components implements the following functions:

1. Creation of cellular ensemble geometry (Fig. 3), i.e. assignment of the initial parameters of ensemble formation in dialog with the user. The following parameters are defined: the number of the cells in ensemble, the shape of the ensemble, and also the cell geometry in the selected ensemble, that depends on topology of intercellular connections, definition of cell types, boundary cells, etc.

2. Marking of cellular ensemble, i.e. definition the fields of non-geometrical cell parameters and assignment values for them. The next parameters are definable: pressure inside a cell, levels of morphogens, osmotically active molecules, etc. These parameters determine model of intracellular dynamics.

3. The important function of interface component – Editing of the cellular ensemble (Fig. 5) during simulation. This function allows the user to change the parameters of the cells: intracellular pressure, levels of morphogens and other molecules concentrations, lengths of relaxed cell walls, vectors of growth and division.

4. The next module implements the interface with the process of simulation. (Fig. 4). It allows the user to execute scripts, specifying cellular ensemble dynamics using of various operators, and to track the results after each simulation step.

5. And the last function is the graphical representation of results of the current simulation: visualization of cellular ensemble parameters in dynamics, and static images of parameter fields in cellular ensemble. Useful options are supplied in the block, for example, representation of a chosen cell parameter in pseudo-colors (intracellular amounts of molecule of interest, cell wall tensions, thickness, permeability, etc.).

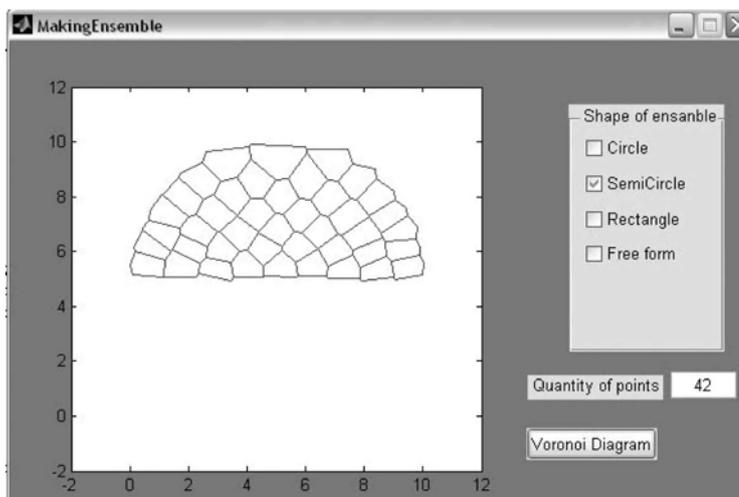


Figure 3. Initial geometry specification.

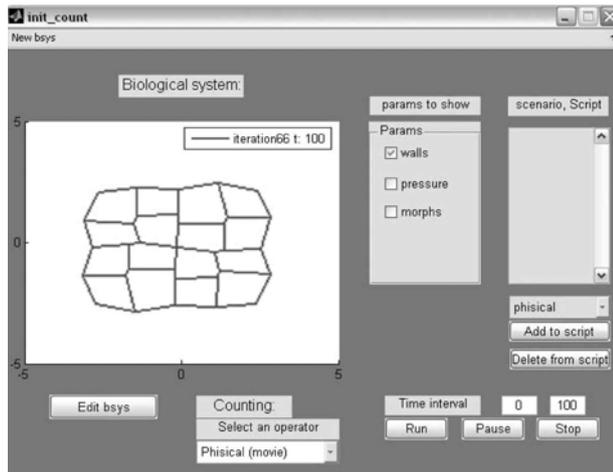


Figure 4. Model calculation

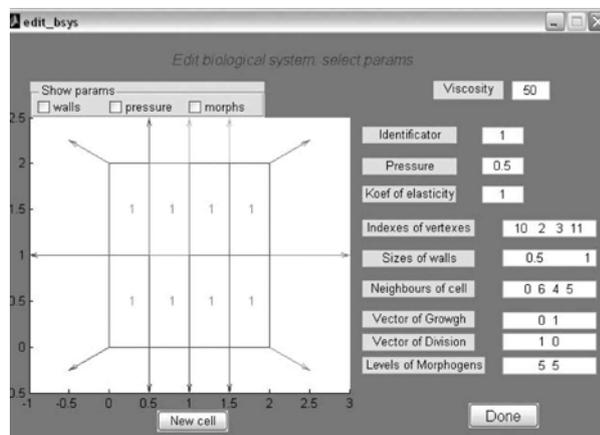


Figure 5. Modification of a cellular ensemble

ACKNOWLEDGEMENTS

This work was supported by the NSF: FIBR EF-0330786 grant, Innovation project of Federal Agency of Science and innovation IT-CP.5/001 “Development of software for computer modeling and design in postgenomic system biology (system biology *in silico*)”.

REFERENCES

- Cosgrove D. (1986) Biophysical control of plant cell growth. *Annu. Rev. Plant. Physiol.*, **37**, 377–405.
 Nelson C.M., Jean R.P., Tan J.L., Liu W.F., Sniadecki N.J., Spector A.A., Chen C.S. (2005) Emergent patterns of growth controlled by multicellular form and mechanics. *Proc. Natl. Acad. Sci. USA*, **102**, 11594–11599.
 Nobel P. (1973) Plant cell physiology: a physicochemical approach. Mir, Moscow. (In Russ.).
 Prigogine I., Defay R. (1966) Chemical thermodynamics. Nauka. (In Russ.).

