

# BGRS\SB-2018

The 11<sup>th</sup> International Conference



## Mathematical Modeling and High-Performance Computing in Bioinformatics, Biomedicine and Biotechnology, MM-HPC-BBB-2018

The 3rd International Symposium

MM-HPC-BBB-2018  
NOVOSIBIRSK, RUSSIA  
21–24 AUGUST, 2018

[CONF.ICT.NSC.RU/MM-HPC-BBB-2018](http://CONF.ICT.NSC.RU/MM-HPC-BBB-2018)

Institute of Computational Mathematics and Mathematical Geophysics,  
Siberian Branch of Russian Academy of Sciences

Novosibirsk State University

Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences

Sobolev Institute of Mathematics, Siberian Branch of Russian Academy of Sciences

MATHEMATICAL MODELING  
AND HIGH-PERFORMANCE COMPUTING  
IN BIOINFORMATICS, BIOMEDICINE  
AND BIOTECHNOLOGY (MM-HPC-BBB-2018)

The 3rd International Symposium

Abstracts

21–24 August, 2018  
Novosibirsk, Russia

Novosibirsk  
ICG SB RAS  
2018

**Mathematical Modeling and High-Performance Computing in Bioinformatics, Biomedicine and Biotechnology (MM-HPC-BBB-2018)** : The 3rd International Symposium (21–24 Aug. 2018, Novosibirsk, Russia); Abstracts / Institute of Computational Mathematics and Mathematical Geophysics, Siberian Branch of Russian Academy of Sciences; Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences. – Novosibirsk: ICG SB RAS, 2018. – 80 pp. – ISBN 978-5-91291-039-5.

## Program Committee

### Chairs

S.I. Kabanikhin, Professor, Corresponding Member of the RAS, Institute of Computational Mathematics and Mathematical Geophysics of SB RAS

N.A. Kolchanov, Professor, Full Member of the RAS, Institute of Cytology and Genetics of SB RAS

S.S. Goncharov, Professor, Full Member of the RAS, Sobolev Institute of Mathematics of SB RAS

Aulchenko Yu.S. (Institute of Cytology and Genetics of SB RAS)

Bektemesov M.A. (Al-Farabi Kazakh National University, Kazakhstan)

Bocharov G.A. (Marchuk Institute of Numerical Mathematics of RAS, Moscow)

Cheng M. (Zhejiang University, China)

Chernykh I.G. (Institute of Computational Mathematics and Mathematical Geophysics of SB RAS)

Chupakhin A.P. (Lavrentyev Institute of Hydrodynamics of SB RAS)

Demidenko G.V. (Sobolev Institute of Mathematics of SB RAS)

Fadeev S.I. (Sobolev Institute of Mathematics of SB RAS)

Fedoruk M.P. (Novosibirsk State University)

Fedotov A.M. (Institute of Computational Technologies of SB RAS)

Glinksiy B.M. (Institute of Computational Mathematics and Mathematical Geophysics of SB RAS)

Golushko S.K. (Institute of Computational Technologies of SB RAS)

Hofstaedt R. (University of Bielefeld, Germany)

Ilyin A.I. (Scientific Center of Anti-Infective Drugs, Kazakhstan)

Koptiug I.V. (International Tomography Center of SB RAS)

Krebs O. (Heidelberg Institute for Theoretical Studies, Heidelberg, Germany)

Kulikov I.M. (Institute of Computational Mathematics and Mathematical Geophysics of SB RAS)

Kuramshina G.M. (Moscow State University, Moscow)

Lashin S.A. (Institute of Cytology and Genetics of SB RAS)

Likhoshvai V.A. (Institute of Cytology and Genetics of SB RAS)

Makeev V.Yu. (Vavilov Institute of General Genetics of RAS)

Marchuk A.G. (A.P. Ershov Institute of Informatics Systems of SB RAS)  
Moshkin M.P. (Institute of Cytology and Genetics of SB RAS)  
Nurseitov D.B. (K.I. Satpaev Kazakh National Technical University, Kazakhstan)  
Orlov Yu.L. (Institute of Cytology and Genetics of SB RAS)  
Penenko A.V. (Institute of Computational Mathematics and Mathematical Geophysics of SB RAS)  
Rodionov A.S. (Institute of Computational Mathematics and Mathematical Geophysics of SB RAS)  
Rzhetsky A.Y. (University of Chicago, USA)  
Snytnikov V.N. (Boreskov Institute of Catalysis of SB RAS)  
Shokin Y.I. (Institute of Computational Technologies of SB RAS)  
Tsoumpas Ch. (University of Leeds, UK)  
Vityaev E.E. (Sobolev Institute of Mathematics of SB RAS)  
Yagola A.G. (Moscow State University, Moscow)  
Zhang Sh. (Tianjin University of Finance and Economics, China)

## Organizing Committee

### Chair

Kabanikhin S.I.

### Co-Chair

Krivorotko O.I., Marchenko M.A.

### Scientific Secretary

Latyshenko V.A.

Yermolenko D.V.

Kluychinskiy D.V.

Novikov N.S.

Kondakova E.A.

Shishlenin M.A.

Kulikov I.M.

Chernyh I.G.

Pogkolodnyi N.L.

Zvonareva T.A.

## Contacts

Institute of Computational Mathematics and  
Mathematical Geophysics of SB RAS  
630090 Novosibirsk, Lavrentyeva, 6

Tel: +7 (383) 330-83-53

Fax: +7(383) 330-66-87

URL - ICMMG SB RAS:

<https://icmmg.nsc.ru/en/>

Organizing committee:

MM-HPC-BBB@yandex.ru

# Organizers

- Institute of Computational Mathematics and Mathematical Geophysics of SB RAS
- Novosibirsk State University
- Institute of Cytology and Genetics of SB RAS
- Sobolev Institute of Mathematics of SB RAS
- Russian Foundation for Basic Research

# Sponsors



Russian Foundation  
for Basic Research  
Grant No. 18-01-20052 Г



Federal Agency  
for Scientific  
Organizations  
FASO Russia



Ministry of Education and  
Science of the Russian Fed-  
eration (Minobrnauki  
of Russia) Grant No. ДННТ  
28.12487.2018/12.1

## GOLD SPONSORS



## SILVER SPONSORS



MP Biomedicals



АЛЬБИОГЕН

Albiogen



Roche Diagnostics Rus Ltd.



Skygen, Ltd



Bioline, Llc



Khimexpert Ltd.



DIA-M, Ltd

ХИМЭКСПЕРТ

## BASIC SPONSORS



geneXplain GmbH



IOS Press

## The Institute of Computational Mathematics and Mathematical Geophysics SB RAS

The Institute of Computational Mathematics and Mathematical Geophysics SB RAS (ICM&MG SB RAS), former Computing Center of the Siberian Branch of the USSR Academy of Sciences, was founded by the RSFSR Council of Ministers (order no. 1693-p, May 4, 1963) and the Presidium of the USSR Academy of Sciences *наук СССР* order no. 455, May 24, 1963.

Basic research directions of the Institute are:

- Computational Mathematics
- Mathematical Modeling and Methods of Applied Mathematics
- Parallel and Distributed Calculations
- Information Systems

ICM&MG SB RAS has 17 scientific laboratories. The personnel of the Institute is 298 workers (2016), with one Academician of RAS, three Corresponding Members of RAS, 44 Doctors of Science, and 84 Candidates of Science (PhDs).

ICM&MG SB RAS is a known leader in the development of direct and inverse problems of mathematical physics, numerical statistical simulation (Monte Carlo methods), geophysics, physics of the atmosphere, ocean, and environment, chemistry, and electrophysics. The developed algorithms and programs are used to solve important problems of environmental management, explore for oil and gas deposits, predict natural and technogenic disasters and estimate their consequences, perform Earth's sounding from space, and develop efficient supercomputer equipment, in medicine, nanoindustry, and information security.

ICM&MG SB RAS holds 7 scientific seminars. It is a basic institution for 4 departments of Novosibirsk State University and 2 departments of Novosibirsk State Technical University. The institute performs postgraduate teaching in 6 education programs (specialties). ICM&MG SB RAS has 2 Dissertation Councils.

ICM&MG SB RAS has a Super Computing Center SB RAS and Collection of Algorithms and Programs SB RAS.

# Contents

Single-molecular fluorescence spectroscopy in protein folding: a theoretical modeling of multi-color experiments. <i>V.A. Andryushchenko, A.Yu. Palyanov, S.F. Chekmarev</i>	10
Pseudo one-compartment models. Methods for assessing the peripheral compartment for them. <i>N. Asmanova, A.I. Ilin</i>	11
Finding epistasis in high-throughput experimental data. <i>L. Aviño Esteban, N.S. Bogatyreva, F.A. Kondrashov, D.N. Ivankov</i>	12
Identifiability analysis of nonlinear dynamical system. <i>Zh. Bektemessov</i>	13
Application of monte carlo simulations in nuclear medicine imaging. <i>J. Cal-Gonzalez</i>	14
On the construction of the cerebral hemodynamics model based on clinical data. <i>A.A. Cherevko, M.A. Shishlenin, A.K. Khe, E.E. Bord, V.V. Berestov, K.Y. Orlov, V.A. Panarin</i>	15
Siberian supercomputer center as a service for bioinformatics research. <i>I. Chernykh, B. Glinskiy, N. Kuchin, S. Lomakin</i>	16
Fighting celiac disease: improvement of pH-stability of Cathepsin L by computational design. <i>A. Chugunov, D. Nolde, V.F. Tereshchenkova, E.A. Dvoryakova, I.Yu. Filippova, E.N. Elpidina, R. Efremov</i>	17
Inverse problems in tomography: an evolutionary approach. <i>V. Dedok</i>	18
Methods of mathematical modeling in modern diagnostic nuclear medicine. <i>N. Denisova</i>	19
Principal Component Analysis for any type Sequences (PCA-Seq). <i>V. Efimov, K. Efimov, V. Kovaleva</i>	20
Estimates from evolutionary algorithms theory applied to gene design. <i>A. Ereemeev, A. Spirov</i>	21
HEDGE: Highly accurate GPU-powered protein-protein docking pipeline. <i>T. Ermak, A. Shehovtsov, P. Yakovlev</i>	22
Revealing the research institutes and their interactions: a case study of miRNA research. <i>A. Firsov, I. Titov</i>	23
Method of reconstruction of a sequence of non-ribosomal peptides from mass spectra with noise. <i>E. Fomin</i>	24
The performance improvement of the permutation test algorithm for GSEA. <i>M. Grishchenko, A. Yakimenko, M. Khairtdinov, A. Lazareva</i>	25
An inverse problem in modelling of a symmetric gene network regulated by negative feedbacks. <i>V. Golubyatnikov, V. Gradov</i>	26
On cycles in models of asymmetric circular gene networks. <i>V. Golubyatnikov, N. Kirillova</i>	27

On existence of a piecewise smooth cycle in one asymmetric gene network model with piecewise linear equations. <i>V. Golubyatnikov, L. Minushkina</i>	28
Investigation of stopping criterion for OSEM algorithm with application to nuclear medicine. <i>N.V. Denisova, O. Krivorotko</i>	29
A numerical algorithm of parameter identification in mathematical model of tuberculosis transmission with control programs. <i>S.I. Kabanikhin, O.I. Krivorotko, V.N. Kashtanova</i>	30
Simulation and image reconstruction of the combined Siemens PET/CT and PET/MRI systems. <i>H. Kertesz, A. Renner, I. Rausch, T. Beyer, J. Cal-Gonzalez</i>	31
Creation of a modular model of metabolic processes in skeletal muscles during moderate physical load using BioUML platform. <i>I.N. Kiselev, V.I. Baranov, F.A. Kolpakov</i>	32
Population-based mathematical modeling antihypertensive drugs effect using BioUML platform. <i>I.N. Kiselev, A.F. Kolpakova, F.A. Kolpakov</i>	33
Assessment of software for somatic single nucleotide variant identification using simulated whole-genome sequencing data of cancer. <i>W. Kittichotirat, P. Khongthong, K. Kusonmano, S. Cheevadhanarak</i>	34
Spatial heterogeneity influences evolutionary scenarios in microbial communities explained by ecological stratification: a simulation study. <i>A.I. Klimenko, Yu.G. Matushkin, S.A. Lashin</i>	35
Different effects of agroclimatic factors on time to emergence and time to flowering in nine soybean accessions. <i>K. Kozlov, L. Novikova, I. Seferova, S. Nuzhdin, M. Samsonova</i>	36
The optimal control of stochastic differential equations arising in biology, economy and finance. <i>E. Kondakova, O. Krivorotko, S. Kabanikhin</i>	37
Supercomputer analysis of social, epidemiological and economic processes. <i>O. Krivorotko</i>	38
High performance computing in astrophysics. The organic formation in protostellar disc. <i>I. Kulikov</i>	39
Genome-scale modeling of carbon assimilation in <i>Geobacillus icigianus</i> . <i>M. Kulyashov, I. Akberdin, A. Rozanov, S. Peltek</i>	40
Agent-based modelling of genetic deafness propagation under various sociodemographic conditions. <i>S.A. Lashin, Yu.G. Matushkin, A.A. Smirnova, G.P. Romanov, O.L. Posukh</i>	41
Identifiability analysis of mathematical models of immunology and epidemiology. <i>V. Latyshenko, O. Krivorotko, S. Kabanikhin</i>	42
Parameters sensitivity of pharmacokinetics model parameters. <i>V. Lifenko, D. Voronov</i>	43
Bayesian approach to big data processing: problems and perspectives. <i>M.A. Marchenko</i>	44
The <i>Multiplex Phase Interlocker</i> : a novel and robust molecular design synchronizing transcription and cell cycle oscillators. <i>T.D.G.A. Mondeel, C. Linke, S. Tognetti, W. Liebermeister, M. Loog, H.V. Westerhoff, F. Posas, M. Barberis</i>	45



Developing FoldGO, the tools for multifactorial functional enrichment analysis. <i>A.M. Mukhin, D.S. Wiebe, I. Grosse, S.A. Lashin, V.V. Mironova</i>	46
Mathematical modeling of medicinal preparations diffusion process in tissues of the person. <i>A. Nafikova</i>	47
The possibilities of a Universal computer model in the readiness assessment of the Russian regions resource to epidemics of especially dangerous infectious diseases. <i>L. Nizolenko, A. Bachinsky</i>	48
The 2D coefficient inverse problem of the ultrasound waves propagation. <i>N. Novikov, M. Shishlenin</i>	49
The optimal feedbacks in the mathematical model of chemotherapy for a nonmonotonic therapy function. <i>N. Novoselova</i>	50
Mathematical phantoms development for computer simulation of the patient examination procedure by a positron emission tomography method. <i>M. Ondar, N. Denisova</i>	51
DEPPDB v.3: a portal to study electrostatic and other physical properties of genome DNA and its elements. <i>A. Osypov, G. Krutinin, E. Krutinina, P. Beskaravayny, S. Kamzolova</i>	52
Complex information system to study common energy metabolic deficiency under neurodegenerative diseases. <i>A. Osypov, I.Yu. Popova</i>	53
An algorithm for tracking <i>C. elegans</i> body movement and muscular activity in $\text{Ca}^{2+}$ dynamics video for tuning and validation of its locomotion simulation. <i>A.Yu. Palyanov</i>	54
Inverse modelling of diffusion-reaction processes with image-type measurement data. <i>A. Penenko, Z. Mukatova, S. Nikolaev, U. Zubairova</i>	55
The use of Kirlian photography in preventive medicine and the education. <i>L.A. Pesotskaya, T.V. Lakiza, N.V. Glukhova, T.O. Tretiak</i>	56
Computer system for reconstructing and analyzing random structural models of protein-protein interaction networks. <i>N.L. Podkolodnyy, D.A. Gavrilov, O.A. Podkolodnaya</i>	57
Circadian rhythms: data analysis and mathematical modeling. <i>N.L. Podkolodnyy, N.N. Tverdohkleb, O.A. Podkolodnaya</i>	58
Digital heart: personalized medicine and inverse problems. <i>A. Prikhodko, M. Shishlenin</i>	59
Mathematical model of membrane potential formation at <i>E. coli</i> growth on nitrite. <i>N.A. Ree, V.A. Likhoshvai, T.M. Khlebodarova</i>	60
The uniqueness of the solution of the two-dimensional direct problem is the propagation of the action potential along the nerve fiber. <i>A.J. Satybaev, G.S. Kurmanalieva</i>	61
Mathematical models of p53–microRNA and their applications. <i>S.D. Senotrusova, O.F. Voropaeva</i>	62
An effective subgradient method for simultaneous restoration and segmentation of blurred images. <i>T. Serezhnikova</i>	63
The software and database for Vertebrate imperfect mtDNA repeats annotation. <i>V.A. Shamanskiy, K.Yu. Popadin, K.V. Gunbin</i>	64
Inverse and Ill-Posed problems for nonlinear PDE: applications to life and social sciences. <i>M. Shishlenin, D. Lukyanenko</i>	65

Deep bioinformatics expert system of analysis, modeling and interpretation of omics BigData of the human genome. <i>A. Shlikht, N. Kramorenko</i>	66
Asymptotic stability of solutions in one model of disease. <i>M.A. Skvortsova</i>	67
Algorithm for solving the inverse problem of pharmacokinetics to determine the transition coefficients. <i>A. Takuadina</i>	68
Comparison of quality of automated gene network reconstruction using connectivity of random and functional networks. <i>E. Tiys, P. Demenkov, V. Ivanisenko</i>	69
Chaos theory as a bioinformatics promissory instrument for a human organism systemic response in-depth study. <i>B.G. Vainer, A.V. Shepelin</i>	70
ARGO_CEL: GPU based approach for potential composite elements discovery in large DNA datasets. <i>O. Vishnevsky, A. Bocharnikov, N. Kolchanov</i>	71
Teaching medicine and biology through systems biology. <i>H.V. Westerhoff</i>	72
FoldGO for functional annotation of transcriptome data to identify fold-change-specific GO categories. <i>D.S. Wiebe, A.M. Mukhin, N.A. Omelyanchuk, V.V. Mironova</i>	73
Investigation and numerical solving of a mathematical model of intracellular HIV dynamics: from ODE to PDE. <i>D. Yermolenko, O. Krivorotko, S. Kabanikhin</i>	74
Inverse problems for mathematical models in social networks: from PDE to SDE. <i>Sh. Zhang, S. Kabanikhin, O. Krivorotko, Yu. Wang</i>	75
Gene network analysis of complex diseases using GenCoNet. <i>O. Zolotareva, A. Shoshi, R. Hofestädt, A. Maier, V. Ivanisenko, V. Dosenko, E. Bragina</i>	76
Inverse problem for partial differential equations in social networks. <i>T. Zvonareva, O. Krivorotko, S. Kabanikhin</i>	77
Agent-based modelling of genetic deafness propagation under various sociodemographic conditions. <i>S.A. Lashin, Yu.G. Matushkin, A.A. Smirnova, G.P. Romanov, O.L. Posukh</i>	78
Author index	79

# Single-molecular fluorescence spectroscopy in protein folding: a theoretical modeling of multi-color experiments

V.A. Andryushchenko<sup>1,2</sup>, A.Yu. Palyanov<sup>1,3</sup>, S.F. Chekmarev<sup>1,2\*</sup>

<sup>1</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup> Institute of Thermophysics, SB RAS, Novosibirsk, Russia

<sup>3</sup> Institute of Informatics Systems, SB RAS, Novosibirsk, Russia

\* e-mail: chekmarev@itp.nsc.ru

**Key words:** protein folding, single-molecular fluorescence spectroscopy, molecular dynamics, collective variables, free energy surfaces

*Motivation and Aim:* The single-molecular fluorescence spectroscopy methods, such as the Förster Resonance Energy Transfer (FRET) and Photoinduced Electron Transfer (PET), have become a powerful tool to study protein folding. Currently, the donor and acceptor are typically positioned at the ends of the protein chain [1, 2]. The results of the measurement are presented in the form of one-dimensional (1D) free energy profile along a reaction coordinate connecting the unfolded and native states of the protein, which allows one to see how the protein folds. However, since the protein usually follows a variety of essentially different folding pathways, in which case the folding kinetics are often very complex, such 1D profiles do not give a reasonably complete description of the folding process. At the same time, the donor and acceptor can be placed not only at the ends of the protein but also within the protein chain, so that a multi-color signal coming from different mutual positions of a set of donors and acceptors can be recorded [3]. In this case, two-dimensional (2D) free energy surfaces (FESs) can be constructed, which provide incomparably richer information about the folding process than the 1D profiles do. In the present work, using molecular dynamics simulations, we examine what information can be obtained if the fluorescence signal is monitored for two sets of donors and acceptors, and how the picture of folding thus obtained is complete in comparison to an “ideal” choice of collective variables to characterize the folding process.

*Methods and Algorithms:* Since our goal was to understand the situation in general, we used a coarse-grained protein representation, i. e., each protein residue was represented by a bead placed at the position of the C $\alpha$ -atom. The simulations were performed with molecular dynamics methods. Using the commonly employed collective variables, such as the radius of gyration and the RMSD from the native state, a “theoretical” FES was constructed for each protein, which was supposed to give a best representation of the folding process. To construct the corresponding “experimental” FES, two pairs of donors and acceptors were chosen, for which the characteristics relevant to the FRET and PET experiments were monitored. In the former case, the distances between the donors and acceptors were calculated, and in the latter case, the moments when the donors and acceptors approached each other at the Van-der-Waals radius. The donors and acceptors were represented by selected residues (beads).

*Results:* Folding of a set of proteins, including BBL domain in application to the PET experiments [1], and GB1 protein and WW domain in application to the FRET experiments [2], has been studied. The comparison of the “experimental” and “theoretical” FESs has shown that even in the case of the simplest, two-state folding kinetics, a 1D free energy profile does not necessarily distinguish the protein states. In the case of more complex kinetics, at least two pairs of donors and acceptors placed in specific positions are required to reveal the essential states and represent the kinetics.

*Acknowledgement:* This work was supported by Russian Foundation for Basic Research, grant No. 18-04-00013.

## References

1. Neuweiler H., Johnson C.M., Fersht A.R. (2009) Proc. Natl. Acad. Sci. USA 106:18569.
2. Chung H.S., McHale K., Louis J.M., Eaton W.A. (2012) Science 335:981.
3. Lerner E., Cordes T., Ingargiola A., Alhadid Y., Chung S.Y., Michalet X., Weiss S. (2018) Science 359:288.

## Pseudo one-compartment models. Methods for assessing the peripheral compartment for them

N. Asmanova\*, A.I. Ilin

JSC "Scientific center for anti-infection drugs", Almaty, Kazakhstan

\* e-mail: asmanova@inbox.ru

**Key words:** pseudo one-compartment model, first order absorption, direct and inverse problems in pharmacokinetics

*Motivation and aim:* If  $k_a = k_{21}$ , then the equation of PK curve for two-compartment model with first order absorption (2ev) is transformed from eq. (1) into (2). In solving inverse problems, it is identified only as one-compartment model-1ev (3), where  $k_a > k_{10}$  [1, 2]:

$$C_1 = A_1 e^{-\alpha t} + A_2 e^{-\beta t} - (A_1 + A_2) e^{-k_a t} \quad (1)$$

$$C_1 = A_2 (e^{-\beta t} - e^{-\alpha t}) \quad (2)$$

$$C_1 = A (e^{-k_{10} t} - e^{-k_a t}) \quad (3)$$

So, actual loss of the term  $-(A_1 + A_2) e^{-k_a t}$ , associated with absorption, turns into a fictitious disappearance of the distribution phase. The reason for this problem is the ambiguous role of the volume of distribution of the drug, in practice of pharmacokinetics it is not measured, but is calculated.

*Methods:* The analysis of eqs. (1–3) and the solution of direct and inverse problems for them.

*Results:* It is shown that an approximate evaluation of the peripheral compartment can be obtained with the help of the parameters of the inversion line [1], relation (4) and their comparison with intravascular (iv) bolus data (5). Here  $AUC_1$  and  $AUC_2$  are the areas under the PK curves of the central and peripheral compartments,  $k_a$  and  $k_{10}$  are absorption and elimination rate constants respectively.

$$AUC_{2(k_a < k_{10})} > AUC_{2(k_a = k_{10})} > AUC_{2(k_a > k_{10})} \quad (4)$$

$$AUC_2 / AUC_1)_{ev} \approx AUC_2 / AUC_1)_{iv} \quad (5)$$

*Conclusion:* The inverse problem for eq. (2) is solvable only for a known volume, but there are no methods for determining it. Equation (3) is its pseudo one-compartment version, so the calculation of the drug dosage regimen on its basis is incorrect.

### References

1. Asmanova N., Ilin A.I. (2016) Futures of solution of some inverse problems in PK models with first order absorption. Abstracts: International Conference Mathematical modeling and High Performance Computing in Bioinformatics, Biomedicine and biotechnology MM-HPC-BBB-2016, Novosibirsk, Russia, 29 August – 2 September, p. 15.
2. Asmanova N., Ilin A.I. (2016) Flip-flop phenomenon in two-compartment model with first order absorption, *ibid*, p. 16.

## Finding epistasis in high-throughput experimental data

L. Aviño Esteban<sup>1</sup>, N.S. Bogatyreva<sup>1, 2, 3</sup>, F.A. Kondrashov<sup>4</sup>, D.N. Ivankov<sup>3, 4\*</sup>

<sup>1</sup> *Universitat Pompeu Fabra (UPF), Barcelona, Spain*

<sup>2</sup> *Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG), Barcelona, Spain*

<sup>3</sup> *Laboratory of Protein Physics, Institute of Protein Research of the RAS, Pushchino, Moscow region, Russia*

<sup>4</sup> *Institute of Science and Technology, Klosterneuburg, Austria*

\* e-mail: ivankov13@gmail.com

**Key words:** epistasis, fitness, higher-order epistasis, multi-dimensional epistasis

*Motivation and Aim:* Epistasis is one of the most important factors of molecular evolution. Epistasis in its simplest form stands for a phenomenon when the fitness of double mutant differs from the fitness expected from the two single mutants [1]. For higher-order epistasis, we look for the deviation between the fitness of multiple mutant and the fitness expected from all the mutants of lower order [2]. Another concept in protein fitness landscapes is multi-dimensional epistasis. This is the type of epistasis when experimental data cannot be fitted by a monotonic function of fitness potential, the linear combination of contributions from single amino acid substitutions [3]. To analyze epistasis, we have to find hypercubes either in two-dimensional space or in a higher-dimensional space. Different designs of experiments can produce combinatorially complete datasets of genotypes [2] or much bigger datasets where nucleotide variants are generated randomly [1].

*Methods and Algorithms:* Three algorithms were designed and implemented to obtain the results.

*Results:* First, in the presented work we find all hypercubes in the random mutagenesis dataset of yeast protein HIS3 [4]. For more than 700 thousand measured phenotypes we found more than 170 millions hypercubes, the biggest dataset available so far. Next, we realize here an idea that genotypes can be searched at any distance. Thus, we can investigate epistasis in hyperrectangles, not only in hypercubes. Using this approach, we found much more rectangles in genotype space than squares. And last, we present here a completely new type of multi-dimensional epistasis when two groups of four genotypes fit unidimensional picture individually but not simultaneously. In the presented work we elucidated all >20000 cases when the multi-dimensional epistasis of that kind can occur in the experimental data of GFP [1].

*Conclusion:* Overall, the methods presented here have practical importance for the analysis of fitness landscapes.

*Acknowledgements:* Supported by the HHMI International Early Career Scientist Program [55007424], the MINECO [BFU2015-68723-P], Spanish Ministry of Economy and Competitiveness Centro de Excelencia Severo Ochoa 2013-2017 [grant SEV-2012-0208], Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat's AGAUR [program 2014 SGR 0974], and the European Research Council under the European Union's Seventh Framework Programme [FP7/2007-2013, ERC grant agreement 335980\_EinME].

### References

1. Sarkisyan K.S. et al. (2016) Local fitness landscape of the green fluorescent protein. *Nature* 533:397- 401.
2. Poelwijk F.J. et al. (2016) The context-dependence of mutations: a linkage of formalisms. *Plos Comp. Biol.* 12:e1004771.
3. Kondrashov F.A., Kondrashov A.S. (2001) Multidimensional epistasis and disadvantage of sex. *PNAS* 98:12089-12092.
4. Pokusaeva V. et al. (2017) Experimental assay of a fitness landscape on a macroevolutionary scale. *BioRxiv*:

# Identifiability analysis of nonlinear dynamical system

Zh. Bektemessov

al-Farabi Kazakh National University, Almaty, Kazakhstan

\* e-mail: jolaman252@gmail.com

**Key words:** practical identifiability, dynamical system, parameter estimation, inverse problems

*Motivation and Aim:* As it is known mathematical modeling plays a huge role in the research of various scientific areas of our life, so ordinary differential equations are a powerful tool for modeling the dynamic processes of biomedicine, especially in modeling the processes of pharmacokinetics, epidemiology and immunology. In practice, it is necessary to determine unknown parameters in the ODE models on the basis of experimental data. Identifiability analysis is the first step in determining them.

