RUSSIAN ACADEMY OF SCIENCES
SIBERIAN BRANCH

INSTITUTE OF CYTOLOGY AND GENETICS

THE SEVENTH INTERNATIONAL CONFERENCE ON BIOINFORMATICS OF GENOME REGULATION AND STRUCTURE\SYSTEMS BIOLOGY

Abstracts

BGRS\SB'10
Novosibirsk, Russia
June 20–27, 2010

Novosibirsk
2010

INTERNATIONAL PROGRAM COMMITTEE

Nikolay Kolchanov, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia (Chairman of the Conference)
Ralf Hofestadt, University of Bielefeld, Germany (Co-Chairman of the Conference)
Konstantin Skryabin, "Bioingineering" Center, RAS, Moscow, Russia (Co-Chairman of the Conference)
Dagmara Furman, Institute of Cytology and Genetics SB RAS, Novosibirsk, (Conference Scientific Secretary)
Dmitry Afonnikov, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Lev Beloussov, Moscow State University, Moscow, Russia
Sergey Deyev, Shemyakin/Ovchinnikov Institute of Bioorganic Chemistry RAS, Moscow, Russia
Roman Efremov, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry RAS, Moscow, Russia
A. Fazel Famili, University of Ottawa, IIT/ITI - National Research Council Canada, Ottawa, Canada
Igor Goryanin, Edinburgh Centre for Bioinformatics, University of Edinburgh, Edinburgh, United Kingdom
Charlie Hodgman, Multidisciplinary Centre for Integrative Biology, Centre of Plant Integrative Biology, University of Nottingham, UK
Vladimir Ivanisenko, Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
Alexis Ivanov, Institute of Biomedical Chemistry RAMS, Moscow, Russia
Yoichiro Iwakura, Center for Experimental Medicine and Systems Biology, Institute of Medical Science, University of Tokyo, Tokyo, Japan
Alexey Kochetov, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Fyodor Kondrashov, Centre for Genomic Regulation, Barcelona, Spain
Vladimir Kuznetsov, Division of Genome and Gene Expression Data Analysis, Bioinformatics Institute, Singapore
Mikhail Lavrentiev, Sobolev Institute of Mathematics, Institute of Automation and Electrometry, Novosibirsk, Russia
Andrey Lisitsa, Institute of Biomedical Chemistry of RAMS, Moscow, Russia
Luciano Milanesi, National Research Council - Institute of Biomedical Technologies, Milan, Italy
Viatcheslav Mordvinov, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Mikhail Moshkin, Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
Timo Nevalainen, Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
Ludmila Ogorodova, SibNMU of Russian Health Deparment, Tomsk, Russia
Youri Pavlov, Eppley Institute for Cancer Research, University of Nebraska Medical Center, Omaha, USA
Vladimir Poroikov, Laboratory for Structure-Function Based Drug Design, Institute of Biomedical Chemistry of RAMS, Moscow, Russia
Egor Prokhortchouk, Center "Bioengineering" of the Russian Academy of Sciences, Moscow, Russia
Igor Rogozin, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, USA
Alexei Romanyukha, Institute of Numerical Mathematics RAS, Moscow, Russia
Nikolay Rubtsov, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Maria Samsonova, St.Petersburg State Polytechnic University, St.Petersburg, Russia
Dmitry Sherbakov, Limnological Institute SB RAS, Irkutsk, Russia
Vladimir Shumny, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Banchob Sripa, Khon Kaen University, Thailand
Georges St. Laurent, St. Laurent Institute, Providence, USA
Evgenii Vityaev, Sobolev Institute of Mathematics, Novosibirsk, Russia
Valentin Vlassov, Institute of Chemical Biology and Fundamental Medicine of SB RAS, Novosibirsk, Russia

LOCAL ORGANIZING COMMITTEE

Ilya Akberdin, Institute of Cytology and Genetics, Novosibirsk, Russia (Chairperson)
Victoria Mironova, Institute of Cytology and Genetics, Novosibirsk, Russia (vice-chairperson)
Galina Orlova, Institute of Cytology and Genetics, Novosibirsk, Russia (secretary)
Svetlana Zubova, Institute of Cytology and Genetics, Novosibirsk, Russia (financial manager)
Sergei Amstislavsky, Institute of Cytology and Genetics, Novosibirsk, Russia
Kirill Bezmaternykh, Institute of Cytology and Genetics, Novosibirsk, Russia
Luba Glebova, Institute of Cytology and Genetics, Novosibirsk, Russia
Nadezda Glebova, Institute of Cytology and Genetics, Novosibirsk, Russia
Tatyana Karamisheva, Institute of Cytology and Genetics, Novosibirsk, Russia
Andrey Kharkevich, Institute of Cytology and Genetics, Novosibirsk, Russia
Galina Kiseleva, Institute of Cytology and Genetics, Novosibirsk, Russia
Anna Onchukova, Institute of Cytology and Genetics, Novosibirsk, Russia
Maxim Ree, Institute of Cytology and Genetics, Novosibirsk, Russia
Natalia Sournina, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

ISBN 978-5-91291-009-8

Organizers

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

Chair of Information Biology

Siberian Branch of the Russian Academy of Sciences

PBSoft Ltd.

The Vavilov Society of Geneticists and Breeders

Russian Scientific Centre “Kurchatov Institute”

Laboratory of Theoretical Genetics

Center “Bioengineering” of the Russian Academy of Sciences

Novosibirsk State University

Sponsors

Federal Agency for Science and Innovation

Russian Foundation for Basic Research

International Science and Technology Center

GOLD SPONSOR

Applied Biosystems

SILVER SPONSORS

Intel

Roche Diagnostics

BRONZE SPONSORS

Hewlett Packard

Bio-Rad

GE Healthcare Life Sciences

Carl Zeiss

Genoanalytica

Contents

REDUCING OF FALSE POSITIVES FOR VARIANT CALLING: SECOND ILLUMINA BASE CALL ESTIMATION

Irina Abnizova*, Steven Leonard, Tom Skelly and Tony Cox.
Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK
Email: ia1@sanger.ac.uk

Key words: Illumina sequencing, conditional error probability, second call conditional error probability, second call

Motivation: The new generation of short read sequencing technologies still requires both accuracy of data processing methods and reliable measures of that accuracy. The measures of accuracy, quality values, are very important for many biological application, especially for variant calling. Not only probability of sequencing error is important, but also the conditional probability of 'correcting' this error, so called 'second best call' probability. It allows to reduce an amount of false positively called SNPs significantly.

Results: We developed the simple and reliable way to predict an error probability for Illumina sequencing. Based on the quality value for each called base, we estimate possible second best call, and its conditional probability. We also found some context dependency of error, and incorporated it into resulting probability. We used several phX runs, Genome Analyser 2, to validate our method.

Conclusion: Surprisingly, that around 70% of errors can be 'corrected' with the second call.

Availability: The recalibration and second call pipeline are implemented into Sanger Institute production pipeline. The source codes and software for the prelease 4.0 are freely available for scientific community from WTSI repository.

QASMOD: PROTEIN MODEL QUALITY ASSESSMENT SUITE

A.A. Adzhubei*, A.V. Vlasova, P.V. Mazin
Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia
e-mail: alexei.adzhubei@eimb.ru
**Corresponding author*

Key words: *protein structure, modelling, quality assessment, ontology, software*

Motivation and Aim: Computer models of protein structure are less accurate compared to the experimental structures, though they constitute a much faster growing part of all available protein structures, and can be used when experimental data is not available. However the fundamental problem of protein models application is the absence of standard and easily interpretable information on the completeness of modelled structural data and the accuracy (reliability) of the modelled structures. The accuracy of model predictions varies sharply, creating difficulties in defining the possible areas where models can be applied. The project aims to solve this problem. The goal of this project is to standardise model assessment scores, classify models according to their quality and possible areas of application and develop software that will provide model quality linked to possible model application assessment. This would open access to the models data for a wider community of researchers. Ontologies of modelled structures, and models quality and application are being created as part of this effort, based on the MDB-mmCIF modelling data format [1].

Methods and Algorithms: Existing model assessment software has been analysed, seven programs have been selected and used in the Qasmod metaserver: Bala, Modcheck, ProQv, Prosa, Solvx, Victor, Procheck. The selection criteria used were that the programs should be (1) stable and supported by the authors, (2) non-commercial and available for downloading, (3) produce model quality indicators that correlate with the CASP GDT_TS (Global Distance Test Total Score) score [2]. Protein Model Ontology has been developed in OWL format using Protege ontology editor.

Results: Qasmod – protein structure models quality and application assessment software has been developed, and is available as a server and as program suite for users who wish to run it locally. An ontology describing models applications and modelling methods has been developed. Qasmod allows to submit models for assessment, view results interactively and receive by email, and download full assessment results.

Conclusion: Protein structure models quality assessment has been combined with ontology-based classification of possible applications – linked to model quality, and implemented as the Qasmod server.

Availability: Qasmod is available at http://monster.eimb.relarn.ru/testmod/Qasmod.html

Acknowledgments: This work was supported by the RFBR 08-04-12217-ofi grant.

References:

1. A.A. Adzhubei, E. Miglivacca. (1999) MDB Modelling Extension Dictionary, *Available from:* http://mmcif.pdb.org/dictionaries/.
2. A. Kryshtafovych, C. Venclovas, K. Fidelis, J. Moult. (2005) Progress over the first decade of CASP experiments, *Proteins: Structure, Function, and Bioinformatics*, **61(S7)**: 225-236.

SEMANTIC RELATIONSHIPS DERIVED FROM EXPERIMENTAL ANALYSIS EXPERIENCE HELP TO PROCESS AND VISUALIZE EXPERIMENTAL DATA.

D.G. Alexeev*, N.A. Bazaleev, V.M. Govorun
Research Institute for Physico-Chemical Medicine, Federal Medico-Biological Agency, Moscow, Russia
e-mail: exappeal@gmail.com
**Corresponding author*

Key words: *semantic, database, proteogenomic, visualization*

Motivation and Aim: The current state of art in proteogenomic related experimental approaches implies a variety of experimental data to be integrated in one project, including sequencing information, annotation attempts, shot-gun and gel proteomic data, RNA expression information. In case of one laboratory producing all the data described, there usually exists an ad hoc experimental database, linked to the international biological databases. Keeping in mind variety of approaches in semantic integration [1] and existing techniques of semantic data queries [2], we developed set of database add-ins allowing to integrate pre-collected data without changing the database structure and making it readily available for WC3 promoted RDF format [3] export.

Methods and Algorithms: The database scripting part was done in PL/SQL language (version for Oracle database). Visualization part was programmed in ASP format.

Results: Implementation of semantic add- ins to existing database allowed for tight integration of various datasets (also with outer data sources), inferring new dataset relationships and human friendly visualization interfaces. Add-ins were tested on a few projects currently running, including *Spiroplasma milliferum* proteogenomic profiling, comparison of *Mycoplasma gallisepticum* and *Acholeplasma laidlawii* proteomes grown in similar conditions and exposed to stress. We acquired inferred relationships in all the datasets as well as easily readable data relation maps.

Conclusion: Application of database add-ins to existing database stored information allows viewing dataset semantic inter relationships without changing the overall structure and therefore with a little effort. Over more such an add-ins makes the classical relation stored database data readily available for export in WC3 approved RDF format.

Availability: available on request from the authors

References:

1. A Steve Pettifer et al (2009) Visualising biological data: a semantic approach to tool and database integration, *BMC Bioinformatics* 2009, **10** (Suppl 6)**:**S19
2. Perez J et al. (2003) Semantics and Complexity of SPARQL, In: *Lecture Notes in Computer Science*. 30-43 (Springer).
3. D. Beckett. (2004) RDF/XML Syntax Specification (Revised). http://www.w3.org/TR/rdf-syntax-grammar

CRYOBANKING AND REPRODUCTIVE TECHNOLOGIES FOR MAMMALIAN SPECIES CONSERVATION

S.Ya. Amstislavsky
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: amstis@bionet.nsc.ru

Key words: *embryo cryopreservation, mammalian species, genome resource bank*

Motivation and Aim: the aim of this research was to estimate applicability and current status of embryo cryobanking and associated reproductive technologies such as intra- and interspecies embryo transfer for cryobanking of the laboratory (mice), some farm (cattle, sheep, horses) and some endangered/exotic mammalian species.

Methods and Algorithms: presentation is based on the own published experimental study of reproductive biology and embryo technology in mustelids (polecats/ferrets, endangered European mink) and the comprehensive literature reviewing.

Results: Embryo cryopreservation has become a routine practice for maintaining the diversity of mouse genetic resources, as the number of strains increased drastically recent decades; thus the traditional methods of managing the collections of laboratory mouse genotypes are not sufficient nowadays. Moreover, embryo cryobanking, associated with embryo transfer and superovulation, is an important tool for managing such farm animals as cattle and sheep, but horse species represents the challenge for reproductive biologists and cryobiologists. Despite a demand, currently there is no embryo cryopreservation practice in horse breeding programs. Numerous attempts have been made to apply assisted reproductive technologies (ART) including embryo transfer and embryo cryobanking to some exotic/companion/endangered species of *Felidae, Canidae*, and *Mustelidae* families. Our own experimental results including world first successful attempt of embryo cryopreservation in mustelids [1]; thorough investigation of preimplantation embryo development in endangered European mink [2]; and study of reproductive biology of European mink [3], as well as some recent unpublished results on interspecies embryo between European mink and European polecat demonstrate potential for the use GRB concept in the future program of conservation *Carnivora* species.

Conclusion: Different aspects of cryobanking are thoroughly studied for some laboratory and farm animal species; some other mammalian species of mustelids, felids, canids e t c, however, represent a challenge for the application of the GRB concept.

Acknowledgements: This study was supported by the grant Nr. 08-04-00147, Russian Foundation for Basic Research.

References:

1. *H. Lindeberg et al.(2003)* Surgical recovery and successful surgical transfer of conventionally frozen-thawed embryos in the farmed European polecat (*Mustela putorius*). Theriogenology **60**:1515-1526.
2. *S. Amstislavsky et al* (2006) Embryo development and embryo transfer in the European mink (Mustela lutreola), an endangered Mustelidae species. Reprod. Fert. Develop. **18:** 459-467.
3. S. Amstislavsky et al. (2009) Reproduction in the European mink, Mustela lutreola: Oestrous Cyclicity and Early Pregnancy. Reprod. Dom.Anim. **44**: 489-498.

HIGH-THROUGHPUT GENOTYPING USING SECOND GENERATION SEQUENCING TECHNOLOGIES

V. Amstislavskiy*, T. Borodina, D. Parkhomchuk, H. Lehrach, A. Soldatov
Max Planck Institute for Molecular Genetics, Berlin, Germany
e-mail: amstisla@molgen.mpg.de
**Corresponding author*

Key words: *next generation sequencing, genotyping, pwd mouse strain.*

Motivation and Aim: High density genotyping is an important genetic tool. However, current high-throughput genotyping platforms are expensive, labour- and time-consuming. They require preselection of genetic markers and designing of specific oligonucleotides for each marker. We describe an application of a second generation sequencing platform for high-throughput genotyping.

Methods and Algorithms: Using the Illumina sequencing system we were able to perform genetic mapping of hybrids of C57BL6/J and PWD/Ph mouse strains. We have developed a barcoding system which allows to genotype several mice on one lane of Illumina flowcell. Depending on the multiplexing scale (from one to ten mice per lane) it was possible to reveal from ~220,000 to ~25,000 markers.

Results: ~3,7mln PWD/B6 markers were obtained previously by chip-based resequencing [1]. We have considerably enlarged the markers collection by resequencing PWD genome on Illumina 1G sequencing system. About 300,000,000 sequencing reads were obtained. ~50% of them (150,083,706 reads) were uniquely mapped to the reference genome and resulted in ~130,000,000 nucleotide substitutions.

The majority of mismatches were related to the sequencing mistakes. To find reliable markers we take into account only sequencing fragments longer then 26 nucleotides and have applied three groups of selection criteria (high, medium and low).

Applying the same filtering for RNA PWD data reveals about 13.5 thousands markers.

Common PWD database includes perlegen, DNA and RNA data.

Conclusion: The largest PWD markers database was created. It contains more than 20 millions markers, 38% from which are high quality. The quality of the database improves continuously. Using this database several PWDxB6 mouse were genotyped. Interactive web-based tool was developed for visualization of genotyping maps created using second generation sequencing data.

Availability: By authors request.

References:

1. Frazer KA, Wade CM, Hinds DA, Patil N, Cox DR, Daly MJ. (2004) Segmental phylogenetic relationships of inbred mouse strains revealed by fine-scale analysis of sequence variation across 4.6 mb of mouse genome, *Genome Res.,* **14**: 1493-1500.

BIFUNCTIONAL ENZYME DESTABILASE – LYSOZYME. MUTANT FORMS OF RECOMBINANT PROTEIN.

L.L. Zavalova, N.V. Antipova*, [1]M.S. Pavljukov, [2]I.I. Artamonova, [1]I.P. Baskova
Institute of Bioorganic Chemistry RAS, Moscow, Russia e-mail: lz@ibch.ru
[1]Biological faculty of Moscow State university, Moscow, Russia e-mail: saliva1@yandex.ru [2]Vavilov Institute of General Genetics RAS, Moscow, Russia e-mail: irenart@gmail.com **Corresponing author*

Key words: *Destabilase-isopeptidase, lysozyme, bifunctional enzyme, active sites, mutant forms of recombinant protein, moonlighting proteins, domain organisation, medicinal leech salivary gland secretion.*

Motivation and Aim: Destabilase-Lysozyme (Dest-Lys) of the medicinal leech salivary gland secretion defines its thrombolytic properties, depended on endo-ε-(γ-Glu)-Lys-isopeptidase, and bactericidal properties, depended on 1-4β-glycosidase. When previously isolated bifunctional recombinant protein is considered as the basement for the creation of novel medicine with thrombolytic and antimicrobial properties, the fundamental task is the identification of amino acids of its two active sites. The comparative analysis of Dest-Lys primary structure with the sequences of the known i-lysozymes permits us to consider E14 and D26 as participants of glycosidase, and H92 as a participant of peptidase (isopeptidase) active sites.

Methods and Algorithms: Mutant forms of recombinant Dest-Lys with replacements E14A, D26A and H92A were prepared after expression in *E.coli* of special their coding cDNA forms, which was included into transforming pQE30 vector. 4 primers were synthesized and amplified by PCR for synthesis of cDNA, first the matrix and then mutant one. PCR products were separated in agarose gel. The band correspondng to the required fragment was cut out and by use of "Promega"- kit mutant cDNA was prepared. In this cDNA tripled, coding the required amino acid, was replaced by GCC coding alanin. Resulting cDNA was used for restriction and following ligation into pQE30 vector on restriction sites Shall and BamHI. Vector DNA was purified and used for of *E.coli* strain BL21DE3. After the transformation the cell extract was purified with metal-affinity chromatography and renaturated.

*Results:*Mutant recombinant Dest-Lys H92A showed lysozyme activity (cell walls *M. lysodeikticus)* and isopeptidase function (substrate D-dimer). It may evidence of the non-participation of H92 in functioning of peptidase (isopeptidase) active site of Dest-Lys.

The bioinformatics - analysis of the domain organisation of the enzyme has demonstrated that the destabilase (isopeptidase) domain (PFAM-id PF05497) forms the main part of the protein, to the exception of the signal peptide. The corresponding InterPro-family (IPR008597) includes 118 proteins from various eukaryotic organisms. Based on the similarity of 115 N-terminal amino acids the protein can be assigned to the lysozyme-like superfamily (SSF53955) in the SCOP structural classification.

References:

1. L.L.Zavalova, I.P.Baskova, E.V.Barsova, E.V.Snezhkov, S.B.Akopov, S.A.Lopatin (2004) Recombinant Destabilase-Lysozyme: Synthesis *de novo* in *escherichia coli* and Action Mechanism of the Enzyme expressed in *spodoptera frugiperda, Biochmistry (Moscow)* **69,** 952 – 958.
2. I.P.Baskova, L.L.Zavalova (2008) Polyfunctionality of Destabilase, a Lysozyme from the Medicinal Leech, *Russian Journal of Bioorganic Chemistry,* **34**, 304–309
3. L.L.Zavalova, V.N.Lazarev, S.A.Levitsky, T.G.Yudina, I.P.Baskova (2010) Destabilase-Lysozyme from the medicinal leech. Poly-functionality of recombinant protein, *Biochemistry (Moscow),* **75** (in press).

PolyCTLDesigner – SOFTWARE FOR CONSTRUCTING POLY-EPITOPE CYTOTOXIC T-CELL IMMUNOGENS.

D.V. Antonets*, A.Z. Maksyutov, S.I. Bazhan
State Research Center of Virology and Biotechnology "Vector", Koltsovo, Russia
e-mail: antonec@yandex.ru
**Corresponding author*

Key words: *immunoinformatics, polyepitope, T-cell, antigen, antigen processing*

Motivation and Aim: Design of the artificial polyepitope immunogens capable of eliciting high levels of CD8+ CTL responses to is a promising approach in creation of an efficient vaccines. When designing such immunogens, it is necessary to optimize the processing and presentation of contained epitopes. The aim of this work was the creation of software for designing efficient polyepitope immunogens capable of taking into account major steps of MHC class-I dependent processing and presentation of antigens.

Methods and Algorithms: To predict proteasomal and/or immunoproteasomal cleavage our PolyCTLDesigner program utilizes predictive models developed by Toes et al. [1]. To predict peptide binding to TAP (Transporters Associated with Processing) it uses models developed by Peters et al. [2]. For predicting T-cell epitopes PolyCTLDesigner uses TEpredict program [3]. To assess population coverage by selected set of peptides our program uses known genotypic HLA allele frequencies data taken from (http://www.ncbi.nlm.nih.gov/gv/mhc/). Graph theory approach is used by the program to select appropriate spacer sequences between epitopes and to construct optimal aminoacid sequence of desired polyepitope immunogen. Program was written in Python programming language (http://python.org).

Results: Created PolyCTLDesigner software constructs polyepitope CTL immunogen selecting superior spacers for every pair of selected epitopes (to provide (immuno)proteasomal release of the epitopes and optimize binding of generated peptides to TAP), choosing appropriate epitope matchings and optimal arrangement of epitopes within designed construction, thus increasing efficiency of polyepitope processing and favoring presentation of target epitopes. It also tries to minimize the number of „non-target" epitopes within desired polyepitope immunogen and is able to assist in collecting the set of peptides covering selected HLA repertoire with desired rate of redundancy using known genotypic HLA allele frequencies data together with either known or predicted specificity of selected peptides towards different allotypes of HLA class I molecules.

Conclusion: Developed software realizes the rational approach to designing highly immunogenic poly-CTL-epitope immunogens. In our Center PolyCTLDesigner was used to construct novel HIV-1 polyepitope antigens which are now under experimental evaluation. Program could be freely downloaded at http://tepredict.sourceforge.net.

References:

1. B. Peters et al. (2003) Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *J Immunol*, **171**:1741-1749.
2. R.E. Toes et al. (2001) Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. *J Exp Med.*, **194**:1-12.
3. D.V. Antonets, A.Z. Maksyutov (2010) TEpredict: software for T-cell epitope prediction. *Mol. Biol.*, **44**:119-127.

MODELLING LIGAND-RECEPTOR COMPLEXES OF VARIOLA AND COWPOX VIRUS CRMB PROTEINS WITH MOUSE AND HUMAN TNFS.

D.V. Antonets*, T.S. Nepomnyashchikh, I.P. Gileva, S.N. Shchelkunov
State Research Center of Virology and Biotechnology "Vector", Koltsovo, Russia
e-mail: antonec@yandex.ru
**Corresponding author*

Key words: *TNF, immunomodulation, orthopoxvirus, molecular modelling*

Motivation and Aim: Orthopoxviral TNF-binding proteins and especially variola virus (VARV) CrmB may be used to develop novel medications for treatment of rheumatoid arthritis, Chron's desease and other pathologies driven by TNF overproduction. The aim of this study was the theoretical analysis of molecular mechanisms underlying interaction of orthopoxviral TNF-binding CrmB proteins with their ligans.

Methods and Algorithms: Models were built using Modeller software (http://salilab.org/modeller) and validated with ProCheck [1]. Models of ligand-receptor complexes of VARV and cowpox virus (CPXV) CrmBs with hTNF (1TNF) and mTNF (2TNF) were produced by superimposing corresponding molecules onto the crystall structure of human TNF receptor I (p55) complex with lymphotoxin (1TNR). All constructed models were then energy minimized using either NOC (http://noch.sourceforge.net) or FoldX (http://foldx.crg.es). Stability of ligand-receptor complexes was predicted either with FoldX or using residue-level pairwise potentials BETM990101 [2].

Results: Models of VARV- and CPXV-CrmB and their complexes with m- and hTNF were constructed. Analysis of these models with either FoldX or with BETM990101 pair potentials revealed that mTNF should bind to CPXV-CrmB with higher affinity than hTNF. VARV-CrmB was predicted to bind both cytokines with higher affinity than CPXV-CrmB; CPXV-CrmB was predicted to bind hTNF(R31Q) with significantly higher affinity than wild type hTNF. Using FoldX both CrmBs were predicted to less efficiently bind to hTNF(E127Q), than to the wild type hTNF. All these findings were then approved by experimental evaluation of VARV- and CPXV-CrmB proteins ability to inhibit cytotoxic action of mTNF, hTNF, hTNF(R31Q) and hTNF(E127Q) on L929 murine fibroblast cells.

Conclusion: Predicted stability of modelled ligand-receptor complexes of both CrmBs with selected TNFs was found to be in good qualitative agreement with experimental data. Produced models will be used for designing mutant forms of VARV-CrmB with higher affinity towards hTNF.

Acknowledgements This work was supported by Russian Foundation for Basic Research (grant #09-04-00055a).

References:

1. R.A. Laskowski et al. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, **26**:283–291.
2. M.R. Betancourt, D. Thirumalai (1999). Pair potentials for protein folding: choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Protein Sci.* **8**:361–369.

EVOLUTION OF EUKARYOTIC MRNA NON-CODING REGIONS.

K.S. Antonez[1], A.F. Saifitdinova[1,2]
[1]*Dept. of Genetics, St-Petersburg State University, St-Petersburg, Russia*
[2]*St-Petersburg Branch of Vavilov Institute of General Genetics, St-Petersburg, Russia*
e-mail: kir_ant@mail.ru

Motivation and Aim: The crucial role of the non-coding part of genomes is well known. In particular, mRNA untranslated regions are involved in many regulatory pathways that control mRNA localization, stability and translation efficiency. Information derived from certain functional studies on genes belonging to different organisms have been collected in separate specialised databases and some out-of-date information resources. In order to study systematically the strucrural and functional properties of eukariotic mRNA untranslated regions we have designed the relational database.

Algorithms and Methods: To work with information from databases and analyze the data we used SQL Server and its components. The program was written in C# language. Data about homology of genes was obtained from FTP server NCBI (ftp://ftp.ncbi.nih.gov/pub/HomoloGene). Sequence information was downloaded from GenBank with usage of SOAP web services (http://eutils.ncbi.nlm.nih.gov/entrez/eutils/soap/v2.0/DOC/esoap_help.html). Information about functions was taken from GeneOntology database.

Results and Conclusion: The relational database RNAdb store information about non-coding sequences of eukariotic mRNA. It was used for analysis of structral features of untranslated regions and exploration of consevation of these sequences. A detailed analysis of the evolutionary dynamics of mRNA untranslated regions as compared to the corresponding coding sequences reveals that eukaryotic non-coding mRNA in vertebrates sufficiently differ from those in invertebrates, plants and fungi.

Acknowledgements: The work was supported by FTP “Academic and teaching staff of innovative Russia” of The Russian Ministry of Education and Science.