AGNS (ARABIDOPSIS GeneNet SUPPLEMENTARY DATABASE), RELEASE 3.0

Omelianchuk N.A.^{*1}, Mironova V.V.¹, Poplavsky A.S.¹, Pavlov K.S.¹, Savinskaya S.A.², Podkolodny N.L.¹, Mjolsness E.D.³, Meyerowitz E.M.⁴, Kolchanov N.A.¹

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ² Novosibirsk State University, Novosibirsk, 630090, Russia; ³ Institute for Genomics and Bioinformatics; University of California, Irvine CA 92607, USA; ⁴ Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

* Corresponding author: e-mail: nadya@bionet.nsc.ru

Key words: *Arabidopsis*, gene expression, phenotype, database

SUMMARY

Motivation: Arabidopsis GeneNet Supplementary DataBase (AGNS) provides an integrated view of genetic data for *Arabidopsis thaliana* (Omelyanchuk *et al.*, 2006). AGNS contains the description of the functions of the known Arabidopsis genes at the levels of mRNA, protein, cell, tissue, organs and the organism in both wild type and mutant backgrounds. AGNS annotates published papers on gene expression and function and by using a special format and interface integrates, systematizes, and classifies this heterogeneous, disparate, and scattered information.

Results: The AGNS structure, formats and content have now been dramatically revised in order to ensure the efficient use of the records, rapid growth of the content and integration with other databases. The Arabidopsis ontology used in AGNS has been brought up to international standard. A new section, SD, which collects data on mutation localization in gene nucleotide sequences and provides descriptions to transgenic constructs, has been added. A new web interface with data submission and database navigation capabilities has been designed and implemented. Major updates have been made to all the AGNS sections.

Availability: <http://wwwmgs2.bionet.nsc.ru/agns>.

INTRODUCTION

To study development in plant, access is required to the various and scattered information in publications found as descriptions of gene expression patterns and morphological abnormalities in different mutant and transgenic lines, to compare these data with those for the wild type. Analysis of this body of data and its reduction to a single format would considerably facilitate analysis of the information. The different parts of this information started to be annotated, which is indicative of growing interest to the systematization of such data. TAIR (<http://www.arabidopsis.org/>) and Geneinvestigator (<https://www.geneinvestigator.ethz.ch/>) keep collecting information on gene expression in Arabidopsis obtained from large-scale experiments, AREX systematizes evidence-based (microarrays, in situ, promoter::reporter constructs, etc.) data on Arabidopsis gene expression in the root (<http://www.arexdb.org/database.jsp>). WatDB (<http://www.watdb.nl/>) and SeedGenes (<http://www.seedgenes.org>) systematize data on phenotypic aberrations in mutants. However, a complete understanding of the function of a gene can be reached only when the data about localization of mutations in nucleotide sequences and

information on expression patterns of genes in wild type, mutant and transgenic plants will be associated with detailed description of phenotypes of the wild type, mutant and transgenic plants (including double mutants and transgenic plants in the mutant background). AGNS was developed as a tool, which allow to trace the whole history from change in the nucleotide sequence of a particular gene, through further changes in the expression of that or associated genes, to phenotypic changes. In two previous releases the structure and format of three sections (Expression, Phenotype and Reference Databases) and two controlled vocabularies have been developed (Fig. 1). Expression Database describes gene expression in wild type, mutant, and transgenic plants. Phenotype Database contains information on phenotypic abnormalities of particular organs at particular stage in mutant and transgenic plants. Reference Database includes references to the papers and description of plant growth conditions with an indication of the ecotypes used as controls in the experiments. Detailed controlled vocabularies on growth stages and morphology were developed around the annotated data.

The rationale for AGNS release 3.0 was the need for improvements to the available data presentation format and user interface. The improved format serves to collect the results from different types of experiments on gene expression; to describe and systematize data on phenotypic abnormalities in single mutants, double mutants or combinations of mutant genes with transgenes; to provide cross database querying on Arabidopsis. The user interface in its current version let AGNS be updated easily; besides, it features new queries, which are required for data mining, creation of new hypothesis and modeling of gene expression regulation and developmental processes in Arabidopsis.

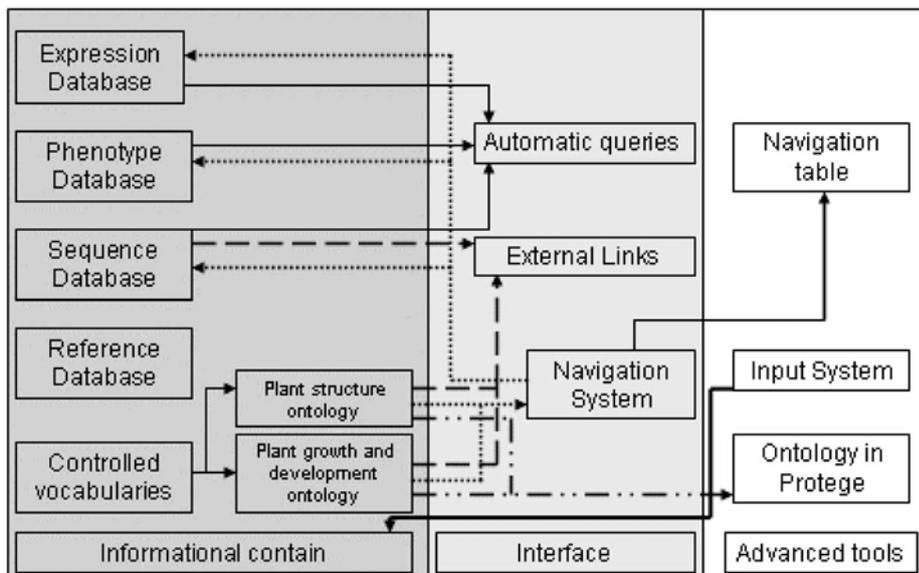


Figure 1. Structure and internal links in AGNS release 3.0.

METHODS AND ALGORITHMS

Under the AGNS project, a toolkit was developed for the modeling of new types of data, data mining and presenting the complex types of data. The software developed to serve the purpose is a set of CASE (Computer Aided Software Engineering) tools, which provide language and system support as follows:

1. a library for data format conversion between Java objects and XML documents stored in the database: resolves references to entries by their system ID, creates proxy-objects; supports separate indexation for objects in different classes.
2. a JSF (Java Server Faces)-based library for specialized GUI components.
3. a library for data management and the creation of skeleton WEB applications.