*Methods and Algorithms:* There was considered different approaches of identifiability analysis like drawing contours of the cost functions of least squares ( $J_{lsq}$ ) or (log-) likelihood functions ( $J_{llk}$ ) by pairs of parameters. This will help to determine strong correlation between parameters, non-identifiability for some parameters if the contours extend to infinity. Another approach is the Crammer-Rao inequality based on the relationship between so called Fisher Information Matrix and the covariance matrix. The robust identifiability gives the Monte Carlo based sampling method, which simulates the possibility of performing hundreds of replicates of the same experimental scheme for a given experimental error. Also to solve the inverse problem and restore unknown parameters by the additional information such as experimental data, the algorithm of differential evolution was used.

*Results:* For the complex two-chamber kinetic model of the C-peptide model with four observables and 8 unknown parameters the mentioned above methods were applied and the next results like lack of identifiability for some of parameters, presence of optimal solutions and good restoration of parameters were obtained.

*Conclusion:* The results obtained in the model suggest that only two parameters showed practical identifiability, while other parameters were structural and two others illustrated strong correlation and weak identifiability.

*Acknowledgements:* Supported by the grant of the Ministry of Education and Science of the Republic of Kazakhstan (project No. AP05134121 “Numerical methods of identifiability of inverse and ill-posed problems of natural science”)

## References

1. Воронов Д.А., Гроздь А.А. (2014) Идентифицируемость динамических систем на примере моделей фармакокинетики и иммунологии. Новосибирск: Сибирские электронные математические известия. 11:94-104.
2. Cobelli C., Romanin-Jacur G. (1976) Controllability, observability and structural identifiability of multi input and multi output biological compartmental systems. IEEE Trans Biomed Eng 93-100 pp.
3. Kabanikhin S.I., Voronov D.A., Grodz A.A., Krivorotko O.I. Identifiability of mathematical models in medical biology. Vavilovskii Zhurnal Genetiki i Selektcii = Vavilov Journal of Genetics and Breeding. 19(6):738-744. DOI 10.18699/VJ15.097 (in Russian)

## Application of monte carlo simulations in nuclear medicine imaging

J. Cal-Gonzalez

*QIMP team, Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Austria*

*e-mail: jacobocalgonzalez@meduniwien.ac.at*

**Key words:** nuclear medicine, Monte Carlo simulation, PET, SPECT

Monte Carlo (MC) methods consist of a very broad area of science, in which many processes, physical systems and phenomena are simulated by statistical methods. Nowadays, MC methods are widely used to solve complex physical and mathematical problems, particularly those involving multiple independent variables where more conventional numerical methods would demand formidable amounts of memory and computer time. In this context, nuclear medical imaging techniques, such as Single-Photon Emission Computed Tomography (SPECT) or Positron Emission Tomography (PET), are ideal for MC methods due to the stochastic nature of radiation emission, transport and detection processes.

This presentation will provide an overview on the different applications of MC simulation techniques in PET and SPECT imaging; from the characterization of existing imaging systems to the design and optimization of new scanners and the evaluation of advanced image reconstruction and data processing techniques.

*Acknowledgements:* Supported by the FWF (I3451-N32) and by RFBR (grant 17-52-14004).



# On the construction of the cerebral hemodynamics model based on clinical data

A.A. Cherevko<sup>1,4\*</sup>, M.A. Shishlenin<sup>2,3,4</sup>, A.K. Khe<sup>1,4</sup>, E.E. Bord<sup>4</sup>, V.V. Berestov<sup>5</sup>, K.Y. Orlov<sup>5</sup>, V.A. Panarin<sup>5</sup>

<sup>1</sup>*Lavrentyev Institute of Hydrodynamics of SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Sobolev Institute of Mathematics of SB RAS, Novosibirsk, Russia*

<sup>3</sup>*Institute of Computational Mathematics and Mathematical Geophysics, Novosibirsk, Russia*

<sup>4</sup>*Novosibirsk State University, Novosibirsk, Russia*

<sup>5</sup>*Meshalkin national medical research center, Novosibirsk, Russia*

\* e-mail: cherevko@mail.ru

**Key words:** hemodynamics, neurosurgery, arterial aneurism, nonlinear oscillator, inverse problem, gradient method

Currently, the monitoring of the hemodynamics of the brain is being implemented by neurosurgeons of the National Medical Research Center of Academician E. Meshalkin in collaboration with colleagues from the Lavrentyev Institute of Hydrodynamics.

This material, which is unique in the world practice, made it possible to approach the construction of a mathematical model of hemodynamics.

As the model, the nonlinear oscillator equation was chosen. In this equation, the velocity is a “governance” function (the right-hand side of the equation), and the second-order differential operator acting on the pressure. The “blood flow – vessel – brain substance” system is nonlinear and has both elastic and damping properties, for this reason the model of generalized Van der Pol–Duffing equation was suggested to identify the characteristic behavior of hemodynamic parameters in the surroundings of vascular pathologies. Equation coefficients characterize individual living system of the patient, the measurement location, the presence of anomalies. The coefficients of this equation are individual for each patient.

We solve coefficient inverse problem to determine the coefficients of this model by known clinical intraoperational data. This model adequately describes the behaviour of hemodynamic parameters.

We investigate and construct numerical method for solving the coefficient inverse problem for essentially nonlinear ODE by clinical data of neurosurgical operation.

We recover the coefficients by clinical data obtained during neurosurgical operation in vicinity of arterial aneurysm, that a pathological enlargement (dilation) of the artery.

The proposed model and the method for solving the inverse problem together allowed us to restore the behavior of pressure in the vicinity of intracranial vascular pathology, having data on the blood flow velocity in the “real” time. Investigation of the dependence of pressure on velocity in blood vessels is of great practical importance, since there are currently non-invasive methods for measuring speed (tomography, ultrasound), but no non-invasive methods of measuring pressure. At the same time, information about pressure is important.

We study the relationship between the properties of this equation and the state of the vascular bed.

*Acknowledgements:* Supported by RFBR (projects No. 17-08-01736) and MSC RK grant AP05134121.



## Siberian supercomputer center as a service for bioinformatics research

I. Chernykh, B. Glinskiy, N. Kuchin, S. Lomakin

*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*

\* e-mail: [chernykh@ssd.sccc.ru](mailto:chernykh@ssd.sccc.ru)

**Key words:** high performance computing, GPU, bioinformatics and life sciences

*Introduction:* Sequencing and protein docking are very compute-intensive tasks that see a large performance benefit by using the latest HPC hardware. At this moment there are a lot of bioinformatics codes which are optimized for the latest Intel HPC hardware [1]. Siberian Supercomputer Center (SSCC) has new HPC cluster with total peak performance ~91 TFlops. This system is well designed for bioinformatics researches due to the using Intel Xeon Phi (KNL architecture) CPUs as well as Intel Optane technology for extending memory size on Intel CPU node.

*Siberian Supercomputer Center resources:* SSCC offers computer resources for bioinformatics researches to its users. Our main system NKS-1P consists of 40 Intel Xeon E5-2697v4 (2.6 GHz, 16 cores) and 16 Intel Xeon Phi 7290 KNL (1.5 GHz, 72 cores, 16 GB MCDRAM) CPUs. Intel Xeon E5-2697v4 CPU nodes have 128 GB DRAM, Intel Xeon Phi 7290 nodes have 96 GB DRAM. For bioinformatics problems, we have 2x 375GB Intel Optane memory which is working as IMDT on Broadwell node. HPC nodes and 200TB Intel Lustre PFS are working on Intel OmniPath 100 Gb/s interconnect. We also have the supercomputer with a hybrid architecture and consists of NKS-30T (platform BL2h220c hp) system with 576 Intel Xeon processors E5450/E5540/X5670 (2688 cores) and hybrid cluster that based on 40 servers HP SL390s G7 (80x CPU X5670 – 480 cores) with 3x NVidia Tesla M2090 GPU on each node. All cluster nodes are connected via Infiniband QDR network interface. Cluster file system IBRIX (4 servers, 32 TB of available disk space) is also connected by Infiniband interface for NKS-30T. The newest part of SSCC resources is based on [2]. This architecture is well suited for open source packages like MUMmerGPU: High-through DNA sequence alignment using GPUs [3], Parallel-META: a GPU- and multi-core-CPU-based open-source pipeline for metagenomic data analysis, which enabled the efficient and parallel analysis of multiple metagenomic datasets [4], and Molecular Dynamics packages like GROMACS [5], LAMMPS [6]. All these resources are available to all organizations that are operated by the Federal Agency of Scientific Organizations Russia.

### References

1. <https://www.intel.com/content/www/us/en/healthcare-it/solutions/genomicscode.html>
2. <http://www.sccc.icmmg.nsc.ru/hardware.html>
3. <https://sourceforge.net/projects/mummergpu/>
4. Su X., Xu J., Ning K. (2011) Parallel-META: A high-performance computational pipeline for metagenomic data analysis. 2011 IEEE International Conference on Systems Biology (ISB), Zhuhai. P. 173-178.
5. <http://www.gromacs.org/>
6. <http://lammmps.sandia.gov/>
7. Kzantsev F. et al. (2008) Proc. of the 6<sup>th</sup> International Conference on BGRS. P. 113.
8. Likhosvai V. et al. (2001) Generalized chemokinetic method for gene network simulation. Mol Biol. 35:1072-1079.

## Fighting celiac disease: improvement of pH-stability of Cathepsin L by computational design

A. Chugunov<sup>1,2\*</sup>, D. Nolde<sup>2</sup>, V.F. Tereshchenkova<sup>3</sup>, E.A. Dvoryakova<sup>4</sup>,  
I.Yu. Filippova<sup>3</sup>, E.N. Elpidina<sup>4</sup>, R. Efremov<sup>1,2</sup>

<sup>1</sup> National Research University Higher School of Economics, Moscow, Russia

<sup>2</sup> M.M. Shemyakin & Yu.A. Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia

<sup>3</sup> Chemical Faculty and <sup>4</sup> A.N. Belozersky Institute of Physico-Chemical Biology of M.V. Lomonosov Moscow State University, Moscow, Russia

\* e-mail: batch2k@yandex.ru

**Key words:** cathepsin L, pH-stability, molecular dynamics, protein design, *in silico* mutations

**Motivation and Aim:** Celiac disease is genetically predisposed autoimmune disorder that is caused by inflammatory response to prolamins – storage proteins of cereal seeds. Several prolamins peptides, resistant to proteolysis by human digestive enzymes, cause chronic diarrhea, abdominal distention, and even cancer and early death in susceptible human population. The common treatment is a strict wheat-, rye- and barley-free diet, known as gluten-free, which is costly and difficult to maintain.

We suggest to help celiac patients by oral treatment with enzyme that is able to effectively hydrolyze the toxic prolamins peptides – cysteine peptidase cathepsin L from a beetle *Tribolium castaneum* (TcCathL). However, this enzyme is active at pH > 3, while the use in human stomach requires it to be active at pH's as low as 2. In this work, we aimed to improve TcCathL pH-stability by *in silico* mutagenesis and computational assessment of candidate mutant variants.

**Methods and Algorithms:** We built a 3D homology model of TcCathL and its point mutants, and assessed their stability and dynamic features by molecular dynamics (MD) simulations in water at pH values 2 and 7, modeled as different ionization states of particular amino acid residues. Total MD time for all systems exceeded 5  $\mu$ s. Processing of MD data included RMSD/RMSF calculations, analysis of intermolecular contacts, secondary structure elements stability, rotameric states of catalytic residues, etc.

**Results:** The major feature that distinguished TcCathL in acidic/neutral medium was structure and dynamics of the “catalytic triad”: Cys-138, His-275 and Asn-295, namely – the rotameric state of His-275, which reproducibly “turned away” from the active site in multiple MD trajectories at pH 2. This peculiarity may be the cause of the loss of the activity at acidic conditions.

Next, we introduced several *in silico* point mutations in the vicinity of His-275 in order to fix its side chain in the “active” conformation by introduction of the novel hydrogen bond, and assessed these enzyme variants by MD. Several “designed” mutants of adjacent to His-275 residues exhibited the intended behavior, and were passed to the experimental verification.

**Conclusion:** By the computational design we suggested TcCathL mutant variants that may possess increased activity at pH 2. If so, these bioengineered enzymes become a basis for prototypic celiac disease treatment.

**Acknowledgements:** This work was supported by the Molecular and Cell Biology Program of the Russian Academy of Sciences, by RFBR-National Intellectual Development grant No. 17-34-80158 mol\_ev\_a and within the framework of a subsidy by the Russian Academic Excellence Project “5-100”. Access to computational facilities of the Supercomputer Center “Polytechnical” at the St. Petersburg Polytechnic University is greatly appreciated.

# Inverse problems in tomography: an evolutionary approach

V. Dedok

*Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*

*e-mail: dedok@math.nsc.ru*

**Key words:** tomography, inverse problems, genetic algorithms

*Motivation and Aim:* A lot of inverse problems in tomography may be reduced to inverse kinematic problem. In this kind of inverse problem, we assume to know a wave travel time between each pairs of points in the boundary of discovered domain. If a discovered domain with unknown internal structure has a cube form with  $n^3$  elementary cubes we have  $O(n^5)$  traces. This large amount of source data makes the problem too hard to solve. Moreover, in practice a wave travel time is unknown, we deal with the phaseless intensity of scattered wave. In this paper we present an effective method of solving of the inverse kinematic problem based on evolutionary genetic algorithms.

*Methods and Algorithms:* Mathematically the inverse problem is formulated the following way. Consider a domain of cube form divided into  $n^3$  elementary cubes with constant refractive index. The problem is to find unknown refractive index in each elementary cube using travel time  $\tau_*(x, y)$  between any points on the board of the domain. To get rid of phaseless data we use the method, introduced in [1].

Numerically we need to construct a set of refractive indexes which corresponds to the minimum of residual functional  $E = (\tau(x, y) - \tau_*(x, y))^2$ . We use a genetic algorithm to find this minimum.

Genetic operations are:

- crossover – average genetic code between two items;
- mutation – random change of genetic code.

The termination condition is a combination of minimum criteria and limited number of generations.

*Results:* We test our numerical method on computationally simulated data. Numerical studies of the genetic algorithm show its effectiveness on model cases. For the test cases we use homogeneous medias with some spherical heterogeneities with different refractive indexes. The method demonstrates pretty well reconstruction of unknown media.

*Conclusion:* We show that the genetic algorithms may be an effective method for inverse problem solving. It shows its effectiveness in discovered tomography problem. Unlike traditional optimization methods the genetic algorithm requires fewer computations than the gradient methods. Also, it allows to use undifferentiable functionals like  $|\tau(x, y) - \tau_*(x, y)|$  and find solution in different metrics.

*Acknowledgements:* Supported by the RFBR (No. 17-01-00120), by interdisciplinary integration project «Inverse problems of mathematical physics and their applications» SB RAS (No. 0314-2018-0003).

## References

1. Romanov V.G. (2017) The problem of recovering the permittivity coefficient from the modulus of the scattered electromagnetic field. *Siberian Mathematical Journal*. 58(4):711-717.

## Methods of mathematical modeling in modern diagnostic nuclear medicine

N. Denisova

*Institute of Theoretical and Applied Mechanics SB RAS, Novosibirsk, Russia*

*\* e-mail: NVDenisova2011@mail.ru*

**Key words:** nuclear medicine, positron emission tomography (PET), single photon emission computer tomography

*Motivation and Aim:* The methods of positron emission tomography (PET) and Single Photon Emission Computer Tomography (SPECT) are widely used for diagnostics in a modern medicine. The aim of this work is a developing of the mathematical modeling method in diagnostic nuclear medicine. The mathematical modeling and computer simulation are playing an increasingly important role in nuclear medicine.

*Methods:* Modeling of SPECT and PET imaging includes three basic components: 1) mathematical models of the activity distribution and attenuation map; 2) data acquisition models; 3) reconstruction algorithms and methods. In this work, the examples of modelling in nuclear cardiology, oncology and neurology are presented. Mathematical models describing the distribution of radiopharmaceuticals in a torso (cardiology), in a brain (neurology) and in a liver (oncology) were developed and used in numerical simulations.

*Results:* The results of numerical simulations in cardiology allowed us to understand the causes of apical artifacts in reconstructed images of myocardial left ventricle. The results of numerical modeling in oncology and neurology have demonstrated the possible directions for improving reconstruction algorithms and methods.

*Conclusion:* Mathematical modeling and computer simulations can effectively add clinical researches.

*Acknowledgements:* The work is supported by RFBR (grant No. 17-52-14004).

# Principal Component Analysis for any type Sequences (PCA-Seq)

V. Efimov<sup>1, 2, 3, 4\*</sup>, K. Efimov<sup>5</sup>, V. Kovaleva<sup>2</sup>

<sup>1</sup> Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

<sup>2</sup> Institute of Systematics and Ecology of Animals SB RAS, Novosibirsk, Russia

<sup>3</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>4</sup> Tomsk State University, Tomsk, Russia

<sup>5</sup> Moscow Institute of Physics and Technology (State University), Moscow, Russia

\* e-mail: efimov@bionet.nsc.ru

**Key words:** time series, PCA, PCo, SSA, molecular sequences

**Motivation and Aim:** In the 40s of the last century, Karhunen and Loève proposed a method for processing a one-dimensional numerical time series by a multidimensional method of principal components. In the 1980s, Takens showed in fact that this method makes it possible to obtain an attractor and, accordingly, phase portraits of the dynamic system from observing only one variable of this system [1]. The method was independently arisen and applied in practice, including by us for the analysis of the animals abundance dynamics [2, 3], and other [4]. The method can be extended for a sequence of any type elements, including numbers, symbols, figures, etc. and, as a special case, for molecular sequences. It is the point of this abstract.

**Methods and Algorithms:** Let there be a sequence  $X = \{x_1, x_2, \dots, x_N\}$  of any type elements. Choose a lag  $L$ ,  $N > L > 1$ . Denote by  $X_i$  the fragment  $X$  of length  $L$  terminated by the element  $x_i$ ,  $X_i = (x_{i-L+1}, x_{i-L+2}, \dots, x_{i-1}, x_i)$ ,  $N \geq i \geq L$ . Compute the matrix of Euclidean distances  $D = (d_{ij} = d(X_i, X_j))$  between all fragments (this is always possible, for example, using the number of unmatched elements, but not only). Apply the method of principal coordinates to the  $D$  and obtain the principal components of it [5]. Call this method PCA-Seq.

**Results:** The amino acid sequence of the *Homo sapiens* *Cytb* gene (AFJ22730.1, GenBank) was processed by PCA-Seq with parameters  $N = 380$ ,  $L = 8$ . The root of the  $p$ -distance is used as the Euclidean distance. The first component (18.2 % of the common variance) clearly reflects the content of Leucine in each fragment and manifest the evident cyclicity, which is most likely determined by the secondary structure of the *Cytb* protein. Jacobi 4 package was used for calculations [6].

**Conclusion:** PCA-Seq is promising for processing molecular sequences, but not only.

**Acknowledgements:** Supported by budget project (No. 0324-2018-0017).

## References

1. Takens F. (1981). Detecting strange attractors in turbulence. In Dynamical systems and turbulence, Warwick 1980 (pp. 366-381). Springer, Berlin, Heidelberg.
2. Efimov V.M., Galaktionov Y.K. (1983) On the possibility of predicting cyclic changes in the abundance of mammals. Zh. Obshch. Biol. (3):343-352. (in Russian)
3. Efimov V.M., Galaktionov Y.K., Shushpanova N.F. (1988). Analysis and prediction of time series by the principal component method. Novosibirsk: Nauka. 70p. (in Russian)
4. Golyandina N., Nekrutkin V., Zhigljavsky A.A. (2001) Analysis of time series structure: SSA and related techniques. Chapman and Hall/CRC.
5. Gower J.C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika, 53(3/4):325-338.
6. Polunin D.A., Shtaiger I.A., Efimov V.M. (2014) Development of software system JACOBI 4 for multivariate analysis of microarray data, Vestnik NSU. Information Technology. 12(2):90-98. (in Russian)

# Estimates from evolutionary algorithms theory applied to gene design

A. Ereemeev<sup>1,2\*</sup>, A. Spirov<sup>1,3</sup>

<sup>1</sup> *The Institute of Scientific Information for Social Sciences RAS, Moscow, Russia*

<sup>2</sup> *Omsk Branch of Sobolev Institute of Mathematics SB RAS, Omsk, Russia*

<sup>3</sup> *The I.M. Sechenov Institute of Evolutionary Physiology and Biochemistry RAS, St. Petersburg, Russia*

\* e-mail: [eremeev@ofim.oscsbras.ru](mailto:eremeev@ofim.oscsbras.ru)

**Key words:** runtime analysis, SELEX procedure, Royal Road function, binding site, promoter, in silico gene design, synthetic biology

*Motivation and Aim:* The field of evolutionary algorithms (EAs) emerged in the area of computer science as a transfer of ideas from biology and developed independently for several decades, enriched with techniques from probability theory, complexity theory and optimization methods. Our aim is to consider how some recent results in theory of EAs may be transferred back into biology.

*Results:* It has been noted in [1] that the EAs optimizing Royal Road fitness functions may be considered as models of evolutionary search for the gene promoter sequences “from scratch”. Here we consider the main known approaches to design the synthetic promoters from the EAs methodology viewpoint. This is the problem to find a tight cluster of the supposedly unknown motifs from the initial random (or partially random) set of DNA sequences using SELEX-type approaches. On the positive side, we apply the upper bounds from [2] on expected hitting time of a target area of genotypic space by EA (the EA runtime) to upper-bound the expected time to finding a sufficiently efficient series of motifs (e.g. binding sites for transcription factors) in a SELEX-type procedure. On the negative side, the pessimistic results from [3] yield upper bounds on expected proportion of the DNA sequences with sufficiently high fitness at a given iteration of SELEX-type procedure.

*Conclusion:* Our results suggest that some of the theoretically provable EA runtime bounds may be used, at least in principle, for a-priori estimation of efficiency of SELEX-based approaches. Further research is required to find out the properties of fitness landscape around the peaks of fitness function corresponding to separate conserved motifs in biologically meaningful fitness functions of Royal Road type.

*Acknowledgements:* Supported by the Russian Science Foundation (grant No. 17-18-01536).

## References

1. Spirov A., Holloway D. (2012) New approaches to designing genes by evolution in the computer. In: *Real-World Applications of Genetic Algorithms* (ed. by O. Roeva) InTech. pp. 235-260. DOI 10.5772/36817.
2. Corus D., Dang D.-C., Ereemeev A.V., Lehre P.K. (2017) Level-based analysis of genetic algorithms and other search processes. *IEEE Transactions on Evolutionary Computation*, Published online. DOI 10.1109/TEVC.2017.2753538.
3. Ereemeev A.V. (2017) On proportions of fit individuals in population of genetic algorithm with tournament selection. *Evolutionary Computation*. Published online. DOI 10.1162/EVCO\_a\_00210.



# HEDGE: Highly accurate GPU-powered protein-protein docking pipeline

T. Ermak\*, A. Shehovtsov, P. Yakovlev

BIOCAD, Saint Petersburg, Russia

\* e-mail: timofei.ermak@live.com

**Key words:** protein-protein complexes prediction, docking, GPU, HPC, *in silico* drug design

**Motivation and Aim:** protein-protein interactions play key roles in living systems functioning: cell signaling, immune system reactions, microelements transport and many other processes are based on protein-protein complexes functions. Thus, protein-protein complexes prediction is very important task especially in terms of drug discovery. For example, *in silico* optimization stages of antibody-based drug development process requires to solve the problem hundreds of times. To perform *in silico* optimization and increase drug candidates' quality, the docking problem must be solved with high accuracy in short time ranges. But it is one of the hardest structural bioinformatics problems due to large solution space (possible molecules orientations), big sizes of protein systems and infinite space of molecules conformations.

**Methods and Algorithms:** the pipeline of algorithms in our tool called HEDGE can be described as follows: 1) scanning translational solution space using FFT correlation theorem; 2) calculation of Gibbs free energy change ( $\Delta G$ ), we use own highly optimized implementation of OPLS [1] force field. 3) minimization of a complex energy, Polak-Ribière-Polyak conjugate gradient method [2] is used to solve optimization problem. Each step of the pipeline above is well-parallelizable, so, we utilize the full power of GPUs (graphics processing units), that allows to scan huge solution space and select best with solid metric of Gibbs free energy change. Moreover, different rotations of molecules can be processed independently, therefore, multi-GPU mode is supported to scale linearly and achieve maximal performance on multi-GPU supercomputers.

**Results:** HEDGE was tested on a subset of CAPRI [3] dataset showing 80 % of correct predictions for different types of proteins. Time required for prediction of one complex in rigid mode is about 7 minutes on Tesla V100 GPU, while other well-known tools (e. g. PIPER [4]) require about 90 minutes on 8 CPUs. Flexible mode requires much more calculations and takes about 1.5 hours on Tesla V100. Thus, our tool is one of the world's fastest in the field.

**Conclusion:** we developed highly accurate highly performant protein-protein docking tool called HEDGE, that successfully used in modern drug discovery pipelines.

## References

1. Robertson M.J., Tirado-Rives J., Jorgensen W.L. (2015) Improved peptide and protein torsional energetics with the OPLS-AA force field. *Journal of chemical theory and computation*. 11(7):3499-3509.
2. Polak E., Ribiere G. (1969) Note sur la convergence de méthodes de directions conjuguées. *Revue française d'informatique et de recherche opérationnelle. Série rouge*. 3(16):35-43.
3. Janin J. (2002) Welcome to CAPRI: a critical assessment of predicted interactions. *Proteins: Structure, Function, and Bioinformatics*. 47(3):257-257.
4. Kozakov D. et al. (2006) PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins: Structure, Function, and Bioinformatics*. 65(2):392-406.

## Revealing the research institutes and their interactions: a case study of miRNA research

A. Firsov<sup>1\*</sup>, I. Titov<sup>2</sup>

<sup>1</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup> Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

\* e-mail: [artemijfirsov@mail.ru](mailto:artemijfirsov@mail.ru)

**Key words:** affiliation disambiguation, institution network, KOFER, K-Mer, miRNA

*Motivation and Aim:* A lot of digital libraries appeared with the growth of the Internet, thus, format of representation of many scientific articles changed. That way, we got a possibility to query articles metadata, gather some statistics, etc. This includes understanding the institutions' activity, their interactions, and other characteristics. However, to do that, one should identify affiliation in order to know in which articles the true underlying organization is mentioned. Issue of affiliation disambiguation is complex if you consider the dataset consisting of  $2 \times 10^7$  articles, such as PubMed database. It becomes more complicated when you consider errors in affiliation made either by the author, or the editor. Moreover, sometimes institution name might be changed, or the affiliation from the papers metadata may have mixed institution names for different authors. E.g. if Author1 has "Institute of Cytology and Genetics, Novosibirsk, Russia" institution and Author2 has "Institute of Mathematics, Novosibirsk, Russia" institution, their resulting affiliation for paper might be "Institute of Cytology and Genetics, Institute of Mathematics, Novosibirsk, Russia". Moreover, affiliation can contain email, postal address and other artifacts.

*Methods and Algorithms:* In this work, we propose the method of the affiliation disambiguation based only on affiliations from papers metadata. The solution consists of 2 stages: preprocessing stage and clustering stage. At the preprocessing stage normalization and splitting of affiliation is performed. At the clustering stage the *DBSCAN* clustering is performed upon K-Mer features extracted from separated affiliations. Also, we proposed another clustering algorithm based on K-Mer Boolean feature vector sorting – *KOFER*. Parameters of the algorithm are trained on the Novosibirsk affiliation dataset consisting of 1000 samples.

*Results:* We show that *DBSCAN* method gives 0.81 v-measure score on the Novosibirsk affiliations dataset, while *KOFER* gives 0.9 v-measure score. We also present how affiliation grouping can be used to provide some statistics about institutional interactions, and provide institutions interaction network for Novosibirsk institutions and institutions in the miRNA science field gathered from PubMed database.

*Conclusion:* The results obtained show that institution from the miRNA conform network with *small-world* properties and that the proposed *KOFER* algorithm performs better than *DBSCAN* on the affiliations names data.

### References

1. Titov I.I., Blinov A.A. (2014) Exploring the structure and evolution of the Novosibirsk biomedical co-authorship network. *Vavilovskii Zhurnal Genetiki i Seleksii* = Vavilov Journal of Genetics and Breeding. 18(4/2):939-944. (in Russian)
2. Fortunato S., Bergstrom C.T., Börner K., Evans J.A., Helbing D., Milojević S., Petersen A.M., Radicchi F., Sinatra R., Uzzi B., Vespignani A., Waltman L., Wang D., Barabási A.-L. Science of science. [Online] [Cited: 5 2, 2018.] <http://science.sciencemag.org/content/359/6379/eaao0185.full>.