EXON-INTRON STRUCTURE OF FIRST CHROMOSOME MONODELPHIS DOMESTICA

A.S. Asheulov*
Kazakh National University named after al-Farabi, Almaty, Kazakhstan
e-mail: laxanier@mail.ru
**Corresponding author*

Key words: *exon, intron, genome, computing analysis.*

Motivation and Aim: Introns perform many functions in genes, and the analysis of theirproperties is necessary for the further understanding of their biological role. Variability ofexon and intron lengths is rather great in genes of eukaryotic organisms. In genomes of thehuman, nematode, arabidopsis and rice the links between exon and intron length, and the sum of exon lengths and intron number in genes have been established [1, 2]. It is obviously important to clarify if there are such links in genes of completely sequencing genomes of other eukaryotic organisms.

Methods and Algorithms: The genes of chromosome 1 Monodephis domestica were arranged in

samplings with 1, 2, 3, 4, 5, 6-9, 10-14, 15 and more introns in a gene.

Result: 4840 genes have been analysed. The average length of genes for groups with the maximum density made 1304 nucleotides, in groups with average density 866 nucleotides, and in groups with the minimum density equaled 961 nucleotide. Also it is necessary to notice that at quantity lengthening introns in a gene average relation C/G – A/T decreased. If to take group with 1-2 introns that this size makes 0,499 and if to take this size with 15 and more introns that it already 0,069. Average GC the maintenance in a chromosome 1 has made 41 %. It is necessary to notice that in group with smaller quantity introns size C/G-A/T it is essential more than in groups with a considerable quantity introns. This phenomenon is marked in all groups of sample with the minimum, average and maximum density. Than more long the sum of lengths introns in a gene that best effect. Also the change size gidropatichnosti in genes in groups with 1-2, 3-5,intron almost identical also equals 0,2. In groups with 6-9 introns the size fluctuates from 0,1-0,3. In groups from 10-14 introns and 15 and more changes in the field of 0,09-0,16. This data can testify to a different role introns in change gidropatichnosti genes with various density. It is necessary to add that the big role in change nucleotides structure introns bring in genes with the minimum density in groups with 3-5 and 6-9 introns in a gene, in genes with about average density in groups with 3-5 and 15 and more introns in a gene, in genes with the maximum density in groups with 1-2, 6-9 and also 15 and more intron in a gene.

References:

1. S.Atambaeva, V.Khailenko, A.Ivashchenko (2008) Changes of introns and exons length in genes of arabidopsis, rice, nematode and human, *Mol. Biol.* (Russ), **42:** 1-10.
2. A.Ivashchenko, S.Atambaeva (2004) Variation in lengths of introns and exons in genes of he *Arabidopsis thaliana* nuclear genome, *Russ. Journ. Genetics*, **40:** 1179-1181.

INVESTIGATION OF THE ROLE OF BRCA1 ASSOCIATED PROTEIN 2 (BRAP2) IN THE REGULATION OF THE UBIQUITIN SYSTEM AND MALE INFERTILITY.

Z.A. Astakhova*, R. Commander, K.D. Wilkinson
Emory University, Atlanta, GA, USA
e-mail: zastakh@emory.edu
**Corresponding author*

Key words: *(BRAP2, the ubiquitin system, the ubiquitin ligase, nuclear-cytoplasmic shuttling, HSFY1 transcription factor, male infertility)*

Motivation and Aim: The ubiquitin-proteasome pathway is a major pathway that regulates a wide array of cell processes and is implicated in many deseases. E3 ubiquitin ligases play an important role in ubiquitination as they serve as specific substrate-recognition elements of the system. BRAP2 is a cytoplasmic E3 ubiquitin ligase that modulates the sensitivity of MAP kinase cascade. In addition, BRAP2 binds to the NLS motif of BRCA1, SV40 large T antigen, mitosin, p21, and various viral proteins. Importantly, BRAP2 functions as a cytoplasmic retention protein for many proteins. We hypothesize that BPAP2 has a dual function in a cell: the cytoplasmic sequestration of proteins that contain NLS by ubiquitination, and the control of the amount of free ubiquitin chains. BRP2 is a *S. cerevisiae* putative homologue of BRAP2 that shares 40.8% identity with BRAP2. Studies from our laboratory indicate that *BRP2* has the same functional domains, as does BRAP2, which allows using a yeast model system to study the function of BRAP2. The aim of this study is to investigate the physiological role of BRAP2 using both yeast and mammalian model systems.

Methods and Results: Both human and yeast BRAP2 orthologues have the same domain structure featuring RING domain, ZnF UBP domain, and Coil-Coiled domain. We have shown that these domains are functional for both proteins. First, BRAP2 forms the ubiquitin chains *in vitro,* which points to a functionality of the RING domain. Moreover, ubiquitin lysine 48 is required for the ubiquitin chain formation catalyzed by BRP2 and BRAP2 in *in vitro* autoubiquitination assays. Second, both recombinant proteins were shown to bind linear ubiquin chains, which identifies ZnF UBP as a functional domain. Importantly, in yeast, the depletion of BRP2 causes the decrease in the total amount of the ubiquitin chains. In addition, we have identified the HSFY1 protein as one of the binding parters of human BRAP2 in the Yeast-Two-Hybrid analysis using BRAP2 as a bait.

Conclusion: Based on the ubiquitin chain topology (K48) BRAP2 assembles, the fact that BRAP2 has both RING and ZnF UBP functional domains, and the impact of the deletion of BRP2 on the amount of total ubiquitin chains, we predict that BRAP2 may be the ubiquitin sensor in a cell, controlling the pool of free ubiquitin chains. In addition, HSFY1 transcription factor, which is deleted in infertile males, may be a substrate of BRAP2 ubiquitin ligase.

Acknowledgements: This work is supported by NIH grant R01 GM30308 to KDW and by DOD training grant to ZAA.

PECULARITIES OF U2 SNRNA AND INTRON INTERACTION IN SPLICING

Feodor Goncharov, Vladimir Babenko*
Institute of chemical biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
e-mail: fedor@mcb.nsc.ru
**Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

Key words: *Splicing, U2 snRNA, BPS.*

Motivation and Aim:

Splicing is a process preceding protein translation from mRNA matrix, excising out non-coding parts (introns) from the pre-mRNA. It is typical for all eukaryotes and commonly it requires the assembly of the spliceosome. Spliceosome is a complex of more than hundred proteins and five snRNA employed on the subsequent stages. We investigated the interaction of U2 snRNA, which is involved in the splicing initiation and further on. Its pairing with intron in the area of Branch Point (BP) features the formation of bulged active adenosine. We undertake the analysis of assessing the selection strength in relation to this pairing and identify the additional binding regions besides BP region.

Methods and Algorithms:

In order to identify areas of U2 snRNA putative intron binding sites we used the scanning algorithm of introns with a sliding window length of 6 nucleotides. We used statistical comparison of complementary intron and U2 snRNA regions which allows efficient identification of a preferred sequence for branch site in organisms with a moderate number of introns (fungi).

Results:

We show that U2 snRNA has a high probability of interaction with the intron in several areas that are conserved in related species. Also, we have shown that there is evolutionary pressure preserving the integrity and selective interaction of RNA intron and U2 snRNA at BPS.

Conclusion:

The developed technique allows to identify regions of U2 snRNA that are likely to actively interact with the intron during splicing stages based on pairing potential, as well as highlight the motifs under selection pressure. The study confirmed the previously verified U2 snRNA – intron interaction sites and identified possible sites which enhance the efficiency of splicing.

Availability: Data and programs are located at: ftp://ftp.mcb.nsc.ru/user/fedor/U2

Acknowledgements:

Molecular and Cellular Biology Program of RAS N22.4

Government contract “Rosnauka” N02.740.11.0099

MOLECULAR DATING IN THE EVOLUTION OF VERTEBRATE POXVIRUSES

I.V. Babkin*[1], I.N. Babkina[2]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*SRC VB "Vector", Koltsovo, Novosibirsk reg, Russia*
e-mail: babkin@bionet.nsc.ru
**Corresponding author*

Key words: *DNA virus, Poxviridae, viral evolution, Bayesian relaxed clock*

Motivation and Aim. The molecular dating of the evolutionary history of the viruses with extended DNA genomes, such as poxviruses, is still vague due to a low rate of accumulation of nucleotide substitutions. The goal of present work was to study the evolutionary history of the vertebrate poxviruses with AT-rich genomes applying Bayesian relaxed clock method to a large set of highly conserved vitally important genes of these viruses. This approach is based on a probabilistic model of the changes in evolutionary rates and utilizes the Markov chain Monte Carlo. This provides for estimation of an *a posteriori* distribution of the evolutionary rates and divergence time estimates. An expanded set of data on the gene nucleotide sequences made it possible to perform the evolutionary analysis with a higher reliability.

Methods and Algorithms. When analyzing the genome primary structures of 18 strains belonging to different vertebrate poxvirus genera, we used the sequences of 35 highly conserved genes. These sequences were aligned into 35 nucleotide alignments and then concatenated the individual alignments. Phylogenetic analysis was performed by ML method using the program Paup4.0b10. The model of evolution was preliminary determined using the Modeltest v.3.7. The dating method of Bayes, realized in the Multidivtime v.1.5 program, was used in the work.

Results. In this work, the rate of accumulation of nucleotide substitutions is estimated as $1–6\times10^{-6}$ nucleotide substitutions per site per year. Applying the Bayesian method for time estimates, we inferred that the modern viruses of the genus *Avipoxvirus* diverged from the ancestor approximately 283±102 Tya. Analysis of the chronogram suggests that the ancestor virus of the modern mammalian poxviruses had a wide host range. During the evolution, this ancestor virus had specialized to different host organisms. The ancestor of the genus *Orthopoxvirus* was the first to diverge approximately 171±55 Tya. It was followed by separation of the ancestor of the genus *Leporipoxvirus,* which took place about 136±44 Tya. The next to diverge was the ancestor of the genus *Yatapoxvirus;* members of this genus induce benign tumors in primates. The ancestor of three virus genera—*Capripoxvirus*, *Suipoxvirus,* and deerpox virus, recently discovered and yet unclassified—diverged 107±36 Tya.

Conclusion. The evolutionary analysis based on the historical data and utilizing the Bayesian relaxed clock allowed us to determine the molecular evolution rates of the genomes of the vertebrate poxviruses with AT-rich genomes and assess the time of their emergence. Three time constrains were used in this analysis. It has been found that the main vertebrate poxvirus genera diverged 100–300 Tya. The rate of mutation accumulation in the genomes of these viruses is about 10^{-6} nucleotide substitutions per site per year. Involvement of a large set of the conserved genes controlled by stabilizing selection allowed us to perform molecular dating of the vertebrate poxvirus history.

Acknowledgements. This work was supported by the Russian Foundation for Basic Research (project no. 08-04-00443-a).

INVESTIGATION OF SEX DIFFERENCE IN SUSCEPTIBILITY TO HEPATOCARCINOGENESIS IN MICE.

N.V. Baginskaya*, S.I. Ilnitskaya, V.I. Kaledin
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: bagin@bionet.nsc.ru
**Corresponding author*

Key words: *mice, hepatocarcinogenesis, sex differences, insulin resistance*

Motivation and Aim. Exploration of hepatocarcinogenesis mechanisms is one of actual problem of biology and medicine. Existence of interstrain and sex differences on liver tumour development in mice makes it convenient model for such researches. The aim of this work was to establish connections between metabolic imbalance and development of liver tumors in male and female C57Bl/6J mouse strain in chemical hepatocarcinogenesis.

Methods and Algorithms. Male and female C57Bl/6J mice at 2 wk of age were injected ip with a single dose of N,N-diethylnitrosamine (DEN, 50 mg/kg). For tumor promotion the half of experimental animals starting from 2 month of age were injected ip of 1,4-bis[2-(3,5-dichloropyridyloxy)] benzene (TCPOBOP, 3 mg/kg) every 2 wk for 2 and 6 months. The blood levels of glucose, cholesterol, triglycerides (TG), insulin and free thyroxine were estimated by corresponding Vector-Best and Immunotech detection kits. TCPOBOP-induced liver damage characterized by alanine aminotransferase (ALAT) blood level. The number of liver tumors and preneoplastic nodules the were estimated in 8-month's animals.

Results. The glucose, TG, and cholesterol blood levels were higher in 4 month's old intact males as compared with females. After carcinogen injection essential decrease in thyroxine blood level was observed at males. Application TCPOBOP within 2 months led to considerable liver damage both in males and females. These animals showed signs of a metabolic syndrome development: increase in glucose, TG and insulin blood levels as compared with both control mice received the only DEN injection and intact animals. Besides, considerable decrease in free thyroxine blood levels both in males and females after TCPOBOP application was observed. The average number of liver tumors in 8 month age males received only DEN injection in early ontogenesis, was 27±9 per animal that is essentially higher than in females (3±1 per animal). Application of TCPOBOP during 6 months increased the number of tumors both of males and females in 2 and 25 times respectively. Intact animals have no liver tumors up to this age.

Conclusion On the data obtained it is possible to make following conclusions:

1. DEN injection does not change free thyroxine blood level and insulin sensitivity in female C57Bl/6J mice are resistant to it hepatocarcinogenic action.
2. TCPOBOP application decreases free thyroxine blood level and insulin sensitivity and considerably increases the number of liver tumors in both males and females.

Thus, sex differences of mice liver tumor sensitivity may be connected with sex differences in insulin sensitivity and thyroid status of animals.

SIMULATION AND COMPUTATIONAL STUDIES OF CARVEDILOL AND OTHER HEART DRUGS: DENSITY FUNCTIONAL CALCULATION

K. Bakhshi*
Department of chemistry, Miyaneh Branch, Islamic Azad University, Miyaneh, Iran
**Corresponding author: e-mail: kamalbakhshi@hotmail.com*

Key words: *Density Functional Theory, Carvedilol, Propranolo, Atenolol*,

Motivation and Aim: Nowadays, simulation technologies and computational studies are widely used in design of new drugs. Carvedilol as important heart failure drug has been used for many years. Because of experimental difficulties, computer simulation became important alternative source biologically relevant information for such drugs systems. Therefore, we have investigated some properties of carvedilol by density functional theory (DFT) calculation theoretically.

Methods and Algorithms: we used a (density functional theory) DFT calculation for simulation of the carvedilol. In this study we utilized B3LYP, BLYP, BP86, BPW91 and LSDA methods. Gaussian 98 code was used in our calculations and optimization. 6-31G, 6-31G* and 6-31G** as standard basis sets were used.

Results: Gaussian 98 code was used in order to predict some thermodynamics properties of carvedilol and other heart drugs. The results of carvedilol and other drugs were compared each other and these results indicate that there is good agreement between experimental and theoretical ways.

Conclusion: we can simulate and develop several drug systems and predict some of the properties theoretically.

ION CHANNELS: COMPUTER SIMULATION OF QUANTUM EFFECTS

M. Batukov*, G. Isaeva, P. Isaev
Taganrog Pedagogical Institute, Taganrog, Russia
e-mail: traintoamsterdam@gmail.com
* *Corresponding author*

Key words: *ion channels, the electrostatic potential, membrane*

Motivation and Aim: Ion channels research is under scrutiny of pharmaceutical industry for several decades. Ion channels play a vital role in basic physiological functions such as generation of electrical nerves activity and muscles that control cardiac excitability, intracellular signaling, hormone secretion, cell proliferation, and many other biological processes. Recent discoveries in the field of determining the atomic structure of important membrane channels defined objective of the study - to obtain macroscopic, measurable properties of these channels (e.g., ion current) on the basis of their atomic structure and the electrostatic potentials thus that can help to understand the physical mechanisms of their action [1]. *Methods and algorithms:* The limited capacities of even modern supercomputers do not allow simulating the actual process of charge transport through the membrane in the channels, using all the atomic models. To solve this problem we use an integrated hierarchical approach. It combines the simulation of molecular dynamics and calculation of the electrostatic profile of the barrier free energy, which is then used for stochastic modeling of charge transport through the channel. To find the energy eigenvalues in the Schrodinger equation used the step method.

Results: The simulation shows that in all cases the ion transport depends on the electrostatic effects, namely the relationship between the magnitude of change in the charge self-energy and energy of its interactions with protein polar groups, and stochastic models of proton transport is reflected can be used in the study of ion transport in gramicidin channel. The results of the step method for finding the energy eigenvalues give a small error, and therefore can be used in the study of complex shapes and even those that are intractable analytically [2]. *Conclusion:* Current results allow us to simulate ion currents through the channel, as well as its selective properties based on the atomic structure and the electrostatic potential. We believe that the findings can be applied to the practical construction of artificial nanofilters on the basis of membrane permeability.

Availability: Available upon request.

References:

1. Nimigean, C. M. and C. Miller (2002). Na+ Block and Permeation in a K+ Channel of Known Structure. J. Gen. Physiol. **120**: 323-335.
2. Andres Udal, Reeno Reeder (2006). Comparison of methods for solving the Schrödinger equation, Proc. Estonian Acad. Sci. Eng.,**12**: 246–261.

REGULATORY VOLUME DECREASE. ROLE OF WATER CHANNELS.

G.S. Baturina*, L.E. Katkova, A.V. Ilyaskin, E.I. Solenov.
Institute of Cytology and Genetics SB RAS, Novosibirsk State University, Novosibirsk, Russia
e-mail: baturina@yandex.ru
**Corresponding author*

Key words: *aquaporin, water transport, regulatory volume decrease, collecting duct.*

*Motivation and Aim***.** Cell volume perturbation elicits a wide array of signaling events, leading to activation of volume regulatory osmolyte and water transport. Kidney medullar cells are exposed to low extracellular osmolarity during diuresis. After acute swelling, cell volume is regulated by the process of regulatory volume decrease (RVD). It is known, that osmotically swollen cells release KCl, organic osmolytes, and water. In mammalian cells high water permeability reflects the presence of aquaporins (AQPs), and it has been shown that the presence of AQP2 in collecting duct cells dramatically increases the rate of initial cell swelling following hypotonic exposure. Role of aquaporins in RVD is currently not known. In the present study the effect of induction of AQP2 and AQP3 by water deprivation during 36 hours and dDAVP on time course of the volume-regulatory response (RVD) during acute swelling in hypotonic medium was measured.*Methods.* Renal medullar collecting duct suspension was prepared as described earlier (Solenov et al., 2003). The solutions in incubation chamber were changed with characteristic time 60 ms.

Hypotonic shock was created by PBS diluted with 50% of water. Changers in cell volume were measured with calcein quenching method (Solenov et al., 2004). The content of aquaporins in the kidney outer medulla slices was analyzed with a Western blotting. Increased expression of AQP2 and AQP3 mRNA in plasma membranes was shown by RT PCR.

Results.

Principal cells of micro dissected collecting duct fragments swelled with the characteristic time of swelling was 0.65 ± 0.05 s. Cell volume increased more than 60% (92.9 ± 5.6 and 151.3 ± 9.8 μm3 control and peak volume correspondently, $p<0.01$). After cell volume reached the peak of swelling the RVD began without lack period. The characteristic time of this decreasing was 9.4 ± 2.1 s. Agonist V2 receptors (dDAVP 10 nM) increased the rate of RVD (characteristic time decreased to 5.6 ± 0.1 s, $p<0.05$). Restoration of the medium osmolality to normotonic lead to cell volume stabilization on significantly low level in comparison with control level (71.4 ± 6.1 μm3, $p<0.05$). It was shown that the hypotonic shock induced a rapid increase in the cell volume with the characteristic time that depended on plasma membrane water permeability.

*Conclusion***.** Finally, our studies indicate, that water channels besides their important role for cell swelling might be also involved in cell volume regulation mechanism.

Acknowledgements.

The work was supported by the Russian Foundation for Basic Research (projects nos. 09-04-00197-a and 08-04-00541-a).

References:

1. Solenov E., Watanabe H., Manley G. T., Verkman A. S.// Am. J. Physiol. Cell. Physiol. 2004. V. 286. P. C426–C432.

ADAMANTANE DERIVATIVE: THE SEARCH OF THE DRUGS INCREASING ABILITY.

T.A. Bednaya*, G.A. Isaeva, P.P. Isaev
Taganrog State Pedagogical Institute, Taganrog, Russia
e-mail: bednayat@mail.ru
**Corresponding author*

Key words: *drugs design, adamantane, QSAR, electronic constant of the substitutes,*

Motivation and Aim: The method of quantitative structure activity relationship (QSAR) is used for the most effective drugs design. The goal of a QSAR model is to extract information from a set of numerical descriptors characterizing molecular structure and to use this information to develop a relationship between structure and property [1].

The presence of psychoactivating properties was shown in a line of derivative adamantane. These properties are demonstrated in a usual condition of animals and human at mental and neuropsychiatric frustrations on a background of action of psychodeprimating drugs. This provides the detectability among given compounds effective activating means increasing mental and physical capability, including work in severe environment [2].

Methods and Algorithms: The value A=LD50 is chosen as an index of biological activity. At performance comparative QSAR analysis as one of descriptors molecular refraction MR is used, describing energy of interaction of substance from a surface of a biomembrane. It is necessary to take into account transport of compounds, therefore second descriptor is value of lipophilicity logP, where P – the octanol/water partition coefficient.

Results: The equation for interpretation physiological activity of derivative adamantane is: $lgA = -0{,}145 \cdot logP + 0{,}0485 \cdot MR + 3{,}088$ (n=15; r2=0,272; s=0,801; F=2,242), where LD50 = concentration (mM) 50% of killed mice in a specified time, n = number of chemicals, r = correlation coefficient, and s = standard error of the estimate, F – criterion of Fisher's. Statistical analysis shows an inefficiency of this equation and necessity to choose other descriptors. As an additional descriptor an electronic constant of the substitutes is chosen (σ^+), which describes electronic property of molecules. In opposition to earlier used descriptors: molecular refraction and lipophilicity, it concerns to a separate group of atoms.

The equation for interpretation physiological activity of derivative adamantane is:
$lgA = -1{,}867\sigma^+ - 0{,}09 \cdot logP + 0{,}01 \cdot MR + 6{,}71$ *(n=15; r=0.85; S=0.5; F=9,76).*

Conclusion: The value of lipophilicity not significant in this equation, therefore it is wise to remove it from the equation, then the QSAR-equation accepts a view: $lgA = -1{,}877\sigma^+ - 0{,}018 \cdot MR + 7{,}04$ *(n=15; r=0.85; S=0.49; F=15,7).*

Thus, the correct choice of descriptors is of great importance for the creation of QSAR equation having a high degree of trust.

References:

1. J. C. Dearden et al. (2009) How not to develop a quantitative structure–activity or structure–property relationship (QSAR/QSPR), *SAR and QSAR in Environmental Research* **20**: 3–4.
2. Morozov I.S. et al. (2001) Proizvodnie adamanntana kak sredstva povisheniya rezistentnosti organizma k deictviu exrtimalnih factorov sredi obitaniya i deyatelnosti, in *Farmokologiya adamantanov,* 172-221 (Volgograd).

IDENTIFICATION OF EUKARYOTIC TRANSCRIPTION REGULATORY REGIONS USING PROTEIN-PROTEIN INTERACTION DATA

A.A. Belostotsky*, I.V. Kulakovskiy, V.J. Makeev
State Research Institute for Genetics and Selection of Industrial Microorganims, GosNIIgenetika, Russia
e-mail: alexbel.composite@gmail.com
**Corresponding author*

Key words: *protein-protein interaction, transcription factors, composite element, transcription factors binding site*

Motivation and Aim: Protein-protein interaction is very important along with protein-DNA interaction for DNA sequence recognition. But only protein-DNA interaction is taken into account by all available methods for prediction of regulatory elements in genome. We used protein-protein interaction data in the most general form to predict such elements.

Methods and Algorithms: Protein-protein interaction data are used to identify all pairs of transcription factors (TF) observed to interact with each other. From the training set of regulatory segments we extracted pairs of co-localized transcription factors binding sites (TFBS) for interacting TFs only (composite elements - CE). From the set of hypoxia inducible genes we extracted CEs containing HIF1alpha, the primary hypoxia mediator. Distance preferences between TFBS in HIF1alpha CE were also studied.

Results: For known regulatory regions our algorithm outperforms all currently available CE-discovery methods. We have identified several TFs that are very likely to form CEs with HIF1a, including the most important factors EP300, p53, Sp1, HNF4. We also have detected several CEs in the vicinity of HIF1alpha dependent genes.

Conclusion: CE prediction is quite important for better understanding of hypoxia genes regulation and revealing of hypoxia regulatory network. Successful usage of protein-interaction data for identification of CEs with HIF1alpha suggests that the method can be used to analyze transcription regulation in other systems.

Acknowledgements: This work has been partially supported by Russian Fund of Basic Research project 07-04-01623 and Russian State Agency of Science and innovations contract 02.740.11.5008

FURTHER ELABORATION OF A STRESS HYPER-RESTORATION (HR) MORPHOGENETIC MODEL

L.V. Beloussov
Laboratory of Developmental Biophysics, Department of Embryology, Faculty of Biology Moscow State University Moscow 11992.
E-mail: morphogenesis@yandex.ru

Motivation and aim

Our main aim is to reveal generative laws of morphogenesis permitting to derive each next stage of development from a preceded one.

Methods and Algorithms

We combine methods of experimental embryology and in situ hybridizatrion studies of gene expression patterns with mathematical modeling. The main algorithm used in our studies is a so-called hyper-restoration (HR) model which implies that morphogenesis of multicellular organisms is driven forth by an intrinsic tendency of embryonic cells and tissues to restore with overshoots their mechanical stress values altered by external forces. According to HR model, the morphogenesis of multicellular organisms is driven forth by an intrinsic tendency of embryonic cells and tissues to restore with overshoots their mechanical stress values altered by external forces.

Results

Our recent results related to HR model fall into the following categories:

- As regarding morphogenesis per se, a statistically significant increase of the tangentially contracted and closely packed cells in mechanically relaxed epithelia, as compared with intact ones was shown. Also, we demonstrated that an embryonic epithelial sheet is able to actively increase a curvature imposed by an external force.
- In embryonic tissues which underwent primary induction the abovementioned active curvature increase affects space arrangement of the main cell differentiations. In particular, the areas of the neural genes expression are located at the concave, while those of the muscle genes expression at the convex sides of the curved tissue explants.
- HR model was shown to reproduce several structurally stable evolutionary important Bauplans.
- From the model point of view, there are Van der Pol equations (in trigger modification) which look as one of the most adequate representations of HR-mediated feedbacks. Within this model framework, the genetic and epigenetic factors affecting HR-mediated feedbacks can be adequately represented.

Conclusion

HR model looks as an adequate tool for reproducing a wide range of developmental processes as self-organized events.

GENOME SEQUENCING AND MIRNA DISCOVERY IN THE REGENERATING FLATWORM MACROSTOMUM LIGNANO

E. Berezikov*[1], D. Simanov[1], P. van Zon[1], E. de Bruijn[1], S. Linsen[1], K. de Mulder[1], E. Cuppen[1], A. Canela[2], G.J. Hannon[2], D.B. Vizoso[3], L. Schärer[3], P. Ladurner[4]
[1]Hubrecht Institute and University Medical Center Utrecht, Utrecht, The Netherlands; [2]Cold Spring Harbor Laboratory, Watson School of Biological Sciences and Howard Hughes Medical Institute, Cold Spring Harbor, NY; [3]University of Basel, Zoological Institute, Basel, Switzerland; [4]University of Innsbruck, Innsbruck, Austria
e-mail: e.berezikov@hubrecht.eu
**Corresponding author*

Motivation and Aim: Macrostomum lignano is a free-living flatworm with high regeneration capacity facilitated by neoblasts, the stem cell system of the worm [1-3]. Due to its small size, short generation time, amenability to genetic manipulation and easy maintenance in laboratory conditions, *M. lignano* is a potent invertebrate experimental model for stem cell research. Yet limited molecular biology resources and tools are currently available for this emerging model organism.