The system features the following main capabilities:

1. data storage and indexation
2. optimized advanced querying (WEB)
3. tools for creating and editing entries (WEB)
4. various navigation options (WEB)
5. an option to use custom data types for the benefit of external users (databases not linked originally to the system can, if so desired, be integrated into the system).

For efficient data storage, the system in its basic release uses Berkeley DB XML, an embedded XML database, with W3C XQuery as the basic query language.

The extensibility of the system was demonstrated on reflexive and model-based Java and XML technologies. This approach provides the maximum versatility of the system's main components and keeps it applicable to a semantically broad spectrum of tasks at the expense of a minor reduction in throughput. This approach (MDA, Model Driven Architecture) is believed to be absolutely adequate to most bioinformatics disciplines.

RESULTS AND DISCUSSION

The main problem concerned with extending and upgrading AGNS so far has been the tediousness of the annotation process. To address the problem, a software toolkit was developed (see Methods and Algorithms), which allowed us to create a specialized, "smart" data submission system for AGNS and an easy-to-use navigation system. The new interface makes AGNS upgrade rapid and more accurate.

Ontology is one of the backbones to AGNS. The AGNS ontology consists of two controlled vocabularies on plant developmental stages, anatomy and morphology. These vocabularies are used in the database format for annotation of data. The structure of the AGNS ontology is different from those of other databases (Jaiswal *et al.*, 2005): what makes it unique is that any plant organ or developmental stage can be described to the highest detail and, as a result, it is strongly extendable.

Now AGNS ontology has been brought up to international nomenclature, which will enable cross database querying with other databases on arabidopsis.

A new section, Sequence DataBase, which contains references to the nucleotide (genomic, mRNA and protein-encoding) sequences of the genes for morphogenesis and the indications of where the mutations which disrupt morphogenesis are located within these sequences; descriptions to the transgenic constructs built using these genes and transgenic modules included in transgenic constructs with the intact coding sequence of the gene (for example, another promoter or untranslated mRNA regions). Depending on what data type, one of three data formats is provided in Sequence DataBase: gene description, mutant description and transgene description. An external link to TAIR through the AGI gene code improves AGNS cross database querying capabilities.

In order to enhance the efficiency of the use of data contained in AGNS, the formats were revised in other AGNS sections, too. The strategy for describing double mutants was changed in Phenotype DataBase: if the genes interact epistatically or additively, this information is entered, in a particular format, as comments to the main gene; if the phenotype of the double mutant is not identical to those of single mutations, this piece of information is annotated separately and the double mutant is necessarily indicated in the "allele/transgene" field.

Today the most thoroughly examined in AGNS data are the phenotypes of mutant and transgenic plants and the expression of the genes in wild type, mutant and transgenic plants for the genes expressed in the apical shoot meristems and involved in miRNA biogenesis. These data were used for the reconstruction of the gene network for the functioning of the apical shoot meristem of *Arabidopsis* (Mironova, Omelyanchuk, 2004), for the creation of the ontology of *Arabidopsis* development and developmental abnormalities in the Protégé system (Ponomyrov, 2006), for the development of cell automaton models (Akberdin *et al.*, 2006) and a mathematical model of meristem functioning (Nikolaev *et al.*, 2006).

Thus, the current release has altogether with a two times higher content (Table 1), the improved format, database structure, database model, web interface, data submission system and navigation system.

Table 1. AGNS content

Releases	Expression patterns	Genes	Phenotype abnormalities	Mutant alleles or transgenes	Anatomical elements	Developmental stages	Papers annotated
Release 2	514	44	192	526	243	169	197
Release 3	1154	103	229	759	417	230	259

The progress made allows us, in the nearest future, to rapidly increase the volume of the AGNS content, to crosslink AGNS with other databases on the *Arabidopsis* genome, to extensively use AGNS references for the development of computer-based systems, which analyze AGNS data, and, following these analyses, to create models for plant organ development.

ACKNOWLEDGEMENTS

This work was supported in part by Russian Federal Agency of Science and innovation (IT-CP.5/001), Russian Foundation for Basic Research (grant No. 05-07-98012), Russian Academy of Sciences (grant No. 10.4 and Program “Origin and evolution of biosphere”), Siberian Branch of Russian Academy of Sciences (Integration Project No. 115) and the US National Science Foundation (FIBR EF-0330786 Development Modeling and Bioinformatics).

REFERENCES

- Akberdin I.R. *et al.* (2006) A cellular automation to model the development of shoot apical meristem of *Arabidopsis thaliana*. *This issue*.
- Jaiswal P. *et al.* (2005) Plant Ontology (PO): a controlled vocabulary of plant structures and growth stages. *Comparative and Functional Genomics*, **6**, 388–397.
- Mironova V.V., Omelyanchuk N.A. (2004) Gene network of the *Arabidopsis* developing shoot meristem and its description in the GeneNet computer system. *Proceedings of the Fourth International Conference on Bioinformatics of Genome Regulation and Structure (BGRS'2004)*. ICG, Novosibirsk, **2**, 101–104.
- Nikolaev S.V. *et al.* (2006) A one dimensional model for the regulation of the size of the renewable zone in biological tissue. *This issue*.
- Omelyanchuk N. *et al.* (2006) AGNS-a database on expression of *Arabidopsis* genes. In Kolchanov N., Hofstaedt R., Milanesi L., (eds), *Bioinformatics of Genome Regulation and Structure II*. Springer Science+Business Media, Inc. 433–442.
- Ponomyrov D.K. *et al.* (2006) Semantically rich ontology of anatomical structure and development for *Arabidopsis thaliana*. *This issue*.

SEMANTICALLY RICH ONTOLOGY OF ANATOMICAL STRUCTURE AND DEVELOPMENT FOR *ARABIDOPSIS THALIANA* L.