# Method of reconstruction of a sequence of non-ribosomal peptides from mass spectra with noise

E. Fomin

*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*  
*e-mail: fomin@bionet.nsc.ru*

**Key words:** algorithms, mass spectroscopy, sequences

*Motivation and Aim:* An important fraction of the peptidoma of bacteria is non-ribosomal peptides (NRP), representing a class of secondary peptide metabolites, usually produced by bacteria and fungi, and having an extremely wide range of biological activity and pharmacological properties. In the overwhelming majority of cases (73 %), NPFs have a complex nonlinear structure [1]. The monomers that make up the NRP have a wide variety of types (~500) and include, apart from 20 proteinogenic amino acids, non-proteinogenic amino acids and modified proteinogenic forms (methylated, glycosylated, D-forms) [2]. In connection with their biosynthesis from the non-bryosomal path, the identification of NPF by classical methods of bioinformatics and genomics is impossible, and is carried out only on the basis of mass spectrometry. At present, the possibilities of de novo reconstruction of the structure of complex NRF from mass spectra are limited. Thus, the development of new bioinformatic methods for the reconstruction of bacterial non-ribosomal peptides is very relevant.

*Methods and Algorithms:* Previously, we proposed a new method for solving the problem of reconstruction of a sequence of cyclic peptides from mass spectra, based on the removal of redundancy from the spectra [1,2]. We made a computer implementation of the method on the assumption that there were no noises or omissions in the spectra. The high efficiency of the proposed method was shown.

*Results:* In this work, the next step in de novo reconstruction of a sequence of cyclic peptides from mass spectra is made. A generalization of the previously proposed method was constructed by using continuous integral transformations. It is shown that the method makes it possible not only to significantly reduce the additive noise, that is, independent of the signal, in the initial data, but also to restore the omissions in the data.

## References

1. Caboche S., Pupin M., Leclère V., Fontaine A., Jacques P., Kucherov G. (2008) Norine: a database of nonribosomal peptides. *Nucleic Acids Res.* 36:D326-D331.
2. Caboche S., Leclère V., Pupin M., Kucherov G., Jacques P. (2010) Diversity of Monomers in Nonribosomal Peptides: towards the Prediction of Origin and Biological Activity. *Journal of bacteriology*, 192(19): 5143-5150
3. Fomin E. (2016) A Simple Approach to the Reconstruction of a Set of Points from the Multiset of  $n^2$  Pairwise Distances in  $n^2$  Steps for the Sequencing Problem: I. Theory. *J. Comput. Biol.* 23(9):769-75;
4. Fomin E. (2016) A Simple Approach to the Reconstruction of a Set of Points from the Multiset of  $n^2$  Pairwise Distances in  $n^2$  Steps for the Sequencing Problem: II. Algorithm *J. Comput. Biol.* 23(12):934-942.

# The performance improvement of the permutation test algorithm for GSEA

M. Grishchenko<sup>1\*</sup>, A. Yakimenko<sup>1,2</sup>, M. Khairtdinov<sup>1,2</sup>, A. Lazareva<sup>2</sup>

<sup>1</sup>*Institute Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Novosibirsk State Technical University, Novosibirsk, Russia*

\* e-mail: mikhail.grishch@gmail.com

**Key words:** resampling, randomization, permutation test, GSEA

*Motivation and Aim:* Processing of genetic data for the analysis genetic determination of traits is very important problem for modern biology. Resampling method are widely used to solve this problem. Resampling methods combine three different approaches: permutation test, “jack-knife” method and bootstrap [1]. In this work, permutation test method is considered. The basic idea of this method is to randomly permute rows or columns of observed values table [2]. It is important that the size of the table and the number of samples do not change during permutations. It allows analyzing multiple hypotheses simultaneously without correction of the statistical significance level. However, permutation test method requires much computational resources. The aim of this paper is to determine a minimal number of iterations of the permutation test algorithm to calculate steady p-value depending on the input data.

*Methods and Algorithms:* Permutation test algorithm allows us to calculate the p-value simultaneously for all characteristics of the gene sequence. The process of computing p-value is an iterative, in which the values of the computed statistics gradually converge to the stable value of the neighborhood of a certain value  $p^*$ . The average number of iterations was estimated to achieve a stable p-value, with a given confidence interval. It was shown that the average number of iterations is 27500–28500 iterations and in most cases, it does not depend on the amount of input data. It could be used this number of iterations. However, this approach has two drawbacks: 1) not all p-values achieved their stable values; 2) are cases when this number of iterations is not enough. Another approach is to use the maximum number of iterations, when all p-values reach their stable values.

*Results:* We investigated the permutation test algorithm aimed at finding statistically significant overrepresented gene characteristics under different external and/or internal conditions. It was obtained that the necessary number of iterations does not depend on the number of genes in the input data, but depends on the number of properties of the genes. In addition, we replace algorithm of random permutations to Fisher-Yates shuffle algorithm [3].

*Acknowledgements:* This work was supported by the Russian foundation for basic research (Grant No. 16-37-00240)

## References

1. Efron B. (1988) Nontraditional methods of statistical analysis. Moscow: Finansy i statistika. 263 p.
2. Yakimenko A.A., Gunbin K.V., Khairtdinov M.S. (2014) Search for the Overrepresented Gene Characteristics: The Experience of Implementation of Permutation Tests Using GPU. Optoelectronics, Instrumentation and Data Processing. 50(1):123-129.
3. Knuth D.E. (1969) Seminumerical algorithms. The Art of Computer Programming. 2. Reading, MA: Addison-Wesley. pp. 139-140. OCLC 85975465.

# An inverse problem in modelling of a symmetric gene network regulated by negative feedbacks

V. Golubyatnikov<sup>1,2\*</sup>, V. Gradov<sup>2</sup>

<sup>1</sup> Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

\* e-mail: vladimir.golubyatnikov1@fulbrightmail.org

**Key words:** Negative feedbacks, piece-wise linear dynamical systems, inverse problems

*Motivation and Aim:* We study one piece-wise linear dynamical system which describes functioning of a gene network regulated by negative feedbacks in order to find conditions of existence and uniqueness of periodic regimes of its functioning and show existence and uniqueness of solution of an inverse problem of identification of parameters of this system.

*Methods and Algorithms:* The approaches to modelling of similar gene networks, description of phase portraits of corresponding dynamical systems and detection of their periodic trajectories (cycles) are presented in [1, 2]. For some other non-linear dynamical systems, similar constructions were described in [3].

*Results:* For positive parameters  $A, m, \alpha$ , where  $A > \alpha$ , we consider symmetric piece-wise linear 3D dynamical system

$$\frac{dx}{dt} = L(z) - mx; \quad \frac{dy}{dt} = L(x) - my; \quad \frac{dz}{dt} = L(y) - mz. \quad (1)$$

Here  $L: [0, \infty) \rightarrow [0, \infty)$  is monotonic step-functions which corresponds in gene network to negative feedback,  $L([0, \alpha)) = A$ ,  $L([\alpha, \infty)) = 0$ . We show that the cube  $Q = [0, A] \times [0, A] \times [0, A]$  is invariant and decompose it to 8 blocks by hyperplanes  $x = \alpha$ ;  $y = \alpha$ ;  $z = \alpha$ . Note, that the system (1) is symmetric with respect to cyclic permutation of the variables  $x \rightarrow y \rightarrow z \rightarrow x$ .

**Theorem 1.** For the system (1), there exists unique piece-wise linear cycle  $C$  symmetric with respect to that cyclic permutation. This cycle  $C$  travels through six blocks of the decomposition of the invariant domain  $Q$ .

Let  $\tau$  be the period of this cycle  $C$  which can be measured in experiments, and let the parameters  $A$  and  $m$  be known as well. Also, we assume that we can measure the time  $t_1$  between two consecutive peaks of the graphs of the piece-wise linear functions  $x(t)$ ,  $y(t)$ ,  $z(t)$ . At the same time these three functions are not assumed to be known.

**Theorem 2.** Let the parameter  $A$  and the times  $\tau$ ,  $t_1$  for the system (1) be known, and  $\alpha \in (0, A)$  be unknown. Then the inverse problem of determination of the parameter  $\alpha$  has unique solution.

*Conclusion:* The main reason of our studied is the fact that the time measurements  $\tau$  and  $t_1$  of the oscillations in the gene network can be realized in non-invasive way. Similar inverse problem can be formulated for asymmetric dynamical systems of other dimensions as well.

*Acknowledgements:* Supported by RFBR, (18-01-00057) and by complex program of basic research of SB RAS (0314-2018-0011).

## References:

1. Ayupova N.B., Golubyatnikov V.P. (2014) On the uniqueness of a cycle in an asymmetric three-dimensional model of molecular repressilator. *Journal of Applied and Industrial Mathematics*. 8(2):1-6.
2. Golubyatnikov V.P., Kalenykh A.E. (2016) On structure of Phase Portraits of Some Nonlinear Dynamical systems. *Journal of Mathematical Sciences*. 215(4):475-483.
3. Glass L., Pasternack J.S. (1978) Stable oscillations in mathematical models of biological control systems. *Journal of Math. Biology*. 6:207-223.

# On cycles in models of asymmetric circular gene networks

V. Golubyatnikov<sup>1, 2\*</sup>, N. Kirillova<sup>2</sup>

<sup>1</sup> Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

\*e-mail: vladimir.golubyatnikov1@fulbrightmail.org

**Key words:** Circular gene networks, equilibrium points, cycles

*Motivation and Aim:* We consider nonlinear dynamical systems as models of functioning of asymmetric circular gene networks more complicated and general than analogous models studied in [1–3]. Our main aim here is to find conditions of existence of oscillating trajectories (cycles) of these systems.

*Methods and Algorithms:* Our constructions and studies of circular gene networks models and description of geometric and combinatorial structures of their phase portraits are based on our previous results, see [3]. In our numerical experiments we used the soft STEP elaborated in the Sobolev institute of mathematics.

*Results:* For positive parameters  $k_j$  and  $\mu_s$  and positive monotonically decreasing smooth functions  $f_m$ ,  $m = 1, 5, 8$ , which describes negative feedbacks in the gene network, we consider 9D-dynamical system

$$\begin{aligned} \frac{dx_1}{dt} &= f_1(x_9) - k_1 x_1; \quad \frac{dx_j}{dt} = f_j(x_{j-1}) - k_j x_j; \quad j = 5, 8; \\ \frac{dx_s}{dt} &= \mu_s(x_{s-1}) - k_s x_s; \quad s = 2, 3, 4, 6, 7, 9. \end{aligned} \quad (1)$$

Here  $x_1, x_5, x_8$  are concentrations of mRNA's, and all the other variables denote concentrations of proteins which are "intermediate" stages of this gene network functioning. Here, in contrast with [1–3], several intermediate stages can appear between each pair of mRNA's with consecutive indices, not just one.

We show uniqueness of equilibrium point  $S_0$  of the system (1) and find conditions of existence of a cycle  $C$  of this system, and describe an invariant polyhedral domain  $W$  of this system in the positive octant of 9-D space and contains  $C$ . These conditions are formulated in terms of matrix of linearization of the system (1) at the point  $S_0$ : the non-diagonal non-zero terms of this matrix should be sufficiently large with respect to the parameters  $k_j, k_s$ . The invariant domain  $W$  is composed by 18 adjacent parallelepipeds and retracts to  $C$ . Our numerical experiments illustrate and correspond to the theoretical results. We show non-uniqueness of the cycles in some higher-dimensional dynamical systems of the type (1).

*Conclusion:* In contrast with [2], where the particular case  $m_1 = m_2 = m_3 = 1$  symmetric with respect to cyclic permutations of the variables was studied, the shifts along trajectories of the system (1) are not described by equations with delayed arguments. The cycle  $C$  is not symmetric with respect to this permutation.

*Acknowledgements:* Supported by RFBR (18-01-00057) and by complex program of basic research of SB RAS (0314-2018-0011).

## References

1. Elowitz M.B., Leibler S. (2000) A Synthetic Oscillatory Network of Transcriptional Regulators. *Nature*, 335-338.
2. Kolesov A.Yu., Rozov N.Kh., Sadovnichii V.A. (2016) Periodic Solutions of Travelling-Wave Type in Circular Gene Networks. *Izvestiya RAN: Ser. Mat.* 80(3):67-94.
3. Ayupova N.B., Golubyatnikov V.P., Kazantsev M.V. (2017) On the Existence of a Cycle in an Asymmetric Model of a Molecular Repressilator. *Numerical Analysis and Applications*. 10(2):101-107.

# On existence of a piecewise smooth cycle in one asymmetric gene network model with piecewise linear equations

V. Golubyatnikov<sup>1,2\*</sup>, L. Minushkina<sup>2</sup>

<sup>1</sup> Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

\* e-mail: vladimir.golubyatnikov1@fulbrightmail.org

**Key words:** Negative feedbacks, piecewise linear dynamical systems, invariant domains, cycles, state transition diagram

**Motivation and Aim:** We construct a simple piecewise linear dynamical system which simulates one gene network regulated by negative feedbacks in order to find conditions of existence of periodic regimes (cycles) of its functioning and to describe location of these cycles in the phase portrait of the system.

**Methods and Algorithms:** Some approaches to modelling of similar gene networks and description of combinatorial structures of discretizations (State Transition Diagram) of the phase portraits of corresponding nonlinear dynamical systems are presented in [1–3].

**Results:** For positive parameters  $m_j$ ,  $A_j$  and  $\alpha_j$ , where  $A_j > \alpha_j$ ,  $j = 1, 2, 3$ , we consider 3D-dynamical system

$$\frac{dx}{dt} = L_1(z) - m_1x; \quad \frac{dy}{dt} = L_2(z) - m_2y; \quad \frac{dz}{dt} = L_2(y) - m_3z. \quad (1)$$

Here  $L_j$  are non-negative step-functions which correspond in gene network to negative feedbacks:  $L_j([0, \alpha_j)) = A_j$ , and  $L_j([\alpha_j, \infty)) = 0$ . We show that trajectories of the system (1) are piecewise smooth, and that the polyhedral domain  $Q = [0, A_1] \times [0, A_2] \times [0, A_3]$  is positively invariant with respect to shifts along these trajectories. Let us decompose this domain  $Q$  to 8 smaller parallelepipeds by hyperplanes  $x = \alpha_1$ ;  $y = \alpha_2$ ;  $z = \alpha_3$ .

**Theorem.** There exists a piecewise smooth cycle  $C$  of the system (1) which passes through union  $U_6$  of 6 of these parallelepipeds  $B_k$ . The angle points of this cycle are located on the common faces of the parallelepipeds  $B_k$ .

So, this union  $U_6$  is an invariant domain of the dynamical system (1) as well, it does not contain two parallelepipeds containing the origin and the “opposite” point  $(A_1, A_2, A_3)$ . The theorem follows from the analysis of linearization of the system (1) in each of the parallelepipeds  $B_k$  near their common point  $(\alpha_1, \alpha_2, \alpha_3)$ . The existence of the cycle  $C$  is shown with the help of the Brouwer fixed point theorem.

**Conclusion:** In contrast with [2], where the particular case  $m_1 = m_2 = m_3 = 1$  was studied, the shifts along trajectories of the system (1) are not described by projective transformations of the faces of adjacent blocks  $B_k$  which contain  $C$ . Thus, the uniqueness of this cycle does not follow from the geometric arguments used in [2, 3].

**Acknowledgements:** Supported by RFBR, No. 18-01-00057.

## References:

1. Likhoshvai V.A., Golubyatnikov V.P. et al. (2008) Theory of gene networks. In: System computerized biology. Novosibirsk, SB RAS, 397-480.
2. Ayupova N.B., Golubyatnikov V.P. (2014) On the uniqueness of a cycle in an asymmetric three-dimensional model of molecular repressilator. Journal of Applied and Industrial Mathematics. 8(2):1-6.
3. Ayupova N.B., Golubyatnikov V.P. (2015) On two classes of nonlinear dynamical systems: the 4-dimensional case. Siberian Mathematical Journal. 56(2):231-236.

# Investigation of stopping criterion for OSEM algorithm with application to nuclear medicine

N.V. Denisova<sup>1</sup>, O. Krivorotko<sup>2,3</sup>

<sup>1</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup> Khristianovich Institute of Theoretical and Applied Mechanics, Novosibirsk, Russia

<sup>3</sup> Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia

\* e-mail: lijyu1234@yandex.ru

**Key words:** inverse problem, SPECT, PET, OSEM, optimization, regularization

*Motivation and Aim:* The OSEM (Ordered Subset Expectation Maximization) algorithm [1, 2] is studied in this work. A diagnostically acceptable image is obtained by interrupting (stopping) of the iterative process because the OSEM algorithm is developed on the basis of an unregularized approach. In fact, the interrupt is a “rough regularization”. The iteration number of the “stop of the algorithm” is determined in most cases empirically in preliminary studies and recorded in the patient examination protocol for a particular type of installation. The doctor must follow the appointed protocol. However, patients differ in their anatomical constitutions therefore the requirements of the protocol do not always correspond to the obtaining of the optimal image.

*Methods and Algorithms:* It was suggested to use the Pearson statistical criterion Chi-square as the stopping rule [3]. However, this proposal was not implemented on commercial installations. A theoretical analysis of regularization of OSEM based on stochastic properties of process and mathematical analysis of misfit function is carried out [4].

*Results:* In this work, studies of the OSEM image reconstruction algorithm are performed in the context of applications to positron emission tomography (PET) and single-photon emission computed tomography (SPECT)]. It is shown theoretically and in numerical simulation that if the source function is stochastic and includes regions with very different levels of statistics of emitted gamma quanta, the Pearson criterion gives incorrect values of “stopping”. Our research has shown that the reason is that regions with different statistics behave differently in the iterative process and give different values for the stopping criterion.

*Acknowledgements:* Supported by the Russian Foundation for Basic Research (No. 17-52-14004).

## References

1. Shepp L.F., Vardi Y. (1982) Maximum Likelihood reconstruction for Emission Tomography IEEE Trans. Med. Imag. 1(2):113-121.
2. Hudson H., Larkin R.S. (1994) Accelerated Image Reconstruction Using Ordered Subsets of Projection Data IEEE Trans. Imag. Imag. 13(4):601-609.
3. Veclerov E., Llacer J. (1987) Stopping Rule for the MLE Algorithm Based on Statistical Hypothesis Testing IEEE Trans. Med. Imag. 6(4):313-319.
4. Kabanikhin S.I. (2008) Definitions and examples of inverse and ill-posed problem. Journal of Inverse and Ill-Posed Problems. 16(4):317-357.



# A numerical algorithm of parameter identification in mathematical model of tuberculosis transmission with control programs

S.I. Kabanikhin<sup>1,2</sup>, O.I. Krivorotko<sup>1,2</sup>, V.N. Kashtanova<sup>2\*</sup>

<sup>1</sup>*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Novosibirsk State University, Novosibirsk, Russia*

\* e-mail: vikakashtanova@ya.ru

**Key words:** model of tuberculosis transmission, reconstruction of model parameters, system of ordinary differential equations, parameter identification, inverse problem, optimization approach, fast simulate annealing, gradient descent method

*Motivation and Aim:* The development of an individual mathematical model describing the process of the propagation of Tuberculosis (TB) infection in the population is one of the most effective methods for prediction of the epidemic spread in a particular region. Such models are described by systems of nonlinear ordinary differential equations (ODE) with the coefficients that characterize the features of population and disease spread. Consequently, it is necessary to qualitatively evaluate parameters of model (or their combinations) [1] for specification model for special population.

*Methods and Algorithms:* The purpose of this work is the construction and investigation of the numerical algorithm for determining the coefficients of nonlinear ODE system which describes TB transmission processes with treatment and drug resistance [2] using additional information about a special population according to statistical data for the previous few years (namely, the number of healthy, latently infected and infectious diseases individuals). The numerical algorithm is based on combination of very fast annealing and gradient approaches for minimization of least squares objective function [3].

*Results and Conclusion:* The results of numerical calculations show that above approach determines the set of more sensitive parameters to a particular region that differs significantly from its widely used standard values. The numerical results are analyzed and discussed.

*Acknowledgments:* This work is supported by the Scholarship of the President of RF No. MK-1214.2017.1. and by the grant No. 18-71-10044 of Russian Scientific Found (RScF).

## References

1. Kabanikhin S.I. (2011) *Inverse and Ill-Posed Problems: Theory and Applications* (Berlin: de Gruyter).
2. Trauer J.M., Denholm J.T., McBryde E.S. (2014) Construction of a mathematical model for tuberculosis transmission in highly endemic regions of the Asia-Pacific, *Journal of Theoretical Biology*, 358:74-84.
3. Banks H.T., Hu Sh., Thompson W.C. (2014) *Modeling and Inverse Problems in the Presence of Uncertainty* (Chapman and Hall/CRC press).

## Simulation and image reconstruction of the combined Siemens PET/CT and PET/MRI systems

H. Kertesz<sup>1</sup>, A. Renner<sup>2</sup>, I. Rausch<sup>1</sup>, T. Beyer<sup>1</sup>, J. Cal-Gonzalez<sup>1</sup>

<sup>1</sup>QIMP group, Center for Medical Physics and Biomedical Engineering

<sup>2</sup>Digital Image Processing Laboratory, Center for Medical Physics and Biomedical Engineering Medical University of Vienna, Vienna, Austria

\* e-mail: hunor.kertesz@meduniwien.ac.at

**Key words:** Monte Carlo simulation, performance evaluation, GATE simulation, NEMA protocol

*Motivation and Aim:* The objective of this work is to validate a Monte Carlo (MC) simulation model for two commercially-available, whole-body PET systems. The MC models will be used to evaluate the performance of different image reconstruction methodologies at low count rates.

*Methods and Algorithms:* GATE (GEANT4 Application for Tomographic Emission) was used as the MC toolkit for the modeling of the Siemens Biograph 64 TruePoint TrueView PET/CT (TPTV) and the Siemens Biograph PET/MR (mMR) systems. In both cases, we included detailed models of the detector electronics, system geometry and the physical processes involved in the data acquisition. The performance of both system models was validated following the NEMA (National Electrical Manufacturers Association) NU 2-2012 protocol. We compared the simulation results with the measured values for sensitivity, count rate (CR), and noise equivalent count rate (NECR). Moreover, three voxelized NEMA IQ phantom was simulated. The simulated data was reconstructed with the STIR framework using the standard OSEM algorithm.

*Results:* The calculated (reference value from measurements) sensitivity for the mMR was 13.8 (15.0) kcps/MBq and 14.4 (13.9) kcps/MBq at the center of the field-of-view (FOV) and at 10 cm radial offset, respectively. The NECR peak was 189 kcps @ 23.8 kBq/ml (184 kcps @ 23.0 kBq/ml) and the scatter fraction at the NECR peak was 29.0 (37.9) %. For the TPTV, the sensitivity was 8.0 (8.1) kcps/MBq and 7.9 (8.2) kcps/MBq at the centre of FOV and at 10 cm radial offset, respectively. The NECR peak was 151 kcps @ 27 kBq/ml (161 kcps @ 31 kBq/ml) and the scatter fraction at the NECR peak was 24.8 (32.5) %.

*Conclusion:* Both PET/CT and PET/MRI models showed a good agreement (< 10 %) with the measured reference values. The application of these models for the evaluation of different image reconstruction algorithms in simulated numerical phantoms is work in progress.

*Acknowledgements:* The financial support of the Austrian FWF Project I3451- N32 is gratefully acknowledged. The computational results presented have been achieved using the Vienna Scientific Cluster (VSC).



# Creation of a modular model of metabolic processes in skeletal muscles during moderate physical load using BioUML platform

I.N. Kiselev<sup>1,2\*</sup>, V.I. Baranov<sup>3</sup>, F.A. Kolpakov<sup>1,2</sup>

<sup>1</sup>*Institute of Computational Technologies, SB RAS, Novosibirsk*

<sup>2</sup>*LLC «BIOSOFT.RU» Ltd., Novosibirsk*

<sup>3</sup>*Institute of Physiology and Basic Medicine, Novosibirsk*

\* e-mail: axec@dote.ru

**Key words:** mathematical model, modular modeling, skeletal muscles, metabolism, physiology, BioUML

*Motivation and Aim:* Global aim of this project is studying molecular mechanisms in muscles along with gene expression regulation. First step in this direction is creation of mathematical model of metabolic processes in muscle which can be further extended and linked with genetic expression in skeletal muscle under different influences.

*Methods and Algorithms:* Software platform BioUML ([www.biouml.org](http://www.biouml.org)) provides graphical representation and automatic generation of Java code for numerical modeling of the systems dynamics, it utilizes modular approach which implies creation of models as a set of interconnected parts (modules) each of modules is a mathematical model itself and describes particular subsystem. Modules can be modular itself, creating nested hierarchy of models. Modular representation facilitates understanding and consequent work with the model, which can be updated by adding new modules, improving existing and combining mathematical models obtained from different sources. It allows mathematical modeling of wide range of biological systems using different mathematical formalisms.

*Results:* We have implemented model of metabolic processes in muscles [1] as a modular model in BioUML. Model consists of 5 main modules: arteries, veins, blood flow through capillary, transport of metabolites from muscle fiber and muscle fiber. Muscle fiber module is a modular model itself. It consists of cytosol, mitochondria and block representing transport of metabolites between them. Such decomposition leads the way to further addition of new parts and/or replacing of existing blocks with more complicated and improved versions. For example modular version of this model from the same authors [2] can be obtained by duplicating muscle fiber block and initializing of two fibers with different parameters (representing red and white muscle fibers). Similarly we can construct models with other types of muscles in arms, legs, back, etc. Other ways to improve the model is adding new blocks describing: – heart, lungs, liver, etc.; – different types of training; – molecular mechanism of gene expression regulation during physical load.

*Conclusion:* We have shown decomposition into modules and creation of a modular model with BioUML platform on the example of the muscle metabolism model. Created modular model is initial point for further improvement by adding new blocks and improving of existing blocks.

*Availability:* Created model is freely available as a part of BioUML platform at [http://wiki.biouml.org/index.php/Muscle\\_metabolism](http://wiki.biouml.org/index.php/Muscle_metabolism).

*Acknowledgements:* Supported by RFBR, research project No. 17-00-00296 KOMFI.

## References

1. Li Y., Dash R.K., Kim J., Saidel G.M., Cabrera M.E. (2009) Role of NADH/NAD<sup>+</sup> transport activity and glycogen store on skeletal muscle metabolism during exercise: in silico studies. *Am J Physiol Cell Physiol.* 296:25-46.
2. Li Y., Lai N., Kirwan J.P., Saidel G.M. (2012) Computational Model of Cellular Metabolic Dynamics in Skeletal Muscle Fibers during Moderate Intensity Exercise. *Cell Mol Bioeng.* 5(1):92-112.

# Population-based mathematical modeling antihypertensive drugs effect using BioUML platform

I.N. Kiselev<sup>1,2\*</sup>, A.F. Kolpakova<sup>1,2</sup>, F.A. Kolpakov<sup>1,2</sup>

<sup>1</sup> Institute of Computational Technologies, SB RAS, Novosibirsk

<sup>2</sup> LLC «BIOSOFT.RU», Novosibirsk

\* e-mail: axec@dote.ru

**Key words:** mathematical model, population modeling, cardiovascular system, arterial hypertension, antihypertensive drugs, BioUML

*Motivation and Aim:* Mathematical modeling can be used to predict effect of antihypertensive and other drugs on the modeling organism. Experimental data for drugs is usually given as population data (mean±SD) thus in order to reproduce this data we need to transit from single mathematical models to populations of such models.

*Methods and Algorithm:* Population approach implies generating set of models reflecting real population of modeled organisms. Each particular model of this set represents distinct organism with certain parameters. We use metropolis algorithm to generate population of models with given distribution of observed values (systolic and diastolic blood pressure). When population is generated we extend the model by adding blocks describing pharmacokinetics and pharmacodynamics of selected drugs to each virtual patient thus describing treatment course (which can be monotherapy or therapy with several drugs) of given population. Finally we can compare results of treatment course for virtual and real populations.

*Results:* As a structural model we used mathematical model of the human cardiovascular system created in the BioUML platform and comprising modules (blocks) of different subsystems (heart, arterial system, salt-water balance, renin-aldosterone-angiotensin system and others). Using this model we have generated virtual population of nearly 1000 virtual organisms imitating patients with arterial hypertension (systolic pressure  $148 \pm 15.5$  mmHg, diastolic pressure  $89 \pm 10$  mmHg). Additionally to core model we have implemented detailed models of pharmacokinetics and pharmacodynamics of aliskiren [1] and losartan [2] and pharmacodynamics blocks of bisoprolol, enalapril and amlodipine. Created modules were used to simulate 8 different treatment courses for virtual population using said drugs including 4 different daily doses of aliskiren (37.5 mg, 75 mg, 150mg and 300 mg), daily dose of 100 mg losartan [3], 5 mg of amlodipine [4] and 5 mg of enalapril [5], all courses have duration of 4 weeks. Separately treatment with bisoprolol was tested in combination with physical stress [6]. Changes in blood pressure under different drugs were compared to results of published clinical studies and showed good agreement with experimental data except for diastolic pressure under effect of bisoprolol drug which indicates needed improvements in the model.

*Conclusion:* Implemented models and software provides effective platform for reproducing results of clinical trials.

*Acknowledgements:* Supported by RFBR, research project No. 16-01-00779 A.