Methods and Algorithms: We have initiated *de novo* genome sequencing of *M. lignano* using hybrid 454 and Solexa/Illumina approach. We are also using second-generation sequencing technologies to characterize transcriptome of the worm, including small RNAs.

Results and Conclusions: The estimated genome size of *M. lignano* is 600 Mb. Initial 8x coverage of the genome by 454 technology produced ~300Mb of assembled contigs with N50 size of 1.8 kb. While this low-coverage fragmented assembly is not productive for annotation of protein-coding genes, it is already efficient for discovery of microRNAs – small RNAs that regulate gene expression at the posttranscriptional level and are involved in various cellular processes, including maintenance and differentiation of stem cells. Sequencing of small RNAs from irradiated (=neoblast-depleted) and non-irradiated worms allowed annotation of more than 50 miRNA genes. Levels of several miRNAs are decreased in the irradiated worms, suggesting neoblast-specific expression of these miRNAs.

The progress on the genome sequencing, transcriptome and small RNA analysis in *M. lignano* will be presented. Detailed information on *Macrostomum lignano*, including draft genome assemblies, are available at http://www.macgenome.org.

References:

1. Ladurner et al. (2005). A new model organism among the lower Bilateria and the use of digital microscopy in taxonomy of meiobenthic Platyhelminthes:Macrostomum lignano, n. sp. (Rhabditophora, Macrostomorpha). *JZS* **43:** 114-126.
2. Pfister et al. (2008). Flatworm stem cells and the germ line: developmental and evolutionary implications of macvasa expression in Macrostomum lignano. *Dev. Biol*. **319**:146-159.
3. De Mulder et al. (2009). Stem cells are differentially regulated during development, regeneration and homeostasis in flatworms. *Dev Biol* **334**:198-212.

PLANT RETRO-TRANSPOSON SEQUENCE MINING FROM THE NCBI DATA BASE & SEQUENCE DIVERSITY ANALYSIS

Brijmohan Singh Bhau
Plant Genomics Laboratory, MAEP Division, North-East Institute of Science & Technology (CSIR), Council of Scientific & Industrial Research (CSIR), Jorhat 785 006, Assam, India
e-Mail: bsbhau@gmail.com; bhaubs@rrljorhat.res.in

Motivation and Aim: Retro-transposons are well established ubiquitous and highly prevalent components of plant genomes in general and also play a major role in plant gene and genome evolution These elements are nevertheless often referred to as "junk DNA" implying that they are inert in comparison with the genes required for cellular function.

Methods and Algorithms: This present study was performed using sequence assembly approaches. Four hundreds and seventy seven different plant retro-elements sequence assembly producing 156 cotigs that comprises 92.20 % of total sequence length. The average number of retro-elements core nucleotide sequences per assembly was 2.0. Highly assembled contig sequences were mapped by using BLASTN search approach.

Results: Above outcome indicates that Copia like, Ty-1 Copia like, BARE, SIRE reto-elements fraction are present in non retro-element plant database sequences. These retro-element fractions sequences show structurally as well as functionally similarity with non reto-element sequences. Assembly of retro-elements will be of major importance as studies of retro-element homology and variation with the different plant genome. It will provide the mapping information that has been detecting consensus of sequence in retro-element as well as plant genomes, and it should stimulate experiments in areas as different as evolution and genome dynamics.

Conclusion: Finally, genomics is the ultimate inter disciplinary approach, as it covers the entire spectrum from DNA sequencing to field-based research. Only through the integrated endeavor of genetics, biology, bioinformatics, molecular biology, engineering, microbiology and related fields will the extensive benefits of genomics to mankind become reality.

MODELING GENOMIC REGULATION: TRANSCRIPTIONAL MODES AND EPIGENETIC SWITCHES IN ARTIFICIAL GENOMES

Hans Binder[1*], Lydia Hopp[1,2], Henry Wirth[1,3], Jörg Galle[1]
[1] *Interdisciplinary Centre for Bioinformatics of Leipzig University, Leipzig, Germany*
[2] *University Greifswald, Germany*
[3] *Helmholtz-Zentrum für Umweltforschung, Leipzig, Germany*
[*] *Corresponding author: binder@izbi.uni-leipzig.de*

Key words: *artificial genome model, activation and repression of transcription, expression spectrum,*

Motivation and Aim: The present study addresses transcriptional regulation in a simple model genome. Our whole genome view is motivated by the idea that many aspects of gene functioning cannot be understood at the level of single genes but require a systemic approach which considers the manifold of an ensemble of genes, their possible microstates of activity and their mutual interactions.

Methods and Algorithms: We presented a statistical thermodynamics model of whole genome transcriptional regulation which combines the random genome model of gene regulatory network organization with a biophysical description of gene activity. The model characterizes the response of the global expression pattern to the changing properties of basal regulatory building blocks. Random genomes are generated which express and bind transcription factors according to the appearance of short motifs of coding and binding sequences. Transcriptional regulation is supplemented by a simple model of gene activation/silencing via epigenetic switches based on the interaction of regulators (protein complexes) to DNA-response elements.

Results: Our model predicts a power law distribution of gene activity. Repressors and activators of gene activity give rise to increasing and decreasing tails of the distribution with a maximum in between. This maximum is assigned to the basal transcriptional activity of unregulated promoter states. The decay exponents of the power laws are inversely related to the network connectivity and the average strength of regulation. The position of the expression spectra and the slopes of their increasing and decreasing flanks provide a simple framework for the interpretation of experimental gene expression spectra. We apply these rules to experimental expression spectra which were calculated for a series of microarray measurements taken from public data repositories. These examples comprise different biological issues and samples ranging from embryonic development and cell differentiation to mutants and oncogenic de-regulation in different tissues and organisms.

Conclusion: The chosen examples show that characterization of the global expression pattern in terms of the distribution of expression values provides valuable information about transcriptional regulation which complements conventional searches for differentially expressed genes.

References:

1. Binder, H., Wirth, H., Galle, J. Gene expression density profiles characterize modes of genomic regulation – theory and experiment. *Journal of Biotechnology* **in press** (2010). (preprint: http://www.izbi.uni-leipzig.de/izbi/Working%20Paper/2010/WP19.pdf)

IDENTIFICATION OF SOURCES OF ERROR AFFECTING BASE CALLING IN NEXT GENERATION ILLUMINA/SOLEXA SEQUENCING

R. te Boekhorst[1]*, I. Abnizova[2], I. Sabir[1], S. Brar[1] and S. Beka[1]
[1] School of Computer Science, University of Hertfordshire, [2] Wellcome Trust Sanger Institute, Hinxton, United Kingdom
e-mail: r.teboekhorst@herts.ac.uk
**Corresponding author*

Key words: *Next generation sequencing, Quality confirmation, Base-calling, Statistics*

Motivation and Aim: In Illumina/Solexa sequencing clusters of single stranded cloned DNA fragments are attached to a flow cell (divided and sub-divided in "lanes" and "tiles"). The template DNA strands are complemented at each position by fluorescently labelled nucleotides in separate chemistry cycles. The fluorescence intensity of the clusters is measured for each of the four nucleotides and, ideally, at each cycle only one of the four signals is displayed. However, due to a number noise factors ("Phasing" and "Cross-talk", [1, 2]) this is mostly not the case and measures are needed to capture the (un-) ambiguousness of base-calling. These measures can be used in calibration, if they correlate well with the probability by which called bases are found back after aligning the read fragments (in which they are contained) to a reference genome. We investigated: 1) which measure best predicted the correctness of base-calling and 2) how much of the variance of this measure is due to lanes, tiles, cycles, identity of the called-base-call and its neighbours ("sequence context").

Methods and Algorithms: Eight metrics routinely used in ecology as measures of species diversity were adapted to assess the signal diversity of a base call. The metrics were computed for data obtained from the genome of the phage FX174, and sequenced by Illumina's Genome Analyzer GA2, release 1.4, run 3259. They were then correlated to the log odds of correct base calls by means of logistic regression. ANOVA was applied to the best performing measure to test for differences between tiles, lanes, type of nucleotide and cycle number. A second ANOVA was carried out to assess the effects of sequence context on the signal diversity of the middle nucleotide of a trimer.

Results: "Purity" (= maximum intensity/sum of the intensities) and a version of it (the relative sum of the two highest intensities) correlated best with the proportion of back-aligned base pairs. Lanes, cycles, and bases on explained up to 9% of the total variance of purity. The interaction of all four factors was significant and accounted for 11% of the overall variance. When T is the middle base of a trimer, preceded by G, average purity is lower than for all other combinations.

Conclusion: Illumina's original metric (purity) is a good measure for sequencing accuracy, but its value depends significantly on the combined effect of lane, tile, cycle and nucleotide.

References:

1. J.C. Dohms et al. (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing, *Nucleic Acids Research,* **36:** 1-10.
2. Y. Ehrlich et al. (2008) Alta-cyclic: a self-optimizing base caller for next-generation sequencing. *Nat Methods,* **8**: 679-82.

QUANTIFICATION OF LONG-RANGE MEMORY EFFECTS IN PROTEINS BY RETURN INTERVAL STATISTICS

A.R. Kayumov[1], M.I. Bogachev[2*], E.O. Mikhailova[3]
[1]Kazan State University, [2]St. Petersburg State Electrotechnical University, [3]Kazan State Technological University
[*]Corresponding author, e-mail: rogex@mail333.com

Key words: *protein sequences analysis, return interval statistics*

Motivation and Aim: in the post-genomic era, the proteins structure characterization became an actual problem. In the last decade, there is an increasing evidence of non-random distribution of amino acids in proteins – the so-called long-range memory effect in the polypeptides [1], following similar findings also in the DNA [2]. Long-range memory properties of proteins have been mainly studied in the context of (multi)fractal models, a powerful class of statistical models widely used to describe biological systems.

Methods and Algorithms: we introduce the return interval approach to protein sequences analysis. We have shown, that return interval statistics are a powerful tool to analyze nonlinear complex systems exhibiting (multi)fractal behavior and to elucidate the long-range memory in biological systems [3]. The major indicator of long-range memory, the distribution of the return intervals P(r), changes its shape drastically in the presence of long-range memory. For the sequence of random samples, return intervals distribution is a simple exponential P(r)~exp(-r), for linearly long-range correlated sequence it is significantly broader and can be well approximated by a stretched exponential P(r)~exp(-r^γ). The major advantage of this approach, in contrast to the (multi)fractal models, is its capability to reveal particular components in the protein sequence that are responsible for the long-range memory effect and thus making one step forward towards relating these statistical properties to the protein functional organization.

Results: the distributions of return intervals between amino acids in 700 bacterial proteins, divided later to 5 functional groups, were analyzed. In the averaged-over-all proteins analysis, only for 5 amino acids (C,H,W,E,R) return interval distributions appeared significantly broader than a simple exponential and could be well approximated by a stretched exponential with $\gamma<0.7$. The return interval distributions for different amino acids varied among protein functional groups, indicating that different amino acids contribute to the long-range memory in each of the studied cases. In particular, following amino acids demonstrated long-range memory: in the intracellular enzymes C, H and W, in extracellular proteases – C,F,H,N and W, in cellulases – C and H, in the transmembrane transport proteins C,D,E,H,W, in the cell-wall associated proteins – C,F,H,M,R,W,Y, and finally for histidine kinases – C and W.

Conclusion: the return interval statistics are a powerful tool to quantify the contribution of amino acids to the long-range memory effect in the proteins related to their function.

Acknowledgements: we like to acknowledge the support of this work from the Federal Programme "Scientific and Academic Staff of Innovational Russia".

References:

1. Yu et al., Phys. Rev. E 68 (2003); Yu et al., J. Theor. Biol. 226 (2003); Moret et al., Phys. Rev. E 71 (2005); Yang et al., Chaos 40 (2009).
2. C.-K. Peng et al., Nature 356 (1992); Chatzdimitriou-Dreismann et al., Nature 361 (1993); C.-K. Peng et al., Phys. Rev. E 49 (1994).
3. Bogachev et al., Phys. Rev. Lett. 99 (2007); Bogachev et al., Eur. Phys. J. - Spec. Topics 161 (2008); Bogachev et al., New J. Phys. 11 (2009).

COMPUTER SUPPORT SYSTEM DEVELOPMENT FOR BIOLOGICAL OBJECTS RESEARCH IN MICROSCOPY

A.G. Bogomolov[1]*, N.L. Podkolodnyy[1,2], N.B. Rubtsov[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*
e-mail: mantis_anton@bionet.nsc.ru
**Corresponding author*

Key words: *LIMS, electronic microscopy, database, metadata, ontology*

Motivation and Aim: Researches in microscopy are used for analysis of different types of biological objects: chromosomes, cellular compartments, cells, tissues and etc. Such tasks as getting high-quality images, isolating and recognizing of biological objects and measuring of different type features on images, have come into being. It is relevant to carry a massive series of microscopic experiments, when there is a necessity to analyze many images of biological objects with different descriptions of their state. In this case, as a rule, typical scenarios of image processing are used. However, such problems as experiment description and store of this description in the database, a search problem for experiment results after completion of the experiment, a problem of joint analysis and reuse of the experiment results are arising. The purpose of this project is computer support system development for biological objects research in microscopy, which is based on metadescription of microscopical experiment.

Methods and Algorithms: The created system includes the following components:

1) The libraries of image analysis algorithms for various scenarios of research.
2) The computer support module for description microscopical experiment with metadata.
3) The module for search, comparison and analysis for microscopical experiment results.

As a basis for metadescription the protocoling formats OME XML and OME TIFF specially developed for microscopy were chosen. Both formats were developed in the frames of OME project (http://www.openmicroscopy.org/site).

Results and conclusion: The project of computer support system development for biological objects research in microscopy, including automatic image analysis, the database management of microscopical experiment results, metadata and images with the possibility of forming metadata, which describe images in semiautomatic mode are described. It is assumed that the developed system will be used in the Multi Access Center of biological objects analysis in microscopy SB RAS. (http://www.bionet.nsc.ru/microscopy/index.html).

References:

1. Ilya G Goldberg et al. (2005) The Open Microscopy Environment (OME) Data Model and XML file: open tools for informatics and quantitative analysis in biological imaging http://genomebiology.com/2005/6/5/R47

COMPUTER ANALYSIS OF CONFORMATIONAL PEPTIDES IN PROTEIN FAMILIES

A.O. Bragin*, P.S. Demenkov, V.A. Ivanisenko
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
e-mail: ibragim@bionet.nsc.ru

Motivation and Aim:

It is known that properties of the protein determined by most conservative parts of protein surface. Such parts may be represented as so called conformational peptides. Representation of protein surface as a set of conformational peptides allows us to compare these peptides with sequence fragments. We suggest that similarities between protein sequence and conformational peptides may lead to common function. Thus, the aim of this work was to create a database of matches between conformational peptides and linear peptides from SWISS-PROT and evaluate their similarities.

Methods and Algorithms:

Conformational peptides are patches on protein surface which satisfied the following conditions: a) the distance between C-alpha atoms of amino acid residues in a conformational peptide should not exceed distance between C-alpha atoms of covalently bound amino acids; b) the amino acid number in conformational peptide was taken 8; c) solvent accessibility of a residue should not be less than 10%.

A set of linear peptides was created by using a sliding window with length 8 amino acid residues. We considered all protein sequences from SWISS-Prot for calculation of a set of linear peptides. For calculations of a set of conformational peptides we used 3d structures from non redundant PDB. Then identical peptides from these two sets were selected.

Results:

Comparisons of conformational and linear peptides allowed us to find about 938830 matches from 48210 different SWISS-Prot proteins. A database with matches between conformational and linear peptides was created. It is interesting that enzymes from phosphotransferase family more often have similarity with conformational peptides from other type of proteins.

Conclusion:

Thus we showed that there are similarities between protein surface conformational peptides and protein sequences. Such similarities may be reason for the same function of not related proteins. Created database may be useful for analysis of relationships between protein structure and function.

Acknowledgements :

State contract FASI №02.514.11.4123, Interdisciplinary integration project SB RAS №111, №94, Program RAS 22. “Molecular and cellular biology”.

COMPUTING POWER OF MICROBIAL GENE LIBRARIES: GENOME ENGINEERING THROUGH THE GENETIC SELECTION

A.V. Bryksin*, I. Matsumura
Emory University School of Medicine, Atlanta, GA, USA
e-mail: anton.bryksin@emory.edu

Key words: *pathway optimization, protein expression, gene regulation, gene libraries.*

Motivation and Aim: The present study addresses the question of the efficient genome engineering for the production of the desirable macromolecules. Metabolic engineers strive to increase the productivity of economically valuable pathways using "divide and rule" process. In this process all target genes participating in the pathway are removed, then tweaked for overexpression and reinserted; and all known shunts, revealed through metabolic network models, are deleted. All pathways in a cell are interconnected, so it is difficult to know *a priori* which mutations in a genome will increase the flux through any given pathway. To add the complexity, some of those interconnectors are among the proteins with unknown functions. We have created a generic functional genomic approach to optimize the flux through a model pathway using the process of genetic selection.

Methods and Results: Our approach is fueled by the notion that functions of many annotated genes in the genome of even well studied microorganisms such as *E. coli* are not known. The experimental design is rather simple and can be applied to any model. Use of the approach could make it easy to engineer the desirable genome. The *lux* operon of *Photorhabdus luminescens* was employed as a model; it encodes enzymes that use exchangeable cofactors (ATP, NADPH and $FMNH_2$) to catalyze the reduction of oxygen to water and the concomitant release of light. We have screened the ASKA Open Reading Frame (ORF) collection and identified 70 multicopy enhancers and 57 suppressors of *lux* activity. Apparently, many *E. coli* enzymes supply metabolites to this foreign pathway or divert resources from it.

Conclusion: Metabolic network models are currently limited in their accuracy. Functional genomics experiments will continue to elucidate the biological roles of uncharacterized Open Reading Frames (ORFs). We have already identified a number of uncharacterized "hypothetical proteins" that affect the flux through a foreign pathway.

ENDOSYMBIONT *WOLBACHIA* IN HOKKAIDO POPULATIONS OF GRASSHOPPER *PODISMA SAPPORENSIS* SHIR.

A.G. Bugrov[1,2]*, Yu.Yu. Ilinsky[3]
[1]*Institute of Systematics and Ecology of Animals, SB RAS, Novosibirsk, Russia* [2]*Novosibirsk State University, Novosibirsk, Russia*
e-mail: bugrov04@yahoo.co.uk
[3]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
e-mail: paulee@bionet.nsc.ru
[*]*Corresponding author*

Key words: *Wolbachia, Podisma sapporensis, populations.*

Motivation and Aim. The brachypterous grasshopper *P. sapporensis* consists of two main chromosome races in Hokkaido. The western group of isolated populations: Shimokawa, Naganuma, Yotei and Disengen has the X0/XX sex chromosome determination (X0 race). The eastern group of populations: Akan, Tanno has the XY/XX sex chromosome determination (XY race). The X0 and XY chromosome races are geographically isolated by the mountainous system consisting of the Mts Daisetsu and Hidaka range, occupying the central part of the island. (Bugrov et al., 2001).

In crossings of XY-Akan x X0-Teine and XY-Akan x X0-Naganuma the zygotic barrier have been found. The vast majority of embryos of these type crosses are non-vital haploid/diploid parthenogenetic. A zygotic barrier was found also between the X0-Shimokawa and XY-Akan populations. On the contrary the crossing between of the XY-Tanno and X0-Teine subraces gave viability F1 and F2 generation in spite of many chromosome differences (Bugrov et al., 2004; Warchalowska-Sliwa et al., 2008).

The results of the crossings do not support the hypothesis that chromosomal differences play a key role in restricting gene flow between X0 and XY races. Genetic effects of reproductive isolation between crossed populations of *P. sapporensis* were similar to those for infection with endosymbiotic bacteria *Wolbachia. Wolbachia* in arthropods can induce cytoplasmic incompatibility, partenogenesis, feminization or male-killing, i.e. causes of partial reproductive isolation in their host. So we decided to examine the samples from each population of *P. sapporensis* for infection with *Wolbachia.*

Methods and Algorithms. For identification of *Wolbachia* strains the nucleotide sequence of the 16SrRNA and wsp genes were used.

Results. Nucleotide sequence of 16SrRNA and wsp genes *Wolbachia* in *P. sapporensis* populations corresponds to the supergroup B. Analysis of *P. sapporensis* populations revealed a polymorphism on *Wolbachia* infection. In populations Naganuma, Yotei Akan, and Tanno the nucleotide sequences of the gene *Wolbachia*-wsp belonging to the strain I were detected, while the Disengen population is infected with *Wolbachia* belonging to the strain II. In Shimokawa population *Wolbachia* not found.

Conclusion and Availability. The polymorphism in studied populations of *P. sapporensis* on the infection with endosymbiont *Wolbachia* may be a reason of reproductive barriers between populations.

Acknowledgements. Granted by the Federal Program "Scientific and Pedagogical Staff of Innovative Russia (project 02.740.11.0277).

MORPHOGENESIS OF *DROSOPHILA MELANOGASTER* MACROCHAETES: A GENE NETWORK DESCRIBING THE ESTABLISHMENT OF BRISTLE PREPATTERN

T.A. Bukharina[1, 2]*, D.P. Furman[2]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, RUSSIA*
[2]*Novosibirsk State University, Novosibirsk, Russia*
e-mail: bukharina@bionet.nsc.ru
[*]*Corresponding author*

Key words: *neurogenesis, prepattern, gene network, Drosophila melanogaster*

Motivation and Aim: The external sensory organs (macrochaetes) are located on the fly body is a strictly determined order forming a constant pattern, which is determined by the positions of the groups of presumptive cells (proneural clusters) in the ectoderm of the wing imaginal disc. It has been experimentally shown that the sites for development of proneural clusters (prepattern) are specified by a local combination of particular transcription factors (or prepattern factors), while characteristic of the cells forming these clusters is an increased content of the Achaete and Scute proteins. Formalized description and logical analysis of the relevant experimental data require bioinformatics approaches, in particular, the methodology of gene networks.

Methods and Algorithms: The GeneNet technology [1] was used for reconstructing the gene network "Neurogenesis (prepattern)".

Results: The constructed gene network so far contains the data extracted from 91 scientific papers and comprises the information about 137 components (27 genes, 32 proteins and protein complexes, 2 RNAs, and 2 processes) and 74 links between them. The main elements and regulatory circuits providing for the function of this gene network have been detected.

Conclusion: The gene network "Neurogenesis (prepattern)" describes the stage of prepattern development preceding the separation of proneural clusters in the ectoderm of wing imaginal discs. The functioning of this gene network is determined by relationships between expressions of two groups of genes—the genes encoding transcription factors (initiating the achaete-scute complex genes) and the genes determining the structuring of wing imaginal discs into longitudinal domains. Establishment of the domain structure in imaginal discs leads to a localized expression of certain genes encoding transcription factors, namely, U-shaped, Pannier, the protein of the *iroquois* gene complex (Araucan, Caupolican, and Mirror), and several proteins of the EGFR signaling pathway. When interacting with enhancers in the regulatory region of *achaete-scute* complex, they induce its transcription in various regions of the ectoderm, thereby determining the prepattern and final precise positioning of proneural clusters.

Acknowledgements: The work was supported by the Russian Academy of Sciences (programs 15 and 22.8); Siberian Branch of the Russian Academy of Sciences (program "Genomics, Proteomics, and Bioinformatics: Hi-Tech Research into Molecular Biological and Molecular Genetic Systems and Processes" and interdisciplinary integration project no. 119); Federal Agency for Science and Innovations (contract no. 02.512.11.2332); and President of the Russian Federation (program for support of leading scientific schools, grant no. NSh-2447.2008.4).

References:

1. E.A. Ananko et al. (2005) GeneNet in 2005, *Nucleic Acids Res.,* **33**: D425–D427.

THE RELATION BETWEEN BIOLOGICAL COMPLEXITY OF EUKARYOTES AND EVOLUTIONARY CHANGES OF NOTCH CASCADE PROTEIN FEATURES

K.V. Gunbin, T.A. Bukharina*
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: bukharina@bionet.nsc.ru
**Corresponding author*

Key words: *Notch cascade, protein family, evolution, phylogenetic analysis*

Motivation and Aim: Notch cascade is a highly conserved signaling system. It's important for specifying cell fates and for defining boundaries between different cell types. Proteins, which are the main parts of Notch cascade (Notch, Delta, Serrate, Fringe) form large protein families and are involved in a great number of early-event developmental processes: differentiation, proliferation, and apoptosis. In different species of eukaryotes the numbers of different protein family belonging to Notch family are different. The aim of our work was to search the relations between the complexity of extant organism and evolutionary changes of Notch cascade protein features.

Methods and Algorithms: We developed a novel method (SAMEM) which implements the permutational test for comparing the molecular evolution model of the Notch, Delta, Serrate, Fringe protein families with the real data. The SAMEM was used for finding rare amino acid substitutions on the Notch's, Delta's, Serrate's, Fringe's phylogenetic trees branches. We performed the correlation analysis of the cumulative evolutionary changes of the 531 amino acid physicochemical properties (PCP) [1] with the organism complexity [2] using the Spearman statistic.

Results: The phylogenetic trees were reconstructed for Notch, Delta, Serrate and Fringe proteins separately. We found bursts of rare amino acid substitutions for the tree branches on which of Notch, Delta, Serrate and Fringe paralogs diversified. It is consistent with the functions of these proteins as receptor, ligands and glycosyltransferase, the function of their paralogs are noninterchangeable [3, 4]. The correlation analysis between the separated Notch, Delta, Serrate, Fringe proteins evolution and the organism complexity allows us to discover the most significant correlation for Delta-4 (93.6% PCP, $p<0.05$) and Serrate-2 (88.1% PCP, $p<0.05$) genes. Also some correlations were found for the Notch-1 (0.8% PCP, $p<0.05$).

Conclusion: This result may reflect the necessity of additional types of Notch receptors or its ligands in vertebrates and especially in mammals for creating and supporting the great variety of cell types and tissues. Besides, the main reason of such effect may be the necessity of supporting structures which can realize new functions.

Acknowledgements: This research was supported in part by the RFBR (№09-04-01641-a), Nsh-2447.2008.4.

References:

1. S. Kawashima et al. (2008) AAindex: amino acid index database, progress report 2008, *Nucleic Acids Res.,* **36**: D202-D205.
2. C. Vogel, C. Chothia. (2006) Protein family expansions and biological complexity, *PLoS Comput Biol.,* **2**: e48.
3. K. Katsube, K. Sakamoto (2005) Notch in vertebrates--molecular aspects of the signal, *Int. J. Dev. Biol.*, **49**: 369-374.
4. U. Marklund et al. (2010) Domain-specific control of neurogenesis achieved through patterned regulation of Notch ligand expression, *Development*, **137**: 437-445.

REVEALING REGULAR ORGANIZATION IN THE CODING REGIONS OF GENES AND STRUCTURE OF THE PROTEINS

M.B. Chaley[1]*, V.A. Kutyrkin[2]
1Institute of Mathematical Problems of Biology RAS, Pushchino, Russia
2Moscow State Technical University n.a. N.E. Bauman, Moscow, Russia
e-mail: maramaria@yandex.ru
**Corresponding author*

Key words: *latent profile periodicity, gene structure, protein structure*

Motivation and Aim: A new type of latent periodicity in biological sequences which expands on the notion of approximate tandem repeats was called profile periodicity or profility [1]. In the present work the original methods have been developed in order to identify the latent profile periodicity in DNA and biological meaning of the latent profility has been analyzed.

Methods and Algorithms: The methods elaborated for profility recognition in DNA sequences [2] are based on analysis of statistical spectra, one of which is called characteristical.

Results: Regularity with the three nucleotides in the picks of characteristical spectrum is observed in the coding regions (see figure 1). Such a peculiarity is caused by the first level in organization of encoding – the genetic triplet code. The latent profility of region (which is different from 3-profility) defines the second level in organization of genetic encoding. The clear-cut maxima of the characteristical spectrum single out the level (see figure 1).