Ponomaryov D.^{*1}, Omelianchuk N.², Kolchanov N.², Mjolsness E.³, Meyerowitz E.⁴

¹ Institute of Informatics Systems, Novosibirsk, 630090, Russia; ² Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ³ Institute for Genomics and Bioinformatics, University of California, Irvine CA 92607, USA; ⁴ Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

* Corresponding author: e-mail: ponomaryov@ngs.ru

Key words: plant, anatomy, morphology, development, phenotype, ontology

SUMMARY

Motivation: An ontology of a subject domain is a set of statements that are true for every possible situation in this subject domain. The aim of our work is a formal ontological description of anatomical structures in development of *Arabidopsis thaliana* wild-type plant, which serves as the target subject domain. The reason for developing this type of ontology in presence of TAIR and Plant Ontology was to support analysis and processing of expression and phenotypic data from AGNS database. For this task, it was necessary to capture the fact that the hierarchy of anatomical elements changes within the developmental process. This means that we should have been able to express not only development of one structure into another, but also changes in the number of contained entities, as well as containers for anatomy elements.

Results: We have developed an ontology that consists of a core structure for representing statements of the kind “*anatomy element X at stage Sx exists in anatomy element Y at stage Sy*” (where Sx and Sy are development stages of X and Y respectively) and is filled in with facts from Arabidopsis development studies, which have been extracted from publicly available articles. In general, the problem of describing anatomical structures in development (without taking spatial orientation into account) was reduced to ordering and inclusion of developmental stages.

Availability: The ontology was built as a wrapper around the AGNS database, which is available at <http://wwwmgs2.bionet.nsc.ru/agns/>

INTRODUCTION

AGNS (Arabidopsis GeneNet Supplementary database) is an Internet-available resource that provides access to description of the functions of the known Arabidopsis genes at various levels—the levels of mRNA, protein, cell, tissue, and ultimately at the levels of organs and the organism in both wild type and mutant backgrounds (Omelianchuk *et al.*, 2006). At present, the data are collected, assembled, and curated using the formats developed in two AGNS modules: expression, and phenotype databases. In addition, while annotating, it appeared necessary to built two controlled vocabularies around contents of annotated data and an ontology of development of the cell types, tissues, organs based on the data presented by authors of publications on gene expression patterns and mutant phenotypes and by the publications describing

developmental processes. While controlled vocabularies were developed for paper curators, the aim in constructing ontology was to apply it in algorithms of analysis of expression and phenotypic data. The benefits of use of formal ontology in processing of experimental data are recognized in bioinformatics (Karp, 2000; Bard, 2004). In this paper, we explain the results of formalizing plant anatomy in development by example of navigating in the AGNS database. We believe this use case to be rather simple, yet very illustrative for our work. We show that by having a concrete practical task to be solved, the choice of formalization becomes well-founded and easier to evaluate, than in some abstract case.

METHODS AND ALGORITHMS

We formulate the task of navigating the database via ontology as follows:

- query the database in ontological terms (concepts);
- use relations between ontological terms to broaden/narrow data extraction from the database.

In practice, AGNS consists of two databases: ED (expression database) and PD (phenotype database) and the AGNS facts, which we denote by predicates with corresponding names *ED* and *PD*, have the following form:

$ED(Gene, Anatomy_Element, Stage, Express_Level, IsAbnormal),$

$PD(Phenotype_ID, Anatomy_Element, Stage, Abnormality).$

Here *Anatomy_Element* is an organ, tissue or cell, *Stage* is a developmental stage of *Anatomy_Element* and *Abnormality* is a textual name for the phenotypic abnormality. Our first step is to navigate in the expression database ED. A typical example of navigation in this case would be like this: if we are interested in expression of gene *AG* in some organ *X*, then obviously we would like to know its expression in sub-organs or tissues of *X*. Or we would like to restrict ourselves to only certain parts of *X* or certain development stages of *X*. This implies that the ontology should have the necessary relations between anatomy elements and developmental stages. It turns out that in order to formulate queries and manage extraction of data from the database, one needs to query the ontology itself. In our case, two typical queries, the answers which ontology should provide are:

(Q1) *X* is an anatomy element, find all elements *Y*, belonging to *X*

(Q2) *S* is a developmental stage, find all stages earlier/later, than *S*.

The main problem in representing plant anatomy in development may be formulated as follows: if *X*, *Y* are two anatomy elements, then *X* belongs to *Y* at developmental stage *S_i* does not necessary imply that *X* belongs to *Y* at another stage *S_k*. The number of organs in a plant may change within development and moreover, anatomy elements may have different direct containers (i.e. elements to which it directly belongs) at different stages. With respect to the task of navigation, this directly affects what anatomy elements are considered and what expression data are extracted for a queried plant/organ/tissue developmental stage. From the facts given above one can notice that the relation “belongs to” is temporal. Fortunately, Arabidopsis development is well studied, “what-where-when” is known, and stages are discrete. Eventually, the information we would like to be expressed in the ontology are statements of the kind:

“Anatomy element *X* at stage *S_X* exists in anatomy element *Y* at stage *S_Y*” (I)

We only need to represent this concept by a combination of suitable binary predicates.

From a theoretical point of view, we consider formal ontology as a set of sentences in First Order Logic language. But for practical purposes we restrict ourselves to a decidable fragment of FOL and to formulas of a special kind, which will be reflected by the choice of the OWL-DL language for implementation. In a formal ontology we distinguish a signature that is a set of predicate, functional, constant symbols and axioms that restrict

possible interpretations of these symbols. Throughout this paper, we will simultaneously use names for binary predicates and the term “relation” to denote the same things.

The core elements in the ontology are the predicates listed below and axioms that will be introduced further in this section.