## References

1. Hong Y., Dingemans J., Mager D.E. (2008) Pharmacokinetic/Pharmacodynamics Modeling of Renin Biomarkers in Subjects Treated With the Renin Inhibitor Aliskiren. *Clinical Pharmacology Therapeutics*. 84(1):136-43.
2. Nguyen H.Q., Lin J., Kimoto E., Callegari E., Tse S., Obach R.S. (2017) Prediction of Losartan-Active Carboxylic Acid Metabolite Exposure Following Losartan Administration Using Static and Physiologically Based Pharmacokinetic Models. *J Pharm Sci*. 106(9):2758-2777.
3. Stanton A., Jensen C., Nussberger K., O'Brien E. (2003) Blood pressure lowering in essential hypertension with an oral renin inhibitor, aliskiren. *Hypertension*. 42(6):1137-1143.
4. Guo C., Pei Q., Tan H., Huang Z., Yuan H., Yang G. (2015) Effects of genetic factors on the pharmacokinetics and pharmacodynamics of amlodipine in primary hypertensive patients. *Biomedical reports*. 2:195-200.
5. Ruddy M.C., Mroczek W.J. (1993) Comparison of Ramipril and Enalapril in Patients With Essential Hypertension. *Pharmacotherapy*. 13(3):224-228
6. Leopold G., Ungethum W., Pabst J., Simane Z., Buhring K.U., Wiemann H. (1986) Pharmacodynamic profile of bisoprolol, a new beta1-selective adrenoceptor antagonist. *Br. J. clin. Pharmacol* 22:293-300.

# Assessment of software for somatic single nucleotide variant identification using simulated whole-genome sequencing data of cancer

W. Kittichotirat<sup>1\*</sup>, P. Khongthon<sup>1</sup>, K. Kusonmano<sup>2</sup>, S. Cheevadhanarak<sup>2</sup>

<sup>1</sup> *Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi, Bangkok, Thailand*

<sup>2</sup> *School of Bioresources and Technology, King Mongkut's University of Technology Thonburi, Bangkok, Thailand*

\* e-mail: weerayuth.kit@kmutt.ac.th

**Key words:** analytical pipeline, somatic single nucleotide variants, whole-exome sequencing

*Motivation and Aim:* Next-generation sequencing is an important tool for identifying disease-causing mutations in human. However, it can be relatively difficult to identify many true somatic single nucleotide variants (sSNVs) because they may not be supported by enough sequencing reads to pass the minimal criteria [1]. This could be caused by the contamination of normal cells, cell population heterogeneity, or sample preservation. As a result, some sSNVs calling software that performs well in one sample may perform poorly in another [2].

*Methods and Algorithms:* The reference human genome sequence and known variations were used to build a model for generating germline and somatic mutations. This was then used to simulate 2x101 bp paired-end whole-genome sequencing reads at 50x coverage for both normal and cancer samples. The contaminated cancer samples were also constructed at different levels of purity. Four software (VarScan2, SomaticSniper, Strelka, and MuTect) were used to identify sSNVs. The accuracy of variant calling software was assessed by comparing to the known variants using sensitivity and specificity analysis. Moreover, the assessment was also conducted for different combinations of variant calling software to search for the most optimal combination for identifying sSNVs.

*Results:* VarScan2 had the lowest accuracy in identifying sSNVs in the low purity sample. However, VarScan2 also outperforms all other software in identifying sSNVs in high purity sample. On the other hand, MuTect excelled at identifying sSNVs in low purity sample with sensitivity greater than VarScan2 by 20 folds. Interestingly, the combination of MuTect, Strelka, and VarScan2 provided the highest sensitivity for identifying sSNVs in both pure and contaminated cancer samples.

*Conclusion:* This study provides an assessment of software for sSNVs identification using simulated cancer and matched normal datasets. The results suggested that combination of outputs from multiple software can help to improve the prediction accuracy.

*Acknowledgements:* Supported by the Bioinformatics and Systems Biology Program, KMUTT and BIOTEC, NSTDA, Thailand.

## References

1. Fang L.T. et al. (2015) An ensemble approach to accurately detect somatic mutations using SomaticSeq. *Genome Biology*. 16(1):197-209.
2. Ewing A.D. et al. (2015) Combining tumor genome simulation with crowdsourcing to benchmark somatic single-nucleotide-variant detection. *Nature Methods*. 12(7):623-630.

# Spatial heterogeneity influences evolutionary scenarios in microbial communities explained by ecological stratification: a simulation study

A.I. Klimenko\*, Yu.G. Matushkin, S.A. Lashin

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

*Novosibirsk State University, Novosibirsk, Russia*

\* e-mail: klimenko@bionet.nsc.ru

**Key words:** microbial communities, ecological modelling, evolutionary modelling

*Motivation and Aim:* There are two evolutionary trends in genome organization among bacteria inhabiting microbial communities – towards genome amplification and towards genome reduction. Which scenario overcomes depends on the environmental conditions and the complexity of underlying gene networks. Facing changes of environmental conditions, a population might either adapt or utilize the available reserves more efficiently, or the third way is to migrate pursuing optimal environmental conditions, intruding new habitats and engaging in competition with local species. The impact of such adaptive individual behavior on a developmental scenario of the whole community is of current interest for acquiring a more profound comprehension of interactions between various evolutionary mechanisms.

*Methods and Algorithms:* The models of evolution of microbial communities were created using the HEC software package [1], which is based on the agent-based approach and super-individual concept [2]. The key simulation unit in the HEC is a genetically homogeneous population of cells that interacts with its environment and other populations. The classification by ecogroups has been proposed that takes into account, the role of organisms in their mutual interactions, and involves their function in the community structure estimating the complexity of their metabolism.

*Results:* In this simulation study, we have shown that spatially structured habitats allow subdivision of evolutionary trends depending on the distance of a sub-habitat to the nutrient source. We considered ecogroups distributions dynamics according to model parameters. Motile cells typically exhibit the dominance of either aedificators or quasi-commensals throughout the whole environment, whereas immobile cells show spatial stratification of dominant ecogroups in line with the distance towards nutrient source. Aedificator biomass prevails in ecosystems consisting of immobile organisms while quasi-mutualist biomass is higher in ecosystems of motile organisms. Active migrations have been shown to impede introduction of new species and to decrease total biomass of the community that is caused by a rapid expansion of well-established populations resulting in the chain reduction of available ecological licenses.

*Conclusion:* The simulations have shown that ecological patterns of self-organization of microbial communities causes sustainability of different strategies underlying the antagonistic evolutionary scenarios. Different evolutionary trends sustain in habitats with contrasting ecological conditions due to nutrient gradients that structure the environment spatially.

*Acknowledgements:* The study was supported by the Budget Project No. 0324-2018-0017.

## References:

1. Klimenko A.I. et al. (2015) Modeling evolution of spatially distributed bacterial communities: a simulation with the haploid evolutionary constructor. BMC Evol. Biology. 15:S3.
2. Scheffer M., Baveco J.M., DeAngelis D.L., Rose K.A., van Nes E.H. (1995). Ecol Modell. 80:161-170.

## Different effects of agroclimatic factors on time to emergence and time to flowering in nine soybean accessions

K. Kozlov<sup>1\*</sup>, L. Novikova<sup>1,2</sup>, I. Seferova<sup>2</sup>, S. Nuzhdin<sup>1,3</sup>, M. Samsonova<sup>1</sup>

<sup>1</sup> *Peter the Great St.Petersburg Polytechnic University, St.Petersburg, Russia*

<sup>2</sup> *Federal Research Center the N.I. Vavilov All-Russian Institute of Plant Genetic Resources, St. Petersburg, Russia*

<sup>3</sup> *University of Southern California, Los Angeles, CA, USA*

\* e-mail: kozlov\_kn@spbstu.ru

**Key words:** soybean, time to emergence, time to flowering, agroclimatic factors, grammatical evolution

*Motivation and Aim:* The effect of climate change on agronomically important crops can be assessed with regression models that connect agronomic traits to climatic factors. Accounting for interactions between genotype and environment will provide valuable insights into phenological characteristics of cultivars across different geographical locations and genotypes.

*Methods and Algorithms:* We build a regression model for time periods from sowing to emergence and emergence to flowering as a linear combination of  $N$  control functions  $F_n$  that describe the dependence of phenotype on climatic factors. The interaction between genotype and environment is modeled by additional term in regression function that has the form of a weighted sum of pairwise products between a control function and each accession. We use a combination of Grammatical Evolution, LASSO and Differential Evolution [1] to recover analytic form of control functions, find regression coefficients and determine the set of significant climatic factors, respectively. Due to a stochastic nature of the procedure it is repeated several times to obtain an ensemble of models. Consequently, the effect of a factor on accession is estimated with the coefficient of determination averaged over the ensemble of models, from which the terms that do not depend on the factor were excluded. To reveal factors that affect accessions differently, their effects on accession pairs were compared with Wilcoxon test.

*Results:* The method was applied to predict the time periods from sowing to emergence (coefficient of determination  $R^2 = 0.65$ ) and from emergence to flowering ( $R^2 = 0.65$ ) in a dataset that comprises 379 plants of 9 soybean accessions of different origin phenotyped at Pushkin and Kuban VIR stations in 1999–2013 [2].

*Conclusion:* Our analysis allowed us to reveal pairs of accessions with quantitatively different reaction on temperature and precipitation. The analysis of relative difference in the sum of squares for a model with and without a weighted sum of pairwise products between a control function and each accession allows us to conclude that genotype-environment interaction accounts for about 6.3 % and 15.9 % of variation in time periods from sowing to emergence and emergence to flowering, respectively.

*Acknowledgements:* Supported by the Federal Targeted Program (Agreement No. 14.575.21.0136). Calculations were performed in Supercomputer Center of Peter the Great St.Petersburg Polytechnic University and University of Southern California CBB computer cluster.

### References

1. Kozlov K. et al. (2018) A Mathematical Model of the Impact of Climatic Factors on Soybean Development. *Biophysics*. 63(1):175-176.
2. Seferova I., Novikova L. (2015) Climatic factors that impact development of fast ripening accessions of soybean in Northwest Russia. *Applied Botany, Genetics and Selection*. 176:88-97.



# The optimal control of stochastic differential equations arising in biology, economy and finance

E. Kondakova<sup>1\*</sup>, O. Krivorotko<sup>1,2</sup>, S. Kabanikhin<sup>1,2</sup>

<sup>1</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup> Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia

\* e-mail: ekondak95@mail.com

**Key words:** stochastic differential equations, inverse problems, optimal control, mathematical models in economy, Fokker-Plank equation, optimization.

*Motivation and Aim:* Real biological, social and economic processes are exposed to external influences that can't always be clearly described (for example, hormonal fluctuations, variations in blood pressure, respiration in biological problems or political changes, the "human factor", natural disasters in social and economic problems). In this connection, there is a growing need to expand deterministic models to models that are described by stochastic differential equations (SDE), where the relevant parameters are modeled as suitable random processes, or stochastic processes are added to the equations of the motion system. The problems of controlling stochastic dynamical systems are widely encountered in practice and are the subject of deep mathematical research. Inverse problem [1] for SDE that consists in determination of the right-hand side function (control function) for the Merton equation with the Wiener process [2] is numerically investigated.

*Methods and Algorithms:* A numerical regularization algorithm to ensure optimal control is developed. To determine the control function in SDE, various methods are used. The first is to reduce SDE to partial differential equations (the Fokker-Planck equation) [3]; the second is based on the method of dynamic programming. In this paper, the numerical algorithm relies on the principle of dynamic programming (applied to the Merton problem).

*Results:* A numerical algorithm for solving the inverse problem on the basis of dynamic programming, which determines the optimal control function, is developed.

*Conclusion:* The developed numerical algorithm allows one to determine the optimal function on the right-hand side in SDE describing real biological, social and economic processes or interactions.

*Acknowledgements:* This work was partially supported by the President Grant of Russian Federation (No. MK-1214.2017.1), by the Ministry of Education and Science of Russian Federation and by the grant No. 18-71-10044 of Russian Scientific Found (RScF).

## References

1. Kabanikhin S.I. (2008) Definitions and examples of inverse and ill-posed problem. Journal of Inverse and Ill-Posed Problems. 16(4):317-357.
2. Soner H.M. (2004) Processes Stochastic Optimal Control in Finance.
3. Dunker F., Hohage Th. On parameter identification in stochastic differential equations by penalized maximum likelihood. Inverse Problems. 30(9):095001.

# Supercomputer analysis of social, epidemiological and economic processes

O. Krivorotko

*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia  
Novosibirsk State University, Novosibirsk, Russia*

*\* e-mail: krivorotko.olya@mail.ru*

**Key words:** inverse problems, mathematical modelling, social processes, epidemiology, economy, optimization, Hamilton-Jacobi-Bellman equation, Fokker-Plank equation, dynamic programming

*Motivation and Aim:* The specifics of the dissemination of information in society, the development of socially significant diseases (tuberculosis, HIV / AIDS) and economic processes depend on the region. One of the most effective methods is the development and identification of mathematical models that describe the processes of information dissemination in social networks [1], infections in the population [2] and economic processes [3]. Such models are described by systems of differential equations, the coefficients of that characterize the distribution of information, population, disease development and economic processes in the country. To control information in social networks, epidemics in individual regions and economic processes, it is necessary to refine the model coefficients by some additional information (the inverse problem) [4].

*Methods and Algorithms:* One way to solve the problem of improving the coefficients is to reduce the inverse problem to a variational formulation, where the functional satisfies to the Hamilton-Jacobi-Bellman partial differential equation, or characterizes the quadratic deviation of the model data from the experimental ones for systems of stochastic differential equations or partial differential equations. The Tikhonov regularization, gradient methods and stochastic approach (genetic algorithm, simulated annealing) are used for solving ill-posed inverse problems. To reconstruct the control function the Pontryagin maximum principle and dynamic programming are used.

*Results:* The reconstructed coefficients and control functions in mathematical models of social, epidemiological and economic processes allows one to refine the prognosis of dynamic of processes and give the recommendation for control its.

*Conclusion:* A prototype of the digital Earth globe will be created with the possibility to visualize social, epidemiological and economic processes in various countries and continents based on mathematical modeling dynamic plug-in.

*Acknowledgements:* Supported by the Ministry of Education and Science of Russian Federation, by the President Grant of Russian Federation (No. MK-1214.2017.1) and by the grant No. 18-71-10044 of Russian Scientific Found (RScF).

## References

1. Dai G., Ma R., Wang H., Wang F., Xu K. (2015) Partial differential equations with robin boundary condition in online social networks. *Discrete and Continuous Dynamical Systems Series B*. 20(6):1609-1624.
2. Kabanikhin S.I., Krivorotko O.I., Kashtanova V.N. (2018) A combined numerical algorithm for reconstructing the mathematical model for tuberculosis transmission with control programs. *Journal of Inverse and Ill-Posed Problems*. 26(1):121-131.
3. Herowitz J. (2014) Ill-posed inverse problems in economics. *Annual Review of Economics*. 6:21-51.
4. Kabanikhin S.I. (2008) Definitions and examples of inverse and ill-posed problem. *Journal of Inverse and Ill-Posed Problems*. 16(4):317-357.

## High performance computing in astrophysics. The organic formation in protostellar disc

I. Kulikov

*Institute of Computational Mathematics and Mathematical Geophysics SB RAS*

*\* e-mail: kulikov@ssd.ssc.ru*

**Key word:** high performance computing, computational astrophysics, protostellar disc, chemodynamics

*Abstract:* In this talk, a novel computation technique for numerical simulations of astrophysics structures at the Peta- and Exascale supercomputers is described. The co-design of parallel numerical algorithms for astrophysical simulations is described in detail. The hydrodynamical numerical model for the astrophysical simulation, numerical methods for solving the hyperbolic equations and brief description of parallel implementation of the gooPhi code are described. The results of numerical experiments of simulations of organic formation in protostellar disk are presented.

*Acknowledgements:* The research work was supported by the Grant of the President of Russian Federation for the support of young scientists number MK - 1445.2017.9, RFBR grants 18-01-00166 and 18-07-00757.



# Genome-scale modeling of carbon assimilation in *Geobacillus icigianus*

M. Kulyashov<sup>1\*</sup>, I. Akberdin<sup>1,2,3</sup>, A. Rozanov<sup>2</sup>, S. Peltek<sup>2</sup>

<sup>1</sup>*Novosibirsk National Research University, Novosibirsk, Russia*

<sup>2</sup>*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

<sup>3</sup>*Biology Department and Viral Information Institute, San Diego State University, San Diego, USA*

\* e-mail: m.kulyashov@mail.ru

**Key words:** genome-scale modeling, *Geobacillus icigianus*, metabolic pathways

**Motivation and Aim:** In the last decade, genome-scale modeling became a powerful and useful tool to investigate whole-cell microbial metabolism [1]. For model microbes, such as *E. coli*, a huge amount of experimental data on different hierarchical levels of cellular organization is a basis for the most detailed and complete metabolic reconstruction of the organism [2], while advances in high-throughput techniques and in theoretical approaches pave the way to explore metabolic potential of the organism for non-model microbial species. Here, we represent recently obtained theoretical and experimental results for *Geobacillus icigianus*, a thermophilic bacterium isolated from a hot spring three years ago.

**Methods and Algorithms:** *G. icigianus* was cultivated under the conditions of a bioreactor at pH = 7 with the constant rate of air supply, and in microbiological shakers at pH = 8. Measurements of culture growth were made using a spectrophotometer PE 5400, using a 1 ml plastic cuvette to measure culture from a shaker and a 10 ml glass cuvette for measuring culture from a bioreactor. To build the genome-scale model and its subsequent modifications, we harnessed a sequenced genome [NZ\_JPYA00000000.1] and a set of tools [3–5].

**Results:** We conducted experiments on the cultivation of *G. icigianus* on various substrates and under different conditions. As a result, it was shown that the most effective for cultivation are the substrates: glucose and glycerine, and the optimal conditions: a temperature of 62 °C and pH = 7. To enable the growth of culture we needed to add salts of iron, magnesium, calcium, vitamins, nitrilotriacetic acid and trace elements into the media. The growth rate in experiments with bioreactor was equal to 0.5 h<sup>-1</sup>, while the theoretically predicted value of the growth rate was 0.66 h<sup>-1</sup>.

**Conclusions:** Initial results allowed us to determine the optimal growth parameters that will facilitate further experiments and also demonstrated that developed genome-scale model qualitatively corresponds to the data observed in the experiments. However, it still requires both refinement of the biomass equation and consideration of transcriptomic data for different growth conditions in order to quantitatively reproduce measured growth rate and make reliable predictions on the essentiality of diverse metabolic pathways.

## References

1. Orth J.D., Thiele I., Palsson B.Ø. (2010) Nature biotechnology. 28(3):245.
2. Monk J.M. et al. (2017) Nature biotechnology. 35(10):904.
3. Overbeek R. et al. (2013) Nucleic acids research. 42(D1):D206-D214.
4. Arkin A.P. et al. (2016) bioRxiv, p.096354.
5. Kelley J.J. et al. (2017) Bioinformatics. 33(16):2596-2597.

# Agent-based modelling of genetic deafness propagation under various sociodemographic conditions

S.A. Lashin<sup>1,2\*</sup>, Yu.G. Matushkin<sup>1,2</sup>, A.A. Smirnova<sup>1</sup>, G.P. Romanov<sup>3,4</sup>, O.L. Posukh<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup> *MK Ammosov North-Eastern Federal University, Yakutsk, Russia*

<sup>4</sup> *Yakut Science Centre of Complex Medical Problems, Yakutsk, Russia*

\* e-mail: lashin@bionet.nsc.ru

**Key words:** agent-based modeling, genetic deafness, model “Population Genetics of Deafness”

*Motivation and Aim:* Genetic deafness is a heterogeneous disability with different inheritance patterns, the most common form of which is a recessive deafness caused by mutations in connexin-26 gene (*Cx26*, *GJB2*). High frequency of *Cx26*-associated deafness in European and American populations was previously assumed to be caused by combined effects of assortative mating tradition among deaf people based on linguistic homogamy (sign language) and relaxed selection against deafness [1, 2]. Although there were some attempts to study interrelations between social and genetic factors in spreading of genetic deafness using mathematical modelling and computer simulations [2], there is still lack of both methodological basis and software tools for this task. In this study, we present an agent-based model “Population Genetics of Deafness” (PGD) along with its software implementation and biologically relevant simulation results.

*Methods and Algorithms:* Agent-based model “Population Genetics of Deafness” (PGD) describes propagation of genetic deafness in human populations through generations. The model takes into account social factors of mating, such as: choice-by-phenotype and sign language. Model allows simulating either simple mendelian (one-, two-, or multi-locus, bi- or multi-allelic) or multifactorial genotype-phenotype interactions. The software implementation of PGD is based on Diploid Evolutionary Constructor framework [3].

*Results:* We have considered series of simulations (20-fold for each scenario) of deafness propagation varying the following parameters: initial population size (including frequencies of each cohorts), sign language assortativity and phenotypic assortativity. Simulation results are counter-intuitive versus [1, 2]: increase of assortative mating in deaf subpopulation led to decrease of the frequency of “deaf” alleles after 25 generations. It is especially true for the cases when increased assortativity is associated with relatively high significance of sign language proficiency and low number of its carriers in initial population.

*Conclusion:* We have presented an agent-based model “Population Genetics of Deafness” (PGD) and its software implementation. Simulations performed with PGD have shown counter-intuitive behavior of genetic deafness propagation associated with assortative mating and other social factors like sign language expansion in populations.

*Acknowledgements:* The study was supported by the Budget Project 0324-2018-0017.

## References

1. Nance W.E. et al. (2000) Relation between choice of partner and high frequency of connexin-26 deafness. *The Lancet*. 356(9228):500-501.
2. Nance W.E., Kearsey M.J. (2004) Relevance of connexin deafness (DFNB1) to human evolution. *The American Journal of Human Genetics*. 74(6):1081-1087.
3. Lashin S.A., Matushkin Yu.G. (2013) Multi-layer computer models of gene network evolution in diploid populations. *FEBS journal*. 280:563-563.

# Identifiability analysis of mathematical models of immunology and epidemiology

V. Latyshenko<sup>1,2\*</sup>, O. Krivorotko<sup>1,2</sup>, S. Kabanikhin<sup>1</sup>

<sup>1</sup>*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Novosibirsk State University, Novosibirsk, Russia*

\* e-mail: Latushenko\_varia@mail.ru

**Key words:** ordinary differential equations, inverse problems, practical identifiability, sensitivity-based analysis, immunology, epidemiology

*Motivation and Aim:* Lately ordinary differential equations (ODEs) became the predominant tool in the field of biology (immunology, epidemiology), medicine (pharmacokinetics, tomography), sociology, economics, *etc.* Mathematical models are based on systems of ODEs (regular, nonlinear) and their coefficients characterize, for example, individual characteristics of patient and population in epidemiology field and should be identify for construction of individual treatment plan and the best forecasting of epidemic.

*Methods and Algorithms:* Before determining unknown parameters of models (inverse problem [1]), we should understand: whether is there a solution? how many parameters can we determine from the available data? how many measurements (additional information about solution of ODEs in fixed times) need to be taken to determine the required set of parameters? These questions are answered by analysis of the identifiability [2]. The identifiability is the ability to uniquely determine the values of parameters with sufficient data volume. Analysis of the identifiability represents a study for a deep understanding of the model.

In this work we analyze several methods of identifiability analysis. For practical identifiability we use methods of correlation matrix and Monte Carlo simulation, for sensitivity-based analysis we use orthogonal and eigenvalues methods.

*Results:* Methods of identifiability was applied to different mathematical models of biology (model of dynamic HIV-infection [3], model of spread of co-infection HIV and tuberculosis [4]). Sequences of identified parameters were obtained for each mathematical model.

*Acknowledgements:* The authors were supported by the grant No. AP05134121 of the Ministry of Education and Science of Republic of Kazakhstan, Ministry of Education and Science of Russian Federation and by the Scholarship of the President of RF No. MK-1214.2017.1. and the grant No. 18-71-10044 of Russian Scientific Found (RScF).

## References

1. Kabanikhin S.I. (2008) Definitions and examples of inverse and ill-posed problems. *Inverse Ill-posed Probl.* 16:317-357
2. Miao H., Xia X., Perelson A.S., We H. (2011) On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Rev. Soc. Ind. Appl. Math.* 53(1):3-39.
3. Adams B.M., Banks H.T. et al. (2005) HIV dynamics: Modeling, data analysis, and optimal treatment protocols. *Journal of Computational and Applied Mathematics.* 184:10-49.
4. Roeger L.W., Feng Z., Castillo-Chavez C. (2009) Modelling TB and HIV Coinfections. *Mathematical biosciences and engineering.* 6(4):815-837.

## Parameters sensitivity of pharmacokinetics model parameters

V. Lifenko<sup>1</sup>, D. Voronov<sup>2</sup>,

<sup>1</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup> Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia

\* e-mail: lera.lifenko@gmail.com

**Key words:** identifiability, pharmacokinetics, systems of ordinary differential equations, parameters sensitivity, inverse problem, singular value decomposition

*Motivation and Aim:* Systems of ordinary differential equations are an essential tool for modeling pharmacokinetic processes. In practice it is important not only to determine unknown parameters of the observed process but also to make sure that the solution is stable to errors in the experimental data [1].

*Methods and Algorithms:* Identifiability analysis is the first step in determining unknown parameters. Investigating parameters sensitivity to the data gives the necessary information for obtaining more stable solutions. A new technique for sensitivity analysis using singular value decomposition of the sensitivity matrix is shown. It was shown the influence of the stability of the parameters on the convergence of the gradient method. In addition, an overview of the methods for identifiability in nonlinear mixed effects models was given [2].

*Results:* A brief historical review of identifiability analysis methodologies for dynamic systems is presented. Methods of sensitivity analysis are reviewed in more details. Finally, some examples and numerical results for pharmacokinetic models of the kinetics and secretion of the C-peptide, the glucose-insulin control model, and the pharmacokinetics model of the Digoxin are presented.

*Conclusion:* Qualitative analysis of the sensitivity of mathematical models allows to avoid receiving unhelpful data and unnecessary expenses for experiments. Therefore, by investigating dynamical systems using the methods of identifiability analysis, we find sensitive parameters. Applying bounding conditions on parameters, or fixing them, we get a more stable problem.

*Acknowledgements:* This work was supported by RFBR grant No. 17-31-50060.

### References:

1. Hongyu M. et al. (2011) On Identifiability of Nonlinear ODE Models and Applications in Viral Dynamics.
2. Lavielle M., Aarons L. (2016) What do we mean by identifiability in mixed effects models?

## Bayesian approach to big data processing: problems and perspectives

M.A. Marchenko

*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*

*Novosibirsk State University, Novosibirsk, Russia*

\* e-mail: marchenko@sscc.ru

**Key words:** Big Data, Bayesian inference, Markov Chain Monte Carlo, high-performance computations

*Motivation and Aim:* We desire to apply Bayesian inference (BI) methods for analysis and processing of Big Data arising in system biology, genomics, phylogenetics, etc. The BI methods have already proved their efficiency in many applications, such as nuclear physics, economics, genetics, etc. [1, 2]. In the BI approach, the Markov Chain Monte Carlo (MCMC) methods are used for simulation of the underlying random processes to get the statistical output, which is the evaluation of the model parameters [3, 4].

Unfortunately, standard MCMC methods can scale poorly to big data settings due to the need to evaluate the likelihood at each iteration [5]. There have been a number of approximate MCMC algorithms that use sub-sampling ideas to reduce this computational burden, but with the drawback that these algorithms no longer target the true posterior distribution.

*Methods and Algorithms:* We developed the PARMONC (PARallel MONte Carlo) solver for distributed stochastic simulation on clusters with massive-parallel and hybrid architectures [6]. A core of the PARMONC is the well-tested, fast and reliable 128-bit parallel random numbers generator, which is in intensive use for more than a decade.

*Results:* The BI approach and the MCMC methods began to apply on all architectures of computing systems with parallel and distributed computing. We propose the complex methodological approach: the parallel random number generators, libraries, data processing, control programs, etc., i.e. all the stages of creating a “digital product”.

*Conclusion:* New theoretical study and high-performance MCMC methods are needed to make the BI approach a highly effective technique to process Big Data.

*Acknowledgements:* Supported by the RFBR (18-01-00599, 18-41-540017, 16-01-00755, 16-01-00530).

### References

1. Wong K.-C. (2016) Big data analytics in genomics. Springer. ISBN 9783319412788
2. G.M. Allenby et al. (2014) Perspectives on Bayesian Methods and Big Data. Cust. Need. and Solut. 1:169-175.
3. Ahmed S.E. (2016) Big and complex data analysis, contributions to statistics. Springer. ISBN 9783319415727.
4. Insua D. et al. (2012) Bayesian Analysis of Stochastic Process Models. Wiley. ISBN 978-0-470-74453-6.
5. Solonen A. et al. (2012) Efficient MCMC for climate model parameter estimation: parallel adaptive chains and early rejection. Bayesian Anal. 7(3):715-736.
6. Marchenko M. (2014) Efficient Computational Approaches for Parallel Stochastic Simulation on Supercomputers. Parallel Programming: Practical Aspects, Models and Current Limitations. New York: Nova Science Publishers: 117-142.