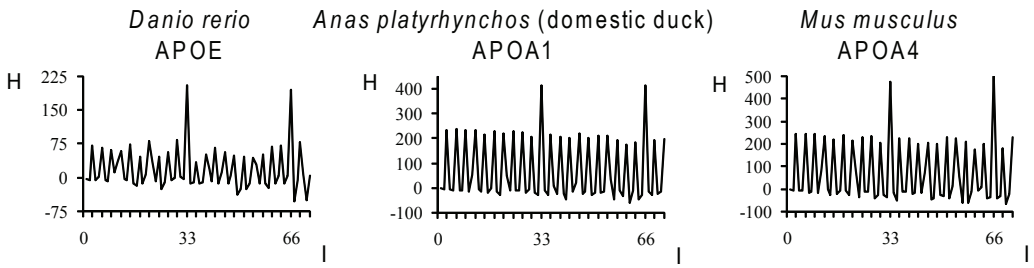


Figure 1. Characteristical spectra for coding regions of the three genes from apolipoprotein family. APOE (GenBank ACESSION AJ236882; 1..846 nt). APOA1 (GenBank ACESSION U86131; 1..795 nt). APOA4 (GenBank ACESSION M64248; 1..1176 nt).

Conclusion: The latent profility in coding DNA regions defines the different regularity levels in genetic information encoding and correlates with peculiarities of the 2D structure of the proteins.

References:

1. M.B. Chaley, V.A. Kutyrkin. (2008) Model of perfect tandem repeat with random pattern and empirical homogeneity testing poly-criteria for latent periodicity revelation in biological sequences, *Mathematical Biosciences*, **211**: 186-204.
2. M.B. Chaley, V.A. Kutyrkin. (2009) [Latent Profile Periodicity as New Type of Periodicity in Genome], In: *Dokladi XIV Vserossiiskoi Konferentsii "Matematicheskie Metodi Raspoznavaniya Obrazov"*, 614-617 (Moskva: Max Press). [in Russian]

FOLDING OF A SH3 DOMAIN: HYDRODYNAMIC DESCRIPTION

I.V. Kalgin[a], M. Karplus[b,c] and S.F. Chekmarev[a,d#]
[a]*Novosibirsk State University, Novosibirsk, Russia*
[b]*ISIS Université de Strasbourg, Strasbourg, France*
[c]*Harvard University, Cambridge, Massachusetts, U.S.A.*
[d]*Institute of Thermophysics SB RAS, Novosibirsk, Russia*
e-mail: chekmarev@itp.nsc.ru
[#]*Corresponding author*

Key words: *SH3 domains, folding, molecular dynamics, folding flows, turbulent phenomena*

Motivation and Aim: The standard analysis of the folding process is typically based on consideration of the mean-force (free energy) surfaces. However, these surfaces determine only the probability for the system to be in a given state and do not show the direction of the motion; that is, the system can either fold or unfold when it visits this state.

Methods and Algorithms: Discrete molecular dynamics has been employed to study folding of fyn SH3 domain with a C_α-based Gō-model. To analyze the folding process, a "hydrodynamic" description of folding flows [1] in 2D and 3D spaces of collective variables has been used.

Results: Two types of folding trajectories (fast and slow) have been observed, which follow essentially different routes in the final stage of folding. The hydrodynamic flows for the fast trajectories are mostly "laminar", whereas those for the slow trajectories are essentially "turbulent", involving many vortices [2]. Comparison of the simulation results with the experimental data (Ref. [3] and papers cited therein) suggests that the two-state kinetics observed for fyn and src SH3 domain folding are associated with the slow trajectories, in which a partly formed N- and C-terminal β-sheet hinders the RT-loop from attaching to the protein core; the fast trajectories presumably are not observed in the stopped-flow experiments because they are in the dead time (1 ms). It has been shown that the folding flows for the slow trajectories have fractal dimension, and the space correlation (structure) functions of the flows are similar to those in the 3D hydrodynamic turbulence [4].

Conclusion: The hydrodynamic description has shown that the folding flows do not follow the free energy surface of the protein. The flows for the slow folding trajectories are essentially "turbulent" and have many properties of turbulent flows in hydrodynamics.

References:

1. S.F.Chekmarev, A.Yu. Palyanov and M.Karplus. (2008) Hydrodynamic Description of Protein Folding, *Phys. Rev. Lett.,* **100:** 018107 (1-4).
2. I.V.Kalgin, M.Karplus and S.F.Chekmarev. (2009) Folding of a SH3 domain: Standard and "hydrodynamic" analyses, *J. Phys. Chem. B,* **113:** 12759-12772.
3. J.Li, M.Shinjo, Y.Matsumura, M.Morita, D.Baker, M.Ikeguchi and H. Kihara. (2007) An alpha-helical burst in the src SH3 folding pathway, *Biochemistry*, **46:** 5072–5082.
4. L.D.Landau and E.M.Lifshitz. (1987) *Fluid Mechanics* (Pergamon, New York).

MORPHOGENETIC GUIDANCE OF THE EVOLUTION OF HAIRS (TRICHOMES) IN PLANTS OF THE *DRABA* GENUS

V.G. Cherdantsev, O.V. Grigorieva, V.A. Scobeyeva
Dept. of biological evolution, Faculty of Biology, Moscow State University, Moscow, Russia
E-mail: arnosnew@mail.ru

Key words: *Morphogenesis, evolution, plants, hair cells, active shells, branching, self-organization*

Motivation and Aim: The bias of modern developmental and evolutionary biology is considering the variation of form as a genetic and (or) environmental impact. Our aim is to show that morphogenesis itself is the main source of morphological variation, which, in spite of low heritability and adaptive value, anticipates the trends of morphological evolution.

Methods and Algorithms: We studied the normal (usually within-individual) variability of branching of unicellular hairs (trichomes) in 11 species of plants of the *Draba* genus basing on quantitative morphological analysis of SEM data. Geometry of shape of the hair cells and its dynamics during their early morphogenesis were reconstructed by the analysis of SEM images with the aid of AUXIOVISION program.

Results: The hair cell surface is an active shell, which means the shell being subject to passive stretching because of the osmotic pressure and, at the same time, capable of generating active stresses by recruiting of new structural elements, which leads to an increase in the cell surface area and origination of lateral pressure inside the shell. For all species, and for all types of the unicellular hairs, we could infer the same cyclic algorithm of morphogenetic growth in which smoothing of the shell shape leading to an increase in the volume/surface ratio alternates with enhancing of the local shape differences leading to a decrease in the volume/surface ratio. Such a cycle provides a single mechanism of growth and branching, which means variability of shape of the active shell being a direct and inevitable consequence of its developmental mechanics. Comparative analysis revealed a strong parallelism between within-individual and taxonomic differences suggesting that variability based almost exclusively on developmental self-organization does afford materials for evolution. Moreover, the evolution of branching of the hairs is obviously directional, leading to a gradual replacement of the radial-bilateral symmetry by the bilateral one, to a gradual attenuation of the secondary branching, and to an appearance of successive developmental stages of the primary branching. One can hardly explain this by directional or stabilizing selection because branching geometry has no adaptive value on its own and does not correlate with the cell size. Rather, the mechanics of morphogenesis turns the growing hair cell into a non-holonomic system capable of selective filtration of random fluctuations. The developmental acceleration of a branch rudiment decreases a probability of its secondary branching and, at the same time, produces in the active shell additional stresses preventing the secondary branching of other rudiments. Developmental retardation of any given rudiment has no effect on the behavior of neighboring rudiments. It follows that, among equally probable opposite types of the fluctuations, only one of them has a chance to be remembered.

Conclusion: The conclusions are, first, that neither low heritability, nor a lack of the adaptive value of a structure, prevent its directional evolution, and, second, that its evolutionary trends are subject to morphogenetic guidance not in the trivial sense of "morphogenetic constraints", but rather because morphogenesis is capable of unidirectional accumulation of bidirectional effects of selection.

THE MODEL OF SPATIAL PATTERN FORMATION IN TRIMERIC FIVE-WHORLED FLOWER

V.V. Choob
Moscow State University, Moscow, Russia
e-mail: choob_v@mail.ru

Key words: *floral morphogenesis*, *spatial pattern formation*, *meristem*, *organotaxis.*

Motivation and aim. Flower morphogenesis is in the focus of modern plant developmental genetics. The mainstream of this investigation is framed by the ABCDE-model of genetic control of flower development. Despite of the increasing knowledge, the mechanisms of spatial control of organ initiation by now are poorly understood. Exploiting the principles (i) of the available space; (ii) of inhibition and (iii) of the space economy for primordia initiation we were aimed to investigate the organotaxis in whorled flowers. Trimeric flower was used as a model for the reasons of geometric simplicity and of its' wide distribution among Monocots and some Dicot families [1].

Methods and algorithms. In previous work [2] we have elaborated the algorithm for analysis of spatial pattern formation, based on the differential equations for the spherical surface. Biological principle of the model: (i) floral meristem is a hemisphere with radius R, growing in time; (ii) perianth organs find their position on the surface of meristem acropetally (toward the pole) in the belt, limited by radii rMin and rMax; (iii) carpels and stamens may find their position on the meristem surface basipetally (downwards from the pole), starting from the cap, which is limited by radius R3, growing with rate dR3; (iv) all the flower organs in the model are characterized by minimal space d0 and inhibition zones around (Dmin/Dmax), in which new organs can not appear. For the details of the mathematical algorithm of the model see [2].

Results. We put forward 4 different hypothesis of the flower pattern formation: (i) acropetal; (ii) acopetal + basipetal patterning of carpels; (iii) acropetal + basipetal patterning of all the stamens and carpels; (iv) acropetal patterning of perianth and uoter stamens + basipetal patterning of inner stamens and carpels [1]. The hypothesis (iii) is in the best correspondence with the ABCDE-model [2]. Calculatory experiment led to the appropriate solutions for the values of basic parameters of the model for (i), (iii) and (iv) hypothesis, but not for (ii). Further investigation of the flower structure, created in the model, revealed, that in the frame of (i) and (iii) hypothesis the position of organs is highly sensitive to the variation of inhibition zone parameters (Dmin/Dmax). On the contrary, within the (iv) hypothesis the geometric solution for flower organs is stable and does not change significantly even at 30-60% variation of Dmin/Dmax.

Conclusion. The comparison of calculatory experiment data and of polymorphism of flower organ position *in vivo* in *Rheum* demonstrated, that (iv) hypothesis gives the best predictions for the structure of abnormal flowers [1]. Thus the most stable mathematical solution could be regarded as an adequate model for the trimeric flower patterning.

Availability. For the algorithm exploitation the software "Phyllotaxis" was developed. By now it is not user-friendly, but available on special request from the authors [2].

References:

1. V.V.Choob, O.V.Yurtseva. (2007) A mathematical model of flower structure of Polygonaceae, Russian Botanical Journal, 92: 114-134 [in Russian].
2. K.G.Skryabin et al. (2006) Type specification and spatial pattern formation of flower organs: a dynamic model of development, Biology Bulletin RAS, 33: 523-535.

MOLECULAR DESIGN OF VIP RECEPTOR (GPCR B) — A PATHWAY TO NOVEL ANTI-INFLAMMATORY THERAPIES

A.O. Chugunov*[1], M. Rooman[2], I. Langer[2], R.G. Efremov[1]
[1] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry RAS, ul. Miklukho-Maklaya 16/10, Moscow, 117997 Russia.*
[2] *Universitŭ Libre de Bruxelles, Brussels, Belgium.*
** Corresponding author. E-mail: batch2k@yandex.ru.*

Motivation and aim. G-protein coupled receptors (GPCRs) are in the limelight of both academia and pharmaceutical industry since they mediate plurality of biological processes from signalling to transport. Availability of 3D-structure of GPCRs is a success factor for drug discovery projects, and the source of these data may be not only experiments but also computation. To date structures of four members of the GPCR A subfamily have been solved, but still there are no complete structures of GPCR B members, which bind in their extracellular domain such biologically active peptides as secretin, glucagon, calcitonin, corticotropin-releasing factor, vasoactive intestinal peptide (VIP), and others. Given the almost non-significant level of homology between amino acid sequences of GPCR A and B subfamilies, pipeline of homology modeling hampers at the very early stage: building pairwise alignment with the structural template — receptor of the GPCR A subfamily.

Methods and Algorithms. Equivocation of conventional algorithms for sequence alignment in case of GPCR B and A families is emphasized by the fact that all alignments found in literature and built with various software tools differ exceedingly. We designed an iterative procedure of alignment "selection" by making independent sequence shifts for individual transmembrane (TM) helices and assessing reasonableness of resulting models with the knowledge-based "membrane score" approach (Chugunov *et al*., 2007), developed for estimation of "quality" of protein chain packing in 3D-models of TM domains. There were several rounds, where the best result from the previous one was an input for the subsequent.

Results. As a part of a project for design an exclusively selective inflammation inhibitor that should act through engineered VIP receptor (VIPR1), a structural model of TM domain of this receptor was constructed, and also mutagenesis study was performed (Chugunov *et al.*, 2010). The results suggest that residues important for VIP binding and/or receptor activation — R^{188}(TM2)–Q^{380}(TM7)–N^{229}(TM3) — draw up in a line, and this proximity is further confirmed by cooperative and anti-cooperative activation effects that appear in double mutants built from reciprocal residue exchanges.

Conclusion. The results obtained suggest that the mentioned residues may play crucial role in VIPR1 activation process, which seems to differ from GPCR A receptors activation.

Refeences:

1. Chugunov A.O., Novoseletsky V.N., Nolde D.E., Arseniev A.S., Efremov R.G. (2007). A method to assess packing quality of transmembrane α-helices in proteins. I. Parameterization using structural data. *J. Chem. Inf. Model.* **47**, 1150–1162 and … II. Validation by "correct *vs.* misleading" test. *J. Chem. Inf. Model.* **47**, 1163–1170;
2. Chugunov A.O., Simms J., Poyner D.R., Dehouck Y., Rooman M., Gilis D., Langer I. (2010). Evidence that interaction between conserved residues in transmembrane helices 2, 3 and 7 are crucial for human VPAC1 receptor activation. *Mol. Pharm.*, submitted.

ECTOCARPUS PORTAL: INTEGRATION OF GENOMICS, TRANSCRIPTOMICS AND DNA PHYSICAL PROPERTIES DATA

A.A. Osypov[1,2,3], K. Billiau[4], L. Sterck[4], S. Dittami[1,2], T. Tonon[1,2], Y. Van de Peer[4], M. Cock*[1,2]

[1]*UPMC Univ. Paris 06, The Marine Plants and Biomolecules Laboratory, UMR 7139, Station Biologique de Roscoff, Place Georges Teissier, BP74, 29682 Roscoff Cedex, France.*

[2]*CNRS, UMR 7139, Laboratoire International Associe Dispersal and Adaptation in Marine Species, Station Biologique de Roscoff, Place Georges Teissier, BP74, 29682 Roscoff Cedex, France.*

[3]*Laboratory of Cell Genome Functioning, Institute of Cell Biophysics of RAS, 142290, Pushchino MR, Russia*

[4]*Bioinformatics & Systems Biology division, Department of Plant Systems Biology (VIB), Ghent University, Technologiepark 927, 9052 Gent, Belgium*

e-mail: cock@sb-roscoff.fr

**Corresponding author*

Key words: *brown algae, Ectocarpus siliculosus, data integration, genomics, transcriptomics, DNA physical properties, electrostatics*

Motivation and Aim: Brown algae is one of only five lineages on the tree of life that independently obtained complex multicellularity, and together with red algae are least studied. To establish a genome resource for the brown algae model organism *Ectocarpus siliculosus* we launched a project to generate a user-friendly database resource for its genome. It is based on 1) the improved large-scale assembly of the genome through assigning supercontigs to linkage groups using a genetic map, 2) improved identification of transcribed regions using a whole genome tiling array approach, 3) a detailed manual annotation of the genome.

The main goal of the DB project is to make internet portal to accumulate, hold and provide all the available information on *Ectocarpus siliculosus* as a brown algae model organism.

Results: At present the portal is based on the BOGAS genome database with all the results of the mentioned analyses and the genome annotation. Genome lists and other aggregations and annotations are added. There is a Transcriptomics section integrated, with data and analytical tools for different transcriptomic experiments. Also graphical and analytical tools were developed to represent and study the genome DNA physical properties, including electrostatic potential profile along the DNA molecule. These properties are important for interactions of specific regions of DNA with functional and regulating proteins such as RNA-polymerase and transcription factors. A Literature section holds the description of useful articles. All the data are integrated and cross-linked within the portal, providing the ability to investigate the genome on the individual genes basis as well as large comparative studies.

Availability: the portal is free available to academic use on the station website at http://www.sb-roscoff.fr

Acknowledgements: the project is funded by the Europol Mer.

ANALISYS OF HUMAN PROTEOME BASED ON PROTEIN STABILITY TO MUTATIONS

P.S. Demenkov*, Korepanova O.A., T.V. Ivanisenko, V.A. Ivanisenko
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
e-mail: demps@bionet.nsc.ru
**Corresponding author*

Motivation and Aim: The characterization of protein interactions is essential for understanding biological systems. Within interaction networks, most proteins interact with few partners, while a small proportion of proteins, called hubs, participate in a large number of interactions and play a central role in biological processes. Thermodynamic stability is an important characteristic of proteins. Changes in thermodynamic stability upon mutations cause many diseases. Mutations can lead to increase or decrease in thermodynamic stability. Our aim was to identify relationships between human protein susceptibility to single mutations and connectivity of human proteins in a molecular-genetic interaction network.

Methods and Algorithms: For every human protein, we calculated the number of single mutations increasing, decreasing or not affecting thermodynamic stability. Protein susceptibility was defined as single mutations percent that decreased protein thermodynamic stability. The updated version of ProtStability program was used to predict changes in protein thermodynamic stability on primary structure. The ProtStability program is based on the «Modified KRAB» method. The updated vertion of ProtStability program used position specific score matrix. The ANDCell system was used to reconstruct associative network. Associative network consists of interactions and associations between human genes, proteins, microRNAs, metabolites, molecular processes, and pathways, cellular components.

Results: Susceptibility to single mutations within the 0.5-0.85 range was observed. Uniform distribution of protein susceptibility to single mutations was found for proteins with connectivity within the 1-200 range. Protein susceptibility depends on the number of connections of protein in the associative network. Marginally essential proteins (hubs) have average susceptibility to single mutations.

Conclusion: Variation in susceptibility to single mutations decreased with the increase in the number of connections of protein in the network. Thus the susceptibility to single mutations for hubs varies in a close range. It may be assumed that for hubs exists the optimal value of susceptibility to single mutations that is necessary for maintenance of the flexible protein structure and noise immunity to mutations.

Work was supported in part by State contract FASI №02.514.11.4123, interdisciplinary integrative project for basic research of the SB RAS № 94, 111, 119, SS-335.2008.1 and Program RAS 22. Molecular and cellular biology.

COMPUTER SIMULATION OF FERREDOXIN-FNR INTERACTION IN SOLUTION

A.N. Diakonova*, I.B. Kovalenko, A.M. Abaturova, G.Yu. Riznichenko
Biophysical Department, Biological Faculty, Lomonosov Moscow State University, Moscow, Russia
e-mail: alex-diakonova@yandex.ru
**Corresponding author*

Key words: *computer simulation, reaction kinetics, ferredoxin, FNR*

Motivation and Aim: In the photosynthetic electron transfer chain ferredoxin reduced by Photosystem I donates an electron to FNR which in turn catalyses $NADP^+$ reduction.Our goal is to develop a computer multiparticle model of complex formation between proteins ferredoxin and FNR.

Methods and Algorithms: In our model several hundreds of proteins diffuse in a viscous medium due to random Brownian and electrostatic forces [1,2]. When two proteins of different species approach each other and acquire certain orientation they can form a complex. The model allows obtaining kinetic curves for the complex formation between two proteins. The electrostatic potentials of the proteins are calculated using Poisson-Boltzmann equation.

Results: Using our model we have obtained second order rate constants for complex formation between wild type (wt) ferredoxin and wt and mutant FNR. Also we have calculated ionic strength dependencies for wt and mutant proteins. We have shown that charged amino acid residues on FNR surface can be divided into two groups in respect to their contribution to electrostatic attraction of ferredoxin. Ionic strength dependencies for these two groups of residues show different behavior: slightly impaired mutant and wt FNR show non-monotonic behavior with maximum value at 200 mM due to formation of tight non-productive complexes at low ionic strength, while second-rate constants of highly impaired mutants decrease monotonically with ionic strength. Results of our simulations are in good correspondence with the experimental data.

Conclusion: Our model allows to simulate kinetics of protein binding taking in account shape of the molecules and electrostatic interactions between them.

Availability: The software is available on request from the authors.

References:

1. 1. A.B. Rubin, G.U. Riznichenko. (2009) Modeling of the Primary Processes in a Photosynthetic Membrane, In: *Photosynthesis in silico: Understanding Complexity from Molecules to Ecosystems,* A. Laisk et al. (Eds.), 151-176 (Springer).
2. 2. Kovalenko et al. (2008) Direct computer simulation of ferredoxin and FNR complex formation in solution, In: *Flavins and Flavoproteins International Symposium,* S. Frago et al. (Eds.), 437-442 (Prensas Universitarias de Zaragoza).

USING THE COMPUTER-BASED IMAGE PROCESSING TECHNIQUE IN GENETIC ANALYSIS OF LEAF HAIRINESS IN WHEAT TRITICUM AESTIVUM L.

A.V. Doroshkov*[1], M.A. Genaev[1], T.A. Pshenichnikova[1], D.A. Afonnikov[1,2]
[1]- *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]- *Novosibirsk State University, Novosibirsk, Russia*
e-mail: ad@bionet.nsc.ru
[*]*Corresponding author*

Key words: *wheat, leaf hairiness, trichome*

Motivation and Aim: Leaf hairiness in wheat is of great importance for protection from pests and for adaptation to environmental factors. For example, this trait is characteristic of a number of drought resistant wheat cultivars referred to the steppe ecological group. Study of the features of leaf hairiness morphology and identification the corresponding genes will allow to obtain varieties resistant to hard climatic conditions and certain pests. To identify the genes responsible for the leaf hairiness, mass analysis of a great number of plants belonging to different hybrid populations is needed, accompanying with a laborious manual job.

Methods and Algorithms: Furthermore, a more accurate description of the morphological properties of the trait for correct determination of phenotypic classes is timely. Using of new computer–based technologies for descriptions of quantitative characteristics of leaf hairiness is the important step in this direction. In the course of the work, we used the LHDetect program for determining the degree of leaf hairiness and its morphological properties on the basis of its microscope image processing. The suggested method appeared to be the effective approach for a large scale analysis of leaf hairiness morphological peculiarities in individual plants. For example in according with genotyping this approach can be useful to quantitative trait loci (QTL) mapping.

Results: In this study we detailed analysis of hairines in wheat as a complex feature. For two different cultivars with similar leaf hairiness was shown differences. The disjoining of hairness trait in F2 generation hybrids was studied for several combinations of parents. This allowed us to qualitatively estimated the possible number of genes that may control the hairness trait in different cultivars. It was shown that this trait for several cultivars is polygenic. We show correspondence between trichome middle length and number of trichomes on several cultivares.

Conclusion: LHDetect is fast and powerful tool for analysing leaf hairiness in wheat.

Availability: The LHDetect system is available at http://www.wheatdb.org/lhdetect

The work was supported by RAS Program 2 «Origin and evolution of biosphere»

EXPERIMENTAL VERIFICATION OF THE PROGNOSIS HUMAN TBP/TATA AFFINITY CHANGE AS A RESULT OF POLYMORPHISMS, ASSOCIATED WITH HEREDITARY PATHOLOGIES

I.A. Drachkova*, P.M. Ponomarenko, T.V. Arshinova, M.P. Ponomarenko, L.K. Savinkova and N.A. Kolchanov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** idrachkova@mail.ru*

Key words: *TBP, TATA-box, affinity, polymorphism, equilibrium equation of TBP\TATA binding*

Motivation and Aim: SNPs represent the most general form of genetic variability of the human. Functionally significant SNPs define a susceptibility of the human to a bacterial and virus infection, sensitivity to medicines etc. Effective methods of a prediction of functionally significant SNPs genes regulatory regions are absent till now. Therefore the purpose of our research consists in creation of a reliable and effective experimentally-computer method for prediction of functionally significant for health of the human of TATA-boxes SNPs.

Methods and Algorithms: In our earlier experimental and in silico studies of eukaryotic TATA boxes SNPs we designed a step-by-step model of TBP binding to TATA boxes and the equilibrium equation of four steps of binding TBP\TATA: nonspecific binding → sliding → recognition→ stabilisation [1]. By means of this equation sequences of TATA-boxes of human gene promoters containing SNPs, associated with predisposition to certain hereditary pathologies are analysed, and forecasts change of affinity TBP/TATA [2] are made.

Recombinant human TBP, was expressed in the *E.coli* strain BL21 (DE3) – plasmid pAR3038-hTBP was kindly presented by proff. Puhg, Center for Gene Regulation, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsilvania, USA. For experimental verification of equilibrium dissociation constants of complexes TBP with TATA-containing oligonucleotides, corresponding to "normal" and "mutant" variants, the traditional approach including titration of fixed quantity TBP by increasing quantities TATA-containing oligos (length 26 bp), is used. Densitometry of the radioautographs carried out by means of the program «Gel-Pro Analyzer 3.1». Values of equilibrium dissociation constants TBP\TATA counted by means of the computer program "OriginPro 8".

Results: In the given work we represent the first results of experimental verification by means of method EMSA in silico prognosis. Results of experimental verification of change TBP/TATA dissociation constants as a result SNPs well agreed at qualitative level with the prognosis.

Conclusion: Experimentally-computer research carried out full volume will allow not only to predict functionally significant SNPs TATA-boxes which can be associated with predisposition to various diseases of the human, but also will improve our understanding of molecular mechanisms of a polymorphic variety of eukariotic gene expression regulation.

Acknowledgments:
This work was supported in part by the Russian Foundation for Basic Research, projects nos. 08-04-01048, 06-04-49556 and 10-04-00462, grant 119 of SB RAS, grant 23.29 "Biological diversity" of RAS.

References:

1. Ponomarenko P.M., Savinkova L.K. et al. Doklady RAN, 2008, **419**,88-92.
2. Ponomarenko P.M, Ponomarenko M.P. et al. Mol.Biol (Russia), 2009, **43**, 512-520.

REFERENCE-INFORMATION SYSTEM FOR MOLECULAR GENETIC CERTIFICATION OF CARP (*CYPRINUS CARPIO* L.)

S.E. Dromashko*, Ya.I. Sheiko, A.Yu. Koneva
Institute of Genetics and Cytology NASB, Minsk, Belarus
e-mail: S.Dromashko@igc.bas-net.by
**Corresponding author*

Key words: *Cyprinus carpio L., molecular markers, MS Excel, RAPD analysis, reference-information system*

Motivation and Aim: Genome registration of living organisms is a high priority task nowadays. As for fish-breeding, biochemical markers were a main method of certification until recently. Meanwhile foreign researchers have revealed high genome variability of some European (Hungary) and Asiatic (China) carp races using DNA typing methods (RAPD markers and microsatellites). Within NIS countries only Russia carries out investigations of Russian and European carp races by multipoint genetic typing [1]. We developed the technology of genetic certification for Belarusian carp races and offsprings using molecular markers. The technology includes reference-information system as its component part.

Methods and Algorithms: RAPD markers were used for genetic characterization of Byelorussian carp. These markers make it possible to reveal a high level of DNA polymorphism and to analyze a large part of genome. The reference-information system is based on MS Excel platform, giving an opportunity to use graphic, table and text information in a file, as well as to organize retrieval system.