*Anatomy_Element*¹
*Development_Stage*¹
*Has_Development_Stage*² (*Anatomy_Element* x*Development_Stage*) (II)
*Before*² (*Development_Stage* x*Development_Stage*)
*Occurs_In*² (*Development_Stage* x*Development_Stage*)

We assume that binary predicates are defined on Cartesian products of those sets, that are defined by unary ones and are mentioned informally in the parenthesis. Let us denote the statement (I) by predicate *ExistsIn*(*X*,*S_X*,*Y*,*S_Y*), i.e. the predicate is true, whenever statement (I) holds for *X*, *S_X*,*Y*, *S_Y*. We define this predicate by the following combination of binary predicates that were introduced above:

ExistsIn (*X*, *S_X*, *Y*, *S_Y*) ↔

Has_Development_Stage (*X*, *S_X*) & *Has_Development_Stage* (*Y*, *S_Y*) & *Occurs_In* (*S_X*, *S_Y*)

An example for Arabidopsis development is the case, when *Y* is Leaf, *S_Y* is LDS4 (the fourth leaf development stage), *X* is midrib and *S_X* is MDS1 (the first midrib development stage). One should notice that there are no direct relations between anatomy elements. Instead, the relation *Occurs_In* serves for inclusion of anatomy elements into each other. The ontology also includes axioms restricting interpretation of the predicates introduced above. We do not mention them here due to paper size limitations. Core concepts defined by unary predicates in (II) together with relations defined by the mentioned binary predicates and these axioms make up a formal ontology, which we consider as an *ontology for anatomical structure and development of plants*.

The facts that help to evaluate the proposed ontology originate from the experience of its instantiation with real data from Arabidopsis development studies. Textual information about Arabidopsis anatomy and development was extracted from 100+ publicly available articles and entered into the ontology in the form defined by formal constructs from the previous section. During this process we have encountered two main problems regarding incompleteness of information, namely:

1. stages were not defined for all anatomy elements we would like to include in the ontology;
2. development of lots of anatomy elements was described in the manner “*X has the following properties, when Y is at stage S*” (e.g. “*LDS2 leaf shows meristematic divisions throughout the mesophyll*”), but these descriptions were given only for some stages of *Y*.

This has led to the need of creation of “artificial” development stages for anatomy elements. Even though such stages may not be distinguished in reality, we need to have them in the ontology to provide a proper unfolding of organ anatomy structure. On the other hand, they also allowed for a more detailed description of developmental processes of Arabidopsis in comparison with the well-known of TAIR ontology.

IMPLEMENTATION AND RESULTS

In the implementation we were guided by two objectives: to have a widely supported expressive ontology language with full reasoning capabilities and to use an ontology editor with rich import-export and visualization functions. This resulted in the choice of the *OWL* language and *Protégé* ontology editor (version 3.1, <http://protege.stanford.edu/>). The problem of representing rule-like axioms was not considered as a significant one, because the number of such axioms in our ontology is small. Thus, it is possible to rewrite them manually into the language of the needed inference engine. In our case we used the *Algernon* engine for *Protégé*. A part of the axioms was implemented as constraints, other axioms – as forward or backward chaining rules. The notion of transitivity is present in

OWL as the *TransitiveProperty* construct, but as far as *OWL* is not yet fully supported by existing reasoners, we decided to implement transitivity axioms both in *OWL* and in the inference engine language.

DISCUSSION

We introduced a formal framework for representing plant anatomy in development by the example of an ontology for *Arabidopsis thaliana* (L.) Heynh. The proposed formalization was developed with orientation to the task of navigating in a gene expression database for the plant. The main problem of representing plant anatomy in development was reduced to ordering and inclusion of developmental stages. The built ontology has a small core, yet with strong expressive capabilities. It presents a richer language for describing anatomical structures in development in comparison to the known controlled vocabularies. This is the reason we call this ontology *semantically rich*. It became possible to define *Develops_From* relation for anatomy elements not as a separate one, but on top of core structures of the ontology. Information about *Arabidopsis* development taken from 100+ publicly available articles was presented in a formal way. Each stage in the ontology was provided with textual description with a reference to the source article. The implemented ontology is a deductive database consisting of facts about plant development and inference rules.

ACKNOWLEDGEMENTS

This work was supported in part by RF Ministry of Education grant “The development of the higher education scientific potential” (project No. 8328), Russian Foundation for Basic Research (grants No. 05-07-98012 and 03-04-48506), Russian Academy of Sciences (grant No. 10.4), Siberian Branch of Russian Academy of Sciences (Integration Project No. 119) and the US National Science Foundation (FIBR EF-0330786 Development Modeling and Bioinformatics).

REFERENCES

- Bard J.B., Rhee S.Y. (2004) Ontologies in biology: design, applications and future challenges. *Nat. Rev. Genet.*, **5**, 213–222.
- Karp P.D. (2000) An ontology for biological function based on molecular interactions. *Bioinformatics*, **16**, 269–285.
- Omelyanchuk N. *et al.* (2006) AGNS—a database on expression of *Arabidopsis* genes. In Kolchanov N., Hofstaedt R., Milanesi L., (eds), *Bioinformatics of Genome Regulation and Structure II*. Springer Science+Business Media, Inc. 433–442.

A PROGRAM METHOD OF CONSTRUCTING ONTOLOGY OF PHENOTYPIC ABNORMALITIES FOR *ARABIDOPSIS THALIANA*

Ponomaryov D.^{*1}, Omelianchuk N.², Mironova V.², Kolchanov N.², Mjolsness E.³, Meyerowitz E.⁴

¹ Institute of Informatics Systems, Novosibirsk, 630090, Russia; ² Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; ³ Institute for Genomics and Bioinformatics, University of California, Irvine CA 92607, USA; ⁴ Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

* Corresponding author: e-mail: ponomaryov@ngs.ru

Key words: development, phenotype, computer analysis, algorithms, ontology

SUMMARY

Motivation: Modern researches prove that studying gene networks and development of organisms at a higher level of abstraction allows for a better understanding of mechanisms of developmental processes and their interactions. We develop a classification of phenotypic abnormalities of Arabidopsis to distinguish and prove the existence of specific functional modules in the plant, to identify key points of abnormal development and to find parallel regulatory pathways leading to abnormalities.

Results: We have developed an algorithm that takes a description of phenotypic abnormality caused by a mutation as an input and generates a graph of relations of this abnormality to other abnormalities of mutant and transgenic phenotypes, basing on data from AGNS database. As *ontology* is usually viewed as a set of terms and relations between them, we call this graph an ontology of phenotypic abnormalities of Arabidopsis.