# The *Multiplex Phase Interlocker*: a novel and robust molecular design synchronizing transcription and cell cycle oscillators

T.D.G.A. Mondeel<sup>1</sup>, C. Linke<sup>1</sup>, S. Tognetti<sup>2</sup>, W. Liebermeister<sup>3</sup>, M. Loog<sup>4</sup>,  
H.V. Westerhoff<sup>1</sup>, F. Posas<sup>2</sup>, M. Barberis<sup>1\*</sup>

<sup>1</sup> Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands

<sup>2</sup> Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain

<sup>3</sup> Institut für Biochemie, Charité - Universitätsmedizin Berlin, Berlin, Germany

<sup>4</sup> Institute of Technology, University of Tartu, Tartu, Estonia

\* e-mail: M.Barberis@uva.nl

**Key words:** network motifs, network dynamics, cell cycle, autonomous oscillations, timing, waves of cyclins, budding yeast, Multiplex Phase Interlocker

**Motivation and Aim:** The eukaryotic cell cycle is robustly designed, with molecules interacting and organized within definite network topologies. A transcriptional oscillator interlocks with waves of cyclin-dependent kinases (cyclin/Cdk) to guarantee execution of a timely cell cycle progression. Although details about transcription of cyclins, the regulatory subunits of these kinases, are available, a lack of understanding exists about network motifs responsible for the precise timing of cyclin/Cdk oscillations. Here we investigate the robustness of molecular designs interlocking the transcriptional and cyclin/Cdk oscillators in budding yeast. We have recently identified a transcriptional cascade that regulates the relative timing of waves of mitotic (Clb) cyclin expression, which involves the Forkhead (Fkh) transcription factors (TF) [1]. Here we aim to unravel the network motif(s) responsible for timely cyclin/Cdk oscillations that interlock Clb waves through Fkh-mediated signaling.

**Methods and Algorithms:** An integrated computational and experimental framework is presented. A kinetic, ODE model of the cyclin/Cdk network is simulated under a quasi-steady state assumption, and fitted to *in vivo* quantitative, time course data of Clb dynamics. Robustness analyses are then performed by testing 1024 possible network motifs for their ability (i) to fit Clb oscillations and (ii) to generate sustained oscillations in the form of limit cycles, on which sensitivity analysis is conducted.

**Results:** A novel regulatory motif, coined as *Multiplex Phase Interlocker*, is unraveled, that timely synchronize Clb oscillations. This motif uniquely describes a molecular timer (TF) that relies on separate inputs (Clb/Cdk) converging on a common target (TF itself). Within the motif, a progressive TF activation may be realized by multiple Clb/Cdk. Experimental validation supports computational analyses, with the Clb/Cdk-Fkh axis being pivotal for timely transcriptional dynamics.

**Conclusion:** Altogether, our integrative approach pinpoints how robustness of cell cycle control is realized by revealing a novel and conserved principle of design that ensures a timely interlock of transcriptional and cyclin/Cdk oscillations.

**Acknowledgements:** Supported by the SILS Starting Grant of the University of Amsterdam, UvA and by the UvA-Systems Biology Research Priority Area Grant.

## References

1. Linke C., Chasapi A., González-Novo A., Al Sawad I., Tognetti S., Klipp E., Loog M., Krobitch S., Posas F., Xenarios I., Barberis M. (2017) A Clb/Cdk1-mediated regulation of Fkh2 synchronizes CLB expression in the budding yeast cell cycle. *Nature NPJ Systems Biology and Applications*. 3:7.



# Developing FoldGO, the tools for multifactorial functional enrichment analysis

A.M. Mukhin<sup>1,2\*</sup>, D.S. Wiebe<sup>1,2</sup>, I. Grosse<sup>2,3</sup>, S.A. Lashin<sup>1,2</sup>, V.V. Mironova<sup>1,2</sup>

<sup>1</sup> Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>3</sup> Martin-Luther University, Halle-Wittenberg, Germany

\* e-mail: mukhin@bionet.nsc.ru

**Key words:** functional enrichment analysis, differentially expressed genes, GO terms, web-service, R, RServe, Java, Spring, JavaScript, Vue.JS

*Motivation and Aim:* Due to development of whole-genome sequencing technologies researchers generate a big amount of tabular data about differentially expressed genes (DEGs), those with statistically significant change of transcription level in response to a stimulus or condition. Further analysis of DEGs is an actual task. DEGs are the genes, that. One of the conventional methods of DEGs' analysis is Functional Enrichment Analysis using Gene Ontology (GO) – a unified semantic system for gene annotation. In particular, this analysis provides for a list of enriched GO terms (molecular and biological function and cell component). There are online services DAVID and agriGO implemented this method. However, if one needs to study multiple connected gene lists, e. g., classified by the fold change, it is impossible to do it correctly in these services. The aim of this work was to develop a web-service implementing multifactorial functional enrichment analysis method.

*Methods:* We developed a new method, which classifies DEGs according to their fold change, performs functional annotation of the multiple gene lists generated on the basis of this classification (with overlaps) and provides for the GO terms, significantly overrepresented for the genes responded to the factor within a narrow fold-change-interval. The tools were developed using several programming languages and technologies:

- R – a programming language for R-package
  - RServe – a library, which provides access to the R by using binary protocol
- Java – a programming language for writing a server, that will transmit R inputs and parameters
- Spring – a framework for developing web-services
- JavaScript – a programming language working in a browser for building a web-client
  - Vue.JS – a library for building the interface

*Results:* The FoldGO tools are implemented as an R package and as a WebService. We tested FoldGO tools on a set of transcriptomes generated for different organisms, including Arabidopsis [1]. We found fold-change-specific GO categories within most of tested datasets. The web-service is deployed to the network of ICG SB RAS and the project is tested by users.

*Acknowledgments:* The work was supported RFBR grant No. 18-34-00871.

## References

1. Omelyanchuk et al. (2017).



## Mathematical modeling of medicinal preparations diffusion process in tissues of the person

A. Nafikova

*Sterlitamak branch of the Bashkir state university, Sterlitamak, Russia*  
*e-mail: albinabikbaeva@gmail.com*

**Key words:** diffusion, medicinal preparations, mathematical modeling, boundary problem

*Motivation and Aim:* The substance transport problem in the diffusing stratified mediums is a subject of researches both theorists, and practitioners of various areas: medicine, ecology, geology and geophysics. So, in the field of medicine problems of medicinal preparations concentration achievement, necessary for treatment, in the tissues of the person having, as a rule, bedded structure in necessary time slices are relevant. The known mathematical models of medicinal preparations transfer at the surface or intramuscular influences have an appearance of a one-dimensional boundary value problem in the patch and homogeneous stratified mediums. But the used models do not consider anisotropy of diffusion properties of tissues or possible local formations of the changed tissues (for example, oncological tumors) [1]. Therefore, the research of substance mass transfer processes in piecewise constant non-isotropic stratified mediums with inclusions on the example of medicinal preparations diffusion process method of mathematical model operation has important theoretical and applied value and can make an essential contribution to carrying out further scientific research in the field of medicine.

*Methods and Algorithms:* Means of the computer Maple system developed the programs realizing numerical algorithms of diffusing liquid concentration distribution function finding in the patch and homogeneous horizontally stratified medium with plane-parallel borders, inversion of integral Laplace transform on the basis of the generalized quadrature formulas of the most advanced stage of accuracy [2].

*Results:* The comparative analysis of computer modeling received results with presented in the considered works of other scientists (for example, [3]) which showed adequacy and reliability of the problem solution offered algorithms is carried out.

*Conclusion:* The received results in the field of medicinal preparations diffusion process mathematical modeling in tissues of the person are important for definition of parameters number, such, for example, as the dose of the injected medicinal preparation sufficient for achievement of tissues internal departments, time necessary for efficient influence of medicine and so on that is of great importance for treatment of diseases number. Work on creation of three-dimensional problem mathematical model about distribution of the diffusing liquid concentration in the multilayer medium modeling tissue of the person and development of its decision algorithms is at the moment conducted.

*Acknowledgements:* Supported by grant of the Bashkir state university Sterlitamak branch.

### References

1. Nafikova A.R. (2015) Mathematical modeling of radon transfer processes in piecewise constant non-isotropic stratified mediums with inclusions. PhD phys. and math. sci. diss. Sterlitamak. 101 p.
2. Matveeva T.A. (2003) Some methods of a Laplace transform inversion and their application. PhD phys. and math. sci. diss. St. Petersburg. 117 p.
3. Stolnits M.M. et al. (2008) Mathematical model of medicinal preparations and immersion liquids diffusion in tissues of an eye of the person. News of the Saratov University. Physics series. 8(1):15-20.

## The possibilities of a Universal computer model in the readiness assessment of the Russian regions resource to epidemics of especially dangerous infectious diseases

L. Nizolenko\*, A. Bachinsky

State Research Center of Virology and Biotechnology Vector, Koltsovo, Russia

\* e-mail: nizolenko@vector.nsc.ru

**Key words:** epidemics, mathematical model, limited resources

*Motivation and Aim:* Since it is impossible to perform epidemiological experiments in human populations, mathematical modeling is the very important tool for studying the effects of particular factors on the dynamics of epidemics. Mathematic approaches can be especially useful to estimate the level of biological safety of the territories.

*Methods and Algorithms:* The tool, developed in the SRC VB Vector, is based on the model simulating epidemics of acute infectious diseases, where the main modes of transmission are from an external source or by casual contact between people. Detailed description is done in [1].

*Results:* The model provides a number of anti-epidemic activities: isolation of the infected persons, contacts, and suspects, vaccination, quarantine, treatment. But a certain amount of resources is needed for their implementation. This feature allows us to use the model to assess the resource preparedness of Russian regions for epidemics of especially dangerous infectious diseases. Of particular relevance is the assessment of the most “expensive” resources, which must be constantly maintained at a certain level, to consider the region prepared for the epidemic, although the epidemic, may be, never begin. Namely: health workers trained to work with quarantine infections; places for strict isolation (for confirmed cases); places in provisional hospitals (for suspects) and places in quarantine departments for contacts.

To assess the level of resource readiness, the model provides the following opportunities: Computation of the epidemic dynamics, depending on the number of people initially infected.

If it turns out that there are not enough resources, it is possible to optimize their reserve. This function allows by repeatedly automatic calculations in accordance with a certain algorithm to select the minimum necessary costs to maximally mitigate the consequences of the epidemic.

Computation “typical” scenarios for epidemic development in the case of a mass infection in real subjects of the Russian Federation. The effect of delay with countermeasures is assessed here in addition to resources.

*Conclusion:* A Universal model of epidemics can be recommended as a tool for assessing of regions preparedness for epidemics of especially dangerous infections.

*Availability:* <http://vector-epimod.ru>

### References

1. Bachinsky A.G., Nizolenko L.P. (2013) A universal model for predicting dynamics of the epidemics caused by special pathogens. BioMed Research International. 2013(467078):7.

# The 2D coefficient inverse problem of the ultrasound waves propagation

N. Novikov<sup>1, 3\*</sup>, M. Shishlenin<sup>1, 2, 3</sup>

<sup>1</sup>*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*

<sup>3</sup>*Novosibirsk State University, Novosibirsk, Russia*

\* e-mail: novikov-1989@yandex.ru

**Key words:** inverse problem, acoustic equations, gradient descent method

*Motivation and Aim:* Ultrasound imaging is intensively developing the past two decades. Various methods have been developed, including magnetic resonance, positron emission, optical tomography, etc. x-ray tomography has become the most popular among these methods both in medicine and in industrial diagnostics. However, the disadvantage of medical x-ray tomography devices is a high dose of radiation that prevents their application in the regular examination. The limitation of exposure of ionizing radiation has become a current trend in medicine, which makes the development of methods of ultrasonic tomography of paramount importance. One of the major problems of ultrasound tomography is the development of methods for solving inverse problems which are nonlinear.

*Methods and Algorithms:* We consider the direct and inverse problem for the hyperbolic system of equations, that describes the two dimensional acoustic wave propagation. The hyperbolic first-order system allows us to propose more realistic model from the physical point of view. On the other hand, we can apply optimization approach for recovering coefficients of such system, like the density of the medium, the speed of the wave propagation or the attenuation coefficient. We consider the numerical algorithm for solving direct problems, based on the S.K. Godunov scheme. In order to solve the inverse problem, we use optimization approach and propose gradient-based method to minimize the cost functional.

*Results:* We present the result of direct problem solution. We obtain the gradient of the functional for recovering the density and the acoustic wave's velocity in the medium. We use it to obtain the solution of inverse problem and study the behavior of the cost functional.

*Acknowledgements:* Supported by the RFBR 16-29-15120 and by MSC RK grant No. AP05134121.

## References

1. Kabanikhin S.I., Shishlenin M.A. (2008) Quasi-solution in inverse coefficient problems. *Journal of Inverse and Ill-posed Problems*. 16(7):705-713.
2. Kulikov I.M., Novikov N.S., Shishlenin M.A. (2015) Mathematical modelling of ultrasound wave propagation in two dimensional medium: direct and inverse problem. *Siberian Electronic Mathematical Reports* 12:219-228.
3. Kabanikhin S.I., Novikov N.S., Oseledets I.V., Shishlenin M.A. (2015) Fast Toeplitz linear system inversion for solving two-dimensional acoustic inverse problem. *Journal of Inverse and Ill-posed Problems*. 23(6):687-700.

# The optimal feedbacks in the mathematical model of chemotherapy for a nonmonotonic therapy function

N. Novoselova

*Krasovskii Institute of Mathematics and Mechanics, Ural Branch of the RAS, Yekaterinburg, Russia*  
*Ural Federal University, Yekaterinburg, Russia*

\* e-mail: n.g.novoselova@gmail.com

**Key words:** therapy function, optimal control, optimal feedbacks, Hamilton–Jacobi–Bellman equation, Cauchy method of characteristics, Rankin-Hugoniot line

*Motivation and Aim:* We investigate a pharmacokinetic problem for a deterministic nonlinear system with piecewise monotonic dynamics describing the process of chemotherapy of a malignant tumor. We consider the case when the therapy function, which describes the effect of the drug on the cell growth rate, has two maxima.

*Methods and Algorithms:* The work presents results of numerical calculation of the optimal result (the value function) and optimal positional strategy of therapy (optimal feedbacks) in a corresponding optimal control problem. The construction use the fact that the value function is the unique minimax (viscosity) solution [1, 2] of the Cauchy problem for the basic Hamilton–Jacobi–Bellman (HJB) equation. By means of the continuous gluing of a finite number of smooth functions obtained by the Cauchy method of characteristics for auxiliary linear HJB equations, the continuous function  $\phi$  is constructed. The paper [3] proves that the constructed function  $\phi$  coincides with the value function.

*Results:* A new element of the construction is the construction of a line of nonsmooth gluing using the Rankin-Hugoniot conditions [4, 5]. This line plays a key role for the optimal feedback strategy, because it determines its discontinuity line. The results of numerical calculations of the Rankin-Hugoniot line are exposed. Comparisons with the results for the case of a single maximum in the therapy function in this model [6] is given.

*Acknowledgements:* Supported by the RFBR (No. 17-01-00074).

## References

1. Crandall M.G., Evans L.C., Lions P.-L. (1984) Some properties of viscosity solutions of Hamilton-Jacobi equations. *Trans. Amer. Math. Soc.* 282(2):487-502.
2. Subbotin A.I. (1995) *Generalized Solutions of First-Order PDEs: The Dynamical Optimization Perspective*. Birkhauser: Boston.
3. Subbotina N.N., Novoselova N.G. (2017) Optimal result in a control problem with piecewise monotone dynamics. *Journal “Proceedings of the Institute of Mathematics and Mechanics”*. 23(4):265-280.
4. Goritskiy A.Yu., Kruzhkov S.N., Chechkin G.A. (1999) *First-Order Partial Differential Equations (Study Guide)*. Moscow: Lomonosov Moscow State University, Faculty of Mechanics and Mathematics.
5. Subbotina N.N., Kolpakova E.A., Tokmantsev T.B., Shagalova L.G. (2013) *The Method of Characteristics for Hamilton Jacobi Bellman equations*. RIO UrO RAN, Yekaterinburg.
6. Chumerina E.S. (2016) Choice of optimal strategy of tumor chemotherapy in Gompertz model. *Journal of Computer and Systems Sciences International*. 48(2):325-331.

## Mathematical phantoms development for computer simulation of the patient examination procedure by a positron emission tomography method

M. Ondar<sup>1, 2\*</sup>, N. Denisova<sup>1</sup>

<sup>1</sup> *Institute of Theoretical and Applied Mechanics SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State Technical University, Novosibirsk, Russia*

\* e-mail: [ondar\\_marina93@mail.ru](mailto:ondar_marina93@mail.ru)

**Key words:** positron emission tomography (PET), radiopharmaceutical, anthropomorphic phantoms

*Motivation and Aim:* The method of positron emission tomography (PET) is the “gold standard” in oncology. A 90 % of PET method studies use a radiopharmaceutical fluorodeoxyglucose (FDG) with a property of accumulating in malignant tumors. Therefore, the PET method makes it possible to distinguish the benign and malignant formations. A three-dimensional image creation is the result of solving the ill-posed problem of reconstruction. It is well known, that in solving such problems there are artifacts related to instability of the solution and they are dependent on the source function. The most effective way is the method of mathematical modeling to investigate the causes of the artifacts appearance. The modeling procedure includes the next steps: 1) model creation – source function, 2) raw data calculation using created model, 3) image acquisition based on the chosen method for solving the ill-posed reconstruction problem, 4) comparison of the obtained image with a specified source function and errors analysis of reconstruction. Because of radiation exposure, it is impossible to conduct research in humans in the field of nuclear medicine. To simulate the PET procedure, it is necessary to develop anthropomorphic phantoms simulating the radiopharmaceutical distribution in the patient’s organs. The aim of this work is the development of anthropomorphic phantoms describing the distribution of the FDG radiopharmaceutical for mathematical modeling of the procedure for acquisition the “whole body image” using a PET method. *Methods and Algorithms:* The anthropomorphic phantoms development is carried out on the basis equations of analytic geometry that describe spatial geometric figures using Boolean logical operators.

*Results:* A mathematical model describing the distribution of the FDG radiopharmaceutical throughout the body is developed to simulate the procedure for examining cancer patients using the PET method to diagnose the presence or absence of metastases.

*Conclusion:* These studies are part of the work of the scientific group of ITAM SB RAS to improve the methods of reconstruction of PET images. These phantoms will be used for the subsequent computer simulation of PET: the raw data calculation recorded by a gamma camera; images reconstruction using standard algorithms.

*Acknowledgements:* The work was supported by grant from RFBR (No. 17-52-14004).

## DEPPDB v.3: a portal to study electrostatic and other physical properties of genome DNA and its elements

A. Osypov<sup>1,2\*</sup>, G. Krutinin<sup>3</sup>, E. Krutinina<sup>3</sup>, P. Beskaravayny<sup>3</sup>, S. Kamzolova<sup>3</sup>

<sup>1</sup>*Institute of Higher Nervous Activity and Neurophysiology RAS, Moscow, Russia*

<sup>2</sup>*Institute of Theoretical and Experimental Biophysics RAS, Pushchino MR, Russia*

<sup>3</sup>*Institute of Cell Biophysics RAS, Pushchino MR, Russia*

\* e-mail: aosypov@gmail.com

**Key words:** DNA electrostatics, transcription regulation, genome evolution

*Motivation and Aim:* DNA is a highly charged molecule and its electrostatic and other physical properties define its shape in the functional space and greatly influence its interactions with different proteins, especially regulating transcription, in particular RNA-polymerases and transcription factors.

*Methods and Algorithms:* The electrostatic potential around the double-helical DNA molecule was calculated by the original method [1] using a specially developed program package [2, 3]. Calculations of other physical properties are based on the di- and trinucleotide content. Different cross-correlation analysis algorithms are applied.

*Results:* DEPPDB contains all completely sequenced archaeal, bacterial, viral, mitochondrial, plastids and eukaryotic genomes according to current release of NCBI RefSeq [9]. Data for promoters, regulation sites, binding proteins etc. are incorporated from different databases and compiled from original literature annotation. All data are fully integrated, several tools are provided to support different forms of analysis. Calculation on the fly of the user-provided sequences and selected GeneBank records is available. DEPPDB can be considered as a portal or collection of databases on the electrostatic and other physical properties of whole genomes and different genome elements in different taxa and organisms: Promoter DB, Regulatory Sites (Transcription Factors, TF) DB, Gene Starts DB, Terminator DB, etc. as well as comprehensive analysis toolbox. Several calculation algorithms, data processing approaches and database architecture solutions were fundamentally revisited and fine tuned since the previous database releases to withstand the considerably increased data abundance.

*Conclusion:* DEPPDB was developed to hold and provide all available information on these properties of genome DNA combined with its sequence and annotation of biological and structural properties of genome elements and whole genomes, organized on a taxonomical basis. DB is available at <http://deppdb.psn.ru>.

*Acknowledgements:* Supported by the RFBR (No. 16-04-01865 and 18-34-00942).

### References

1. Polozov R.V. et al. (1999) Electrostatic potentials of DNA. Comparative analysis of promoter and nonpromoter nucleotide sequences, *J. Biomol. Struct. Dyn.* 16(6):1135-1143.
2. Osypov A.A. et al. (2010) DEPPDB – DNA Electrostatic Potential Properties Database. *Electrostatic Properties of Genome DNA*, *JBCB*. 8(3):413-25.
3. Osypov A.A. et al. (2012) DEPPDB – DNA Electrostatic Potential Properties Database. *Electrostatic Properties of Genome DNA elements*, *JBCB*. 10(2):1241004.



# Complex information system to study common energy metabolic deficiency under neurodegenerative diseases

A. Osypov<sup>1,2\*</sup>, I.Yu. Popova<sup>2</sup>

<sup>1</sup> *Institute of Higher Nervous Activity and Neurophysiology RAS, Moscow, Russia*

<sup>2</sup> *Institute of Theoretical and Experimental Biophysics RAS, Pushchino MR, Russia*

\* e-mail: aosypov@gmail.com

**Key words:** neurodegenerative diseases, complex studies, information systems

*Motivation and Aim:* Hypometabolism, characterized by decreased brain glucose consumption, is a common feature of many neurodegenerative diseases. Initial hypometabolic brain state, created by characteristic risk factors, may predispose the brain to acquired epilepsy and sporadic Alzheimer's and Parkinson's diseases. Deficient glucose metabolism is likely a primary initiating factor, and resulting neuronal dysfunction further promotes the metabolic imbalance, establishing an effective positive feedback loop. Metabolic correction leading to the normalization of abnormalities in glucose metabolism may be an efficient tool to treat the neurological disorders by counteracting their primary pathological mechanisms [1].

*Methods and Algorithms:* Database architecture, business logics, data analysis, and modeling programming.

*Results:* We develop an integrative information system to hold and analyze all the data types that originate from complex biological studies of neurodegenerative diseases model objects. These include data from differential expression and individual genome analysis, metabolic and regulatory pathways and their modeling, metabolic profiling by NMR, enzymology essays and *in vitro* and *in vivo* monitoring, mitochondria studies, electro-physiology *in vivo* and *in vitro* data and their modeling, behavioral and cognitive tests, histological and morphological data *etc &c*.

All data is collected in normal healthy stage as well as under specific diseases such as epilepsy or Alzheimer's disease and several pathology models as well as individuals treated with the energy supply metabolites (i.e. pyruvate) under the normal and pathological conditions.

That will help to reveal the underlying basis of neurodegenerative diseases and the neuroprotective effects of energy metabolic correction and further elaborate the patients treatment strategy. Another planned direction is the analysis of the information of the individual viability to the pathological factors that may reveal the weak and strong parts of the system and the potential targets to individual treatment in the frame of individualized medicine.

*Conclusion:* Here we propose a complex information system to accommodate all possible data from studies of common energy metabolic deficiency under neurodegenerative diseases and provide means to its complex analysis and modeling.

*Acknowledgements:* The work is supported by the fundamental research program of the Presidium of the Russian Academy of Sciences "Fundamental Research for Biomedical Technologies" for 2018.

## References

1. Zilberter Y., Zilberter M. (2017) The vicious circle of hypometabolism in neurodegenerative diseases: Ways and mechanisms of metabolic correction. J Neurosci Res. DOI 10.1002/jnr.24064.



# An algorithm for tracking *C. elegans* body movement and muscular activity in $\text{Ca}^{2+}$ dynamics video for tuning and validation of its locomotion simulation

A.Yu. Palyanov

*A.P. Ershov Institute of Informatics Systems SB RAS, Novosibirsk, Russia*

*Novosibirsk State University, Novosibirsk, Russia*

*e-mail: palyanov@iis.nsk.su*

**Key words:** *C. elegans*, locomotion, muscular activity, biomechanics, validation, simulation, Sibernetic

**Motivation and Aim:** Biologically reasonable computational simulation of the *C. elegans* nematode, including models of its sensory, nervous and muscular systems embedded into a body which is able to operate in a virtual physical environment, becomes closer each year, including the efforts of the OpenWorm project [1]. High-resolution explicit 3D simulation of *C. elegans* swimming and crawling, driven by artificial periodical signals activating its muscles, has been successfully performed using the Sibernetic environment [2]. Real or simulated worm's trajectory is determined by many factors, including neural activity, biomechanical properties of the body and muscles and properties of the physical environment as well. Overall neural activity and particularly the rhythmic patterns generation mechanism is still a challenge, so it seems rational to perform tuning and validation of virtual worm's nervous and muscular systems independently. Recently published experimental data of  $\text{Ca}^{2+}$  dynamics in body wall muscles of wild-type freely moving *C. elegans* [3], available as a video (<https://www.youtube.com/watch?v=x861Pl1jpR8>), provided the capability to use it for validation of simulated muscles, body and environment.

**Methods and Algorithms:** In the present work the video mentioned above was split into a sequence of numbered frames (with the VirtualDub software) and an algorithm for tracking *C. elegans* body and obtaining muscular activity at each frame was designed and implemented in C++ with usage of FreeImage library. Image recognition includes raw detection of worm body outline, its further refinement, detection of head and tail points and calculation of worm body midline. The latter is further split into 100 equal length segments, and then every node is used to build the line orthogonal to body midline, which crosses left and right muscle bundles, which allows taking into account local body bending caused by current muscles stretching and contraction. and information about  $\text{Ca}^{2+}$  concentration is gathered.

**Results:** Muscular activity of real crawling *C. elegans* was extracted from 20 seconds long video with 100 ms interval between frames. Resulting post-processed video is available (<https://www.youtube.com/watch?v=Bs72aNroKx0>). Obtained data has been mapped onto *C. elegans* muscles layout and synchronized with Sibernetic simulation time scale. Resulting simulated movement trajectory has been studied and compared with the real one. Perspectives of their difference minimization via optimization of body, muscles and environment related parameters are discussed.

**Acknowledgements:** This work was supported by the RFBR grant (No. 18-07-00903).

## References

1. Szigeti B., Gleeson P., Vella M., Khayrulin S., Palyanov A., Hokanson J., Currie M., Cantarelli M., Idili G., Larson S. (2014) OpenWorm: an open-science approach to modeling *Caenorhabditis elegans*. *Frontiers in Computational Neuroscience*. 8:1-7.
2. Palyanov A., Khayrulin S., Larson S. (2016) Application of smoothed particle hydrodynamics to modeling mechanisms of biological tissue. *Advances in Engineering Software*. 98:1-11.
3. Lefebvre C. et al. (2016) The ESCRT-II proteins are involved in shaping the sarcoplasmic reticulum in *C. elegans*. *J Cell Sci*. 129:1490-1499.

# Inverse modelling of diffusion-reaction processes with image-type measurement data

A. Penenko<sup>1,2\*</sup>, Z. Mukatova<sup>1,2</sup>, S. Nikolaev<sup>3</sup>, U. Zubairova<sup>3</sup>

<sup>1</sup>*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup>*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

\* e-mail: a.penko@yandex.ru

**Key words:** diffusion-reaction model, inverse problem, image analysis, operator equation

*Motivation and Aim:* Nonlinear diffusion-reaction models can be applied to the study of development [1] and transport [2] processes in biological tissues. The inverse problems for nonlinear diffusion-reaction models with image-type data arise when the unobservable parameters like synthesis rates or model coefficients are of interest and the additional measurement data is obtained in the form of concentration distribution images (e.g. from microscopy or tomography devices). The aim of the work is to develop an efficient highly parallelizable numerical algorithm for the considered inverse problems.

*Methods and Algorithms:* The inverse problem sensitivity operators constructed from the ensemble of the adjoint equations solutions allow us to transform the inverse problem stated as the system of nonlinear ODE or PDE to the family of nonlinear operator equations depending on the given set of orthogonal functions in the space of the measurement results [2, 3]. For the solution of the operator equations, we consider Newton-Kantorovich type methods based on the truncated SVD. The operator form of the inverse problem can be further exploited for the analysis and comparison of the different inverse problem statements with the help of the spectral methods.