Results: The reference-information system contains: (1) description of DNA extration method, (2) description of RAPD analysis method, (3) example of its use for genotyping Lakhvinsky and Tremliansky carp races (both common and mirror carp offsprings), and (4) table for comparison of spectra and frequencies of primer #21 amplicons for certifying these races. This system can be updated by results of genotyping other races and offsprings of Belarusian carp.

Conclusion: The reference-information system proposed contains molecular genetic certificates for Lakhvinsky and Tremliansky carp races of Belarusian breeding. We propose to include the following information into race certificate: (1) primer (name, sequence) with amplicons effective for race certification. (2) molecular weight of PCR fragments(s) as a genetic marker for the race (in base pair), (3) image of gel with amplification products for primer used containing marks of sites with race molecular markers (the latter can simplify comparison of results when testing other individuals).

Availability: The reference-information system let us visualize main race characteristics for markers used as well as offers scope for updating additional data for other carp races or molecular markers.

Refereces:

1. R.I.Ludanny, G.G.Chrisanfova, V.A.Vasilyev et al. (2006) Genetic diversity and differentiation of Russian common carp (*Cyprinus carpio* L.) breeds inferred from RAPD markers, *Genetika,* **42:** 1121-1129.

SUPERVISED LEARNING AND PREDICTION OF PHYSICAL INTERACTIONS BETWEEN HUMAN AND HIV PROTEINS

M.D. Dyer[1*], T.M. Murali[2], and B.W. Sobral[1]
[1]Virginia Bioinformatics Institute and [2]Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA
E-mail: Matt.Dyer@LifeTech.com
**Corresponding author (current address Life Technologies, Carlsbad, CA 92008)*

Key words: *Host-pathogen interactions, protein interaction networks*

Motivation and Aim: Infectious diseases cause millions of deaths each year. Despite enormous effort, mechanisms of infection still remain poorly understood. A potentially powerful application of protein-protein interaction (PPI) networks lies in using them to obtain insights into the molecular mechanisms underlying infectious diseases, especially since interactions between pathogen proteins and host proteins play key roles in initiating and sustaining infection. However, the severe lack of large-scale datasets detailing interactions between host and pathogen proteins is a significant hurdle to progress in host-pathogen systems biology.

Methods and Algorithms: While a number of methods have been proposed for predicting PPIs, they have primarily focused on intra-species PPIs. Here we present a supervised predictor for human-HIV PPIs. We obtain known human-HIV PPIs from a number of small-scale experiments and from manually curated data. We use these data to train a Support Vector Machine (SVM) classifier using different combinations of features, including domain profiles, frequencies of protein sequence *k*-mers, and network characteristics of the human interactors in a human PPI network. We compare the performance of an SVM with a linear kernel on different combinations of features.

Results: We find that using a combination of protein sequence four-mers, protein domains, and PPI network information achieves the best performance, with precision greater than 70% for recall greater than 40%. We use this predictor to identify potentially novel viral interacting partners for human proteins. We focus our attention on those human proteins that are known to play an important role in HIV infection. Many predicted interactions involving these human proteins have considerable support in the literature. These interactions illustrate how the virus has evolved to manipulate host cellular processes during pathogenesis. For example, predicted interactions with human cell surface proteins are known to play a critical in the initial invasion of the cell and subsequent movement of viral material across the nuclear membrane.

Conclusion: It becomes imperative to develop computational methods that can robustly and accurately predict host-pathogen PPIs. Such predictors can guide cost effective experimental strategies to detect host-pathogen PPIs, drive research on how pathogens infect host cells, and help identify potential targets for therapeutics. We have developed such a method and show that we are able to make many predictions which are directly supported by the literature.

Availability: SVM models and predicted interactions are available upon request.

COMPUTER SIMULATION OF ORIGIN AND EVOLUTION OF SIGNALLING SYSTEMS

N.N. Dygalo*, S.A. Lashin
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: dygalo@bionet.nsc.ru
**Corresponding author*

Key words: *computer model, evolution, reproduction, hormones, gonads, thyroid*

Motivation and Aim: Origin of neuroendocrine systems in evolution relay on the signaling functions of molecules produced by specific cells and glands, and each of these molecules "codes" some conditions of internal milieu and environment for the cell and the whole organism[1]. The aim of the work was to create a simulation model which imitates origin of signaling systems that are capable of coordinating reproduction of a subject with its ability to perform successfully reproduction and to produce optimal number of offspring.

Methods and Algorithms: The control populations consist of objects that acquire some random amount of resources and produce genetically determined number of offspring (n) if their resources are larger than (R)x(n), (R – minimal resource of a viable descendent, each of which gets an equal part of the parental resource). Otherwise, if the object's resource is less than (R)x(n), it has now descendents. Total amount of the resource randomly distributed in each iteration (generation) among the objects remains constant.

Results: By varying the initial number the objects, genetic level of their fecundity, the flow of resource and inequality of its distribution among the objects, we have selected the parameters leading to extinction of the model control populations within 500-600 iterations and use these parameters in all subsequent experiments. Objects in experimental populations were allowed to obtain randomly two types of mutations "a" and "b" separately or both at a low probability level (0.001). The a-mutation imitates the signaling function of the thyroid hormone and confers to the carrier the ability to bypass one round of reproduction if it's resource is less than (R)x(n). Mutation "b" imitates effect of sex steroids and confers to the carrier the ability to produce variable number of offspring if this carrier has resources more then (R)x2. The number of b-mutant descendants is determined as integer of (b-mutant's resource)/R. Each a- or b-mutants totally replaced wild-type subjects in all populations in which these mutations were allowed within several tenth of iterations, and 1/3 and 1/2 of these populations respectively were still present by the end of experiments lasting 3000 iterations. Double a/b-mutants replaced all other genotypes from the populations within first hundreds of iterations. All populations of the a/b-mutants exist with a maximum number of objects in a generation as compared to populations of other genotypes throughout the whole experiment.

Conclusion: The use of "coding" molecules for adjusting the organism's reproduction to it's ability to perform this process successfully in accordance with conditions of internal milieu and environment may be under strong pressure of natural selection which evidently "creates" hormones and hormonal systems during early stapes of metazoan evolution.

This work was supported by RFBR grant N 08-0400292.

Дыгало Н.Н. Эволюция гормонов, нейротрансмиттеров, тканевых факторов и их рецепторов. Новосибирск, НГУ, 2009, 112 стр.

REDUCTION OF GENE NET DYNAMIC MODELS USING PROPER ORTHOGONAL DECOMPOSITION

V.M. Efimov*, A.S. Novikov, A.A. Tikhonov, I.R. Akberdin, V.A. Likhoshvai
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: efimov@bionet.nsc.ru
**Corresponding author*

Key words: *plant growth, auxin, gene net, differential equations, model reduction, proper orthogonal decomposition.*

Motivation and Aim: One of the problem in analysis of gene net differential equation models is a large amounts of high-dimensional data (tens and hundreds variables and parameters as well as hundreds and thousands time points). But real model dimension may be much lower as a result of intrinsic system regularity. One may well do a regularity detection with the aid of proper orthogonal decomposition (POD). The POD-transformation is the way of the reduction of system differential equations to the one with much less dimension. Such system may be more simple than initial. POD is rare used in biology, but is usual practice in technical sciences [1-2].

Methods and Algorithms: The essence of POD is the singular value decomposition (SVD) of matrice "time points – variables", i.e. trajectory of dynamic system in multidimensional space, and truncation of axes (factors) with small or null variations. The remain factors are a linear combinations on the whole set of variables and are the new integral variables describing the same dynamic system with minimal waste of information. The reverse is true also, source variables are the linear combinations of new factors. The substitution of these combinations in the source differential equations result to a system of differential equations via new factors.

Results: We investigated a family of the trajectories of the plant root growth under the auxin influence gene net dynamic model with 30 variables and different sets of initial parameters. The first factor had 68% total variation, four first factors – 80% and the ten first factors – 99%. All source variables made basic contributions in first factor except "chloroplast cytolasm cell membrane TOP" and "peroxisome cytolasm cell membrane TOP". These variables made basic contributions in the next three factors. Thus, ten integral variables quite enough for this system formal description. Nevertheless, a biological interpretation of these factors is necessary. It is a difficult but interesting task.

Conclusion: POD-transformation of gene net differential equation models is new perspective tool for such systems studies. Further investigations in this direction are need.

Acknowledgements: Work was supported by grants SS-65520.2010.4 and the Programs of Basic Research of the RAS Presidium ''Molecular and Cell Biology (project №10.7)''.

References:

1. M. Rathinam, L.R.Petzold. (2003) A new look at proper orthogonal decomposition. *SIAM J. Numer. Anal.,* **41(5):** 1893–1925.
2. A.X.Wang, Y.Ch.Ma. (2009) An error estimate of the proper orthogonal decomposition in model reduction and data compression. *Numer. Meth. for Part. Diff. Eqns*, **25(4):** 972-989.

mtPHYL AS A TOOL FOR mtDNA ANALYSIS

N.P. Eltsov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: eltsovnp@bionet.nsc.ru

Key words: *mtDNA, phylogenetic reconstruction*

Motivation and Goals: Human mtDNA sequence variation is characterized by a high amount of homoplasy which does not always allow unequivocal phylogeny reconstruction by conventional methods.

Methods and Algorithms: We propose a novel maximum parsimony-based method for reconstruction of human mtDNA phylogeny. Briefly, the method consists of three successive stages: identification of potential homoplastic mutations; analysis of potential homoplastic mutations and identification of true homoplastic mutations; identification of back and parallel mutations.

Results: The method designed was implemented in mtPhyl. This software package allows rapid and comprehensive analysis oh human complete mtDNA sequences. Apart from maximum parsimony phylogenetic tree reconstruction it performs different types of searches; analyzes mutation features; exports list of particular mtDNAs mutations into Excel table; defines mitochondrial haplotype; calculates coalescence time of clusters; estimates the effect of natural selection; makes reference list and downloads human mtDNA complete sequences from GenBank.

Conclusion: mtPhyl represents a timely advance, since the advent of cheaper sequencing methods has generated an excess of sequence data, and there is an urgent need to perform their automatic analysis.

Availability: Demo version of mtPhyl is available from the authors upon request and at http://eltsov.org/mtphyl.aspxl.

A MODEL OF GENOME STRUCTURE

I.L. Erokhin
National Biotechnological Company, LLC, Moscow, Russia
e-mail: i.erokhin@inbox.ru

Key words: *genome structure, junk DNA*

The coding of transcription factors by functional genes which initiate the transcription of other genes can be considered as the possibility to pass control between the promoters of these genes. It is possible to create complex structures, such as programs, subroutines, cycles and nested loops, with the help of different control passing algorithms and to use these structures to model the functioning of a genome.

1. A genome model of a multicellular organism presented here is Dijkstra's nested loop whose outer shell consists of a set of calling modules (steps) and a set of subroutines to be called (cell programs), and whose inner part (each cell program) is comprised of a set of functional genes and a set of structural genes activated by these functional genes.

2. The set of calling modules (loop steps) is arranged in the form of a binary tree with partially looped leaves. Each set of functional genes which corresponds to a specific cell program is arranged in the form of a network.

3. Outer shell branches of Dijkstra's nested loop are presented by 'step - cell program' pairs. Inner part branches of the loop are presented by 'functional gene – set of structural genes' pairs.

4. A calling module (step) is the first one which takes control from the parent cell, calls a specified subroutine (cell program) only for itself (and, therefore, for the given cell), takes control back from this subroutine and passes control to calling modules (steps) of two daughter cells during the process of cell division.

5. Functional genes of a cell program pass control to each other in the order defined by the network, by using at the same time structural genes in the defined order.

6. Different cells and sets of their generations may use the same steps and cell programs due to the presence of additional connections in the tree of steps and their partial looping.

The overwhelming part of genomes of the multicellular organism is presented by a set of steps, some part of genomes is presented by a set of cell programs, and a very small part of genomes is presented by a set of structural genes, housekeeping genes and their gene networks.

Only one specific step is active in each cell of the multicellular organism. Other steps in this cell are not performed and look like junk DNA. Among all available cell programs, only one program is carried out in a given cell, others are not active and look like junk DNA. Only active expression of different structural genes and housekeeping genes can be well defined by existing methods, and this expression formally comes to the foreground.

A more accurate, short and point transcription of functional genes which is responsible for passing control between steps of the loop and inside the networks remains unnoticed, despite the fact that this part of controlling effects is dominant for the development and shaping of the multicellular organism.

STRUCTURAL AND DYNAMIC CHARACTERIZATION OF TRYP6 FROM LEISHMANIA MAJOR (MRHO/IR/75/ER)

Eslami G*, Frikha F, Salehi R
1-The Department of Parasitology and Mycology, Shahid Sadoughi University of Yazd Medical Sciences, Yazd, 8916188/35, Iran
2-Laboratoire de Biochimie et de Genie Enzymatique des Lipases, Ecole Nationale d'Ingenieurs de Sfax, Route Soukra Km 3.5 B.P W 1173 Sfax, Tunisia
3-The Department of Genetics and Molecular Biology, Faculty of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran
e-mail: eslami_g2000@yahoo.com, eslami_g@ssu.ac.ir
**Corresponding author*

Key words: *Tryparedoxin peroxidase, L. major, homology modelling, molecular dynamics*

Motivation and Aim: Tryparedoxin peroxidase is one of the most important molecules in Leishmania viability and then, it may be a good target for challenging against Leishmaniasis. Recently, we reported LmTRYP6 from L. major (MRHO/IR/75/ER). Homology modeling of the LmTRYP6 was proposed to predict some functional proprety of this protein.

Results: The refined model showed that the core structure consists of a 7 Я stranded Яsheet and 5 a helices which are organized as a central 7-stranded Я2-Я1-Я5-Я4-Я3-Я6Я7 surrounded by 2-stranded Я-hairpin, a helices A and D on one side, and a helices B, C and E on the other side. The peroxidatic active site is located in a pocket formed by the residue Pro45, Met46, Thr49, Val51, Cys52, Arg128, Met147 and Pro 148. The catalytic Cys52, located in the first turn of helix aB, is in van der Waals with a Pro45, a Thr49 and an Arg128 that are absolutely conserved in all known Prx sequences. In this study, an attractive molecular target was studied.

Conclusion: Our results not only provide fundamental knowledge and mechanisms of L. major successful colonization of human macrophages but also could facilitate target molecule selection vaccine development and drug designing against this important and wide spread human pathogen.

WHEN ENVIRONMENTAL CHANGES DO NOT CAUSE GEOGRAPHIC SEPARATION OF FAUNA: DIFFERENT DEMOGRAPHIC RESPONSES OF BAIKALIAN INVERTEBRATES

V. Fazalova* [1,2], B. Nevado[3,4], T. Peretolchina[1], Z. Kuzmenkova[1], D. Sherbakov[1,5]
[1]*Laboratory of Molecular Systematics, Limnological Institute SB RAS, Ulan-Batorskaya 3, 664033 Irkutsk, Russia;* [2]*Evolution and Ecology Program, International Institute for Applied Systems Analysis, Schlossplatz 1, A-2361 Laxenburg, Austria;* [3]*Vertebrate Department, Royal Belgian Institute of Natural Sciences, Vautierstraat 29, 1000 Brussels, Belgium;* [4]*Evolutionary Ecology Group, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerpen, Belgium,* [5]*Faculty of Biology and Soil Science, Irkutsk State University, Sukhe-Batora 5, 664003 Irkutsk, Russia*
*e-mail: fazalova@iiasa.ac.at *Corresponding author*
Keywords: Lake Baikal; Baicalia carinata; B. carinatocostata; B. turriformis; Maackia herderiana; Gmelinoides fasciatus; mtDNA CO1; phylogeography; demographic reconstructions; global climate change; sedimentary cores;

While the impact of climatic fluctuations on the demographic histories of species caused by changes in habitat availability is well studied, populations of species from systems without geographic isolation have received comparatively little attention. Using COI mitochondrial sequences, we analyse phylogeographic patterns and demographic histories of populations of five species co-occuring in the southwestern shore of Lake Baikal, an area where environmental oscillations have not resulted in geographical isolation of habitats. The five species analysed include four gastropod species (rock-dwellers *Maackia herderiana* and *Baicalia turriformis* and sand-dwellers *B. carinata* and *B. carinatocostata*) and one amphipod species (ecologically flexible *Gmelinoides fasciatus*). We found that population sizes of species with strong habitat preference (*B. turriformis*, *B. carinata* and *B. carinatocostata*) remained rather stable through their evolutionary history. Conversely, ecologically flexible *M. herderiana* and *G. fasciatus* were found to have expanded their population sizes starting about 30-50 thousand years ago, a period marked by an increase in diatom abundance as inferred from bottom-lake sedimentary cores. In agreement with previous studies, we find that ecological and life history traits do not predict specific demographic histories. Instead, our results suggest that dispersal abilities and degree of ecological specialization play a significant role in the response of species to environmental changes.

THE APPROACH OF BIOINFORMATICS FOR A DEFINITION OF PHYTOPLANKTON INDICATOR SPECIES

L.L. Frolova, S.S. Firsova*
Kazan State University, Kazan, Russia
e-mail: firsovass@gmail.com

Key words: *phytoplankton, indicators, saprobity, bioinformatics analysis*

Motivation and Aim: Currently used in reservoirs the bioindication of saprobity system claim to be a fast and capacious estimation of reservoir condition depending on the correlation of abundances of different indicator organisms' species. However contradictions in definition of indicator species complicate interpretation of the received results and lead to a different estimation of saprobity the same reservoir such as from pure to dirty. At the same time many authors contend that organisms have single nonspecific reaction in phylogenesis on any external influence. So, research of relationship and evolution of indicator species can help in correct estimation the indicator importance of species and its correct use.

The aim of the work is research the evolutionary relationships of phytoplankton indicator species by methods of the bioinformatics and molecular genetic analyses.

Methods and Algorithms: Our work was based on the nucleotide sequences of 18S rRNA gene of indicator phytoplankton species from V.Sladecek's classification (1973). Nucleotide sequences of 18S rRNA were obtained from GenBank database (http://www.ncbi.nlm.nih.gov). For the analysis of sequences was used software as BLASTn (http://www.ncbi.nlm.nih.gov), ClustalW (http://www.ebi.ac.uk/Tools/ clustalw2), Phylip (parsimony method).

Results: The molecular phylogenetic tree is reconstructed for 109 species of a phytoplankton. The phylogenetic analysis on 18S rRNA gene has shown 15 clusters with the high bootstrap containing indicator organisms of the same saprobity reservoirs zone. Three clusters contain indicator organisms of *oligo* - and *xeno*saprobic reservoirs, four clusters- *oligo* - and *beta*-mesosaprobic reservoirs, five clusters - *beta*- and *alfa*-mesosaprobic reservoirs, three clusters – *beta*-mesosaprobic reservoir conditions. For example first two clusters are formed by indicators of *beta*-saprobity reservoir conditions: *Tetraedron caudatum, Pediastrum tetras, Tetraedron minimum, Polyedriopsis spinulosa* with bootstrap 71.6 and *Strombomonas verrucosa, Strombomonas acuminate, Trachelomonas volvocina, Trachelomonas oblonga* with bootstrap 100. Results of the phylogenetic analysis based on a gene 18S rRNA allow to assume about existence of system of indicator organisms grouping by various saprobity fresh water reservoirs zones.

Conclusion: It is possible to use the method based on the bioinformatics analysis as an effective and fast way for definition of stable indicators of saprobity reservoirs from existing lists and less labour-consuming way for definition of new indicator organisms.

Availability: Used software and databases available free of charge in on-line and off-line mode.

MODELLING LONG-TERM EVOLUTION OF THE GEOBIOSPHERE

S. Franck*, W. von Bloh, C. Bounama
Potsdam Institute for Climate Impact Research, Potsdam, Germany
e-mail: franck@pik-potsdam.de
**Corresponding author*

Key words: *geobiosphere, evolutionary biology, Darwinian adaptation*

Motivation and Aim: When in 6.5 billion years the sun develops from a main sequence star to a red giant, it will swell perhaps to the Earth's orbit and double its luminosity. By this time all life will have long since disappeared from our home planet. By 3.5 billion years from now, our central star – as the German astrophysicist Albrecht Unsöld prophesied as early as 1967 – will shine forty percent more brightly than at present. The Earth's surface will have reached the boiling point of water. By then at latest, no form of organic life can continue to exist on Earth.

Methods and Algorithms: Here we present a minimal model for the global carbon cycle of the Earth containing the reservoirs mantle, ocean floor, continental crust, biosphere, and the kerogen, as well as the combined ocean and atmosphere reservoir. The model is specified by introducing three different types of biosphere: procaryotes, eucaryotes, and complex multicellular life.

Results: During the entire existence of the biosphere procaryotes are always present. 2 Gyr ago eucaryotic life first appears. The emergence of complex multicellular life is connected with an explosive increase in biomass and a strong decrease in Cambrian global surface temperature at about 0.54 Gyr ago. In the long-term future the three types of biosphere will die out in reverse sequence of their appearance. We show that there is no evidence for an implosion-like extinction in contrast to the Cambrian explosion. In dependence of their temperature tolerance complex multicellular life and eucaryotes become extinct in about 0.8-1.2 Gyr and 1.3-1.5 Gyr, respectively. The ultimate life span of the biosphere is defined by the extinction of procaryotes in about 1.6 Gyr [1], [2].

Conclusion: An interesting aspect for the long-term future is this: while at the present time too much carbon dioxide due to the anthropogenic entrance accumulates in the atmosphere, in the long term there could well be an insufficiency of this gas, which plants need for photosynthesis. But even aside from carbon dioxide and its greenhouse effect, the Earth's atmosphere will ultimately warm mercilessly with the increasing luminosity of the sun. As an outlook for future investigations we present ideas how to include Darwinian adaptation into the model.

Availability: www.pik-potsdam.de/PLACES

References:

1. S.Franck et al. (2006) Causes and timing of future biosphere extinctions, *Biogeosciences,* **3:** 85-93.
2. S.Franck et al. (2008) The fate of planet Earth, *Z. geol. Wiss. ,* **36:** 115-122.

FRIQUENT REPEATS IN MAMMAL GENOMES AND ACTIVE RETROPOSONES

Fridman M.V.[1]*, Kulakovskiy I.V.[1], Oparina N.J.[2], Makeev V.J.[1,2]
[1]*Institute of genetics and selection of industrial microorganisms, GosNIIgenetika, Moscow, Russia*
[2]*Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*
** Corresponding author: e-mail: marina-free@mail.ru*

Motivation and Aim: Microsatellites are one of the most dynamic class of genome sequences responsible for many functionally important types of genome variability [1,2]. Our objective was to find interspersed repeats in the vicinity of most frequent types of microsatellites for some mammal genomes.

Methods and Algorithms: We used UCSC GenomeBrowser and Galaxy, RepeatMasker Web Server and algorithm **Chipmunk**.

Results: We find Alu-like consensus in the vicinity of all most frequent types of microsatellites with period 4 in human and mouse genome. For some types we find L-like consensus as well. We obtained also an unknown, Y-specific element in mouse genome.

We investigate cow genome. Microsatellites with period 5, not 4, and very peculiar sequences prevail in it. We find consensuses for specific cow retroposones in the vicinity of such microsatellites.

Conclusion: Algorithm **Chipmunk** allow everyone to reveal consensus of sequences, which are tied together with other sequences. Interspersed repeats, which are tied together with microsatellites, are most abandon in genomes among all, and we suppose that these most active repeats generate satellites during integration. For some classes of microsatellites $(BAAA)_n$ we can propose mechanisms of generation. For others they remain unclear.

Acknowledgements: This study was partially supported with Russian Fund of Basic Research Grant # 07-04-01584-a

References:

1. S.S.Arcot (1995) Alu repeats: a source for the genesis of primate microsatellites. Genomics, 29: 136-144.
2. M.A.Batzer, P.L. Deininger (2002) Alu repeats and human genomic diversity. Nature, 370: 370-389.
3. J. Jurka (1997) Sequence patterns indicate an enzymatic involvement in integration of mammalia retroposons. Proc. Natl. Acad. Sci. USA, 94: 1872-1877.
4. Nadir E. et al (1996) Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. Proc. Natl. Acad. Sci. USA, 93: 6370-6475.
5. Toda Y. et al (2000) Characteristic sequence pattern in the 5- to 20-bp upstream region of primate Alu elements. Mol. Evol., 50: 232-237.
6. Kulakovskiy I.V., Makeev V.U. (2009) Integration of data obtained by different experimental methods to determine the motifs in DNA sequences recognized by transcription-regulating factors (rus.). Biofizika, 54 (6): 965-974.

A NOVEL APPROACH FOR CREATION OF COMPLEX COMPUTATIONAL EXPERIMENTS IN BIOINFORMATICS.

M. Fursov*, A. Varlamov
Novosibirsk Center of Information Technologies 'UniPro' e-mail: mfursov@unipro.ru
**Corresponding author*

Motivation and Aim:

Bioinformatics methods play the dominant role in system biology today. During the past decade there was developed number of unique methods in bioinformatics that played vital roles in technological and scientific breakthroughs. Integrated bioinformatics solutions that combine different statistics, algorithms, automated methods of data retrieval in areas of genomics, transcriptomics and proteomics becomes more valuable today than ever before.

Such systems allow to researcher focusing on an experiment and use more complex computational schemas than it was possible when using separate programs and datasets.

To join different bioinformatics methods into a complex computational workflow there is a need in a new domain specific language that will be easy to use and powerful enough at the same time.

Results, Conclusion and Availability:

We have designed and implemented an interactive software solution called Workflow Designer. The key idea of Workflow Designer is to make the process of automation of routine tasks as simple as possible and make it available and understood by non-programmers.

A user of Workflow Designer draws a computational scheme from a predefined elements or computational blocks. Each block contains a complete textual description of what will be done during its execution. The description is automatically adapted to the parameters and environment this computational block is used in. The final diagram can be read and understood as a usual text by anyone who familiar with bioinformatics methods used in the computation.

Workflow Designer is a part of UGENE Integrated bioinformatics suite [1] and supports all computational methods and data retrieval options available in UGENE. Also Workflow Designer allows running computational workflows on remote computers that makes possible to utilize high performance computing facilities in very simple and convenient way.

A ready to use version of the software as well as the complete source code is freely available under the GPL license from the UGENE web site or as a part of major Linux distributions. The solution is available for Linux, Windows and MacOS X platforms.

References:

1. 1. UGENE web site: http://ugene.unipro.ru

POLYMORPHISM OF GENES GLUTATIONETRANSFERASES AT MINERS OF KUZBASS, ILL A CHRONIC DUST BRONCHITIS

N.I. Gafarov*, V.V. Zaharenkov, N.I. Panyov, T.K. Jadykina,
A.S. Kazitskaja, V.P. Puzyryov+, A.A. Frejdin+

Establishment of the Russian academy of medical sciences a scientific research institute of complex problems of hygiene and occupational diseases of the Siberian branch of the Russian academy of medical sciences Novokuznetsk, Russia, Establishment of the Russian academy of medical sciences, a scientific research institute of medical genetics of the Siberian branch of the Russian academy of medical sciences, Tomsk +
Email: ecologia_nie@mail.ru

**Corresponding author*

Key words: *glutationetransferases GSTT1, GSTM1, chronic dust bronchitis*

Motivation and Aim. In structure of professional diseases of workers of the mining and coal industry of Kuzbass of illness of bodies of breath occupy one of conducting places, and duly diagnostics of occupational diseases cost very sharply. Searches of associations of genetic markers with diseases are rather actual, as their research allows revealing among the population of group of persons of the increased risk, and it enables development of measures of duly preventive maintenance.

Methods and Algorithms. In the present work results of research of GSTT1 and GSTM1 genetic polymorphism at coal miners of the south of Kuzbass, ill are submitted by chronic dust bronchitis. It is surveyed 126 patients and 28 healthy persons working in the same conditions, as ill, but not having occupational disease. The method polymerization chain reaction studied polymorphism of the genes GSTT1 and GSTM1.