Availability: Phenotypic data used by the algorithm are available at <http://wwwmgs2.bionet.nsc.ru/agns/>

INTRODUCTION

AGNS database (Omelianchuk *et al.*, 2006) has a AGNS_PD module that contains information about phenotypes of Arabidopsis in the form of statements: $PD(Phenotype_ID, Anatomy_Element, Stage, Abnormality)$, where *Phenotype_ID* is a name for allele or transgene, *Anatomy_Element* is a name for organ, tissue or cell, *Stage* is a name for developmental stage of the *Anatomy_Element* and *Abnormality* is a name for phenotypic abnormality. For instance, the fact $PD(CLV1-1, Floral_Meristem, FDS3, Enlarged)$ states that the floral meristem at *FDS3* developmental stage is enlarged in mutant plants homozygous for the *clv1-1* allele. Facts of this sort have been extracted from publicly available papers describing separate experimental results on mutant and transgenic plants of *Arabidopsis thaliana* (Omelianchuk, 2006). In our work, we found necessary to summarize these data in order to identify the same abnormalities that are described differently in different research groups and to build a general classification of phenotypic abnormalities of Arabidopsis. By this we aim to distinguish and prove the existence of specific functional modules in the plant, to identify key points of abnormal development and to find parallel regulatory

pathways leading to abnormalities (Ponomaryov, Omelianchuk, 2006). This in turn could allow for a better understanding of mechanisms of developmental processes and their interactions and would be a step to reconstructing the underlying gene networks. Modern researches (Gunsalus *et al.*, 2005; Roy, Morris, 2005) show that studying this problem from a higher level of abstraction can benefit in lots of cases, where analysis of too detailed data could not finally lead to sound models.

METHODS AND ALGORITHMS

Let us consider four sets A , E , S and T , where

1. A is a set of phenotypes (alleles and transgenes) of Arabidopsis.
2. E – is a partially ordered set of anatomical elements (cells, tissues, organs and structural elements, such as specially distinguished layers or zones; the whole plant is also considered as anatomical element), with the order \triangleright , defined in the following way: (for all $e_1, e_2 \in E$) ($e_1 \triangleright e_2$, if and only if e_2 develops from e_1).
3. S – is a partially ordered set of developmental stages of anatomical elements with the order $>$ defined as follows: (for all $s_1, s_2 \in S$) ($s_2 > s_1$, if and only if s_1 is before s_2 in time).
4. T is a set of types of phenotypic abnormalities (*abnormal position, abnormal shape, delayed development, increased number*, etc.).

Define a relation $R \subseteq E \times S$ with the property (for all $e \in E$ and $s \in S$) ($(e, s) \in R$, if and only if s is a developmental stage of e). Introduce also a relation $\succ \subseteq E \times S \times E \times S$ as follows: (for all $e_1, e_2 \in E$ and $s_1, s_2 \in S$) ($(e_1, s_1, e_2, s_2) \in \succ$, if and only if e_1 at the developmental stage s_1 exists in e_2 , when e_2 undergoes stage s_2). Note that under several additional assumptions, the structure $\langle E \cup S, \triangleright, >, \succ \rangle$ can be considered as a model for the logical theory described in (Ponomaryov, Omelianchuk 2006).

Following the level of abstraction, at which abnormalities are presented in the AGNS_PD module, we define a phenotypic abnormality as a 4-tuple: $N = \langle G, e, s, t \rangle$, where $G \subseteq A$, $e \in E$, $s \in S$, $t \in T$. AGNS_PD consists precisely of a collection of such 4-tuples with assigned textual names. These names represent short characterizations of abnormalities that have been extracted from papers describing separate experimental results on mutant and transgenic plants of Arabidopsis. It follows immediately from our definition of abnormality that two different names denote the same abnormality in the AGNS_PD, if they correspond to the same 4-tuples. Note that this is not the only rule to identify the same abnormalities.

Besides the task of name disambiguation, we also distinguish six relations between abnormalities that we aim to extract by analyzing AGNS data. Here we only list their names and give a brief informal explanation.

We have developed an algorithm that takes a representation of abnormality in the form of a tuple $N = \langle G, e, s, t \rangle$ as an input and outputs a graph with vertices referring to abnormalities, which are connected by edges that correspond to the relations above. In other words, vertices in the resulting graph correspond to the tuples from the AGNS_PD and are labeled with textual descriptions of abnormalities, while edges are labeled with the names of the relations. In particular, those vertices that have no incoming edges labeled with “Consequence_of” are considered to be candidates for initial points of abnormality development and are checked against AGNS gene expression data by the algorithm. To resolve ambiguous cases when AGNS data is insufficient, the algorithm

uses additional information about known gene functions, as well as information about known gene interactions, which is available in a separate database.

Table 1. Relations between phenotypic abnormalities – informal definitions

Name of relation between abnormalities	Informal definition by an example
Blocked_by	One abnormality is blocked by another one, if it is not present, when the other abnormality is observed.
Consequence_of	An abnormality is a consequence of another one, if it is the result of development of this abnormality within time.
Specialization_of	An abnormality, to which another one is a specialization, is a stronger abnormality.
Inverse_to	Two abnormalities are inverse to each other, if they are presented by opposite phenotypic changes.
Composite_of	An abnormality is a composite of several others, if it is caused by all of them together.
Alternative_to	One abnormality occurs in another percentage of cases, in comparison to that of the second one.

For developing the algorithm it was necessary to give an explicit formal definition to all of the listed relations. As a result, we defined a set of rules, which are the necessary conditions defining these relations. These rules were produced by analysis of all possible cases of structural differences between two arbitrary abnormalities in the form: $N_1 = \langle G_1, e_1, s_1, t_1 \rangle$, $N_2 = \langle G_2, e_2, s_2, t_2 \rangle$. Clearly, there is a restricted number of cases how these abnormalities (defined as tuples) can differ from each other. They originate from considering all possible set-theoretic relations between the sets G_1, G_2 , as well as relations $\triangleright, >, \succ$ pair-wise between the elements e_i, s_i, t_i , $i = 1, 2$. Due to paper size limitations, we list below only some of the rules used in our algorithm.