*Results:* The inverse problem solution algorithm based on the diffusion-reaction model sensitivity operator and ill-posed operator equations solution methods was developed.

*Conclusion:* The developed algorithm can be used to solve the inverse problems for the diffusion-reaction models with the image-type measurement data.

*Acknowledgements:* Supported by the RSF project No. 17-71-10184 in the part of the algorithm development and by RFBR project No. 15-29-04875 in the part of the experimental images acquisition.

## References

1. Nikolaev S.V. et al. (2013) Model of structuring the stem cell niche in shoot apical meristem of *Arabidopsis thaliana*. Doklady Biological Sciences, 452(1):316-319.
2. Penenko A.V. et al. (2016) Numerical Algorithms for Diffusion Coefficient Identification in Problems of Tissue Engineering. Math. Biol. Bioinf. 11(2):426-444.
3. Penenko A. (2018) Consistent numerical schemes for solving nonlinear inverse source problems with the gradient-type algorithms and the Newton-Kantorovich methods. Siberian J. Num. Math. 21(1):83-99.

## The use of Kirlian photography in preventive medicine and the education

L.A. Pesotskaya<sup>1</sup>, T.V. Lakiza<sup>1</sup>, N.V. Glukhova<sup>2</sup>, T.O. Tretiak<sup>3</sup>

<sup>1</sup> *Dnipropetrovsk medical Academy of Ministry of health of Ukraine*

<sup>2</sup> *National mining University, Dnipro, Ukraine*

<sup>3</sup> *National University name Oles Honchar*

\* e-mail: lpesotskaya23@gmail.com

**Key words:** Kirlian photography, preventive medicine, REK-1

The aim of our work was to study a state of health at children and adults using a method of Kirlian photography before and after influence of improving actions homoeopathic, fito - preparations, minerals.

For this inspections applied device “REK-1”, developed in Dnepropetrovsk. Researches were held on a X-ray film. It’s analysis were held on by diagnostic criteria P. Mandel and results of our observations. A normal luminescence crown around fingers is represented by an inner oval, a middle luminescence layer in the form of evenly located streamers. Inspected the 86 children at the age from 8 till 15 years from families of liquidators of consequences of Chernobyl. The comparative analysis of Kirlian-photos after stay in sanatorium has revealed positive dynamics in an energy- information’s condition of an organism at the majority of children. The best results in the threat and endocrinr zones are received in group of children receiving homoeopathic preparations.

Kirlian-diagnostics have been tried on 57 workers of one of the industrial enterprises. As a Kirlian-luminescence at everyone level reactions of adaptation has been established and therapy by phytopreparations

The Kirlian-diagnostic was on 40 persons before stay during 10–20 minutes in clay or shungite rooms. At the moment of inspection people were almost healthy or had diseases out of an aggravation. The phenomena of emotional liability, dysfunction endocrine regulation equally effectively decreased in both groups.

Kirlian-diagnostic’s method examined 56 students of Junior and senior courses of study at the University. At pupils on the basis of psychological tests types of thinking were defined: figurative, logical, intuitive. The obtained results were compared with the results of the analysis area of the crown’s fluorescence of the fingers in the Kirlian photos after each test. Revealed differences in Kirlian images with different types of thinking. Activation of different reactive systems of the organism was observed in according to the age.

The Kirlianography method of human adaptation reserves functional state research should be used in health-improving and medical practice, ecological population health-improving programmes for the estimation of the negative factors of the environment or the existing disease as well as the choice of a needed therapy.

It is advisable to use kirlianography in the pedagogical process to adapt it to the prevailing type of thinking in the group of students.

# Computer system for reconstructing and analyzing random structural models of protein-protein interaction networks

N.L. Podkolodnyy<sup>1,2\*</sup>, D.A. Gavrilov<sup>3</sup>, O.A. Podkolodnaya<sup>1</sup>

<sup>1</sup>*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*

<sup>3</sup>*Novosibirsk State University, Novosibirsk, Russia*

\* e-mail: [pnl@bionet.nsc.ru](mailto:pnl@bionet.nsc.ru)

**Key words:** protein interaction network analysis, random graph models, circadian rhythm

*Motivation and Aim:* Modern experimental technologies in molecular biology allow the reconstruction of various types of biological networks, including gene and metabolic networks, interaction networks, gene co-expression networks, a network of diseases, etc. In this regard, the development of methods and tools for constructing structural random models, preserving the structural and functional features of the biological network, is an urgent task of computer systems biology. The structural random models are needed to search for patterns of the structural and functional organization of biological networks, to study their influence on the function of biological networks, to test various statistical hypotheses on networks, to search for network biomarkers, and so on.

*Methods and Algorithms:* The Markov chain simulation method was used for generating connected random graphs that preserve the structural features of the biological network. The computational server asynchronously executes a client's (Cytoscape plugin's) request for reconstruction of structural random models of biological networks using the software libraries Random Network Generator [1] and new our method based on GTrie Scanner.

*Results:* We developed the computer system for reconstructing structural random models of biological networks preserving the following structural characteristics: the distribution of vertex degrees, the joint degree distribution, the average degree of neighboring vertices, the clustering coefficient, the clustering spectrum, the frequencies of different sizes structural motifs, etc. The Cytoscape plugin provides the loading the biological network for analyzing, making a request to the remote computational server for the reconstruction of various structural models according to their specification, visualize of reconstructed structural models and their comparative analysis in Cytoscape package [2]. The reconstruction of the structural models of protein-protein interactions (PPI) in mouse liver and their comparative analysis allowed us to identify the circadian dynamics of structural and functional regularities of PPI networks.

*Conclusion:* The system developed by us allows us to build random models of biological networks and to analyze them for searching the patterns of the structural and functional organization of biological networks, to test various statistical hypotheses on networks, to search for network biomarkers, and so on.

*Acknowledgements:* Supported by Presidium of SD RAS (No. 0324-2018-0021) and by RF Government (No. 0324-2018-0017).

## References

1. Random Network Generator: <https://github.com/polcolomer/RandNetGen>
2. Shannon P. et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. In *Genome research*. 13(11):2498-2504.

## Circadian rhythms: data analysis and mathematical modeling

N.L. Podkolodnyy<sup>1,2\*</sup>, N.N. Tverdohkleb<sup>1,3</sup>, O.A. Podkolodnaya<sup>1</sup>

<sup>1</sup>*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*

<sup>3</sup>*Novosibirsk State University, Novosibirsk, Russia*

\* e-mail: pnl@bionet.nsc.ru

**Key words:** circadian rhythm, mathematical modelling, gene expression analysis, GO enrichment analysis, biological processes

*Motivation and Aim:* A large-scale analysis of the circadian dynamics of transcriptomes and translomes together with gene networks and protein-protein interaction network makes it possible to identify the molecular processes with circadian rhythm and to reconstruct the structure of the mathematical model for circadian regulation of biologically important molecular processes. The purpose of this work is to develop such kind methods and to apply them for modeling the mutual influence of the circadian oscillator and the NAD<sup>+</sup> / SIRT1 system.

*Methods and Algorithms:* We used experimental data on the gene expression at the mRNA level and ribosome profiling (GSE67305 and GSE81283) in mouse liver and kidney. Identification of circadian patterns of gene expression was carried out by the methods of correlation, cluster analysis and the principal component analysis. The modeling and simulation were performed in Matlab (ODE solver ode15s).

*Results:* We have identified genes that demonstrate pronounced circadian dynamics of transcription and translation. Based on the cluster analysis of the expression data, we identified the groups of genes with different circadian dynamic pattern of transcription and translation and the groups of genes with different circadian phase characteristics.

For these groups of genes the GO enrichment analysis was carried out. As a result, we have identified molecular processes with circadian regulation and the groups of genes with significantly different temporal phase patterns in mouse liver and kidney. We have reconstructed and analyzed the gene network of mammalian circadian clocks. We identified the central component of circadian rhythm regulation and the functional subsystems interacting with it. This makes it possible to simulate circadian regulation of functionally important molecular processes in a mammalian cell. Using these approaches, a simulation of the mutual influence of the circadian oscillator and the NAD<sup>+</sup> / SIRT1 system (cell energy balance regulator and some kind of “metabolic oscillator”) was carried out. Based on the experimental data on the changes in the activity of SIRT1 and the level of NAD<sup>+</sup> with age, we investigated the effect of these age-related changes on the functioning of the circadian oscillator. Mathematical modeling has shown that the age-related decline in SIRT1 activity may be one of the causes of disturbances in the functioning of the circadian oscillator in suprachiasmatic nuclei, which may also lead to disturbances in the circadian rhythms of the organism as a whole.

*Acknowledgements:* Supported by Presidium of SD RAS (No. 0324-2018-0021) and by RF Government (No. 0324-2018-0017).

## Digital heart: personalized medicine and inverse problems

A. Prikhodko<sup>3\*</sup>, M. Shishlenin<sup>1, 2, 3</sup>

<sup>1</sup> Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia

<sup>2</sup> Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia

<sup>3</sup> Novosibirsk State University, Novosibirsk, Russia

\* e-mail: Prikhodko1997@gmail.com, mshishlenin@ngs.ru

**Key words:** pumping function of the heart, inverse problem

*Motivation and Aim:* Cardiovascular disease (CVD) is the leading cause of death of the worldwide: no other cause kills as many people each year as CVD. The purpose of the study is to obtain personalized information (blood viscosity, pressure in ventricles, pulmonary artery, pulmonary vein, aorta, full vein, right atrium) about the patient by solving the inverse problem, that will help in the diagnosis of the cardiovascular system.

*Methods and Algorithms:* We formulate the mathematical model of the heart in the form of the four-chamber pump [1]. In this pump unit, the Atria perform the function of low-pressure pumps, and the ventricles function as high-pressure pumps. The model is the system of the ordinary differential equations (ODE) and consists of the first and the second ODE.

Let us suppose that we can measure the blood flow and the pressure as the functions of time. The inverse problem consists in finding heart mass, for each patient. We reduce inverse problem to the functional minimization. Due to the extremely nonlinearity of the problem we apply the gradient method of the minimization of the cost functional.

*Results:* Numerical methods for solving the inverse problems are constructed.

*Conclusion:* The solution of inverse problem allows us to find the heart parameters for each patient (personalized medicine).

*Acknowledgements:* the work was supported by the MSC RK grant No. AP05134121.

### References

1. Lebedenko I.S., Novoselova E.S., Rakityanskaya A.S., Efimtseva J.A. (2009). Mathematical model of the heart. *Biotechnosphere*. 3:24-31.
2. Lee J.-H. et al. (2003) Experimental study of sliding mode control for a benchmark magnetic bearing system and artificial heart pump suspension." *IEEE Transactions on control systems technology*. 11(1):128-138.



## Mathematical model of membrane potential formation at *E. coli* growth on nitrite

N.A. Ree\*, V.A. Likhoshvai, T.M. Khlebodarova

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

\* e-mail: kashev@bionet.nsc.ru

**Key words:** nitrite respiration, membrane potential, mathematical model

**Motivation and Aim:** The main component of the respiratory chain, leading to ATP synthesis, in *Escherichia coli* cells, cultivated in anaerobiosis on nitrite, is periplasmic Nrf nitrite reductase. It was shown previously [1], that at low extracellular nitrite concentration, genetic component is not sufficient to describe the experimentally measured dynamic of substrate utilization by *E. coli* cells, in the chemostat. It was assumed that additional mechanism, regulating Nrf activity, is influence of membrane potential on enzyme subunits transport to the periplasm and on formation of active Nrf enzyme. This work is dedicated to verification and justification of this hypothesis.

**Methods and Algorithms:** Generalized Hill functions were used to describe expression of operons (*fdhF*, *hyc*, *hyb*, *hyf*, *nrf* и *nir*), which are involved in nitrite electron transport chain and nitrite metabolism in *E. coli* cells. Rate of reactions were described by Michaelis-Menten equations. Parameters of the model were evaluated from the published data or were estimated during model's adaptation to experimental data.

**Results:** The mathematical model of *E. coli* cells respiration on nitrite at anaerobic conditions was created. The model takes into account molecular-genetic mechanisms of respiratory chain formation, nitrite metabolism regulation and also the kinetic of substrate utilization in stationary growth conditions of the chemostat. Lack of particular data on the mechanism of membrane potential formation, the model describes two hypothetical pathways: with involvement of formatelyase complexes FHL-1 or FHL-2. According to the first pathway, FHL-1 complex, composed of FDH-H formate dehydrogenase and Hyd-3 hydrogenase, oxidizes formate to hydrogen, which diffuse to the periplasm, where is utilized by Hyd-2 hydrogenase, transferring electrons through quinone pool on Nrf reductase. According to the second pathway, FHL-2 complex, composed of FDH-H formate dehydrogenase and Hyd-4 hydrogenase, directly transports protons from the cytoplasm to the periplasm and transfers electrons, generated by oxidation of formate, through quinones on Nrf reductase. The model analysis revealed contribution of membrane potential in resulting nitrite utilizing Nrf activity was about 50 % in the micromolar range on substrate concentration, although adequacy of nitrite dynamic reproduction is independent on the pathway of potential formation.

**Conclusion:** Membrane potential, independent on the way of its formation, is a necessary mechanism, complementing genetic regulation of periplasmic Nrf activity at nitrite concentrations less, then 1 mM.

**Acknowledgements:** The work was partially supported by the RFBR (No. 16-01-00237a) and Project Fundamental Research of SB RAS (No. 0324-2018-0017).

### References

1. Ree N.A. et al. (2015) About mechanisms of nitrite utilization by *Escherichia coli* cells under stationary growth cultivation. *Math. Biol. Bioinf.* 10:193-205.



# The uniqueness of the solution of the two-dimensional direct problem is the propagation of the action potential along the nerve fiber

A.J. Satybaev<sup>1\*</sup>, G.S. Kurmanalieva<sup>1,2</sup>

<sup>1</sup> Osh Technological University, Osh, Kyrgyzstan

<sup>2</sup> Osh Technological University, Osh, Kyrgyzstan

\* e-mail: abdu-satbaev@mail.ru

**Key words:** two-dimensional, parabolic, direct problem, hyperbolic, generalized, action potential, nerve fiber, axon, uniqueness of solution

*Formulation of the problem:* The two-dimensional problem of the process of propagation of the action potential along nerve fibers is modeled by the following problem of an equation of parabolic type [1]:

$$C_m(x, y)v_t'(x, y, t) = \frac{r_a(x, y)}{2\rho_a(x, y)}\Delta v - \frac{v(x, y, t)}{\rho_m(x, y) \cdot l}, \quad (x, t) \in R_+^2, y \in R, \quad (1)$$

$$v(x, y, t)|_{t < 0} \equiv 0, \quad v_x'(x, y, t)|_{x=0} = h(y)\theta(t) + r(y)\theta_1(t) + p(y)\theta_2(t), \quad t \in R_+, y \in R, \quad (2)$$

where,  $h(y)$ ,  $r(y)$ ,  $p(y)$  – are given functions,  $\theta(t)$  – is the Heaviside theta function,

$\theta_1(t) = t\theta(t)$ ,  $\theta_2(t) = \frac{t^2}{2}\theta(t)$ ,  $C_m$  – is the capacitance,  $r_a$  – is the radius of the nerve fiber,  $\rho_a$ ,  $\rho_m$  – are the resistivities of the nerve fiber plasma and the membrane material, respectively,  $l$  – is the membrane thickness,  $v(x, y, t)$  – is the intracellular action potential,  $a$  and  $m$  – indexes of nerve fibers and membranes,  $\Delta v$  – is the Laplace operator.

Using the Laplace transform, the parabolic problem (1)–(2) is reduced to an equivalent problem of hyperbolic type [2].

For this hyperbolic problem, a theorem on the uniqueness of the solution is justified [3]. It is shown that the uniqueness of the parabolic problem is established from the equivalence of problems of hyperbolic and parabolic type.

## References

1. Basharina O.V., Artyuhov V.G. (2007) Biophysics. Voronezh: VSU, 83 p.
2. Kabanikhin S.I. (2009) Inverse and ill-posed problems, Siberian Scientific Publishing. Novosibirsk. p. 458.
3. Satybaev A.J., Anishchenko Yu.V. (2014) Uniqueness of the solution of the two-dimensional direct problem of geoelectrics with instantaneous and corded sources. Science and new technologies. 7:25-29.

# Mathematical models of p53–microRNA and their applications

S.D. Senotrusova<sup>1,2\*</sup>, O.F. Voropaeva<sup>1</sup>

<sup>1</sup>*Institute of Computational Technologies SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Novosibirsk State University, Novosibirsk, Russia*

\* e-mail: senotrusova.s@mail.ru

**Key words:** tumor marker, p53, miRNA, delay differential equations, numerical simulation

*Motivation and Aim:* One of the priorities of modern biomedical research is the search for effective biomarkers for early cancer detection and other serious diseases related with dysfunction of processes of cell death. The p53 protein (tumor necrosis factor), involved in many life and death processes, including the formation of tumors and aging, is expressed in all the cells of the organism. Mdm2 protein is considered to be the key negative p53 regulator [1]. It is known that p53 regulates the class of microRNAs (miRNA), which are characterized as the most important intermediates of p53 in tumor control. Thus, the investigation of the function of the p53 protein and updating of the diagnostic properties of miRNA is paramount both for developing new approaches to cancer treatment and determining the prevention strategy for many diseases, including measures to slow the aging processes.

*Methods and Algorithms:* We propose in this work a hierarchy of mathematical models of the dynamics of network p53–Mdm2–microRNA for miRNA class with a direct positive connection with p53. These models include the systems of three nonlinear equations with delaying arguments.

*Results:* All obtained numerical solutions have a sufficiently clear biomedical meaning. A comparative analysis of models in a wide range of parameters is performed. Stress situations associated with the emergence of an imbalance in the rates of p53 and Mdm2 generation and degradation and also with disturbances in the mechanism of implementation interaction of proteins regulated within the considered models through the constants of dissociation have been studied. Numerical investigation of the microRNAs functioning in conditions of the deregulation of p53 and p53–Mdm2-network is carried out. The deregulation of microRNA in detail is studied. The situations in which p53, its inhibitor Mdm2 and microRNAs exhibit critical properties for the patient's status and can be identified as diagnostic markers of cancer and neurodegenerative disease are studied. The results of numerical analysis are in good agreement with the data of clinical and laboratory studies of known microRNAs.

*Conclusion:* According to the results of these investigations, p53-responsive microRNA of given class can be used to clarify function of p53 as a biomarker of cancer and neurodegenerative diseases [1]. If the network p53–microRNA functions normally, then microRNA duplicates diagnostic properties of p53. The variants of microRNAs deregulation are found, where microRNAs are an even better marker for disease than p53. The most significant differences in the analysis results are due not to the introduction of additional microRNA feedback with Mdm2, as could be expected, but to the refinement of the approximation of the interaction function for the p53–miRNA. We present other examples of the application of the developed models.

## References

1. Voropaeva O.F., Senotrusova S.D., Shokin Yu.I. (2017) Numerical study of diagnostic properties of p53-dependent microRNA. Russ. J. Numer. Anal. Math. Modelling. 32(3):203-213.

## An effective subgradient method for simultaneous restoration and segmentation of blurred images

T. Serezhnikova<sup>1,2</sup>

<sup>1</sup> *Krasovsky Institute of Mathematics and Mechanics UB RAS*

<sup>2</sup> *Ural Federal University, Ekaterinburg, Russia*

\* e-mail: [sti@imm.uran.ru](mailto:sti@imm.uran.ru)

**Key words:** image restoration, denoising, segmentation, subgradient construction

The segmentation of blurred and noise images is of great importance. There have been several recent works to link the problems of image segmentation and image reconstruction. Here we describe the universal subgradient method for simultaneous restoration and segmentation of blurred and noise images. Our method is based on the universal subgradient construction.

Our universal subgradient contains both the brightness function and the brightness function gradient.

In the paper we demonstrate that our method is effective for simultaneous restorations and segmentations of blurred images.

## The software and database for Vertebrate imperfect mtDNA repeats annotation

V.A. Shamanskiy<sup>1\*</sup>, K.Yu. Popadin<sup>1,2</sup>, K.V. Gunbin<sup>1,3</sup>

<sup>1</sup> *School of Life Science, Immanuel Kant Federal Baltic University, Kaliningrad, Russia*

<sup>2</sup> *Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland*

<sup>3</sup> *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

\* e-mail: v.a.shamanskiy@gmail.com

**Key words:** imperfect (degenerate) repeats, mtDNA, Vertebrata

*Motivation and Aim:* It is well known that the vast majority of non-B-DNA conformations is tightly associated with pattern compositions especially with various imperfect repeats. It was also known that the number and nature of mtDNA perfect repeats are under strong selection pressure. Thus, the number and nature of various mtDNA repeats is useful for estimation the organismal fitness-related parameters. Despite this fact, there are no any database annotated imperfect mtDNA repeats.

*Methods and Algorithms:* In order to build the database we selected more than 3800 completely sequenced mtDNAs of Vertebrates and more than 25000 human mtDNAs. We find imperfect repeats in these sequences using our modification of dot-plot analysis: we aligned every 10-bases regions (sliding windows) in mtDNA versus this genome by dynamic programming using various indel-scoring schemes. After that, we linked 10-base imperfect repeats in longer repeats by our hash-algorithm implemented in Python in all cases when neighbor windows marked as a repeat-containing.

In order to estimate the evolutionary relationship between imperfect mtDNA repeats of Vertebrates we mapped these repeats on the 3 alternative multiple mtDNA alignments of Vertebrates generated by MAFFT, PASTA and MISHIMA taking into account predefined phylogenetic tree topology. We mapped human imperfect mtDNA repeats on the alignment of human mtDNA that was reconstructed using BLASR.

*Results:* In order to discriminate different selection modes associated with various Vertebrate taxa evolution we compared taxa-specific numbers of main repeat types with ones from whole sample using nonparametric U-test. We found that Mammals have a significantly large number of imperfect repeats than the number of such repeats in the whole Vertebrata sample set. Another interesting finding is the excess of direct and symmetrical imperfect repeats in hominids that, at the same time, characterized by the lack of complimentary and inverted degenerate repeats. According to our data Ascidias are the record-holder in the number of imperfect repeats, while fishes characterized by relatively small number of degenerate repeats. In addition to the taxa specific description of imperfect repeats number we calculated standard statistics on the length of imperfect repeats, degeneration rate, GC richness, dinucleotide composition, etc. There are also data on tandem degenerate repeats and areas of self-replicating.

*Conclusion:* The software and database for Vertebrate imperfect mtDNA repeats annotation were made. The information in the database allowed us to discriminate taxa-specific selection forces shaping mtDNA repeat composition.

*Availability:* available upon request.

*Acknowledgements:* This study was supported by the 5 Top 100 Project at the I. Kant Baltic Federal University.

# Inverse and Ill-Posed problems for nonlinear PDE: applications to life and social sciences

M. Shishlenin<sup>1, 2, 3\*</sup>, D. Lukyanenko<sup>4</sup>

<sup>1</sup>*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*

<sup>3</sup>*Novosibirsk State University, Novosibirsk, Russia*

<sup>4</sup>*Moscow State University, Moscow, Russia*

\* e-mail: mshishlenin@ngs.ru

**Key words:** inverse and ill-posed problems, Life and social sciences, numerical methods, control problems

*Motivation and Aim:* In mathematical models of life and social sciences, financial mathematics inverse and ill-posed problems are investigated for nonlinear convection-diffusion-reaction  $u_t + (v, \nabla u) - \tilde{N}(k \nabla u) = A(q, u)$ , where right-hand side  $A$  is nonlinear in  $q$  and  $u$ , respectively [1]. We consider the coefficient inverse problems of recovering  $v$ ,  $k$  and  $q$  by nonlocal data (integral over the domain given in the discrete time) [2]. This data characterizes a certain reference group (social survey), medicine (drug absorption) or financial market. Desired coefficients can characterize such important characteristics as labor-power ratio, labor productivity, consumption, interpersonal interactions, etc. Also we consider the control problem: how to find the source or initial function to obtain desired statement (or people behavior) in a final fixed time.

*Methods and Algorithms:* Due to extremely nonlinearity we reduce the inverse and ill-posed problem solution to the optimization problem. We apply the gradient method of minimizing the cost functional. A gradient of the functionals were obtained by solving the corresponding conjugate problems [1, 2].

*Conclusion:* The examples of ill-posedness were constructed. We also mention the theoretical and numerical results for considered problems. Numerical results are presented.

*Acknowledgements:* Supported by the RFBR (18-01-00865, 18-41-540017, 17-51-540004, 16-01-00755).

## References

1. Lukyanenko D.V., Shishlenin M.A., Volkov V.T. (2018) Solving of the coefficient inverse problems for a nonlinear singularly perturbed reaction-diffusion-advection equation with the final time data. *Communications in Nonlinear Science and Numerical Simulation*. 54:1339-1351.
2. Kabanikhin S.I., Shishlenin M.A. (2018) Recovering a Time-Dependent Diffusion Coefficient from Nonlocal Data. *Numerical Analysis and Applications*. 11(1):38-44.

## Deep bioinformatics expert system of analysis, modeling and interpretation of omics BigData of the human genome

A. Shlikht\*, N. Kramorenko

Far Eastern Federal University, Vladivostok, Russia

\* e-mail: schliht@mail.ru

**Key words:** genomics, proteomics, bigdata, bioinformatics, expert systems, data base

*Motivation and Aim:* An important task in omics BigData is to use an effective format and structure for data storage. The traditional sequential data file format does not provide sufficient performance for data access and subsequent automatic analysis, modeling, and interpretation. For this purpose, it is necessary to use the formats of databases and knowledge bases [1].

*Methods and Algorithms:* The transition from file sequential access to data and the use of procedural scripting programming languages to the format of databases and knowledge bases and declarative languages allows us to reach a qualitatively new level of artificial intelligence systems. For this purpose, it is necessary to restructure the primary file data of the world portals into local data stores with a database structure. Next, at the local level, a variety of methods of analysis, interpretation of databases and knowledge bases are used.

*Results:* The deep bioinformatics expert system has been developed, which allows to carry out work on the analysis, interpretation, modeling and diagnostics of omics BigData in an automatic mode at the local level with high efficiency. Depth of analysis, interpretation is provided by numerous associations between mutations, genes, transcripts, proteins (enzymes), reactions, metabolites, metabolic and signaling pathways, up to diseases. The system of associative relations allowed to build a deep subsystem of explanations, which is part of the expert system.

*Conclusion:* On the basis of the developed deep bioinformatics expert system, effective processes for the analysis, modeling, interpretation and diagnosis of omics BigData are carried out. The expert system can be useful for researchers, physicians and students.

### References

1. Shlikht A.G., Kramorenko N.V. (2016) Bioinformatic expert system of analysis and interpretation of omics sequence of the human genome. MM-HPC-BBB-2016: P. 94.

# Asymptotic stability of solutions in one model of disease

M.A. Skvortsova<sup>1,2\*</sup>

<sup>1</sup> Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

\* e-mail: sm-18-nsu@yandex.ru

**Key words:** a model of disease, delay differential equations, asymptotic stability, estimates of solutions, attraction domains, modified Lyapunov–Krasovskii functional

*Motivation and Aim:* We consider a system of delay differential equations describing the spread of a disease [1]:

$$\begin{aligned}\frac{d}{dt} x(t) &= \sigma - \mu_1 x(t) - \beta x(t)z(t), \\ \frac{d}{dt} y(t) &= \beta x(t - \tau)z(t - \tau)e^{-a\tau} - \mu_2 y(t), \\ \frac{d}{dt} z(t) &= \rho y(t) - \mu_3 z(t),\end{aligned}$$

where  $x(t)$  is the concentration of uninfected cells,  $y(t)$  is the concentration of infected cells that produce virus, and  $z(t)$  is the concentration of plasma virus. All the parameters of the system are constant and positive. We study the asymptotic stability of stationary solutions to this system.

*Methods and Algorithms:* When studying asymptotic properties of solutions to systems of nonlinear delay differential equations, in [2] it was proposed a modified Lyapunov–Krasovskii functional. It is important to note that the construction of such functional can be reduced to solving well-conditioned problems and does not require finding roots of quasi-polynomials. We use an analogue of such functional.

*Results:* We obtain estimates of solutions characterizing the stabilization rate at infinity and establish estimates of attraction domains of asymptotically stable stationary solutions.

*Acknowledgements:* The author is grateful to Professor G. V. Demidenko for the attention to the research. The reported study was funded by the Russian Foundation for Basic Research and Government of the Novosibirsk region according to the research project No. 17-41-543365.

## References

1. Herz A.V.M., Bonhoeffer S., Anderson R.M., May R.M., Nowak M.A. (1996) Viral dynamics in vivo: limitations on estimates of intercellular delay and virus decay. Proc. Natl. Acad. Sci. USA (Medical Sciences). 93:7247-7251.
2. Demidenko G.V., Matveeva I.I. (2005) Asymptotic properties of solutions to delay differential equations. Vestnik Novosibirskogo Gosudarstvennogo Universiteta. Seriya: Matematika, Mekhanika, Informatika. 5(3):20-28. (in Russian)



## Algorithm for solving the inverse problem of pharmacokinetics to determine the transition coefficients

A. Takuadina

*L.N. Gumilev Eurasian National University, Astana, Kazakhstan*

*e-mail: alyoka.01@mail.ru*

**Key words:** mathematical model, optimization method, inverse problems of pharmacokinetics

*Motivation and Aim:* In this article, a mathematical model of blood circulation of the human body is considered, representing both a set of chambers. By taking blood from the patient, the concentration of the drug in the first chamber is determined, which we take as additional information for solving the inverse problem. According to the known additional information, it is necessary to determine all transition coefficients (rate constants).