*Results. F*requency of phenotypes GSTT1 (0) and GSTM1 (0) at ill and healthy persons is practically identical, though in the senior age group of ill persons (50-59 years) frequency GSTT1 (0) is twice lower, than at patients in the age of 40-49 years (14,95 % against 7,95 %). Frequency double homozygotes GSTT1 (0)/GSTM1 (0) at ill persons is lower, than at healthy (3, 97 % against 7, 14 %). With increase of age and the experience of work frequency double homozygotes GSTT1 (0) / GSTM1 (0) at ill persons is reduced from 6, 25 % up to 2, 56 % that can specify selection against carriers such homozygotes. Probably, it is connected to selection with selection against homozygotes in adverse conditions of a underground coal mining.

*Conclusion.*Thus, the carried out research has shown, that though frequency both separate homozygotes on zero-allele, and at patients with a chronic dust bronchitis as a whole is lower than their combinations, than at healthy persons, with increase of age and the experience of work frequency double “zero” homozygotes is reduced more than in 2 times. It can specify rigid selection against carriers such homozygotes, as causes their lower frequency at ill persons.

Availability. The received results can find application at a substantiation of criteria of individual risk of development of the professional and professionally caused diseases, the decision of other questions scientifically argued preventive maintenance.

AUTOMATIC GENERATION AND NUMERICAL ANALYSIS OF MATHEMATICAL MODELS FOR MOLECULAR-GENETIC OBJECTS IN AN INTEGRATED SYSTEM OF MGS-GENERATOR AND STEP+ MODULES

I.R. Akberdin[1], F.V. Kazantsev[1], V.A. Likhoshvai[1,3], S.I. Fadeev[2,3], I.A. Gainova[3,*], V.K. Korolev[3], A.E. Medvedev[4]

[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*
[3] *Novosibirsk State University, Novosibirsk, Russia*
[4] *Khristianovich Institute of Theoretical and Applied Mechanics, SB RAS, Novosibirsk, Russia*
e-mail: gajnova@math.nsc.ru
[*] *Corresponding author*

Key words: *computer system, automatic generation, numerical analysis, molecular-genetic objects.*

Motivation and Aim: Mathematical models allow us to naturally integrate the multi-scaled experimental data and analyze the cause-effect relationship between the molecular structure, dynamics and phenotypic features of living systems in terms of the common conceptual scheme. The increasing volume of the accumulated data stimulates the development of computer technologies for their processing, storage and analysis. That is why the development of computer tools allowing one to reconstruct and analyze mathematical models for molecular-genetic objects is one of the key problems in the era of systems biology.

Methods and Algorithms: The paper describes a new computer system for generating and analyzing mathematical models of molecular-genetic objects, based on the integration of two program modules: MGSgenerator [1] and STEP+ [2]. MGSgenerator is designed for automatic generation of mathematical models on the basis of structural and functional organization of gene networks extracted from the GeneNet database [3]. STEP+ is a tool for numerical analysis of mathematical models presented as an autonomous system of ordinary differential equations.

Results: The computer system possibilities are demonstrated by the analysis of dynamic features of two gene networks that control auxin metabolism in a shoot meristem cell for higher and lower plants [4]. Two mathematical models were automatically generated by the MGSgenerator module based on the structural and functional organization of the corresponding gene networks. The results of the numerical analysis with the STEP+ module indicate a higher stability of the stationary intracellular concentration of auxin with respect to the model parameters for higher plants as compared with the lower ones.

Conclusion: Structural and functional organization of the program modules as well as the advantages and the results of testing the computer system are discussed in this paper.

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

References:

1. Kazantsev F.V. et al. (2008) MGSgenerator – the tool for automatical generation of molecular-genetic system mathematical models on basis of gene networks structure. *Proc. of the 6th Intern. Conf. on Bioinformatics of Genome Regulation and Structure*, 114.
2. Fadeev S.I. et al. (2006) The package Step+ for numerical study of autonomous systems arising when modeling dynamics of genetic-molecular systems. *Proc. of the 5th Intern. Conf. on Bioinformatics of Genome Regulation and Structure*. **2:** 118-120.

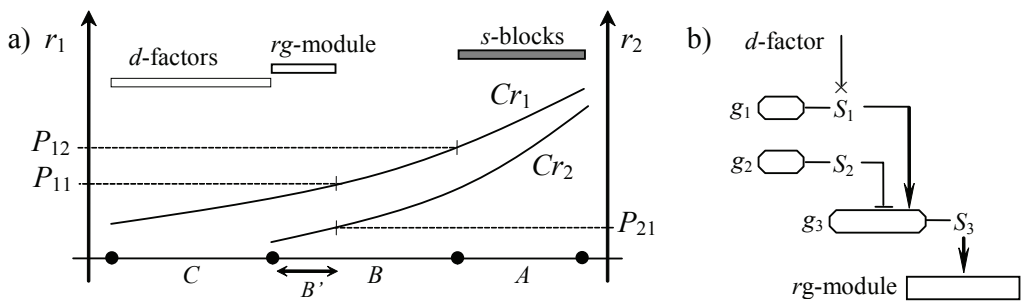
POSSIBLE STRUCTURE OF MOLECULAR-GENETIC SYSTEM ACTIVATING THE TRUNCATED TISSUE RECOVERY

A.V. Galimzyanov
Establishment of the Russian Academy of Sciences
Institute of Biology of the Ufa Research Centre of the RAS, Ufa, Russia
e-mail: galim@anrb.ru

Key words: *diffusion, thresholds, gene networks*

Motivation and Aim: Regeneration is the property inherent in living systems to restore the lost integrity. Tissues (skin), organs (liver), limbs (fins, legs) and organisms (hydra) can be restored. Ability to regeneration assumes existence of the integrity sensor and a competent subsystem to its signals, providing return to structural norm. In this work the qualitative model of formation of a signal to the regeneration, based on regulator substances diffusion and threshold activation of a gene expression is offered.

Results: The hypothetical three-layer cellular mass (HCM-3) with differentiated domains



A, *B* and *C* (Fig. 1) is considered.

Fig. 1. A regeneration context (a) and a gene network providing it (b). In norm Cr_1 and Cr_2 curves of S_1 and S_2 molecules distributions on three zones are stationary.

The block of synthesis of two regulatory proteins S_1 and S_2 functions in *A*-layer cells. The factor of degradation of protein S_1 (*d*-factor) acts in layer *C*. In layer *A* there is a raised concentration S_1, in layer *C* – lowered. As a result of diffusion S_1 flows from *A* in *C*. Zone *B* is a flowing for S_1, and concentration S_1 in it is below the level P_{12}. The area of diffusion of protein S_2 is limited to a *B*-layer. At concentration $r_1 \geq P_{11}$ protein S_1 activates the module of regeneration zones *C* (*rg*-module). Activation occurs in absence of repressors S_2 (at values $r_2 \geq P_{21}$ this protein blocks the activity of the *rg*-module). In norm the similar combination is excluded: in the field of $B' \cup C$ concentration S_1 is below threshold P_{11}, in the field of $A \cup (B \backslash B')$ concentration S_2 is above threshold P_{21}. Truncation of layer *C* leads to that in layer *B* there is a gradual accumulation S_1. At achievement of concentration S_1 of level P_{11} in the field of *B'* the *rg*-module – system of the lateral cell divisions, filling *C* is started. In cells *de novo* built zone *C* the factor of degradation of protein S_1 also is reproduced. The more inserts in *C*, the *d*-factor integrated action in it is more powerful so after some time concentration S_1 in zone *B'* falls, *rg*-functional is turned off, and the system settles into norm.

MODELING OF AMYLOID FIBRIL FORMATION

N.V. Dovidchenko, O.V. Galzitskaya*
Institute of Protein Research RAS, Pushchino, Russia
e-mail: ogalzit@vega.protres.ru

Key words: *amyloid fibril, kinetics, nucleation*

Motivation and Aim: The formation of amyloid fibril has nucleation nature, however it is not clear what processes take priority of the nuclear formation. Earlier, it was mostly believed that for the amyloid formation the protein should be in the unfolded or partly unfolded states. The last experimental data demonstrate that proteins are able to form amyloid fibrils even in native conditions [1]. Our aim was to build a mathematical model of amyloid fibril growth taking into account the last experimental data.

Methods and Algorithms: We constructed a kinetic scheme describing the involvement of monomers in different states: the first stage is transferring of molecules from the pool of the native state to the pool of the native-like state, after that the dynamic equilibrium is established. Then, after accumulation of some amount of monomers in the native-like state, oligomers are formed. Oligomeric particles are able to increase or decrease in size due to inclusion or dissociation of monomers from the oligomers. Oligomeric particles are unstable formations and their instability is maximal when the oligomers become critical in size. However, such critical oligomers can be rearranged in beta-structure oligomers which are already stable and present amyloid fibrils of minimal size. Further growth of amyloid fibrils occurs due to inclusion of native-like monomers into the amyloid fibrils.

Results: A mathematical model of amyloid fibril growth taking into account the last experimental data has been composed. Computer simulation demonstrated that the model appropriately describes the corresponding stages of amyloid formation.

Acknowledgements:

This work was supported by the programs “Molecular and Cellular Biology” (01200959110) and “Fundamental Sciences to Medicine”, by the Russian Foundation for Basic Research (08-04-00561), and by the “Russian Science Support Foundation”.

References:

1. F.Chiti, C.M.Dobson (2009) Amyloid formation by globular proteins under native conditions, *Nature Chemical Biology*, **5**:15-22.

RNA FOLDING NUCLEI PREDICTION

L.B. Pereyaslavets, M.V. Baranov, E.I. Leonova, O.V. Galzitskaya*
Institute of Protein Research, RAS, Pushchino, Russia
e-mail: ogalzit@vega.protres.ru
**Corresponding author*

Key words: *RNA, folding nuclei, folding simulation*

Motivation and Aim: Prediction of RNA folding process and its features such as height of transition state barrier that is connected with mid-transition rate of the folding/unfolding processes, or nucleotides included in folding nuclei is an important task for solving up-to-date molecular biology problems. This is connected with the folding process for recently discovered RNA sequences, stability of RNA structure and even its dependence on mistranslated-induced misfolding and other applications.

Methods and Algorithms: We take into account the method elaborated by us for prediction of protein folding nuclei [1] and adopt it to RNA folding. Energy *E(I)* of intermediate states of RNA molecule including the native state and its transition states obtained from RNA folding/unfolding process has been calculated with the help of the empirical potentials (coarse grained [2] or all-atomic [3]), which implicitly take into account RNA-water interactions. We simulate unfolding and folding processes by the dynamic programming method at mid-transition point considering that the free energy of an intermediate state with a partly unfolded RNA chain is equal to $F(I)=E(I)-T[\sigma N_{free.nucl}+S_{loops}]$, where σ is the entropy parameter of one free nucleotide in the coil state (see details in [1]). This parameter can be calculated from the equality of the free energies of native and coil states at the point of mid-transition.

Results: To estimate folding nuclei in RNA structures, we have calculated the values for each nucleotide which are proposed to be identical to meaning of protein Φ-values. We have estimated the folding nuclei for several RNA structures with known three dimensional structures. Although, there is no experimental data which strictly describe folding nuclei for RNA molecules, our results are in agreement with indirect experimental data for tRNA and ribozyme folding process. Our method also allows us to estimate the mid-transition rate of the folding/unfolding process.

Acknowledgements:

This work was supported by the programs “Molecular and Cellular Biology” (01200959110) and “Fundamental Sciences to Medicine”, by the Russian Foundation for Basic Research (08-04-00561), and by the “Russian Science Support Foundation”.

References:

1. O.V.Galzitskaya, A.V.Finkelstein (1999) A theoretical search for folding/unfolding nuclei in three-dimensional protein structures, *Proc. Natl. Acad. Sci. USA*, **96**:11299-11304.
2. F.Ding et al. (2008) Ab initio RNA folding by discrete molecular dynamics: From structure prediction to folding mechanisms, *RNA*, **14**: 1164-1173.
3. L.B.Pereyaslavets, A.V.Finkelstein (2010) Atomic force field FFSol for calculating molecular interactions in water environment, *Molecular Biology*, **44**:303-316.

WHEATPGE — SYSTEM FOR ANALYSIS OF THE RELATIONSHIPS BETWEEN PHENOTYPE, GENOTYPE AND ENVIRONMENT IN WHEAT

M.A. Genaev[1], A.V. Doroshkov[1], D.A. Afonnikov[1,2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
e-mail: mag@bionet.nsc.ru
[*]*Corresponding author*

Key words: *database, wheat, genotype-phenotype-environment relationships*

Motivation and Aim: New high throughput technologies in genome sequencing/genotyping and modern methods of phenotyping in plants resulted in the possibility of producing large amount of genotype and phenotype data [1,2]. This allows applying statistical approaches to study efficiently genotype-phenotype relationships. The challenge is to maintain large amount of such data, keep this data consistent and useful for analysis. In this work, we developed a WheatPGE system, the web-application for storing and processing of various morphological characteristics, genotype of the wheat plants and various environmental factors.

Methods and Algorithms: The project implemented using Catalyst framework that uses three different components (Model-View-Controller) for the presentation, visualization and processing of user actions. The database server implemented using MySQL. The ORM technology (Object-relational mapping) used for linking database with Catalyst framework. *Results:* As a result, we implemented complex data to describe biological objects; for example, "genotype" object consists of more than 10 related tables. An important feature of the system is the possibility of the automatic extension of the database schema and creating new models of data. This allowed user to add new morphological features without help of a programmer. Current version of the database contains information about 400 plants belonging to more than 100 cultivars.

Conclusion: The WheatPGE system allows analyzing the relationship between genetic and phenotypic traits of plants, as well as environmental conditions.

Availability: The WheatPGE system is available at *www.wheatdb.org*

Acknowledgements: This work supported by SB RAS Integration project 113, RAS programs №22 (project 8) and "Biosphere origin and evolution".

References:

1. Yan Lu, et al. (2008) New Connections across Pathways and Cellular Processes: Industrialized Mutant Screening Reveals Novel Associations between Diverse Phenotypes in Arabidopsis, *Plant Physiology*, **146**: D1482-D1500.
2. Raju Naik Vankadavath, et al. (2009) Computer aided data acquisition tool for high-throughput phenotyping of plant populations, *Plant Methods*, **5**, 18.

BioinfoWF — WEB SERVICES AND WORKFLOW MANAGEMENT FOR BIOINFORMATICS ANALYSIS

M.A. Genaev*[1], D.A. Afonnikov[1,2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
e-mail: mag@bionet.nsc.ru
[*]*Corresponding author*

Key words: *bioinformatics, workflow, grid processing, XML*

Motivation and Aim: The analysis of biological data in bioinformatics usually consists of several steps performed by different programs subsequently. During the analysis progress, the output of one calculation module serves as an input of the other module, etc. Thus, the overall procedure could be organized as a workflow [1, 2]. For example, the calculation of the phylogenetic tree for protein family requires protein sequence extraction from databases, multiple sequence alignment, phylogeny estimation. It should be noted, that most of single steps could be performed using different routines. For example, sequence alignment could be obtained using ClustalW, Mafft, Muscle or T-Coffee programs. The program's choice by user often depends on the data under analysis and the aim of the task.

Methods and Algorithms: To perform workflow data processing for bioinformatics we developed BioinfoWF system. It is written in Perl and based on the XML description of the program options, input and output data for a single step of the workflow. The second part of the system describes the workflow scheme, set the file data, the execution status of each step. The BioinfoWF runs under command line on the UNIX-like systems or as a web-service. The workflow or its part can also perform on the multiprocessor cluster systems under Sun Grid Engine.

Results: We used BioinfoWF to develop Computer System for Analysis of Molecular Evolution Modes of Protein Families (SAMEM) and functionally important SNP detection in the regulatory regions of eukaryotic genes.

Conclusion: The BioinfoWF can be used to organize workflow management for various bioinformatics tasks.

Availability: The BioinfoWF available at *http://pipeline.bionet.nsc.ru*.

Acknowledgements: This work supported by SB RAS Integration project 1, 26, 113, 119, RAS programs №22 (project 8) and "Biosphere origin and evolution".

References:

1. Rhos J., Karlsson J. and Trelles O. (2009) Magallanes: a web services discovery and automatic workflow composition tool, *BMC Bioinformatics*, **10**:334.
2. Oinn T., Addis M., Ferris J., Marvin D., Senger M., Greenwood M., Carver T., Glover K., Pocock M.R., Wipat A., Li P. (2004) Taverna: a tool for the composition and enactment of bioinformatics workflows, *Bioinformatics*, **20**:3045-3054.

PEFF DB: THE MANUALLY CURATED DATABASE OF PROTEIN EVOLUTIONAL AND FUNCTIONAL FEATURES

K.V. Gunbin[1], D.A. Afonnikov[1,2], M.A. Genaev*[,1]
[1]- *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]- *Novosibirsk State University, Novosibirsk, Russia*
e-mail: genkvg@bionet.nsc.ru
[*]*Corresponding author*

Key words: *database, mutations, molecular evolution, mathematical models robustness*

Motivation and Aim: It is well known that genes function and evolve as parts of genetic networks. To study relationships of their function and evolution an enormous amount of data on genes, proteins, and their networks accumulated in various databases. These data, however, lacks the information on network dynamics, that could be very important to study network response to gene mutations in the course of evolution. To integrate the gene evolutionary information and genetic networks dynamics we developed the PEFF database.

Methods and Algorithms: The database consists of three types of information: 1) experimental data on the phenotypic manifestations of the gene mutations extracted from literature; 2) data on the computer analysis of their molecular evolution; 3) data on genetic networks [1, 2] and their parametric robustness based on mathematical models deposited in BioModels database [3]. The evolutionary information includes: multiple alignments of orthologous gene sequences from the KEGG Orthology [2]; phylogenetic trees checked manually for errors; evolutionary changes of amino acid physico-chemical properties from AAindex database [4] in proteins; analysis of positive selection events on specific tree branches and at specific alignment positions by various computational methods. We also developed a computational tool for automatic inference of statistical relation between general or positively selected evolutionary changes of amino acid properties and user defined quantitative phenotypical feature. MySQL 5 was used for relational tables. Perl 5 scripts were used for linking the molecular evolution analysis programs.

Results: Current database version (February, 2010) stores information about key regulatory proteins of eukaryotic cell cycle (4 cyclin families): 92 records of cyclin mutations with corresponding references, robustness of 3 cell cycle mathematical models, 234 cyclin sequences, 12 ancestor reconstructions, 4 manually curated phylogenetic trees.

Conclusion: Thus, PEFF database can be useful for comprehensive analysis of structure-function-evolution relationship of genes functioning as parts of gene networks.

Availability: Available upon request.

Acknowledgements: This work was supported by RFBR grant No. 09-04-01641-a.

References:

1. E.A. Ananko et al. (2005) GeneNet in 2005, *Nucleic Acids Res,* 33: D425-D427.
2. M. Kanehisa et al. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs, *Nucleic Acids Res*, **38**: D355-D360.
3. N. Le Novere et al. (2006) BioModels Database: A Free, Centralized Database of Curated, Published, Quantitative Kinetic Models of Biochemical and Cellular Systems, *Nucleic Acids Res*, **34**: D689-D691.
4. S. Kawashima et al. (2008) AAindex: amino acid index database, progress report 2008, *Nucleic Acids Res*. **36**: D202-D205.

CHARACTERISTICS OF CARBOHYDRATE METABOLISM IN CHILDREN DURING CHUKOTKA ONTOGENETIC DEVELOPMENT

T.V. Godovykh
Northeastern State University, Magadan, Russia
e-mail: tgog@mail.ru

Key words: *ontogeny, carbohydrate metabolism, growth, children, Chukotka.*

Motivation and Aim: Patterns of carbohydrate metabolism boys in extreme conditions of the North in the process of ontogenetic development. *Methods and Algorithms:* The following substrates studied the blood serum of boys are in secondary and vocational educational institutions in Chukotka. Blood sampling was carried out in accordance with the requirements of medical technologies and bioethics. Analytical determination of the substrates was carried out in the laboratory of biological and inorganic chemistry department of Endocrinology of the Institute of Environmental Physiology of Natural Adaptations, Arkhangelsk. All children spent an anthropological survey of the definition of a component of body composition by conventional methods. We examined boys: 25 Aboriginal and 26 of the migrants of the second period of childhood, 50 Aboriginal and 55 migrant adolescents, 39 Aboriginal and 40 migrant adolescents. *Results:* In the process of ontogenetic development revealed a significant decrease in glucose levels (glucose) and stable high values of lactate (LAC) for boys, regardless of population. The level of glucose Aboriginal significantly lower boys at each interval of ontogeny, the values of LAC and pyruvate (PIR) had no differences across population groups. Status of hypoglycemia and hyperlaktatemii found more than half of children in almost every segment of ontogeny. Contents PIR below physiological norms migrants were 2 times more Aboriginal people (23% and 12% respectively). The level of ionized calcium (Ca) did not differ across population groups and segments of ontogeny, but the deviations in the downward normal values of migrants were 2 times more likely than upward. In the group of Aboriginal upward bias is more common than decrease. Mean values of phosphorus (P), Aboriginal were significantly higher in the second childhood and adolescence. Growth length (GL) and body weight (BW) of migrants is much higher than natives, indicating a higher energy metabolism Aborigines. The relative content of bone above the Aboriginal, despite the absolute predominance of basic anthropometric indicators of migrants. *Conclusion:* Peculiarities of carbohydrate metabolism boys Chukotka, regardless of population groups are to reduce energy metabolism, manifested hypoglycemia, the predominance of anaerobic over aerobic oxidation. Aborigines, unlike migrants have higher levels of energy processes, resulting in lower growth rates GL and BW, an increase in the level of P. Increased bone Aboriginal component of the body due to the predominance of hormones that regulate growth and carbohydrate metabolism, with the participation of Ca. Ensemble of hormones and especially the nervous regulation of metabolism determine the ergonomic system, different for each segment ontogeny and population groups.

BMP AND WNT GENE NETWORKS IN MURINE HAIR FOLLICLE HAVE CYCLIC ACTIVITY PATTERNS AFFECTED BY MUTANT GENES

P.K. Golovatenko-Abramov*[1,2], A.P. Nesterova[1], E.S. Platonov[1]
[1] *Vavilov Institute of General Genetics RAS, Moscow, Russia*
[2] *Ariadne Genomics Inc., Rockville, USA*
e-mail: p.golovatenko@gmail.com
[*]*Corresponding author*

Key words: *hair follicle cycle, alopecia, gene networks, BMP, WNT, mutant genes*

Motivation and Aim: Morphogenetic cycle of hair follicle in mammalians is sophisticated self-regulated system that is controlled by coordinated activity of diverse signaling pathways and gene regulatory networks. Determining gene networks necessary for initiation and termination cyclic hair development is key task in this area of research. Ability of direct enhancement or suppression of hair growth depends on the solution of this task. Mutant alleles that cause various alterations in hair coat formation are convenient tool for exploring activity of gene regulatory networks. The aim of present study was to analyze changes in expression of BMP- and WNT-dependent genetic markers during hair cycle in skin of homozygotous mice carrying mutant alleles of *angora* (*Fgf5*$^{go\text{-}Y}$), *hairless* (*Hr*^{hr}), *waved alopecia* (*wal*) and *wellhaarig* (*we*) genes in different combinations.

Methods and Algorithms: Hair cycle stages within 1st and 2nd hair generation in mice were determined using histological approaches. Expression of BMP-dependent genes *Bmp2*, *Id1*, *Sostdc1* and WNT-dependent gene *Mitf* in probes of skin of mice of various genotypes was evaluated using qRT-PCR approach (SYBR Green I/ROX staining, iCycler iQ machine, Bio-Rad). Pathway analysis was performed using ResNet Mammalian Database and Pathway Studio application (Ariadne Genomics Inc.)

Results: In skin of normal genotype mice, elevation of BMP-dependent *Sostdc1* gene mRNA expression entailed downregulation of WNT-dependent *Mitf* gene mRNA expression. In skin of *Hr*^{hr}/*Hr*^{hr} mutant mice, elevation of *Sostdc1* mRNA expression is not associated with downregulation of WNT-dependent gene network. In skin of wild-type mice, BMP gene network retains high activity level at the end of catagen/beginning of telogen stage, and in the middle of anagen stage of hair cycle. In mutant *we*/*we wal*/*wal* mice, expression of BMP-associated *Bmp2* and *Id1* genes was increased, as compared to wild-type mice probes. Mutant *Fgf5*$^{go\text{-}Y}$ gene partially normalizes BMP gene network activity in *Fgf5*$^{go\text{-}Y}$/*Fgf5*$^{go\text{-}Y}$ *we*/*we wal*/*wal* genotype mice.

Conclusion: It is shown for the first time that beginning of anagen, catagen and telogen stages of hair cycle correlates with changes in transcriptional activity of *Bmp2* and *Id1* genes in skin of wild-type mice. Upregulation of BMP gene network leads to alterations of cyclic hair growth and progression of alopecia. Inhibition of WNT-dependent gene network by activity of BMP gene network is required for anagen-to-catagen switch.

Availability: ResNet Mammalian Database and Pathway Studio are the products of Ariadne Genomics Inc. and are available as commercial packages. Evaluation of the software is available upon request on Ariadne website: http://www.ariadnegenomics.com/

INTEGRATION OF CHEMICAL INFORMATION WITH PROTEIN SEQUENCES AND 3D STRUCTURE IN SQL

Golovin A.*, Henrick K., Kleywegt G.
EMBL-EBI/PDBe, Hinxton Hall, Genome Campus, Cambridge, Cams, CB10 1SD, UK
e-mail: golovin@ebi.ac.uk
**Corresponding author*

Key words: *3D structures of proteins, protein kinases, drug research, PDBeMotif*

Motivation and Aim: An understanding of the 3D structures of proteins and how they interact with each other and smaller molecules is important in the fields of medicine and drug research. Whilst there are well-developed tools and algorithms available to examine and perform comparative analysis on the amino acid sequence of proteins, tools for the study of more complex 3D interactions are few and far between. At this point in time PDBeMotif is the only tool that can execute integrated primary, secondary, tertiary and quaternary searches. It facilitates exploration of the Protein Data Bank (PDB) by combining protein sequence, chemical structure and 3D data in a single search. In drug research, protein kinases and GPCRs are important drug targets. PDBeMotif can be used to examine the characteristics of the binding sites of single proteins or classes of proteins such as these and the conserved structural features of their immediate environments either within the same specie or across different species. For example it can highlight a conserved activation loop common to all protein kinases, which is important in regulating kinase activity and is marked by conserved DFG and APE motifs at the start and end of the loop, respectively. The prediction of the effect of modifications to small molecules that bind to the active and/or regulatory sites of proteins on their efficacy can be based on the outcome of analytic work done using PDBeMotif.

Methods and Results: PDBeMotif is an extremely fast, integrated and powerful search tool. It has a finely tuned database architecture and custom algorithms that uses SQL statements to execute complex queries in seconds. It can be ported to all major operating system platforms such as MS Windows, LINUX, Apple Mac and Solaris as it is written in Java and uses Oracle and the free source PostGreSQL database server. PDBeMotif can be used online or downloaded and installed locally where public and private PDB files (including libraries of theoretically derived 3D structures) can be loaded and analyzed. There is also the capability to load protein site annotations, families and domains from Distributed Annotation System (DAS) servers.