Table 2. Some of the rules for inferring relations between abnormalities

Premise	Conclusion
$G_1 = G_2$ and $s_1 < s_2$ and $e_1 = e_2$ and $t_1 = t_2$	Abnormality $N_2 = \langle G_2, e_2, s_2, t_2 \rangle$ is a consequence of abnormality $N_1 = \langle G_1, e_1, s_1, t_1 \rangle$.
$G_1 = G_2$ and $s_1 = s_2$ and $e_1 = e_2$ and $t_1 \neq t_2$	N_1 and N_2 are alternative abnormalities to each other.
$G_1 = G_2 = G_3$ and $s_1 = s_2 = s_3$ and $(e_1, s_1, e_3, s_3) \in \succ$ and $(e_2, s_2, e_3, s_3) \in \succ$	N_3 is a composite of N_1 and N_2 .

IMPLEMENTATION AND RESULTS

We have introduced an axiomatic semantics for the relations considered above by a system of axioms restricting their interpretation. On one hand, it has allowed for constructing a declarative language for description of abnormalities and relations between them. On the other hand, we have used these axioms to define post-conditions for implementation of our algorithm. By this we have proved a total correctness of the implemented algorithm using the Floyd’s method of program proving (the preconditions have been taken according to the formal definition of abnormality in our work).

DISCUSSION

Presently, we are evaluating the developed algorithm on different datasets from the AGNS_PD module of the AGNS database. In particular, we use phenotypic information regarding only certain organs or certain periods of development. The aim is to estimate the adequacy of the output results of the algorithm by checking, whether the inferred relations are correct and also whether theoretically predicted modules are present. As adequacy of the resulting graph potentially depends on the amount of data available for processing, such testing process also helps to discover points of potentially incomplete information in the AGNS database.

ACKNOWLEDGEMENTS

This work was supported by Russian Foundation for Basic Research (grants No. 05-07-98012 and 03-04-48506), Russian Academy of Sciences (grant No. 10.4), Siberian Branch of Russian Academy of Sciences (Integration Project No. 119) and the US National Science Foundation (FIBR EF-0330786 Development Modeling and Bioinformatics).

REFERENCES

- Gunsalus K.C., Ge H. *et al.* (2005) Predictive models of molecular machines involved in *Caenorhabditis elegans* early embryogenesis. *Nature*, **436**, 861–865.
- Omelianchuk N.A. *et al.* (2006) AGNS – A Database on Expression of Arabidopsis Genes. *Bioinformatics of Genome Regulation and Structure II*, Springer Verlag, 433–442.
- Ponomaryov D.K., Omelianchuk N.A. *et al.* (2006) Semantically rich ontology of anatomical structure and development for *Arabidopsis thaliana*. *This issue*.
- Roy P.J., Morris Q. (2005) Network news: functional modules revealed during early embryogenesis in *C. elegans*. *Developmental Cell*, **9**, 307–315.

Indexes

AUTHOR INDEX

- Akberdin I.R., 105, 185
Alfieri R., 145
Ananko E.A., 19, 45, 114
Apasieva N.V., 55
Bazhan S.I., 105, 114
Belavskaya V.V., 218
Bezmaternikh K.D., 109, 156
Bukharina T.A., 190
Dymshits G.M., 128
Fadeev S.I., 97, 118, 121, 194, 199, 213
Furman D.P., 190
Gainova I.A., 114, 118
Grigorovich D.A., 89
Gunbin K.V., 194, 199
Gursky V.V., 125
Ibragimova S.S., 19, 78, 89
Ignatieva E.V., 19, 137
Imaizumi A., 19
Ivanisenko V.A., 109, 148
Kachko A.V., 93
Kalashnikova E.V., 128
Kashevarova N.A., 105
Katokhin A.V., 137, 190
Kel A., 132
Khlebodarova T.M., 19, 49, 55, 89, 93, 105, 109
Kogai V.V., 194, 199, 213
Kolchanov N.A., 19, 213, 218, 223, 227, 231
Kolpakov F.A., 128, 132
Kolpakova A.F., 170
Komarov A.V., 185
Kondrakhin Y.V., 137
Korolev V.K., 118
Korotkov E.V., 165
Korotkov R.O., 148
Kouznetsova T.N., 137
Kozlov K.N., 141, 204
Lashin S.A., 55, 68
Likhoshvai V.A., 13, 105, 109, 121
Matsui K., 19, 25, 30, 40, 45, 68, 78
Matveeva A., 204
Medvedev A.E., 118
Merelli I., 145
Meyerowitz E.M., 223, 227, 231
Milanesi L., 132, 145
Mironova V.V., 185, 209, 223, 231
Mishchenko E.L., 19, 68, 109, 148
Mjolsness E.D., 213, 218, 223, 227, 231
Mordvinov V.A., 137
Nedosekina E.A., 19, 30, 35, 60
Nikolaev S.V., 213, 218
Nishio Y., 19
Omelianchuk N.A., 185, 209, 223, 227, 231
Omelyanchuk L.V., 194, 199
Oshchepkov D.Yu., 137
Ozonov E.A., 185
Paramonova N.V., 174
Pavlov K.S., 223
Penenko A.V., 218
Peshkov I.M., 97
Podkolodnaya N.N., 174
Podkolodnaya O.A., 19, 25
Podkolodny N.L., 19, 174, 223
Ponomaryov D.K., 209, 227, 231
Poplavsky A.S., 19, 209, 223
Poroikov V., 132
Proskura A.L., 19
Ratushny A.V., 13, 25, 35, 40, 45, 49, 73, 151, 156, 160
Reinitz J., 125
Rudenko V.M., 165
Samsonov A.M., 125, 141
Samsonova M.G., 178, 204
Savinskaya S.A., 223
Schwartz Ya.Sh., 114
Shamanina M.Y., 137