*Methods and Algorithms:* To solve the inverse problem, we use the optimization method. The essence of this is to minimize the quadratic residual functional of the observed and calculated state of the drug concentration in the chamber. As the formulation of a direct problem, a system of kinetic equations is used that describes the movement of the preparation from the chamber to the chamber.

*Results:* An auxiliary (adjoint) problem is constructed, with the help of which an explicit form of the gradient of the minimized functional is obtained. A difference analogue of the algorithms for solving the direct and conjugate problem formulated in the discrete formulation [1] is developed.

*Acknowledgements:* to the foreign scientific adviser: Corresponding Member of the Russian Academy of Sciences, Director of the Institute of Computational Mathematics and Mathematical Geophysics, Doctor of ph.-m. sciences, Professor Kabanikhin Sergey Igorevich.

### *References*

1. Kabanikhin S.I. (2009) Inverse and ill-posed problems. Novosibirsk: Siberian Scientific Publishing House. 457 p.

# Comparison of quality of automated gene network reconstruction using connectivity of random and functional networks

E. Tiys<sup>1,2\*</sup>, P. Demenkov<sup>1</sup>, V. Ivanisenko<sup>1</sup>

<sup>1</sup>*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Novosibirsk State University, Novosibirsk, Russia*

\* e-mail: tiys@bionet.nsc.ru

**Key words:** associative gene networks, network reconstruction, random networks, modularity

*Motivation and Aim:* Hartwell in 1999 drew attention to the modularity of biological systems [1]. Later it was shown that at the level of gene networks, modularity manifests itself in more connections between genes taken from one biological process, than from different [2]. Often biological experiment gives a set of genes with the differential characteristics and the task of the researcher is to determine the mechanisms, underlying the flow of the biological process. Such a gene set can be a starting set for reconstruction of the gene network for this process (associative gene network). Knowledge of the modular organization of biological systems can be used to assess the quality of the automated network reconstruction method.

*Methods and Algorithms:* In order to compare methods of gene network reconstruction, we used ROC-curves taken from [2], based on a positive sample of genes sets of size 10 of GeneOntology biological processes and a negative sample including sets of randomly selected genes. As a measure of the quality of network reconstruction, Area under ROC-Curve (AUC) was chosen. As a reconstruction tool, we used ANDSystem. We used two simplest methods of reconstruction, level 0 network, including links only between the starting genes obtained in the experiment and level 1 network, which includes, in addition to the level 0 network, all the vertices adjacent to the level 0 network and all the links between them. P-value of the difference in the AUC of the ROC-curves was estimated using the two-sided unpaired DeLong's test, through the roc.test function of R language.

*Results:* The method of reconstruction of level 0 network showed a significantly higher AUC value than level 1 network (0.88 vs. 0.61,  $p < 1.00e-05$ ) and is estimated as a more promising automated method for reconstruction of associative gene networks.

*Conclusion:* This work is the first attempt to introduce an objective approach to comparing methods of automated reconstruction of associative gene networks.

*Acknowledgements:* Supported by the budget project No. 0324-2018-0017 and by integrational project No. 0324-2018-0021.

## References

1. Hartwell L.H. et al. (1999) From molecular to modular cell biology. *Nature*. 402(Suppl 6761):C47-C52.
2. Tiys E.S., Ivanisenko T.V., Demenkov P.S., Ivanisenko V.A. (2018) FunGeneNet: a web tool to estimate enrichment of functional interactions in experimental gene sets. *BMC Genomics*. 19(Suppl 3):76.

# Chaos theory as a bioinformatics promissory instrument for a human organism systemic response in-depth study

B.G. Vainer<sup>1,2\*</sup>, A.V. Shepelin<sup>2</sup>

<sup>1</sup> *Rzhanov Institute of Semiconductor Physics SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* e-mail: BGV@isp.nsc.ru

**Key words:** chaos theory, organism systemic response, computer analysis, infrared thermography

*Motivation and Aim:* The complexity of biological systems generates a need for involving of various mathematical, statistical and information instruments for their detailed analysis. Each separate research method usually gives new and original results that expand the knowledge about the subject of interest. Since most physiological systems have a nonlinear dynamic nature, an adequate and convenient tool for their study is the chaos theory. In the context of this theory, physiological processes initiated in the living organism in response to sudden external impact meet the requirements of dynamic chaos phenomena [1]. The aim of this study is the chaos theory-based analysis of a human organism systemic response emerged in reply to various transient external interventions that the organism is exposed to.

*Methods and Algorithms:* The dynamic characteristics of respiration, thermoregulation and cardiovascular system were synchronously measured using modern experimental methods, including focal plane array-based infrared thermography. External interventions were realized using local heating of the extremities, short-term clamping of the shoulder vessels, forced modulation of the respiration rate, etc. The heart rate was measured with the 200-Hz signal acquisition rate using microphone-based original pulsometer connected to the Biopac MP 100 measurement system. Skin temperature was measured with the 100-Hz frame rate using TKVr-IFP/SVIT infrared camera. Breathing rate was also measured with the above-mentioned infrared camera using original Sorption-Enhanced Infrared Thermography (SEIRT) method [2]. Computer-assisted quantitative analysis of the synchronously obtained experimental data was made using mathematical statistics and chaos theory algorithms.

*Results:* Drawn on the 2-dimensional or 3-dimensional phase planes, the experimental dependences, reflecting the dynamic change of the human organism physiological characteristics, reveal the hidden features and patterns that are not visible when standard methods of analysis are used.

*Conclusion:* The prospects of using the theory of chaos-based analytical instrument in the biomedical and physiological studies are shown.

*Acknowledgements:* Supported by the RFBR (grant No. 18-08-00956).

## References

1. Vergunov E.G., Vainer B.G. (2016) Bland-Altman plot modification and its application to assessing agreement between heart and breathing rates. *Psychophysiology News* (ISSN: 2227-6157). 4:34-45.
2. Vainer B.G. (2015) Application of the infrared thermography-based adsorption-induced indication method to investigating the human respiration. *Development Prospects of Science and Education, International Scientific-Practical Conference, February 28, 2015. Proceedings. Part 8.* Tambov: Consulting company Ucom Ltd.: 32-34. DOI 10.17117/2015.02.28.08.

# ARGO\_CEL: GPU based approach for potential composite elements discovery in large DNA datasets

O. Vishnevsky<sup>1,2\*</sup>, A. Bocharnikov<sup>1</sup>, N. Kolchanov<sup>1,2</sup>

<sup>1</sup>*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Novosibirsk State University, Novosibirsk, Russia*

\* e-mail: oleg@bionet.nsc.ru

**Key words:** oligonucleotide motifs, composite elements, transcription regulation, ChIP-Seq

*Motivation and Aim:* Composite elements play an important role in the regulation of transcription. Existing methods for the revealing of potential composite elements (PCE) are usually based on assessment of the significance of the mutual presence of the predicted transcription factor binding sites (TFBS). In this case, the recognition of potential TFBSs is performed using weight matrices or other methods trained on samples of binding sites of known transcription factors. Thus, such methods essentially depend on the completeness of training samples and information on existing TFs.

*Methods and Algorithms:* We have proposed a method for *de novo* discovery of PCE, which does not require preliminary information about the localization of potential TFBS. Based on the proposed approach, the Internet-accessible resource [http://argo.bionet.nsc.ru/cgi-bin/ARGO\\_CEL/Argo\\_CEL.cgi](http://argo.bionet.nsc.ru/cgi-bin/ARGO_CEL/Argo_CEL.cgi) was created, which allows the user to obtain sets of mutually present groups of significant motives in the analyzed sample of nucleotide sequences, and annotate them. Such groups of motifs can correspond both to binding sites of known transcription factors, and to certain physico-chemical features of the nucleotide context of regulatory regions of genes. The method developed is based on *de novo* discovery of significant degenerate oligonucleotide motifs of a fixed length, written in a 15-letter IUPAC [1]. Discovered motifs are clustered into groups, corresponding to individual regulatory elements. After that, the significance of the mutual presence of the obtained groups of motives is evaluated and the PCE are identified. The ARGO\_CEL system allows annotation of the constructed groups of motives on the base of Transfac TFBS.

*Results:* The developed approach was used to analyze the results of the FoxA ChIP-Seq dataset [2]. Several groups of significant motifs were obtained using the ARGO\_CEL system. Some of them correspond to known FoxA binding sites, and some to other TFBS. It was shown that some of these groups reliably co-occur in the control set and can correspond to PCEs.

*Conclusion:* We have developed a *de novo* method for discovery of potential composite elements that does not require preliminary information on the TFBS. Using the proposed approach, context signals are identified in the ChIP-Seq dataset, which can correspond to potential composite elements.

*Acknowledgements:* The work was supported in part by ICG SB RAS budget project (0324-2016-0008) and integration project of SB RAS (0324-2018-0021).

## References

1. Vishnevsky O.V., Bocharnikov A.V., Kolchanov N.A. (2017) ARGO\_CUDA: Exhaustive GPU based approach for motif discovery in large DNA datasets. *Journal of Bioinformatics and Computation Biology*, Dec 10:1740012.
2. Wederell E.D. et al. (2008) Global analysis of in vivo Foxa2-binding sites in mouse adult liver using massively parallel sequencing. *Nucleic Acids Research*. 36:4549-4564.

# Teaching medicine and biology through systems biology

H.V. Westerhoff

*Synthetic Systems Biology and Nuclear Organization, University of Amsterdam*

*Molecular Cell Physiology, Vrije Universiteit Amsterdam*

*Systems Biology, The University of Manchester*

*Infrastructure Systems Biology Europe.NL*

*e-mail: h.v.westerhoff@uva.nl*

**Key words:** hands-on modelling, understanding biological complexity, training life scientists

For long, the teaching of the Life Sciences focused on memorization of the metabolites, enzymes and pathways that were known. This gradually led to biochemistry and cell biology text books becoming extensive libraries in excess of half a million of words and to databases with the implementation of genomics. Some courses still require the students to learn these words by heart, and in the right sequence.

This practice is rapidly becoming obsolete, first by the number of facts growing exponentially whereas student brains do not, and second by Wikipedia taking over: students look up the components of biochemistry on their mobile phones faster than the Professor can pronounce them. Should we then refrain from teaching biology altogether? Some may say so, as they see Biology as just a large set of facts; a can of worms; a cacophony of special cases.

We take a rather diametrical position: we think that Biology and Medicine are complexity sciences, where most functions and malfunctions emerge in nonlinear interactions and are thereby hard to understand. And we hold that it is precisely this emergence that students should learn to understand. Yes, it is fine for them to look up the facts in proper databases: We move towards a teaching without facts but full of network mechanisms.

Here we shall demonstrate this teaching mode in hands-on training courses in which students train themselves to recognize and solve some of the paradoxes that abound in biology and medicine. They will be armed with blue-print computational models that they can then populate with facts looked up through Wikipedia. They will then be instructed how by running these models they can discover non-intuitive behaviour. And they will then interrogate the models further and resolve the paradoxes. This will make them realize why the MAP-kinase route should better not be a MAPkinase route, why the most obvious may not be the best drug target, and why GWASH fails to elucidate the key component of most diseases. Networking enables functions that are otherwise thermodynamically impossible, such as the synthesis of ATP, proteins and DNA. We shall here highlight a lesser known function of networking, i.e. diversification. Network diversification followed by selection, sprouted the tree of Life, but that very tree hides a forest of diversity. Early Life on this planet may have benefitted from diversification of the redox network around acetogenesis. Flux Balance Analysis (FBA) of the genome-wide metabolic network of *Cl. Ljungdahlia* reveals carbon fixation at various ATP/acetate stoichiometries. This may have enabled early organisms to survive the erratic environmental conditions by shifting gears. The flexible ATP yield enabled by the Warburg effect may help do so for tumor cells.

Our organs are subject to a drizzle of somatic mutations, which leads to cell diversification with age. We shall review an FBA methodology that simulates this and then demonstrate that this may initially enhance the metabolic versatility of organs such as T-cells and liver, in the absence of adaptation. More somatic mutations would cause loss of function and ageing. And, subpopulations of asocial cells would develop into tumors with Warburg and the new WarburQ (i.e. glutamine dependent) phenotypes.

Likewise, noise such as required by the third law of thermodynamics should diversify cell populations. FISH and deep sequencing experiments show even stronger noise than this, which should thereby be subject to regulation, e.g. through transcription bursting. We shall show that even though such noise varies with time, it may be selectable and may lead to drug resistance of tumor cell populations, either because of nonlinearities or because of the ‘Waddington’ genetic landscape, a remnant of developmental biology.

# FoldGO for functional annotation of transcriptome data to identify fold-change-specific GO categories

D.S. Wiebe<sup>1,2</sup>, A.M. Mukhin<sup>1,2</sup>, N.A. Omelyanchuk<sup>1,2</sup>, V.V. Mironova<sup>1,2\*</sup>

<sup>1</sup>*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Novosibirsk State University, Novosibirsk, Russia*

\* e-mail: victoria.v.mironova@gmail.com

**Key words:** functional annotation, Gene Ontology, RNA-Seq, microarray

*Motivation and Aim:* Scientists extensively generate transcriptome data to study genome response to external factors or internal conditions, thousands datasets are available online. As a rule, the transcriptome data contain much more information than has been extracted and published. Typical scenario of transcriptome data analysis is identification of differentially expressed genes (DEGs), those with significant changes in the number of transcripts, and functional enrichment analysis of the gene set using Gene Ontology [1]. This procedure provides for biological processes, molecular functions and cellular components GO terms that are significantly enriched in up- or down-regulated genes. In this analysis, one ignores the continuous distribution of expression changes and uses a binary classification of genes (DEGs and not DEGs) instead.

*Methods and Algorithms:* We developed a new method FoldGO, which classifies DEGs according to their fold change, performs functional annotation of the multiple gene lists generated on the basis of this classification (with overlaps) and provides for the GO terms, significantly overrepresented for the genes responded to the factor within a narrow fold-change-interval. The application implemented as an R package and as a web service.

*Results:* First, the method has been applied to the transcriptome responded to plant hormone auxin in *Arabidopsis thaliana* roots. In addition to the GO categories, which were significantly enriched for the DEGs, we found many others, more significant if we consider a narrower fold-change-interval [2]. For example, we found that among auxin-responsive genes associated with translation, 90 % only slightly increased their expression in auxin response. We also tested the method on a random set of transcriptomes generated for different organisms. We found fold-change-specific GO categories for most tested datasets.

*Conclusion:* FoldGO allows better describing the effect of an external factor on a gene network and identifying the groups of coordinatively expressed and functionally-related genes.

*Acknowledgements:* The work was partly supported by RFBR 18-04-01130.

## References

1. Ashburner M., Ball C.A., Blake J.A., Botstein D., Butler H., Sherlock G. (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25(1):25-29.
2. Omelyanchuk N.A. et al. (2017) Auxin regulates functional gene groups in a fold-change-specific manner in *Arabidopsis thaliana* roots. *Nat. Sci. Rep.* 7(1):2489.



# Investigation and numerical solving of a mathematical model of intracellular HIV dynamics: from ODE to PDE

D. Yermolenko<sup>1,2\*</sup>, O. Krivorotko<sup>1,2</sup>, S. Kabanikhin<sup>1,2</sup>

<sup>1</sup>*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*

<sup>2</sup>*Novosibirsk State University, Novosibirsk, Russia*

\* e-mail: [ermolenko.dasha@mail.ru](mailto:ermolenko.dasha@mail.ru)

**Key words:** mathematical model of HIV dynamics, ODE, PDE, immunology, epidemiology, parameter specification problem, inverse problem, optimization approach, genetic algorithm, confidence intervals, optimal treatment control.

*Motivation and Aim:* Mathematical models of HIV dynamics are described by systems of nonlinear ordinary differential equations (ODE) and partial differential equations (PDE) and characterized by a set of parameters. These parameters describe virus natural death rate, target cell production rate, treatment efficacy reduction, etc. It is necessary to find the set of parameters for constructing an individual treatment plan. For this, an approach of inverse problem theory [1] in which the unknown model parameters are determined from the available experimental data (observations) is used.

*Methods and Algorithms:* The problem of estimating the HIV-infection parameters and the immune response using additional measurements of the concentrations of T-lymphocytes, free virus, and immune effectors at fixed times for mathematical models of HIV dynamics based on ODE [2] and PDE [3] are investigated numerically. The stability of the inverse problem solution is analyzed using the singular value decomposition for linearized matrix of the inverse problem. The state variable observations are different from each other by orders of magnitude, intuitively, it is critical that the estimation scheme take this into account. One way to do this is by appropriately weighting the states in a least squares cost criterion. A genetic algorithm for solving a least squares minimization problem on iteration of least squares method is investigated. To determine the level of error in the solution of the inverse problem, the confidence intervals of all parameters are obtained and analyzed.

*Results:* The numerical solution of the inverse problem of determining four individual parameters of patient for the mathematical model of HIV dynamics are obtained and analyzed using the combination of least square method and genetic algorithm.

*Conclusion:* It is shown that the common relative accuracy error of the four parameters identification is sufficiently small for a good mathematical model that has a solution quite close to the additional noisy measurements. It is shown that system of PDE describes the HIV dynamic more precisely then the system of ODE.

*Acknowledgements:* This work was supported by the President Grant of Russian Federation (No. MK-1214.2017.1) by the Ministry of Education and Science of Russian Federation and by the grant No. 18-71-10044 of Russian Scientific Found (RScF).

## References

1. Kabanikhin S.I. (2012) Inverse and Ill-Posed Problems: Theory and Applications. Berlin/Boston: de Gruyter.
2. Adams B.M., Banks H.T. et al. (2005) HIV dynamics: Modeling, data analysis, and optimal treatment protocols. Journal of Computational and Applied Mathematics, 184:10-49.
3. Chalub F.A.C.C., Souza M.O. (2011) The SIR epidemic model from a PDE point of view. Mathematical and Computer Modelling. 53:1568-1574.

# Inverse problems for mathematical models in social networks: from PDE to SDE

Sh. Zhang<sup>1\*</sup>, S. Kabanikhin<sup>2,3</sup>, O. Krivorotko<sup>2,3</sup>, Yu. Wang<sup>1</sup>

<sup>1</sup> *Tianjin University of Finance and Economics, Tianjin, China*

<sup>2</sup> *Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*

<sup>3</sup> *Novosibirsk State University, Novosibirsk, Russia*

\* e-mail: shuhua55@126.com

**Key words:** inverse problems, mathematical modelling, partial differential equations, stochastic differential equations, social processes, optimization, Hamilton-Jacobi-Bellman equation, Fokker-Plank equation, dynamic programming

*Motivation and Aim:* Social networks are a developing information environment. Now the following functions are available on the network: access for informing relatives (Facebook's Safety Check service), publishing data about missing friends (the Google Person Finder service) or promptly informing users of an impending threat and their further actions in case of emergencies (Alerts from Twitter). One of the most effective methods of monitoring and managing above processes is the development and identification of mathematical models that describe the processes of information dissemination in social networks [1]. Such models are described by systems of differential equations, the coefficients of that characterize the distribution of information, population and initial data depend on each information type. To control information in social networks it is necessary to refine the model coefficients and initial data by some additional measurements (the inverse problem) [2].

*Methods and Algorithms:* One way to solve the problem of improving the coefficients is to reduce the inverse problem to a variational formulation, where the functional satisfies to the Hamilton-Jacobi-Bellman partial differential equation, or characterizes the quadratic deviation of the model data from the experimental ones for systems of stochastic differential equations or partial differential equations. The Tikhonov regularization, gradient methods and genetic algorithm are used for solving ill-posed inverse problems.

*Results:* The reconstructed coefficients, initial data and control functions in mathematical models of social networks (as an example it is considered Digg.com and Twitter [3]) allows one to refine the information dynamic and give the recommendation for control it.

*Acknowledgements:* Supported by the Ministry of Education and Science of Russian Federation and by the President Grant of Russian Federation (No. MK-1214.2017.1).

## References

1. Dai G., Ma R., Wang H., Wang F., Xu K. (2015) Partial differential equations with robin boundary condition in online social networks. *Discrete and Continuous Dynamical Systems Series B*. 20(6):1609-1624.
2. Kabanikhin S.I. (2008) Definitions and examples of inverse and ill-posed problem. *Journal of Inverse and Ill-Posed Problems*. 16(4):317-357.
3. Wang H., Wang F., Xu K. (2013) Modeling information diffusion in online social networks with partial differential equations. arXiv:1310.0505 [cs.SI].

# Gene network analysis of complex diseases using GenCoNet

O. Zolotareva<sup>1,2</sup>, A. Shoshi<sup>1</sup>, R. Hofestädt<sup>1</sup>, A. Maier<sup>1</sup>, V. Ivanisenko<sup>3</sup>, V. Dosenko<sup>4</sup>, E. Bragina<sup>5</sup>

<sup>1</sup> Bielefeld University, Bioinformatics / Medical Informatics Department, Bielefeld, Germany

<sup>2</sup> Bielefeld University, International Research Group “Computational Methods for the Analysis of the Diversity and Dynamics of Genoms”, Bielefeld, Germany

<sup>3</sup> Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

<sup>4</sup> Bogomoletz Institute of Physiology, Kiev, Ukraine

<sup>5</sup> Research Institute of Medical Genetics, Tomsk NRMС, Tomsk, Russia

\* e-mail: ralf.hofestaedt@uni-bielefeld.de

**Key words:** network analysis, database system, disease

*Motivation and Aim:* Complex diseases affect millions of people worldwide and impose individuals and society a huge socioeconomic burden. Although complex diseases cannot be explained by the Mendelian model of inheritance, certain molecular-genetic components in their pathogenesis are expected. Moreover, since many of complex diseases are comorbid, i.e. demonstrate higher than expected co-occurrence, may be assumed to share molecular players and mechanisms. Identification of these shared molecular players and mechanisms in the pathogenesis of comorbid diseases is necessary for selection of the most appropriate treatment strategy. To address this ongoing need, the Neo4j database “GeCoNet” was developed which integrates various associations between diseases, genes, variants and chemical compounds for four diseases that have a strong molecular-genetic component and demonstrate comorbidity: asthma, essential hypertension, diabetes and Alzheimer’s disease.

GenCoNet is meant to be a qualitative resource that facilitates researchers access the relevant information. Our study [1] showed that the application of methods of reconstruction and analysis of gene networks is a productive tool to rank candidate genes by their importance to complex diseases (e.g. asthma and hypertension).

*Acknowledgements:* Supported by the Volkswagenstiftung grant No. 90335.

## References

1. Saik O.V., Demenkov P.S., Ivanisenko T.V., Bragina E.Y., Freidin M.B., Gonchareva I.A., Dosenko V.E., Zolotareva O., Hofestädt R., Lavrik I.N., Rogaev E.I., Ivanisenko V.A. (2018) Novel candidate genes important for asthma and hypertension comorbidity revealed from associative gene networks. BMC Med Genomics. 11(1):15.

# Inverse problem for partial differential equations in social networks

T. Zvonareva<sup>1\*</sup>, O. Krivorotko<sup>1,2</sup>, S. Kabanikhin<sup>1,2</sup>

<sup>1</sup>Novosibirsk State University, Novosibirsk, Russia

<sup>2</sup>Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia

\* e-mail: zvonareva-tanyushka@mail.ru

**Key words:** inverse problems, partial differential equations, social processes, diffusive logistic model, optimization, gradient method, optimization, tensor train decomposition, regularization

*Motivation and Aim:* A lot of functions are available on the social network: access for informing relatives (Facebook's Safety Check service), publishing data about missing friends (the Google Person Finder service) or promptly informing users of an impending threat and their further actions in case of emergencies (Alerts from Twitter). Such processes can be described by the diffusive logistic mathematical model that characterizes information dissemination in social networks [1]. The type of information is determined by the coefficients of the mathematical model and the initial conditions of the problem. To control and predict the type of information in social networks it is necessary to refine the model coefficients and initial data by some additional measurements (the inverse problem) [2].

*Methods and Algorithms:* One way to solve the inverse problem for partial differential equations is to reduce it to an optimization problem, where the misfit function characterizes the quadratic deviation of the model data from the experimental one. To find the local minimum of misfit function the gradient methods are applied. In most cases, it is necessary to determine the global minimum of the objective function. A new computational method based on tensor train decomposition [3] is applied to solve the optimization problem. The idea of proposed method is to extract the tensor structure of the optimized functional and use it for optimization.

*Results:* The reconstructed coefficients and initial data of mathematical models of social networks (as an example it is considered Digg.com and Twitter [4]) allows one to refine the information dynamic and give the recommendation for control it.

*Acknowledgements:* Supported by the Ministry of Education and Science of Russian Federation, by the President Grant of Russian Federation (No. MK-1214.2017.1) and by the grant No. 18-71-10044 of Russian Scientific Found (RScF).

## References

1. Dai G., Ma R., Wang H., Wang F., Xu K. (2015) Partial differential equations with robin boundary condition in online social networks. *Discrete and Continuous Dynamical Systems Series B*. 20(6):1609-1624.
2. Kabanikhin S.I. (2008) Definitions and examples of inverse and ill-posed problem. *Journal of Inverse and Ill-Posed Problems*. 16(4):317-357.
3. Zheltkova V.V., Zheltkov D.A., Grossman Z., Bocharov G.A., Tyrtshnikov E.E. (2017) Tensor based approach to the numerical treatment of the parameter estimation problems in mathematical immunology. *Journal of Inverse and Ill-posed Problems*. 26(1):51-66.
4. Wang H., Wang F., Xu K. (2013) Modeling information diffusion in online social networks with partial differential equations. arXiv:1310.0505 [cs.SI].

# Agent-based modelling of genetic deafness propagation under various sociodemographic conditions

S.A. Lashin<sup>1,2\*</sup>, Yu.G. Matushkin<sup>1,2</sup>, A.A. Smirnova<sup>1</sup>, G.P. Romanov<sup>3,4</sup>, O.L. Posukh<sup>1,2</sup>

<sup>1</sup> *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

<sup>2</sup> *Novosibirsk State University, Novosibirsk, Russia*

<sup>3</sup> *MK Ammosov North-Eastern Federal University, Yakutsk, Russia*

<sup>4</sup> *Yakut Science Centre of Complex Medical Problems, Yakutsk, Russia*

\* e-mail: lashin@bionet.nsc.ru

**Key words:** deafness, agent based simulations, population genetics, social demography

*Motivation and Aim:* Genetic deafness is a heterogeneous disability with different inheritance patterns, the most common form of which is a recessive deafness caused by mutations in connexin-26 gene (*Cx26*, *GJB2*). High frequency of *Cx26*-associated deafness in European and American populations was previously assumed to be caused by combined effects of assortative mating tradition among deaf people based on linguistic homogamy (sign language) and relaxed selection against deafness [1, 2]. Although there were some attempts to study interrelations between social and genetic factors in spreading of genetic deafness using mathematical modelling and computer simulations [2], there is still lack of both methodological basis and software tools for this task. In this study, we present an agent-based model “Population Genetics of Deafness” (PGD) along with its software implementation and biologically relevant simulation results.

*Methods and Algorithms:* Agent-based model “Population Genetics of Deafness” (PGD) describes propagation of genetic deafness in human populations through generations. The model takes into account social factors of mating, such as: choice-by-phenotype and sign language. Model allows simulating either simple mendelian (one-, two-, or multi-locus, bi- or multi-allelic) or multifactorial genotype-phenotype interactions. The software implementation of PGD is based on Diploid Evolutionary Constructor framework [3].

*Results:* We have considered series of simulations (20-fold for each scenario) of deafness propagation varying the following parameters: initial population size (including frequencies of each cohorts), sign language assortativity and phenotypic assortativity. Simulation results are counter-intuitive versus [1, 2]: increase of assortative mating in deaf subpopulation led to decrease of the frequency of “deaf” alleles after 25 generations. It is especially true for the cases when increased assortativity is associated with relatively high significance of sign language proficiency and low number of its carriers in initial population.

*Conclusion:* We have presented an agent-based model “Population Genetics of Deafness” (PGD) and its software implementation. Simulations performed with PGD have shown counter-intuitive behavior of genetic deafness propagation associated with assortative mating and other social factors like sign language expansion in populations.

*Acknowledgements:* The study was supported by the Budget Project 0324-2018-0017.

## References

1. Nance W.E. et al. (2000) *Lancet*. 356(9228):500-501.
2. Nance W. E., Kearsey M. J. (2004) *American Journal Human Genetics*. 74(6):1081-1087.
3. Lashin S.A., Matushkin Yu.G. (2013) Multi-layer computer models of gene network evolution in diploid populations. *FEBS Journal*. 280:563-563.