Availability: The system is released under GPL and available with the source code from http://sourceforge.net/projects/pdbsam and on line at http://www.ebi.ac.uk/pdbe-site/PDBeMotif/

References:

1. Golovin A., Henrick K.: Chemical Substructure Search in SQL, J.Chem. Inf. Model., 2009, 49 (1), 22-27.
2. Golovin A, Henrick K: MSDmotif: exploring protein sites and motifs. BMC Bioinformatics 2008, 9:312.
3. Golovin A. et al. (2005) MSDsite: A Database Search and Retrieval System for the Analysis and Viewing of Bound Ligands and Active Sites. PROTEINS: Structure, Function, and Bioinformatics 58(1): 190-9.
4. Golovin A. et al. (2004) E-MSD: an integrated data resource for bioinformatics Nucleic Acids Research, 32 (Database issue), D211-D216.

REGULATORY ELEMENTS ANALYSIS INSIDE *VRN1* (VERNALIZATION GENE1) PROMOTER REGION.

K.A. Golovnina*, A.G. Blinov, N.P. Goncharov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: ksu@bionet.nsc.ru
**Corresponding author*

Key words: *vernalization, cold adaptability, wheat, transcription factors, regulatory cascade*

Motivation and Aim: In crops such as common wheat (*Triticum aestivum* L.), a vernalization requirement (a long exposure to low temperatures) distinguishes winter varieties from spring ones. Besides being an important trait for adaptation, the requirement of vernalization is also of a great agronomical importance. *VRN1* gene is dominant for spring growth habit and it is the main initiating factor of the flowering regulatory cascade. It has been demonstrated that damages in either intron or promoter regions are sufficient to accelerate flowering under LD (long day) condition. However less is known about functional significance of regulatory sites and consequently genetic network control vernalization response in wheat.

Methods and Algorithm: Molecular biology methods used in the present study included total DNA isolation, primer design, PCR amplification, cloning, DNA sequencing. For bioinformatic research UniPro Ugene software together with the integrated plugins: Sitecon for TF site recognition, MUSCLE 3,4 and KAlign algorithms for alignment analysis, HMM profiles for transposable elements search was used (http://ugene.unipro.ru). For transcription factors analysis tools of NCBI (http://www.ncbi.nlm.nih.gov), EMBL (http://www.ebi.ac.uk/embl), DBD (http://www.transcriptionfactor.org), Jaspar (http://jaspar.cgb.ki.se) and other databases were used.

Results: Variability of the *VRN1* promoter region of the unique collection of spring polyploid and wild diploid wheat species was investigated. Sequence analysis indicated great variability in the region from -62 to -221 nucleotide positions of the *VRN1* promoter region. Different indels were found within this region in spring wheats. In the present work we carried out searching and analysis of the cis-regulatory elements (recognition sites for transcription factors, TF) inside of the *VRN1* promoter and its homologs in crops and Arabidopsis. Such information helps to investigate the possible consequences of the observed damages in the promoter of the main initiation factor of flowering, and possibly is related to the mechanism of growth habit changing. Some transcription factor recognition sites including hybrid C/G-box for TaFDL2 protein known as the *VRN1* gene upregulator were predicted inside the variable region. Information available in model plants helped to simulate more detailed genetic network responsible for growth habit.

It was also shown that deletions leading to promoter damage occurred in diploid and polyploid species independently. DNA transposon insertions first occurred in polyploid species. At the same time, the amplification of the promoter region was observed in A genomes of polyploid species. All obtained data are useful for deeper insight into regulatory network underlined the origin of spring wheat forms.

AN APPROACH TO THE STREHLER-MILDVAN CORRELATION FROM THE PARAMETABOLIC THEORY OF AGEING

A. Golubev
Research Institute for Experimental Medicine, Saint-Petersburg, Russia
E-mail: lxglbv@rambler.ru

Motivation and Aim: An enigmatic feature of mortality vs. age dependencies in human populations is the so-called Strehler-Mildvan (SM) correlation (Strehler and Mildvan, 1960) or compensational effect of mortality (Gavrilov and Gavrilova, 1991), i.e., lower initial mortality (λ) is associated with accelerated ageing. It was argued (Gavrilov and Gavrilova, 1991; Golubev, 2003) that SM correlation is partly an artifact of treating mortality data conforming to the Gompertz-Makeham (GM) law

$$\mu(t)=C+\lambda\times e^{\gamma\times t}$$

as if they conform to the pure Gompertz model.

Results: This artifact is more significant at greater values and ranges of $\boldsymbol{C}$. However, $\ln\lambda$ and the actuarial ageing rate γ show an astonishingly high correlation even in a comparative 2000-2004 period analysis of data on mortality in 18 developed countries, which was performed with the complete GM model suggesting that, in this case, $\boldsymbol{C}$ may be regarded as negligibly small. Thus, GM correlation remains a challenge for gerontological theory even though its demographic manifestations may seem quantitatively insignificant. According to the parametabolic theory of ageing (Golubev, 2009), GM correlation may be explained by assuming that the partitioning of body resources $\boldsymbol{R}$ between stress resistance $\boldsymbol{E}$ and somatic maintenance including protection from the parametabolic damage $\boldsymbol{P}$ depends on environmental harshness $\boldsymbol{q}$. Let $\boldsymbol{R}$ partitioning between $\boldsymbol{E}$ and $\boldsymbol{P}$ be formalised as $\mathbf{E}/\mathbf{P} = \boldsymbol{k}$ at $\boldsymbol{E}+\boldsymbol{P} =\boldsymbol{R}$, so $\boldsymbol{E} = \boldsymbol{k}\cdot\boldsymbol{R}/(\boldsymbol{k}+1)$ and $\boldsymbol{P}=\boldsymbol{R}/(\boldsymbol{k}+1)$. Let $\boldsymbol{k}$ to increase with increasing $\boldsymbol{q}$, i.e., $\boldsymbol{k} = \delta\cdot\boldsymbol{q}$, where δ captures the ability of an organism to reallocate resources to stress resistance according to environmental challenges. It has been shown that, in this way, $\ln\lambda$ and γ become correlated in a nonlinear fashion:

$$\ln\lambda = -\frac{R_0\,\frac{\gamma\beta}{\alpha}}{q\,\frac{\gamma\beta}{\alpha}+1}$$

Conclusion: Within the narrow range of δ typical of human populations, the present model is very close to the generally accepted linear one; however, its fit to the real data is higher than that of the linear one. It is reasonable to expect that δ may vary not only in different coexisting populations, but also in different existence periods of the same population. For example, improving developmental conditions may result in progressively stouter adults. The unprecedented increase in human stature in the XX century occurred concomitantly with the dramatic increase in life expectancy (Harris, 1997). A puzzling accompanying phenomenon of these gains is the increasing rate of ageing revealed with Gompertzian (Yashin et al., 2001) and GM (Mamaev et al., 2004) analyses. A solution to this paradox may be suggested by the consideration that the partitioning of extra resources towards $\boldsymbol{E}$ is equivalent to increasing δ, which will lead to decreasing λ and simultaneously increasing γ.

MODELING OF GENE NETWORKS REGULATED BY NEGATIVE AND POSITIVE FEEDBACKS

Yu.A. Gaidov*, V.P. Golubyatnikov**, A.G. Kleshchev***
** Novosibirsk State Pedagogical University, Novosibirsk, Russia*
***Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*
**** Ugra Scientific Research Institute of Informatics Technology, Hanty-Mansiisk, Russia*
***e-mail: glbtn@math.nsc.ru*
***Corresponding author*

Motivation and Aim: We study phase portraits of nonlinear 3-dimensional dynamical systems as models of gene networks regulated by combinations of positive and negative feedbacks. Detailed analysis of cycles and other attractors of similar models with multitability properties is an important task both from the biological, and from the mathematical viewpoints, see [1].

Methods and Algorithms: Our description of these nonlinear dynamical systems is based on geometrical and topological methods elaborated in our previous studies ([2,3]).

Results: We obtain sufficient conditions of existence of stable periodic trajectories in some gene networks models regulated by positive and negative feedbacks

$$\frac{dx}{dt} = h_1(z) - x;\ \frac{dy}{dt} = h_2(x) - y;\ \frac{dz}{dt} = h_3(y) - x. \qquad (1)$$

Here the functions are either monotonically decreasing, like the Hill's functions $h(w) = \frac{\alpha}{1+w^{\gamma}}$ or are unimodal, like the Glass-Mackey functions $h(w) = \frac{\alpha w}{1+w^{\gamma}}$ in the chemical kinetics equations, this represents a simple combination of positive and negative feedbacks; $x, y, z \geq 0$.

In the case when one of the functions h_i monotonically decreases and two others are unimodal, we prove stability of one of the stationary points of the dynamical system (1). These models exhibit multistability properties. We construct polyhedral invariant domains of the system (1) near its unstable stationary points with topological index -1, and we find conditions of existence of cycles near these points. All these domains are composed by triangle prisms. Some sufficient conditions of stability of these cycles are obtained as well.

If the functions h_i have the form of the Hill's and Glass-Mackey functions, the verification of these conditions reduces to simple algebraic procedures. Similar approach can be used in mathematical modeling of more complicated and higher-dimensional Gene Networks.

The work was supported by RFBR grant 09-01-00070, by Interdisciplinary project 119 of SB RAS, by the leading scientific schools grant 8526.2006.1, and by ADS-Program of Development of Scientific Potential of Higher School, project 2.1.1/3707.

References:

1. V.A.Likhoshvai, V.P.Golubyatnikov et al. (2008) Theory of gene networks, In: *System computerized biology*, (N.A.Kolchanov, S.S.Goncharov), Novosibirsk, SB RAS, 430–576.
2. Yu.A.Gaidov, V.P.Golubyatnikov (2007) On some nonlinear dynamical systems modeling asymmetric gene networks, *Bulletin of Novosibirsk State University,* **7:** 23-32.
3. Yu.A. Gaidov (2008) On stability of periodic trajectories in some models of gene networks (Russian) // *Siberian Journal of Industrial mathematics*, **11**: 57–62.

STABILITY OF CYCLES IN MODELS OF GENE NETWORKS REGULATED BY NEGATIVE FEEDBACKS.

[1]*Golubyatnikov V.P., [1]Golubyatnikov I.V., [2] Kleshchev A.G.
[1] *Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*
[2] *Ugra Scientific Research Institute of Informatic Technology, Hanty-Mansiisk, Russia*
e-mail: glbtn@math.nsc.ru
[*]*Corresponding author*

Motivation and Aim: We consider periodic trajectories of arbitrary odd-dimensional nonlinear dynamical systems modeling gene networks regulated by negative feedbacks. Qualitative mathematical description of the oscillating functioning regimes of similar gene networks is very important for the system computerized biology.

Methods and Algorithms: Our results on existence and stability of cycles in the gene networks models are based on topological and geometrical methods elaborated in our previous publication [1], and on the approach proposed by R.Smith [2].

Results: We study gene networks models represented by dynamical systems of the type:

$$\frac{dx_1}{dt} = f_1(x_{2k+1}) - x_1; \quad \frac{dx_2}{dt} = f_2(x_1) - x_2; \quad \ldots. \quad \frac{dx_{2k+1}}{dt} = f_{2k+1}(x_{2k}) - x_{2k+1}. \qquad (1)$$

Here the functions f_i, i = 1,2,… 2k+1, are smooth and monotonically decreasing like the Hill's functions $f(w) = \frac{\alpha}{1+w^{\gamma}}$; they correspond to negative feedbacks in gene network. We show that the system (1) has exactly one stationary point S_0 with positive coordinates, and that there is an invariant non-convex polyhedral domain P of the system (1) near this point S_0. This domain is composed by 4k+2 triangle prisms $\Delta \times Q^{2k-1}$, i.e., the products of rectilinear triangles and (2k-1)-dimensional parallelepipeds. Our main results are:

A. If the linearization of the system (1) near the point S_0 has eigenvalues with positive real parts, then the (2k+1)-dimensional domain P contains a cycle of the system (1).

B. If the system (1) is as in A, and for all i=1,2,..2k+1, for all $(x_1, x_2, \ldots x_{2k+1}) \in P$, and for some positive parameter η, the inequality $\left| \eta + \frac{df_i}{dx_{i-1}}(x_{i-1}) \right| < \eta \cdot \sin\frac{2\pi}{2k+1} \cdot \sin\frac{\pi}{2k+1}$ holds, then the domain P contains a stable cycle of the system (1).

We see that the trajectories of the system (1) do not tend to its limit cycle monotonically, so the usual method of the Lyapunov functions in the stability studies can not be applied here.

Conclusion: Similar approach can be used in the cases of more complicated networks.

The work was supported by RFBR grant 09-01-00070, by Interdisciplinary project 119 of SB RAS, by the leading scientific schools grant 8526.2006.1, and by ADS-Program of Development of Scientific Potential of Higher School, project 2.1.1/3707.

References:

1. V.A.Likhoshvai, V.P.Golubyatnikov et al. (2008) Theory of gene networks, In: *System computerized biology*, (N.A.Kolchanov, S.S.Goncharov), Novosibirsk, SB RAS, 397-480.
2. R.A.Smith, (1987) Orbital stability of ordinary differential equations, *Journ. of Diff. Equations*, **69:** 265-287.

MECHANISM OF OBTAINING UNSUPERVISED KNOWLEDGE TO ENRICH CliDaPa APPROACH

S. Gonzalez[1*], J. Veiga[1], V. Robles[1], J.M. Pena[1] and F. Famili[2]
[1] *Department of Computer Architecture and Technology, Universidad Politücnica de Madrid, Spain*
[2] *NRC Institute for Information Technology, Ottawa, Canada*
e-mail:sgonzalez@fi.upm.es, jveiga@fi.upm.es, vrobles@fi.upm.es, jmpena@fi.upm.es, fazel.famili@nrc-cnrc.gc.ca

Keywords: *CliDaPa, unsupervised, clinical, DNA microarrays, data analysis*

Motivation and Aim: Recalling conclusions obtained from [9] when we compare traditional data uses and CliDaPa algorithm applied to two or more information sources (e. g. clinical and gene expression data), the CliDaPa approach improved results on disease classification. However, if we analyze the data used, we appreciate that the only data that is not easily understandable for expert biologists is gene expression data. If we obtain new knowledge from gene information, probably we could use it as new information source and we could improve the CliDaPa executions, improving results too.

Methods and Algorithms: Sample-based clustering can be obtained using an unsupervised method, Quality Threshold [10], with gene expression data. For that, it's necessary to define several features: the distance measure, the threshold and the minimum number of elements in a cluster. Using Euclidean, Manhattan, Pearson correlation and Biweight correlation [11] as distance measures, and using 10 different values of threshold within the range [mean – 2*deviation, mean + 2*deviation], 40 new clusters can be obtained. This new data is injected as new clinical data from the in data. Thus, when we execute the CliDaPa algorithm, where the data can be divided using any of these new data, only if the classification can be improved.

Results: To validate this proposal, several data sets with clinical and gene expression data (from Van't Veer, Van der Vivjer and Brain Cancer datasets) have been used. Several experiments have been carried out, using an external MxN fold cross validation. Results obtained from this proposal show us a 10% of improvement in the classification, if we compare with regular CliDaPa executions.

Conclusions: This new approach demonstrates that new unsupervised knowledge can improve and enrich a supervised classification. Thus, applying to CliDaPa approach, this algorithm gets the fulfillment of the proposed objectives.

Availability: CliDaPa and its improvements are available for the research community. For further information, please, contact the authors.

Acknowledgements: The authors thankfully acknowledge the computer resources, technical expertise and assistance provided by the Centro de Supercomputaciyn y Visualizaciyn de Madrid (CeSViMa) and the Spanish Supercomputing Network.

References:

1. J. Brenton. (2005) Molecular classification and molecular forecasting of breast cancer: ready for clinical application? J. Clin. Oncol., 23:7350–7360
2. P. Larracaga et al. (2006) Machine learning in bioinformatics. Briefing in Bioinformatics.
3. Scott L. et al. Prediction of central nervous system embryonal tumour outcome based on gene expression.
4. M. J. van de Vijveret al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med, 347(25):1999–2009, December 2002.
5. L. J. van 't Veeret al. (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature, 415(6871):530–536, January.
6. T.R. Golub et al. (1999) Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring.
7. S. Paoli, et al. Semisupervised Proling of Gene Expressions and Clinical Data, ITC-irst - Trento, Italy
8. Nathalie L.M. et al. M@CBETH: Optimizing Clinical Microarray Classification, Department of Electrical Engineering ESAT-SCD, Leuven-Heverlee, Belgium
9. S. Gonzőlez et al. CliDaPa: A new approach to combining clinical data with DNA microarrays. IDA Journal 2009.
10. Laurie J. et al.. Exploring expression data: Identification and analysis of coexpressed genes. Genome Research, 9(11):1106–1115, November 1999.
11. Johanna Hardin et al. A robust measure of correlation between two genes on a microarray. BMC Bioinformatics, 8(1):220+, June 2007.
12. Pavel Berkhin. Survey of clustering data mining techniques. Technical report, Accrue Software, San Jose, CA, 2002.
13. Daxin Jiang et al. Cluster analysis for gene expression data: A survey. IEEE Transactions on Knowledge and Data Engineering, 16:1370–1386, 2004.

A FAST ALGORITHM OF BUILDING A SPECIES TREE WITH A SET OF GENE TREES

K.Yu. Gorbunov*, V.A. Lyubetsky
Institute for Information Transmission Problems RAS, Moscow, Russia
e-mail: gorbunov@iitp.ru
**Corresponding author*

Key words: *fast algorithm of building supertrees, novel approach to build supertrees*

Motivation and Aim: The problem of building species supertrees with a set of gene trees is long recognized to be of fundamental and applied value. This problem is NP-hard in terms of algorithmic complexity. Finding effective solutions (algorithms with low degree of polynomial complexity) requires its biologically valid reformulations.

Results: We propose such a reformulation and a fast algorithmic solution to build a supertree, with time complexity typically linear to the number of species and number n of input trees. Simulations show that the algorithm is both fast and accurate. Thus, species trees S with about 200 species were correctly reconstructed on the set $\{G_i\}$ of gene trees (stochastically derived from trees S) in less than one minute on an Intel Core 2 Duo 2.53 GHz, 2 GB RAM PC. In the novel approach, the supertree S is sought among such trees, which have all clades contained in a fixed predefined set P (see bellow). Let V_0 be the set of all species in all input trees, and V_0 be included in P. Define V from P as *basic* if it can be split in *two* sets from P, which can also be split in two, and so on until singlet leaves are obtained. The complexity of our algorithm is $|P|^2 \cdot |V_0| \cdot n$, which is considerably lower than $4n^3 \cdot |V_0|^3$, i.e. its time is at maximum cubic to the natural parameters V_0 and n of the task, for every basic V to accurately infer the supertree $S(V)$ with all leaves from V. According to simulations, the algorithm correctly infers the solution $S(V)$ for leaves from V under the definition of P specified below. The algorithm output is a set of subtrees $\{S(V)$: V is a basic set$\}$ of the final supertree $S(V_0)$. In our experiments the subtrees $S(V)$ are correctly assembled into the final supertree $S(V_0)$ with the algorithm from [1]. Our approach is largely based on the reconciliation algorithm from [2]. The set P is the set of species names from all clades in trees G_i. To account for HGTs, the following sets are added to P for each G_i: (1) both components of the symmetric difference between each two not-nested clades g_1 and g_2 with overlapping species, and (2) the difference between any clade g_1 and its ancestral clade g. The set $\{G_i\}$ was simulated on a given S using the gene evolution model described in [2]. The tree S is divided in time slices with gene event probabilities set for each slice: gene duplication, HGT with or without retaining the donor gene copy. A gene survives in a next time slice or bifurcation with the rest of the probabilities. If a bifurcation is beyond the current slice, the gene survives in the next slice. Otherwise, the gene diverges at a bifurcation with losing one of its descendants.

References:

1. A.Wehe, M.S.Bansal, J.G.Burleigh, O.Eulenstein. (2008) DupTree: A program for large-scale phylogenetic analyses using gene tree parsimony, *Bioinformatics,* **24**(13): 1540–1541.
2. K.Yu. Gorbunov, V.A. Lyubetsky. (2009) Reconstructing genes evolution along a species tree, *Russian Journal of Molecular Biology*, **43**(5): 946–958.

COMPARATIVE PROTEOMICS OF MYCOPLASMS

Govorun V.M.
Shemyakin & Ovchinnikov Institute of Bioorganic Chemistry RAS, RRC Kurchatov Institute, Moscow, Russia
E-mail: govorun@nanorf.ru

Key words: *Mycoplasmas, proteomics, ortholog*

Motivation and Aim: Mollicutes (Mycoplasmas) clade consists of bacteria without a cell wall, possessing very small genomes with a low GC content. It is supposed that mycoplasmas originate from gram-positive bacteria which were a subject of reductive evolution during an adaptation to a parasitic lifestyle. Mycoplasmas are a clade with the largest number of sequenced and annotated genomes to date. That is why they are a very convenient object of comparative studies. Mycoplasmas are the organisms with the smallest genomes known up to date. Their genome sizes vary from 0,5 to 1,5 millions of base pairs. *Mycoplasma genitalium* is an autonomously replicating organism with the smallest known genome which consists of 580 thousands of base pairs comprising 524 genes. Mycoplasmas have very limited number of proteins and metabolic functions as a consequence of their small genome size. Thus they are hugely dependent on their host metabolism or media nutrition content. Thereby mycoplasmas closely resemble the "minimal cell" concept, meaning a cell that is capable of a self-replication on a rich medium having the less possible proteins.

Methods and Algorithms: Proteo-genomic profiling allows us to elucidate functions required for a cell to live and also helps to improve genome annotation. In addition it allows to carry out metabolic reconstruction of the cell. Comparative proteomics helps to find out a core protein content of the cell which is responsible for basic cellular processes and which proteins are involved in adaptation of a given Mycoplasma species to its environment. Our laboratory has studied proteomes of *M. gallisepticum* and *Acholeplasma laidlawii*.

Results: We found out that *M. gallisepticum* and *Ach. laidlawii* express 55% and 58% of encoded proteins respectively. *M. mobile* which proteome is known up to date expresses 88% of encoded proteins.

M. gallisepticum, *Ach. laidlawii* and *M. mobile* are relative species but their genome size varies dramatically. Genome of *Ach. laidlawii* consists of 1497 thousands b.p., *M. gallisepticum* – 996 thousands b.p., *M. mobile* – 777 thousands b.p. In this series *Ach. laidlawii* possesses tens of regulatory proteins, whereas other two have only few. Nevertheless all mycoplasmas show significant changes in protein expression pattern under the stress conditions like a heat shock. This might be an evident that they are capable of some unknown ways of expression regulation. Comparative studies of the three species revealed a constant protein core consisting of 200 proteins responsible for replication, transcription, translation and ATP production. Also core included 4 conserved proteins of unknown function. Strikingly, *M. gallisepticum* and *Ach. laidlawii* are capable of growth on the same medium but *Ach. laidlawii* expresses 250 proteins more than *M. gallisepticum*.

Conclusion: Most of these proteins carry out metabolic functions. It was also found that mycoplasmas express about a hundred of unique proteins with unknown function which do not have any orthologs in other species. One can speculate that these proteins take part in adaptation, improving cell plasticity while evading significant changes in housekeeping protein core.

454/ROCHE SEQUENCING BREAKTHROUGHS: OVERALL NEXT GENERATION SEQUENCING FOR BROADEST VARIETY OF APPLICATIONS

M.G. Gracheva
Roche Diagnostics Rus LLC, Moscow, Russia
e-mail: maria.gracheva@roche.com

Key words: *454 sequencing, GS FLX, GS Junior, de novo, 16S, ultra-deep sequencing*

The Genome Sequencer FLX System is the flagship 454 Sequencing platform. Offering more than 1 million high-quality long reads per run with Q20 read lengths of 400 bases, the system is ideally suited for *de novo* sequencing of whole genomes and transcriptomes of any size, metagenomic characterization of complex samples, resequencing studies and more. Due to broad variety of applications the GS FLX is at the heart of breakthrough scientific discoveries and hundreds of peer-reviewed publications to date. Continuous development of GS FLX chemistry will enable the next leap in performance, with extended read lengths approaching 1000 bases in 2010. Long reads facilitate the use of fewer reads to characterize the genome, characterize alternate gene splice isoforms, fusion transcripts, indels, enable to perform Ultra-Deep sequencing and 16S sequencing of PCR products.

The GS Junior System brings the power of 454 Sequencing technology directly to laboratory bench top. It has the same proven long-read chemistry as the GS FLX System, scaled to suit the needs of individual labs. It helps to quickly proceed from DNA to results to discovery with an easy-to-follow workflow and data analysis at desktop. The system is perfectly sized for rapid sequencing of PCR products (Ultra-Deep sequencing, 16S sequencing, HLA-typing, etc.), targeted human resequencing studies, *de novo* sequencing of microbial and other small genomes and much more.

Both systems enable to complete the whole genome projects, even within a single run, with the combination of long single reads and Long-Tag Paired end reads (>150 base tags from each end of a 3 Kb, 8 Kb, or 20 Kb span). Long-Tag Paired end reads ensure contiguous sequence information and full coverage over the highly repetitive regions of complex genomes. 454 long reads and ability to sequence from single DNA molecules of PCR products enable straightforward identification of low frequency somatic mutations and detect the spectrum of variations in specific disease-associated regions or genes by sequencing PCR products, cDNA libraries, or DNA fragments captured with NimbleGen's Sequence Capture array. Systems offer long reads for haplotyping and straightforward identification of insertions and deletions, without GC bias. They enable to characterize and quantify the viral population within an infected individual with Ultra-Deep 454 Sequencing, enables accurate and quantitative assessment of HIV quasi-species through the sequencing of target genes.

The most important thing that the data analysis is performed without the need for enterprise scale IT solutions with the included easy-to-use software tools - GS De Novo Assembler, GS Reference Mapper, and GS Amplicon Variant Analyzer, De Novo Transcriptome Assembler. Straightforward interpretation of data means faster discovery of biologically meaningful results.

THE POSSIBLE ORGANIZATION OF THE SYNAPTONEMAL COMPLEX CENTRAL SPACE IN NEMATODE *CAENORHABDITIS ELEGANS*

T.M. Grishaeva*, S.Ya. Dadashev, Yu.F. Bogdanov
N.I. Vavilov Institute of General Genetics RAS, Moscow, Russia
e-mail: tmgrishaeva@rambler.ru
**Corresponding author*

Key words: *eukaryotes, meiosis, synaptonemal complex, in silico analysis*

Motivation and Aim: Synaptonemal complex (SC) is a key subcellular structure of meiotic cells (meiocytes) in eukaryotes. SC plays indispensable role in meiotic recombination and homologous chromosome segregation. SC central space (SCCS) serves as the compartment for the complex of recombination proteins and it is structurally organized with several proteins homing these enzymes. Recently new proteins which are the components of SCCS have been reviled and characterized. These are SYCE1, SYCE2 and TEX12 in vertebrates, Corona in *Drosophila*, and SYP-3 and SYP-4 in nematode *Caenorhabditis elegans* in addition to formerly known SYP-1 and SYP-2. The existing model of the SCCS organization in higher eukaryotes turned to be not applicable for nematodes. We aimed on two tasks: 1) to reveal, whether proteins SYCE1, SYCE2 and TEX12 in vertebrates, on the one hand, and four SYP proteins in nematode, on the other one, share some common characteristics; 2) to offer a new model of SCCS organization in the nematode *C. elegans*.

Methods and Algorithms: We have used the NCBI database and software (http://www.ncbi.nlm.nih.gov/), in particular, NCBI CDART (Conserved Domain Architecture Tool) for revealing the functional domains of proteins. As well as we used the program MEME, http://meme.sdsc.edu/meme/website/intro.html), to define a set and sequence of conservative motives in protein molecules. The program "Mobile portal – charge" from package Mobile Pasteur (http://mobyle.pasteur.fr/cgi-bin/MobylePortal/portal.py?form=charge/) was used for studying electrostatic charge distribution along the protein molecules. The probability of coiled coil formation by amino acid sequences of proteins studied was estimated by means of the program "COILS – Prediction of Coiled Coil Regions in Proteins" (http://www.ch.embnet.org/software/COILS_form.html).