Sharipov R.N., 132, 170
Shtokalo D.N., 121
Smirnova O.G., 19, 40, 73, 84, 89
Stepanenko I.L., 19, 93, 174
Surkova S.Yu., 178
Tikunova N.V., 93
Turnaev I.I., 78, 84
Usuda Y., 19, 25, 30, 40, 45, 68, 78

KEY WORDS

Achaete-scute complex (AS-C), 190
Adenine phosphoribosyltransferase, 73
Adipocyte, 137
Algorithms, 231
Anatomy, 227
AP-1, 128
Apoptosis, 170
Arabidopsis, 209, 223
Aromatic amino acid biosynthesis, 45
Attractors, 125
Autonomous system, 97
Bioinformatics, 68, 145
Biomechanics, 218
Biopath, 170
Biosensor, 93
BioUML, 132
Branched chain amino acid, 30
Ca²⁺ signaling, 128
Cancer, 132, 170
Cell cycle, 145
Cell cycle regulation, 132
Cell ensemble, 218
Cell survival, 170
Cell wall, 218
Cellular automaton, 185
Cholesterol homeostasis, 160
Cluster analysis, 174
Computer analysis, 165, 231
Constants, 19
Controlled vocabulary, 209
CREB, 128
Cysteine biosynthesis, 68
cyoABCDE, 49
Data mining, 145
Database, 19, 89, 93, 132, 145, 148, 223
Delay equations, 121
Development, 227, 231
Development of shoot meristems, 185
Differential evolution, 141
D-Lactate dehydrogenase-2, 84
Drosophila, 125, 178, 204
Drug design, 132
Dynamics, 19
E. coli, 60
Endocytosis, 151
Environmental stress, 93
Enzymatic reaction, 156
Enzyme reaction, 13
Escherichia coli, 19, 25, 35, 40, 45, 49, 55, 68, 73, 78, 84
Expression, 19
Feedback loop, 174
Functional parameters, 19
Gene, 49
Gene circuits, 125
Gene expression, 178, 223
Gene network, 13, 25, 30, 35, 40, 45, 55, 60, 73, 78, 84, 97, 105, 160, 174, 190
Gene networks, 19
Genetic net, 165
Genetic systems, 121
Glucose 6-phosphate-1-dehydrogenase, 40
HCV replication, 109
Hepatitis C virus, 148
Hepatitis C virus (HCV), 165
Hill function, 13
Histidine biosynthesis, 25
Human hepatitis C, 109
Hypothesis testing, 199
Image processing, 204
Immune response, 114
Industrially important products, 89
Infection, 114
Inflammation, 170
Influenza virus, 105
Inhibitors of HCV NS3 protease, 109
Intracellular ontogeny, 105
Inverse problem, 97
Isoprenoid and cholesterol biosynthesis, 156
Kinetic parameters, 148
Kinetics, 19
LDL, 151
Limit theorem, 121
Lipid metabolism, 137
LXR α , 137
Mathematical model, 114, 165, 185, 194, 199, 213, 218
Mathematical model(ing), 13, 25, 30, 35, 40, 45, 49, 55, 60, 68, 73, 78, 84, 105, 109, 118, 151, 156, 160
Mathematical models, 121
Matrix process, 121
Metabolism, 19
Microbial biotechnology, 89
Microbial organism, 89
Model(ing), 19
Morphogenesis, 213, 218
Morphology, 227
Mycobacteria tuberculosis, 114
NF- κ B, 170, 174
Numerical analysis, 118
Ontology, 209, 227, 231
Operon, 49
Optimization, 141

Osmotic pressure, 218
Parameter continuation, 118
Parameters of model, 185
Pattern formation, 194, 199
Pentose phosphate pathway, 40
Period of division, 185
Phenotype, 223, 227, 231
Plant, 227
Plant development, 209
Positional information, 194, 199
PPAR γ , 137
Prenyltransferase, 156
Proneural cluster, 190
Purine biosynthesis, 60
Pyrimidine biosynthesis, 35
Pyruvate metabolism, 84
Quantitative information, 204
Receptor, 151
Regulation, 13, 25, 35, 40, 49, 68, 73, 78, 84
Regulatory gene networks, 141
Respiration, 55
Salvage pathways, 73
SAM, 213
Segmentation, 204
Segmentation genes, 178
Serine and glycine biosynthesis, 78
Shoot apical meristem, 213
Sigmoid interaction, 125
Signal pathway, 174
Signal transduction, 128
Signaling pathways, 190
SREBP family, 137
Strain improvement, 89
Systems biology, 145
Tissue, 213, 218
TNF α , 174
Transcription, 19
Transcription factor, 170
Transcription regulation, 93, 137
Treatment regimens, 114
Tyrosine hydroxylase, 128
Variability in expression, 178
Virtual mutagenesis, 199
Viscoelastic deformation, 218
Water potential, 218
Workflow, 204

Научное издание

**Труды пятой международной конференции
“Биоинформатика регуляции и структуры генома”**
Т. 2
на английском языке

**Proceedings of the Fifth International Conference
on Bioinformatics of Genome Regulation and Structure**
V. 2

Отредактировано и подготовлено к печати
в редакционно-издательском отделе ИЦиГ СО РАН
Редакторы: А.А. Ончукова, И.Ю. Ануфриева
Дизайн А.В. Харкевич
Компьютерная графика: А.В. Харкевич, К.В. Гунбин, Т.Б. Коняхина
Компьютерная верстка: А.В. Харкевич, К.В. Гунбин, Н.С. Глазкова

Подписано к печати 21.06.2006 г.
Формат бумаги 70 × 108 1/16. Печ.л. 21. Уч.-изд.л. 22,9
Тираж 250. Заказ 270

Институт цитологии и генетики СО РАН
630090, Новосибирск, пр. акад. М.А. Лаврентьева, 10
Отпечатано в типографии Издательства СО РАН
630090, Новосибирск, Морской пр., 2