## Author index

Akberdin I. 40  
Andryushchenko V.A. 10  
Asmanova N. 11  
Aviño Esteban L. 12

Bachinsky A. 48  
Baranov V.I. 32  
Barberis M. 45  
Bektemessov Zh. 13  
Berestov V.V. 15  
Beskaravayny P. 52  
Beyer T. 31  
Bocharnikov A. 71  
Bogatyreva N.S. 12  
Bord E.E. 15  
Bragina E. 76

Cal-Gonzalez J. 14, 31  
Cheevadhanarak S. 34  
Chekmarev S.F. 10  
Cherevko A.A. 15  
Chernykh I. 16  
Chugunov A. 17

Dedok V. 18  
Demenkov P. 69  
Denisova N.V. 19, 29, 51  
Dosenko V. 76  
Dvoryakova E.A. 17

Efimov K. 20  
Efimov V. 20  
Efremov R. 17  
Elpidina E.N. 17  
Eremeev A. 21  
Ermak T. 22

Filippova I.Yu. 17  
Firsov A. 23  
Fomin E. 24

Gavrilov D.A. 57  
Glinskiy B. 16  
Glukhova N.V. 56  
Golubyatnikov V. 26, 27, 28  
Gradov V. 26  
Grishchenko M. 25

Grosse I. 46  
Gunbin K.V. 64

Hofestadt R. 76

Ilin A.I. 11  
Ivanisenko V. 69, 76  
Ivankov D.N. 12

Kabanikhin S.I. 30, 37, 42, 74, 75, 77  
Kamzolova S. 52  
Kashtanova V.N. 30  
Kertesz H. 31  
Khairtdinov M. 25  
Khe A.K. 15  
Khlebodarova T.M. 60  
Khongthon P. 34  
Kirillova N. 27  
Kiselev I.N. 32, 33  
Kittichotirat W. 34  
Klimenko A.I. 35  
Kolchanov N. 71  
Kolpakov F.A. 32, 33  
Kolpakova A.F. 33  
Kondakova E. 37  
Kondrashov F.A. 12  
Kovaleva V. 20  
Kozlov K. 36  
Kramorenko N. 66  
Krivorotko O.I. 29, 30, 37, 38, 42, 74, 75, 77  
Krutinin G. 52  
Krutinina E. 52  
Kuchin N. 16  
Kulikov I. 39  
Kulyashov M. 40  
Kurmanalieva G.S. 61  
Kusonmano K. 34

Lakiza T.V. 56  
Lashin S.A. 35, 41, 46, 78  
Latyshenko V. 42  
Lazareva A. 25  
Liebermeister W. 45  
Lifenko V. 43  
Likhoshvai V.A. 60  
Linke C. 45  
Lomakin S. 16



- Loog M. 45  
Lukyanenko D. 65
- Maier A. 76  
Marchenko M.A. 44  
Matushkin Yu.G. 35, 41, 78  
Minushkina L. 28  
Mironova V.V. 46, 73  
Mondeel T.D.G.A. 45  
Mukatova Z. 55  
Mukhin A.M. 46, 73
- Nafikova A. 47  
Nikolaev S. 55  
Nizolenko L. 48  
Nolde D. 17  
Novikov N. 49  
Novikova L. 36  
Novoselova N. 50  
Nuzhdin S. 36
- Omelyanchuk N.A. 73  
Ondar M. 51  
Orlov K.Y. 15  
Osypov A. 52, 53
- Palyanov A.Yu. 10, 54  
Panarin V.A. 15  
Peltek S. 40  
Penenko A. 55  
Pesotskaya L.A. 56  
Podkolodnaya O.A. 57, 58  
Podkolodnyy N.L. 57, 58  
Popadin K.Yu. 64  
Popova I.Yu. 53  
Posas F. 45  
Posukh O.L. 41, 78  
Prikhodko A. 59
- Rausch I. 31  
Ree N.A. 60  
Renner A. 31  
Romanov G.P. 41, 78  
Rozanov A. 40
- Samsonova M. 36  
Satybaev A.J. 61  
Seferova I. 36  
Senotrusova S.D. 62  
Serezhnikova T. 63  
Shamanskiy V.A. 64  
Shehovtsov A. 22  
Shepelin A.V. 70  
Shishlenin M.A. 15, 49, 59, 65  
Shlikht A. 66  
Shoshi A. 76  
Skvortsova M.A. 67  
Smirnova A.A. 41, 78  
Spirov 21
- Takuadina A. 68  
Tereshchenkova V.F. 17  
Titov I. 23  
Tiys E. 69  
Tognetti S. 45  
Tretiak T.O. 56  
Tverdohkleb N.N. 58
- Vainer B.G. 70  
Vishnevsky O. 71  
Voronov D. 43  
Voropaeva O.F. 62
- Wang Yu. 75  
Westerhoff H.V. 45, 72  
Wiebe D.S. 46, 73
- Yakimenko A. 25  
Yakovlev P. 22  
Yermolenko D. 74
- Zhang Sh. 75  
Zolotareva O. 76  
Zubairova U. 55  
Zvonareva T. 77



Компания Huawei является ведущим мировым поставщиком ИКТ-решений. Благодаря установлению взаимовыгодных отношений с нашими партнерами и заказчиками компании Huawei удалось добиться существенных преимуществ в сфере операторских сетей, корпоративного и потребительского бизнеса, а также в сфере облачных технологий. Мы стремимся создавать максимальные преимущества для операторов связи, предприятий и потребителей путем разработки конкурентных ИКТ-решений и услуг. Оборудование и решения Huawei используются в более чем 170 странах мира. Компания обслуживает более трети населения земного шара.

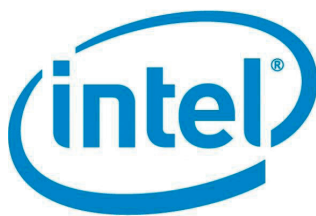
Имея богатый опыт и технические знания в области НИОКР, Huawei придерживается стратегии тесного сотрудничества и интеграции с корпоративными заказчиками и предоставляет им широкий спектр высокоэффективных клиентоориентированных ИКТ-решений и услуг, на базе глубокого понимания их потребностей. Согласно этой стратегии Huawei предлагает широкий выбор передовых ИКТ-решений в сфере государственного управления, общественного сектора, финансов, транспорта, электроэнергетики, крупных предприятий, а также малых и средних предприятий (SME). Эти решения охватывают корпоративные сети, универсальные системы связи и взаимодействия (UC&C), системы облачных вычислений и центры данных, системы корпоративной беспроводной связи, сетевого электропитания, а также инфраструктурные услуги.

ООО «Техкомпания Хуавэй»  
Филиал в СФО:  
630112, Новосибирск, ул. Фрунзе, 242,  
11-й этаж  
Тел.: +7(383) 328 00 70  
Факс: +7(383) 328 00 71  
E-mail: [Kroshin.Fyodor@huawei.com](mailto:Kroshin.Fyodor@huawei.com)  
URL: [e.huawei.com/ru](http://e.huawei.com/ru)

Huawei is a leading global ICT solutions provider. Through our dedication to customer-centric innovation and strong partnerships, we have established end-to-end capabilities and strengths across the carrier networks, enterprise, consumer, and cloud computing fields. We are committed to creating maximum value for telecom carriers, enterprises and consumers by providing competitive ICT solutions and services. Our products and solutions have been deployed in over 170 countries, serving more than one third of the world's population.

By leveraging our strong R&D capabilities and comprehensive technical expertise, Huawei's strategy in the enterprise domain focuses on close cooperation and integration with partners to deliver a wide range of highly efficient customer-centric ICT solutions and services that are based on a deep understanding of customer needs. In line with our strategy, we offer a broad portfolio of innovative ICT solutions that cater to global vertical industry and enterprise customers across government and public sector, finance, transportation, energy, large enterprises and small and midsize enterprises (SMEs). Our portfolio covers enterprise networking, unified communications & collaboration (UC&C), cloud computing & data center, enterprise wireless, network energy and infrastructure services.

HUAWEI Technologies Co Ltd., Russia  
Siberia office:  
630112, Russia, Novosibirsk, Frunze Str., 242  
Business Center "New Height"  
Tel.: +7(383) 328 00 70,  
Fax: +7(383) 328 00 71  
Email: [Kroshin.fyodor@huawei.com](mailto:Kroshin.fyodor@huawei.com)  
URL: [e.huawei.com/ru](http://e.huawei.com/ru)



---

### **Корпорация Intel**

Корпорация Intel была основана в 1968 году Робертом Нойсом и Гордоном Муром. На протяжении 50 лет Intel создает инновационные технологии, открывающие новые возможности для людей.

Корпорация Intel является мировым лидером в области микроэлектроники и информационных технологий. Intel создает технологии для умного мира эпохи больших данных. Основное внимание корпорация уделяет созданию интеллектуальных решений для умного мира, от устройств Интернета вещей и пользовательских ПК до коммуникационной инфраструктуры, технологий для центров обработки данных и суперкомпьютеров.

Штаб-квартира корпорации расположена в г. Санта-Клара, шт. Калифорния. Общий штат Intel насчитывает более 100 тыс. сотрудников в более, чем 60 странах по всему миру. Главным исполнительным директором корпорации является Роберт Свон (Robert Swan).

### **Intel в России**

Первое представительство Intel в России было открыто в 1991 году в Москве. Сегодня в российских офисах Intel в Москве и Нижнем Новгороде работают более 800 человек.

В московском офисе компании представлены отделы маркетинга и развития бизнеса, группы по разработке программного обеспечения, юридический отдел.

В НИОКР центре Intel в Нижнем Новгороде создаются новые и инновационные продукты для разработки ПО. Сегодня он является одним из крупнейших центров исследований и разработок Intel в Европе. Более **700 специалистов и инженеров** разрабатывают программные инструменты и приложения для архитектур Intel. В Нижнем Новгороде также размещаются различные группы поддержки бизнеса (например, административно-хозяйственная часть, финансовый отдел, отдел ИТ, отдел кадров).

### **Центр исследований и разработок Intel в Нижнем Новгороде**

Нижегородский офис Intel был является центром экспертизы корпорации в области высокопроизводительных вычислений, разработки программного обеспечения в области численных методов и беспроводной связи.



✉ MP Biomedicals  
(ООО «МПБА диагностика»)  
Адрес: 109147, г. Москва, ул. Марксистская, д. 3,  
стр. 2, оф. 2.1.20/2  
☎ Тел./факс:  
+7(495)604-13-44  
E-mail rus@mpbio.com

WEB mpbio.com; mpbio.ru

Компания ООО «МПБА диагностика» является дочерней компанией MP Biomedicals, ранее известной как ICN Biomedicals, основанной в 1959 году, признанного лидера в области производства широкого спектра химических реактивов, оборудования для пробоподготовки (система для гомогенизации FastPrep) и наборов реагентов. Каталог продукции компании MP Biomedicals включает более 55000 наименований высококачественных продуктов для проведения биохимических исследований, фармацевтического и биотехнологического производства, для различных отраслей иммунологии и генетики.





ООО «Рош Диагностика Рус» – официальный импортер продукции Roche в России и лицензиат компании F.Hoffmann–La Roche Ltd.

Roche Sequencing Solutions, подразделение Roche, ориентированное на решения для NGS, а в частности на пробоподготовку к NGS, предлагает:

-Наборы **KAPA Biosystems** для приготовления библиотек ДНК (включают баркодированные адаптеры, частицы для очистки, наборы для оценки концентраций ДНК и библиотек методом ПЦР в реальном времени).

-Наборы для направленного отбора генов перед NGS:

**NimbleGen SeqCap EZ** –гибридизационное обогащение панелей генов, экзомов, транскриптомов и метиломов;

**HEAT-seq** - амплификационное обогащение панелей генов, в том числе и панелей онкогенов;

**AVENIO** –гибридизационное обогащение панелей онкогенов из внеклеточной опухолевой ДНК и анализ данных.

ООО «Рош Диагностика Рус» предлагает комплексные решения, включающие в себя не только оборудование и реагенты, но и технический сервис, обучение персонала и постоянную методическую поддержку.



Для научных исследований. Не для диагностики.



ООО «Рош Диагностика Рус»  
115114, Россия, Москва  
Ул. Летниковская, д. 2, стр. 2  
Тел: +7 (495) 229-69-99  
Бизнес-центр «Вивальди Плаза»  
[ras.russia@roche.com](mailto:ras.russia@roche.com)  
[sequencing.roche.com](http://sequencing.roche.com)



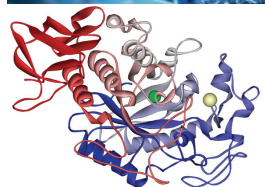
ООО «БиоЛайн» является официальным дистрибьютором компаний **BD Biosciences, Leica Microsystems, PerkinElmer, BioTek**, ведущих мировых производителей приборов и реагентов для биомедицинских исследований.

Благодаря уникальному портфолио продукции и опыту наших специалистов мы выполняем поставки и внедрение комплексных решений для разнообразных задач в области молекулярной и клеточной биологии.



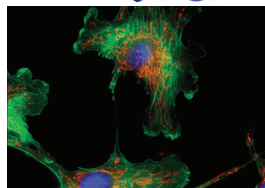
## Молекулярно-генетические исследования

- Системы для выделения и молекулярного анализа одиночных клеток Becton Dickinson
- Станции для выделения ДНК, оборудование PerkinElmer для подготовки и контроля библиотек для NGS
- Наборы Nextflex для подготовки библиотек NGS PerkinElmer: полногеномное и таргетное секвенирование, транскриптомика, эпигенетика, метагеномика



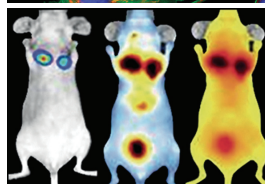
## Протеомные исследования

- Передовые оптические технологии компании BioTek Instruments для биохимических исследований, идентификации и количественной оценки аналитов, исследования взаимодействия биомолекул
- Реагенты и расходные материалы PerkinElmer для протеомных исследований



## Клеточные исследования

- Системы для проточной цитометрии и сортировки клеток компании BD Biosciences
- Оптическая визуализация клеток для моделирования процессов в клеточных культурах и на 3D сфероидсах: решения PerkinElmer и BioTek Instruments
- Системы для конфокальной микроскопии Leica Microsystems



## Исследования на животных

- Приборы для оптической визуализации *in vivo* Spectrum и Lumina, системы для КТ и ПЭТ компании PerkinElmer
- Оборудование для исследований на животных Leica Biosystems

Официальные дистрибьюторы **BD Biosciences, Leica Microsystems, PerkinElmer, BioTek** в России — компания «БиоЛайн»

**ООО «БиоЛайн»**  
Россия, 197101, Санкт-Петербург  
Пинский пер., д. 3, Лит. А  
тел.: +7 (812) 320 49 49  
факс: +7 (812) 320 49 40  
e-mail: [main@bioline.ru](mailto:main@bioline.ru)  
[www.bioline.ru](http://www.bioline.ru)

Москва, тел.: +7 (800) 555 49 40  
Новосибирск, тел.: +7 (383) 227 09 63  
Екатеринбург, тел.: +7 (343) 287 32 49  
Владивосток, тел.: +7 (423) 201 18 08  
Нижний Новгород, тел.: +7 (831) 278 61 47  
Ростов-на-Дону, тел.: +7 (863) 268 99 32  
Казань, тел.: +7 (843) 570 66 88  
Самара: +7 (846) 246-06-54

Единый бесплатный  
номер сервисной  
службы для всех  
регионов России:

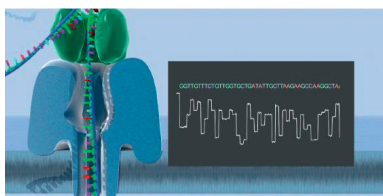
**8 800 333 00 49**





Компания Диаэм – крупнейший поставщик современного лабораторного оборудования на Российском рынке. Каталог компании насчитывает более 500 000 наименований приборов, реагентов и расходных материалов для медицинских и научно-исследовательских лабораторий. В каталоге компании представлена продукция ведущих мировых производителей, как: Abcam, Applied Biosystems, Binder, Bio-Rad, Corning, Eppendorf, Illumina, Ion Torrent, Lexogen, Oxford Nanopore Technologies, Panasonic (Sanyo), Sage Sciences, Sigma-Aldrich, Thermo Fisher Scientific, Qiagen:

- Наборы для подготовки библиотек, для высокопроизводительного секвенирования NGS, для исследовательских работ и, в онкологии, репродуктивной медицине, в изучении наследственных заболеваний, реагенты и наборы для капиллярного секвенирования.
- Секвенаторы капиллярные и высокопроизводительные NGS, оборудование для анализа качества НК для NGS, роботизированные станции для подготовки библиотек и секвенирования.
- Все для ПЦР, реагенты, наборы, пластик, амплификаторы.
- Нанопоровые секвенаторы Oxford Nanopore Technologies, наборы для секвенирования ДНК и РНК.



### Секвенирование теперь доступно каждому!

Диаэм сегодня представляет продукцию [Oxford Nanopore Technologies](https://www.oxfordnanopore.com/) – это секвенаторы третьего поколения – [MiniION](#), [GridION](#), [PromethION](#).

Технология секвенирования [Oxford Nanopore Technologies](https://www.oxfordnanopore.com/) позволяет делать прямое прочтение цепей ДНК или РНК в режиме онлайн, длина рида ограничена только длиной фрагмента, а портативность оборудования и быстрая подготовка библиотек дает возможность секвенировать даже в полевых условиях с минимальными требованиями к генетической лаборатории. С [Oxford Nanopore Technologies](https://www.oxfordnanopore.com/) секвенировать теперь может каждый, даже тот, кто ранее и не задумывался о секвенировании - это просто и доступно.

[Секвенирование третьего поколения](#) не заменяет и не отменяет применение [капиллярных секвенаторов по Сэнгеру](#) или [платформ NGS второго поколения](#), наоборот, сочетание трех поколений генетического анализа открывает новые возможности получения ранее неизвестных данных. Специалисты [Диаэм](#) прошли обучение в [Oxford Nanopore Technologies](https://www.oxfordnanopore.com/), осуществляют профессиональное консультирование и техническую поддержку, помогут спланировать эксперимент и подобрать необходимые наборы реагентов для решения конкретной задачи независимо от бюджета лаборатории.

Обращайтесь в любой из наших офисов или пишите на [sales@dia-m.ru](mailto:sales@dia-m.ru)

## 000 «Диаэм»

[www.dia-m.ru](https://www.dia-m.ru)

Москва  
ул. Магаданская, 7/3  
тел./факс:  
(495) 747-0508  
[sales@dia-m.ru](mailto:sales@dia-m.ru)

Новосибирск  
пр. Акад.  
Лаврентьева, 6/1  
тел./факс:  
(383) 328-0048  
[nsk@dia-m.ru](mailto:nsk@dia-m.ru)

Казань  
ул. Парижской  
Коммуны, д. 6  
тел./факс:  
(843) 210-2080  
[kazan@dia-m.ru](mailto:kazan@dia-m.ru)

Санкт-Петербург  
ул. Профессора  
Попова, 23  
тел./факс:  
(812) 372-6040  
[spb@dia-m.ru](mailto:spb@dia-m.ru)

Ростов-на-Дону  
пер. Семашко, 114  
тел./факс:  
(863) 250-0006  
[rnd@dia-m.ru](mailto:rnd@dia-m.ru)

Пермь  
Представитель  
в УФО  
тел./факс:  
(342) 202-2239  
[perm@dia-m.ru](mailto:perm@dia-m.ru)

Воронеж  
тел./факс:  
(473) 232-4412  
[voronezh@dia-m.ru](mailto:voronezh@dia-m.ru)



АЛЬБИОГЕН

illumina®

AUTHORIZED  
CHANNEL PARTNER

## Компания АЛЬБИОГЕН — официальный дистрибьютор illumina и Lucigen

Компания ООО «АЛЬБИОГЕН» с 2015 года является эксклюзивным (единственным) официальным торговым представителем и дистрибьютором компании [illumina](#) на территории Российской Федерации, Республики Беларусь, Республики Казахстан и Республики Узбекистан.

Нашей задачей является обеспечение полного доступа клиентов к передовым технологиям и сервисам illumina, включая современные системы NGS и анализа ДНК-биочипов, программное обеспечение для биоинформатики и весь спектр реактивов.

ООО «АЛЬБИОГЕН» предоставляет полный комплекс услуг, связанных с продажей, технической поддержкой и сервисным (гарантийным и постгарантийным) обслуживанием продукции компании Illumina, а также обучением пользователей работе на данном оборудовании.

Инновационная и стремительно развивающаяся компания illumina Inc., являющаяся мировым лидером в области геномных технологий, заключила соглашение с компанией АЛЬБИОГЕН, специализирующейся на поставках оборудования и расходных материалов для секвенирования нового поколения (NGS) и анализа на ДНК-биочипах.

Новейшие продукты компании illumina, создаваемые совместно с ведущими мировыми учеными, позволяют изучать геном на очень глубоком уровне и дают возможность для новаторских достижений в науке, медицине, сельском хозяйстве и потребительской геномике. Более 90% научных статей, связанных с технологиями секвенирования нового поколения, сделаны при помощи оборудования Illumina.

Сотрудничество с компанией АЛЬБИОГЕН направлено на то, чтобы сделать технологии NGS и анализа ДНК-биочипов более доступными на территории Российской Федерации и в странах СНГ.

Компания АЛЬБИОГЕН использует свой обширный опыт в области продаж и продвижения продукции, знания передовых технологий и сеть региональных представителей для обеспечения быстрой, эффективной и бесперебойной работы лабораторий клиентов illumina.

Компания АЛЬБИОГЕН также является официальным дистрибьютором компании Lucigen, основными продуктами которой являются ферменты и реагенты для секвенирования нового поколения и молекулярной диагностики.



Компания СкайДжин предлагает к поставке со склада в Москве и под заказ наборы реагентов, оборудование, расходные материалы, реактивы, а также специализируется на сервисном обслуживании и поверке дозаторов, лабораторных весов различных производителей. Мы предлагаем гибкие условия работы и очень большой ассортимент продукции.

Поставляемая нашей компанией продукция широко используется в научно-исследовательских лабораториях и R&D центрах, лабораториях секвенирования, при решении практически любых молекулярно-биологических задач.

Большая часть производителей в нашем портфолио - это прямые, эксклюзивные поставки. Мы являемся первым звеном в поставках для таких компаний как New England Biolabs, Agilent Technologies, Oxford Nanopore Technologies, QIAGEN, 10x Genomics, NIMAGEN, Integrated DNA Technologies, Thermo Fisher Scientific, SIGMA-ALDRICH, BioSan, Gilson.

К флагманским продуктам наших линеек относятся:

- Набор для пробоподготовки образцов от New England Biolabs ULTRA II FS с интегрированной системой фрагментации и другие наборы серии ULTRA для образцов ДНК, РНК и микроРНК;
- Digital NGS: готовые панели и наборы для обогащения на основе ПЦР от QIAGEN с мономолекулярным баркодированием;
- Специализированные наборы для работы с микроРНК и анализа экспрессии от QIAGEN-Exiqon;
- Нанопоровые секвенаторы третьего поколения: портативный секвенатор MinION, высокопроизводительный секвенатор GridION;
- Уникальная система Chromium производства 10x Genomics для автоматической пробоподготовки геномов и транскриптомов единичных клеток.

За дополнительной информацией о производителях, товарах, ценах и условиях поставки обращайтесь к нашим квалифицированным специалистам.

Будем рады ответить на Ваши вопросы и помочь выбрать качественное и недорогое решение для Ваших задач!

ООО «СкайДжин»  
Адрес: 115093, Москва,  
ул. Люсиновская, д. 36, стр. 1  
Тел: 8 (495) 215 02 22  
[info@skygen.com](mailto:info@skygen.com)  
[www.skygen.com](http://www.skygen.com)



## ХИМЭКСПЕРТ Агентство Химэксперт

### Информация о компании:

Компания Химэксперт существует 16 лет и давно зарекомендовала себя, как надежный поставщик приборов, реактивов и расходных материалов для молекулярной биологии. Мы собрали для своих клиентов самые интересные и перспективные бренды, большинство из которых в России можно приобрести только у нас.

Химэксперт предлагает оборудование для анализа ДНК и РНК, в том числе и методами NGS, фундаментальных протеомных и цитологических исследований, фармацевтики и биотехнологий, прикладного тестирования, включая идентификацию личности и установление родства в криминалистике и судебно-медицинской экспертизе.

Наши клиенты выбирают Химэксперт потому что:

- Химэксперт всегда находит самые прогрессивные решения в области Life Sciences. Наша компания постоянно расширяет свое портфолио и в курсе последних веяний в области молекулярной биологии
- Химэксперт осуществляет полную техническую и методическую поддержку наших клиентов: обратившись к нам, вы получаете помощь квалифицированных сотрудников в подборе оборудования и реагентов под поставленные задачи и их последующем использовании
- Химэксперт стремится идти навстречу заказчикам и осуществлять быстрые поставки, так как скорость и четкость исполнения заказов очень важна.

Обратившись к нам, вы можете быть уверены в будущем своего эксперимента.  
Начните сотрудничество с компанией Химэксперт и убедитесь в этом на своем опыте!

ООО «Агентство Химэксперт» 125009, г. Москва, Страстной б-р, д. 4, оф. 101

Тел: +7 (495) 629 28 69, 650 36 66

info@khimexpert.ru,

www.khimexpert.ru



The geneXplain GmbH is glad to welcome you at the BGRS/SB'2018 conference and is proud to introduce you the following software and database solutions for the needs of bioinformatics, systems biology and systems medicine:



[geneXplain platform](#) – is a high-performance tool for multi-omics data analysis, which allows identification of new therapeutic targets and biomarkers. A unique feature of the geneXplain platform is its Upstream Analysis. You can [register](#) and immediately receive access to a free account.



[TRANSFAC database](#) – is a unique collection of transcription factors, their experimentally validated binding sites (TFBS) and a widely known library of positional weight matrices (PWMs). The database has its own integrated methods for TFBS search. It can also be used as an integral part of the geneXplain platform. TRANSFAC is available online or can be downloaded as a set of flat files.



[TRANSPATH database](#) – is one of the biggest and most famous collections of signaling and metabolic pathways, which counts over 489000 reactions. The database can be applied for master-regulators search within the geneXplain platform. TRANSPATH is also available online in one package with HumanPSD database or can be downloaded as a set of flat files.



[HumanPSD database](#) – is a collection of genes, proteins and micro-RNAs, which includes information about disease biomarkers and clinical trials for various diseases. Besides the detailed biomarkers data, the database contains information about drugs.



[BRENDA database](#) – is a comprehensive enzyme and enzyme-ligand information system. Its manually derived core contains over 3 million data points about 77,000 enzymes annotated from 135,000 literature references.



[PASS](#) – is a software tool for evaluating the general biological potential of organic compounds based on their structural formula. This program predicts main and side pharmacological effects, molecular mechanisms of action, specific toxicities, and antitargets, actions associated with the metabolism and transport of pharmaceutical substances and their influence on gene expression.



[PharmaExpert](#) – is a software tool for analysis of the biological activity spectra of substances predicted by PASS and selecting compounds with the desirable set of biological activity, for analyzing the relationships between biological activities, drug-drug interactions and for multiple targeting of chemical compounds.



[GUSAR](#) – is a software tool for analysis of quantitative structure-activity/structure-property relationships (QSAR/QSPR) based on the structural formulas of the compounds and data on their activity/property, and for prediction of activity/property for new compounds. GUSAR can be easily applied to different routine QSAR/QSPR tasks, for building multiple models, and for prediction of the different quantitative values simultaneously.

If you got interested in any of the products, provided by GeneXplain, or you have any questions, please contact us by e-mail [info@genexplain.com](mailto:info@genexplain.com). We will be glad to help you!

Научное издание

**MATHEMATICAL MODELING AND HIGH-PERFORMANCE COMPUTING  
IN BIOINFORMATICS, BIOMEDICINE AND BIOTECHNOLOGY  
(MM-HPC-BBB-2018)**

The 3rd International Symposium

Abstracts

*Printed without editing*

**МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ  
И ВЫСОКОПРОИЗВОДИТЕЛЬНЫЕ ВЫЧИСЛЕНИЯ  
В БИОИНФОРМАТИКЕ, БИОМЕДИЦИНЕ И БИОТЕХНОЛОГИИ  
(ММ-НРС-ВВВ-2018)**

Третий международный симпозиум

Тезисы докладов

*Публикуется в авторской редакции*

Выпуск подготовлен информационно-издательским отделом ИЦиГ СО РАН

---

Подписано к печати 31.07.2018. Формат 70 × 108 <sup>1</sup>/<sub>16</sub>. Усл. печ. л. 7,88.  
Тираж 100 экз. Заказ № 187

---

Федеральный исследовательский центр  
«Институт цитологии и генетики Сибирского отделения Российской академии наук»  
630090, Новосибирск, проспект Академика Лаврентьева, 10

Отпечатано в типографии ФГУП «Издательство СО РАН»  
630090, Новосибирск, Морской проспект, 2