Results: Two groups of proteins (SYCE1, SYCE2 and TEX12 of mouse *Mus musculus* and SYP-1, 2, 3, 4 of *C. elegans*) do not share any conservative motives, as well as have no similarity in distribution of electrostatic charge along the molecules. Similar functional domains were found only in two proteins: SYCE1 (SMC_prok_B) and SYP-1 (SMC). The secondary structure of all the proteins is characterized by coiled-coil configuration of different length that suggests active protein-protein interactions. Possible interactions of SYP proteins were simulated by the original methods described earlier (T.M. Grishaeva et al., 2004. Biomedical Chemistry 50. Suppl. 1. 3-10 (in Russian)).

Conclusions and Availability: A model of organization of the synaptonemal complex central space in nematode *C. elegans* is offered in which four protein molecules (SYP-1, 2, 3, 4) construct half of transversal filament. The model allows to predict new proteins constructing the synaptonemal complex central space in other lower eukaryotes.

MOLECULAR MECHANISMS OF DOPAMINE CONTROL OF JUVENILE HORMONE TITER IN DROSOPHILA

N.E. Gruntenko[1]*, S. Li[2], E.K. Karpova[1], N.V. Adonyeva[1], I.Yu. Rauschenbach[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Shanghai Institute of Biological Sciences CAS, Shanghai, China*
e-mail: nataly@bionet.nsc.ru
**Corresponding author*

Key words: *dopamine, D2-like dopamine receptors, juvenile hormone, corpus allatum, Drosophila.*

Motivation and Aim: A proper hormonal balance is of a paramount importance for normal functioning of any organism. In *Drosophila* a shift of the balance in the direction of either one or another insect gonadotropine (juvenile hormone (JH) and 20-hydroxyecdysone (20E)) leads to dramatic changes in fertility and stress response [1,2]. There is a mediator in JH and 20E interaction and it is dopamine (DA) [2]. The molecular mechanism of this intermediation is under investigation.

Methods and Algorithms: We studied the effect of a considerable decrease in JH titer on reproductive function in *Drosophila*. JH titer was decreased using the UAS-Gal4 system for genetic ablation of *corpus allatum* (*CA*), the gland that produces the hormone. We also used UAS-Gal4 system to reveal the effect of a reduction of *Drosophila* D2-like dopamine receptors (DD2R) number in *CA* on the level of JH-degradation. Immunohistochemical analysis of DD2R expression in wild type flies was performed on fat body where JH-esterase, the enzyme of JH degradation, is produced.

Results: Ablation of a part of *CA* cells in *Aug21>UAS-reaper::UAS-hid* flies results in a dramatic decrease of JH synthesis and significant increase of its degradation. These flies have underdeveloped ovaries and very low fecundity and altered resistance to heat stress. *Aug21-Gal4>UAS-ds-DD2R* females with a decreased DD2R level in *CA* have a significantly decreased level of JH degradation, which indicates an increased JH titer. DA inhibits the activity of JH degrading enzymes in young wild type females and stimulates it in the sexually mature ones. DA affects the enzymes of JH degradation through DD2R: the amount of DD2R in fat body of the young females is significantly greater than in that of mature ones.

Conclusions: (1) Insect fertility and stress resistance depends on JH. (2) DA controls JH titer through both synthesis and degradation of the hormone. (3) A change in the direction of DA effect on the activity of enzymes of JH degradation from the inhibiting one (in young Drosophila females) to the stimulating one (in mature females) occurs due to switching-over on the genome level that results in a change of the number of D2-like receptors in the fat body.

Acknowledgements: The study was supported by the grants of RFBR #09-04-00441, RAS Programme «Biodiversity» #23.25/4, grant from President of Russian Federation #MK-4180.2009.4 and Integration Grant of the SD RAS - AS China #27.

References:

1. M.Soller et al. (1999) Control of oocyte maturation in sexually mature *Drosophila* females, Developmental Biology 208:337-351.
2. N.E.Gruntenko, I.Yu.Rauschenbach (2008) Interplay of JH, 20E and biogenic amines under normal and stress conditions and its effect on reproduction, Journal of Insect Physiology, 54:902-908.

SPATIAL DYNAMICS MODELING OF VIRAL INFECTION IN TWO-DIMENSIONAL CELL ARRAYS

A. Yakimovich*[1,2], H. Gumpert*[3], C.J. Burckhardt[1], V.A. Lütschg[1], A. Jurgeit[1], I.F. Sbalzarini[3], U.F. Greber[1]

** Authors contributed equally to the work, corresponding authors*

Affiliations: Institute of Molecular Life Sciences, University of Zürich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland. [2]Life Science Zürich Graduate School, Molecular Life Science Program. [3]MOSAIC Group, Institute of Theoretical Computer Science and Swiss Institute of Bioinformatics, ETH Zürich, CH-8092 Zurich, Switzerland

e-mail: artur.yakimovich@imls.uzh.ch, hgumpert@student.ethz.ch

Key words: *Cellular Automata, Adenovirus, Computational Modeling*

Motivation and Aim

Adenoviruses (Ads) are non-enveloped icosahedral DNA viruses infecting the respiratory, digestive, excretory or ocular systems. They infect human epithelial cells by receptor-mediated endocytosis, and lytically propagate in many cell types. It is largely unknown, how adenoviruses exit from infected cells and transmit infection.

Methods and Results

Here, we use a 2-dimensional array of cultured human epithelial cells for analysis of the infection dynamics by automated high-throughput time-lapse fluorescence microscopy, and its corresponding computational *in silico* model. We measure cell-cell transmission kinetics of replication competent or incompetent human Ad2 or Ad5 expressing eGFP transgene. Preliminary experiments show that the time of cell lysis at the first round of infection varies significantly between cells, while the onset of gene expression in subsequent rounds of infection remains largely invariant.

We measure a large number of phenotypic parameters, enabling the creation of a computational model of infection dynamics. The model is based on a cellular automaton. Cellular automata are a simple yet powerful modeling paradigm for biology. They consist of a set of discrete computational "cells", each characterized by its time-dependent state and the state of its neighbors. In our case, the state of each cell reflects its infection status, virus load, or lysis probability. These states are constantly updated over time as functions of the states of neighboring cells and biological "rules" (prior knowledge). In addition, we couple the model with a continuum model of virus diffusion in the extracellular medium. The combined model thus exposes biologically relevant parameters, such as dynamic distances between neighboring nuclei, volumes of cells, infection status, and the free virus diffusion constants, that are fully observable and controllable. Iterative comparisons of *in silico* simulation predictions and cell biological experimentations will establish probabilistic infection models, and lead towards the identification of biologically important parameters for viral transmissions between cells.

MOLECULAR EVOLUTION OF THE HYPERTHERMOPHILIC ARCHAEA OF THE PYROCOCCUS GENUS: ANALYSIS OF ADAPTATION TO DIFFERENT ENVIRONMENTAL CONDITIONS

K.V. Gunbin*[1], D.A. Afonnikov[1,2] N.A. Kolchanov[1,2]
[1]- *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]- *Novosibirsk State University, Novosibirsk, Russia*
e-mail: genkvg@bionet.nsc.ru
[*]*Corresponding author*

Key words: *Pyrococcus, positive selection, radical and conservative replacements, substitutional asymmetry*

Motivation and Aim: The increasing number of complete sequences of prokaryotic genomes has provided the basis for studying the molecular mechanisms of their adaptation at the genomic level. The aim was to identify features of molecular evolution of the fully sequenced Pyrococcus species related to adaptation strategy to diverse environmental conditions.

Methods and Algorithms: We apply here a computer-based approach to compare the genomes and proteomes from *P. furiosis, P. horikoshii*, and *P. abyssi*.

Results: Phylogenetic analysis of rDNAs from 27 Pyrococcus strains suggested that the divergence of *P. furiosis, P. horikoshii* and *P. abyssi* might have occurred from ancestral deep-sea organisms. It was demonstrated that the function of genes that have been subject to positive selection is closely related to abiotic and biotic conditions to which archaeobacteria managed to become adapted [1]. Divergence of the *P. furiosus* archaeobacteria might have been due to loss of some genes involved in cell motility or signal transduction, and/or to evolution under positive selection of the genes for translation machinery. In the course of *P. horikoshii* divergence, positive selection was found to operate mainly on the transcription machinery (primarily because of changes in the trophic chain); divergence of *P. abyssi* was related with positive selection for the genes mainly involved in inorganic ion transport. Analysis of radical amino acid replacement rate in evolving *P. furiosis, P. horikoshii* and *P. abyssi* showed that the fixation rate was higher for radical substitutions relative to the volume of amino acid side-chain. The analysis of amino acid substitutional asymmetry was performed also [2]. This allowed us to find the positive correlation between substitutional asymmetry indexes and indexes of amino acid occurrence frequencies in protein β-sheets on the Pyrococcus phylogenetic tree branches. The bulk of β-structures are usually concentrated within a protein globule.

Conclusion: Thus, current results give due credit to the important role of hydrostatic pressure and changes in the trophic chain in the evolving *P. furiosis, P. horikoshii* and *P. abyssi* genomes. Moreover, it can be assume that the dense hydrophobic core of soluble proteins was formed during the Pyrococcus evolution.

Availability: Available upon request.

References:

1. K.V. Gunbin et al. (2009) Molecular evolution of the hyperthermophilic archaea of the Pyrococcus genus. *BMC Genomics*, **10**: 639.
2. J.H. McDonald et al. (1999) Patterns of temperature adaptation in proteins from Methanococcus and Bacillus, *Mol Biol Evol*. **16**: 1785-1790.

MOLECULAR EVOLUTION OF THE *WOLBACHIA* WRI, WMEL AND WPIP GENOMES

K.V. Gunbin*[1], Yu.Yu. Ilinsky[1], D.A. Afonnikov[1,2]
[1]- *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]- *Novosibirsk State University, Novosibirsk, Russia*
e-mail: genkvg@bionet.nsc.ru
[*]*Corresponding author*

Key words: *Wolbachia, positive selection, radical and conservative replacements*

Motivation and Aim: The *Wolbachia*-host associations have become advantageous tools for evolutionary studies of symbiont-host interactions [1]. α-proteobacteria of the Wolbachia genus are widespread infectious agents in arthropods. The molecular mechanisms underlying their partnerships remain unclear [2]. The aim was to clarify wMel and wRi adaptation strategies to the host and diversify those strategies from wPip adaptation.

Methods and Algorithms: Here, we perform a comparative analysis of the evolution mode of more than 600 orthologous protein-encoding genes from the wMel, wRi and wPip Wolbachia strains, which infected *Drosophila melanogaster*, *Drosophila simulans* and *Culex pipiens* respectively. We identified genes under negative selection on the basis of analysis of nonsynonymous nucleotide substitutions and those under positive selection and neutrally evolving based on analysis of radical versus conservative amino acid replacement rates.

Results: The results allowed us to demonstrate features making similar and different the two bacteria in the course of their evolution. We distinguished three functional gene groups subject to the same evolution mode in both wMel and wRi strains: “translation, ribosomal structure and biogenesis” evolving more often under purifying selection; “carbohydrate transport and metabolism” evolving more often under positive selection; “replication, recombination and repair” evolving preferentially in neutral mode. Comparative analysis of the evolutionary changes in the amino acid composition of the wMel and wRi proteins and the number and functions of the positively selected wMel and wRi genes disclosed substantial differences in the wMel and wRi adaptation strategies. These differences were indicative of higher efficiency of positive selection for wMel than wRi.

Conclusion: Our current results suggested that wMel *Wolbachia* strain became more cooperated with the host compared to wRi at the genome level. This suggestion is consistent with the discussed experimental data.

Availability: Available upon request.

Acknowledgements: This work was supported by RFBR grant No. 09-04-01641-a, Biosphere Origin and Evolution program.

References:

1. J.H. Werren et al. (2008) Wolbachia: master manipulators of invertebrate biology, *Nat Rev Microbiol*, **6**: 741-751.
2. L.R. Serbus et al. (2008) The genetics and cell biology of Wolbachia-host interactions, *Annu Rev Genet*, **42**: 683-707.

COMPUTER SYSTEM FOR ANALYSIS OF MOLECULAR EVOLUTION MODES OF PROTEIN FAMILIES (SAMEM): RELATION OF MOLECULAR EVOLUTION WITH THE PHENOTYPICAL FEATURES OF ORGANISMS

K.V. Gunbin[1], D.A. Afonnikov[1,2], M.A. Genaev[1]
[1]- *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]- *Novosibirsk State University, Novosibirsk, Russia*
e-mail: genkvg@bionet.nsc.ru
[*]*Corresponding author*

Key words: *protein molecular evolution, radical to conservative amino acid replacement rate ratio, amino acids change rate*

Motivation and Aim: Recently, great doubts about validity of the nonsynonymous to synonymous nucleotide substitutions rate ratio as a robust criterion of positive selection were raised [1, 2]. Alternative approaches to study the molecular evolution modes of proteins based on the radical to conservative amino acid replacement rate ratio (K_R/K_C) [3] and the rate of change of various properties of amino acids (V_P) [4] in the course of protein evolution. The essential drawback of these approaches is the need to know in advance "positively evolved change" of certain amino acid properties.

Methods and Algorithms: Our novel SAMEM computer system is also based on the K_R/K_C and V_P, but has two crucial differences which allow us to fully overcome the drawback: we take into account all the known properties of amino acids, and we calculate the statistical relation of properties changes with certain user defined adaptive phenotypical features of organisms. We also improve the V_P analysis by implementing the permutational test for comparison of the simulated molecular evolution of sequences from the protein family with real evolution of these proteins.

Results: The opportunity to match evolutionary change of all amino acid properties with positively selected phenotypical features of organisms permits direct attribution of certain protein changes to positive selection events. The results of tests (Cyclin A, Cyclin B, Cyclin D, Cyclin E, Notch, Delta, Serrate, Fringe, TIR1, ARF, etc.) of the improved V_P are consistent with experimental data.

Conclusion: Thus, SAMEM system can be useful for comprehensive analysis of molecular evolution modes of various protein families.

Availability: http://pixie.bionet.nsc.ru/samem/ (web interface available soon; command-line interface available upon request)

Acknowledgements: This work was supported by RFBR grant No. 09-04-01641-a, Biosphere Origin and Evolution program, Integration Projects of SB RAS No. 6, 26, 113.

References:

1. D.A. Drummond, C.O. Wilke (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution, *Cell*, **134**: 341-352.
2. Y. Huang et al. (2009) Selection for minimization of translational frameshifting errors as a factor in the evolution of codon usage, *Nucleic Acids Res*, **37**: 6799-6810.
3. A.L. Hughes (1992) Coevolution of the vertebrate integrin α- and β-chain genes, *Mol Biol Evol,* **9**: 216-234.
4. T. Pupko et al. (2003) Detecting excess radical replacements in phylogenetic trees, *Gene,* **319**: 127-135.

HOW COULD THE METAZOAN COMPLEXITY INCREASED DURING EVOLUTION: THE RESULTS FROM ANALYSIS OF SUPERFAMILY DATABASE INFORMATION

K.V. Gunbin[1*], D.A. Afonnikov[1,2]
[1]- *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]- *Novosibirsk State University, Novosibirsk, Russia*
e-mail: genkvg@bionet.nsc.ru
[*]*Corresponding author*

Key words: *biological complexity, protein superfamily expansion, domain co-occurrence*

Motivation and Aim: Cell types are fundamental units of multicellular life but their evolution is obscure [1]. Cell type number serves as conventional measure of the organism complexity (OC). However, the estimation of cell type number is difficult task. It partially solved for only a several dozens of organisms experimentally. Therefore, the approach suggested to find a correlation between cell numbers and protein superfamily expansions for higher eukaryotes [2]. It was shown for the set of 38 multicellular organisms that the expansion in 194 protein superfamilies correlates with the number of organism cell types. However, this set of proteins superfamilies mainly includes those specific for large taxonomic groups (Metazoa or Protists or Plants). In our work we tested the new measure of the organismal complexity based on domain co-occurrence in the set of housekeeping genes presented in all extant Metazoan and Protists species.

Methods and Algorithms: We used the published data in [2] for selecting the nonspecific protein superfamilies (NPS) from the SUPERFAMILY database 1.73 [3]. We used the estimates of cell type numbers for 34 species (excluding plants) from [2]. We computed the Spearman nonparametric statistic for comparison of the cell type number for these species with total protein numbers in NPS and with number of domain co-occurrence in NPS.

Results: We used 225 NPS from the supplementary information to [2] paper for analysis. The statistical comparison of the cell type number for 34 species with total protein numbers in NPS revealed 177 significant ($p<0.05$) correlations, the mean Spearman statistic was 0.65 ± 0.15 (mean ± standard deviation). The comparison of the cell type number with number of domain co-occurrence in NPS revealed 209 significant ($p<0.05$) correlations, the mean Spearman statistic was 0.72 ± 0.13.

Conclusion: The results obtained showed that the complexity of gene network determining cell types may evolved on the basis of not only protein duplications but on protein functional meshing also.

Availability: Available upon request.

Acknowledgements: This work was supported by RFBR grant No. 09-04-01641-a, Biosphere Origin and Evolution program, Integration Projects of SB RAS No. 6, 26, 113.

References:

1. D. Arendt. (2008) The evolution of cell types in animals: emerging principles from molecular studies, *Nat Rev Genet.,* **9**:868-882.
2. C. Vogel, C. Chothia. (2006) Protein family expansions and biology complexity, *PLoS Comput Biol.*, **2**:e48.
3. D. Wilson et al. (2009) SUPERFAMILY-sophisticated comparative genomics, data mining, visualization and phylogeny, *Nucleic Acids Res.,* **37**:D380- D386.

GTF2I DOMAIN: STRUCTURE, EVOLUTION AND FUNCTION

I.V. Medvedeva[1], K.V. Gunbin[1]*, V.A. Ivanisenko[1], A.O. Ruvinsky[2]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*The Institute for Genetics and Bioinformatics, University of New England, Armidale, NSW 2351, Australia*
e-mail: genkvg@bionet.nsc.ru
[*]*Corresponding author*

Key words: *GTF2I gene family, exon-intron structure, tertiary structure alignment, evolution, gene family, duplication, phylogeny*

Motivation and Aim: The GTF2I gene family consists of *gtf2i*, *gtf2ird1* and *gtf2ird2* genes encoding transcriptional factors and the first two of them involved in Williams-Beuren syndrome if mutated. The main characteristic of this gene family is the presence of several so called GTF2I repeats which number varies in different genes from two to six. The investigation of the exon-intron structure evolution of these genes and the study of the tertiary structure differences of the GTF2I domains, that coded by repeats were our aims.

Methods and Algorithms: We applied the comparative analysis to investigate the exon-intron structure of the genes. We developed novel method (SAMEM) which implements the permutational test for comparing the molecular evolution model of the GTF2I protein family with real data. We used the program package PDB3DScan for structural alignment to find the differences in the 3D structure of GTF2I repeats in *gtf2i* gene and to search the structure homolog of the GTF2I domain.

Results: Comparative sequence analysis identified Activation Domain which is essential for transcription activation and is located between R1 and R2 (*gtf2i* and *gtf2ird2*) and between R2 and R3 (*gtf2ird1*). Comparison of molecular evolution model of the GTF2I family with real data discovered several rare aminoacids substitutions which characterized by greatest physicochemical changes. It is of interest that these aminoacids substitutions caused the 3D structure differences in repeats. We constructed hierarchical classification of the repeats and found that in some cases repeats developed simultaneously in fishes and mammals in evolution. Also we discovered the highly similar 3D structure of the MAP domain (*Staphylococcus aureus*) that is also repeated up to four times in *map protein***.**

Conclusion: Based on our GTF2I hierarchical classification and recent research [1] we conclude that GTF2I gene family have raised from the two genome duplication events early in vertebrate evolution. Also the found statistical rare substitutions in GTF2I repeats could affect their DNA binding properties [2].

Availability: Available upon request

Acknowledgements: This work was supported by RFBR grant No. 09-04-01641-a, Integration Projects of SB RAS No. 6, 26, 113.

References:

1. P. Dehal, J.L Boore. (2005) Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate, *PLoS Biol* **3**: e314.
2. S.J. Palmer et al. (2010) Negative autoregulation of GTF2IRD1 in Williams-Beuren syndrome via a novel DNA binding mechanism, *J Biol Chem.* **285**:4715-4724.

VARIATION OF ITS1, ITS2 AND 5.8S REGIONS IN RDNA LOCUS OF *CHIRONOMUS* SPECIES (DIPTERA, CHIRONOMIDAE)

L.I. Gunderina*, A.V. Katokhin
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: gund@bionet.nsc.ru
**Corresponding author*

Key words: *genus Chironomus, ribosomal DNA, 5.8S rDNA, internal transcribed spacers 1 and 2 (ITS1 and ITS2), variation, phylogeny*

Motivation and Aim: The ribosomal RNA genes, forming a multigene family, are frequently used for species identification and phylogenetic reconstructions of species and genus relationships. The rRNA genes are organized in clusters of tandemly repeated units, each of which consists of rRNA genes – 18S, 5.8S and 28S, separated by internal transcribed spacers (ITS1 and ITS2). Coding regions and spacers differ in their rate of evolution. The aim of this work was to study the variation and divergence of the nucleotide sequences of ITS1, 5.8S and ITS2 regions of rDNA locus in species of genus *Chironomus*, to reconstruct the phylogenetic relationships between species.

Methods and Algorithms: Species of genus *Chironomus* were identified by cytotaxonomic analysis of polytene chromosomes. Genomic DNA was isolated from individual larvae. To amplify ITS1, ITS2 and 5.8S DNA by PCR the primers were designed using the program "Primer3". The purified DNA fragments were sequenced in both directions using a Big Dye Terminator kit and an ABI PRISM 3100 analyzer. The sequences were aligned using ClustalW. The software package *MEGA* 4 was used for phylogenetic and molecular genetic analyses.

Results: ITS1 and ITS2 sequences of *Chironomus* species are similar in the nucleotide composition and demonstrate the G+C content bias (G+C = 31.2%). Nucleotide composition of 5.8S rDNA sequences is unbiased (G+C = 50.7%). ITS1 and ITS2 sequences in *Chironomus* species are more polymorphic as compared with 5.8S rDNA sequences. ITS1 and ITS2 sequences in *Chironomus* species varied in their length due to Indels (nucleotide insertions/deletions) and in the composition due to nucleotide substitutions. Indels were not found in 5.8S rDNA sequences. The patterns of intraspecific variability of ITS1 and ITS2 sequences are similar in different *Chironomus* species. Species-specific fixed nucleotide differences as well as the differences in length and localization of Indels in ITS1 and in ITS2 sequences of *Chironomus* species have been found. The degree of divergence of ITS1 and ITS2 sequences correlates with the taxonomic rank of the compared species groups. NJ-trees for ITS1 and ITS2 sequences recover *Chironomus* species, sibling species groups and cytocomplexes. Topology of these trees indicates that genus *Chironomus* is monophyletic. The NJ-trees for ITS1 and ITS2 sequences of *Chironomus* species are consistent with the phylogenetic trees for another characters (banding sequences in polytene chromosomes, mtDNA genes).

Conclusion: Fast-evolving ITS1 and ITS2 sequences of *Chironomus* species can be used for reconstructing the phylogeny on the species level, whereas conserved 5.8S rDNA sequences – for reconstructing the phylogeny on a higher taxonomic level.

Acknowledgements: This work was supported by the Programs of the Presidium of the Russian Academy of Sciences "Biological diversity" (No. 23.30) and "Origin and Evolution of Geobiosystems". The work was conducted with the use of equipment of the CCU "DNA Sequencing" SB RAS (http://sequest.niboch.nsc.ru)

INTEGRAL EVALUATION OF PROFILE OF NATRIURETIC PEPTIDES SYSTEM'S mRNA EXPRESSION IN CULURED CARDIOMYOCYTES DURING ANOXIA-REOXYGENATION

V.L. Gurianova*, V.E. Dosenko, V.B. Pavlyuchenko, A.A. Moybenko
Bogomoletz Institute of Physiology, Kiev, Ukraine
e-mail: nika.biph@gmail.com
natriuretic peptides, gene expression, cardiomyocyte culture, anoxia-reoxygenation

Motivation and Aim: Natriuretic peptides system (NPS) is a powerful cardioprotective mechanism in mammalians and consists of atrial (ANP), brain, or type B (BNP) and type C (CNP) NPs, and their receptors: NPR-A (for ANP and BNP), NPR-B (for CNP) and NPR-C (clearance receptor eliminating all types of NPs and generating additional signal). This system is in the limelight from the moment of its finding and a lot of studies are carried out in the field of genetic mechanisms underlying the NPs' regulation and their role in cardiac physiology and pathology but more data bear much more questions because of their contradictoriness and uncomplexity. This fact caused our interest in integral evaluation of members of NPS expressed by cardiomycytes on genetic level in different conditions in cultured cardiomyocytes including such important pathological process as anoxia-reoxygenation (AR) as a model of ischemia-reperfusion.

Methods and Algorithms: The study was carried out on neonatal cardiomyocytes isolated from ventricular myocardium of 2-day-old Wistar rats. Anoxia (A) was attained with an airtight jar from which the O_2 was flushed with gas mixture containing 5% CO2 and 95% Ar for 30 min. Reoxygenation (R) was realized by exchanging fresh medium and by its aeration with the standard gas mixture for 60 min (acute AR) or 24 hours (prolonged AR). Total RNA was isolated from cells (Trizol RNA-prep kit) with the followed by reverse transcription and quantitative Real-Time PCR (SYBR Green PCR or TaqMan Gene Expression) performing to determine the level of mRNA expression of NPPA (ANP), NPPB (BNP), NPR1 (NPR-A) and NPR3 (NPR-C) genes in control (incubation for 48 hours in normoxic conditions), acute AR and prolonged AR groups. Except the valuables of obtained results in themselves we tried to find the most optimal coefficient to estimate the NPS's general activity in every individual case and in dynamics of changes. So, we assume that all changes of mRNA level are realized in adequate peptides and proteins changes, the probability of ligand-to-receptor binding is in direct proportion to ligand-to-receptor affinity (or in inverse proportion to dissociation constant (B.Bennett, 1991)) and that the force of effect depends on the level of ANP, BNP and the index NPR1/NPR3 that points at the state of receptor apparatus and NPR3 number also points on the state of clearance potential of the system and is denominator of coefficient. So, NPS coefficient = S Ч NPPA Ч ((0.53 Ч NPR1) / (0.38 Ч NPR3)) + S Ч NPPB Ч ((0.14ЧNPR1) / (0.08 Ч NPR3)).

Results: Received data show that NPPA, NPPB and NPR1 mRNA expression in all studied by us cases changed in the same manner: extremely augmented during the acute AR and backing to the control values (and even lower) during the prolonged R. As opposed to this, NPR3 gene mRNA expression changed in the opposite direction that forms the additive contribution to general tendency of changes of NPS's general activity that we express in form of relevant coefficient (in control group it is equal 0.79 ± 0.2, after acute AR it was increased to 16.03 ± 7.89 ($P < 0.05$, compared to control) and after prolonged AR came back to control values and not significantly lower by them – 0.41 ± 0.14 ($P > 0.05$, compared to control).

Conclusion: We obtained the data about the changes of mRNA expression of members of natriuretic peptides system expressed by cardiomyocytes in conditions of acute and prolonged AR compared to control and are able to conclude that these changes have similar manner for natriuretic peptides end effector receptor and opposite reaction for clearance receptor that is additive for the change (increased or decreased) of NPS's general activity on genetic level.

PREDICTION OF GENOME-WIDE FUNCTIONAL LINKAGES IN MYCOBACTERIUM TUBERCULOSIS USING GENOME CONTEXT METHODS AND GENE EXPRESSION DATA

Shubhada R Hegde*, Chandrani Das and Shekhar C Mande
Centre for DNA Fingerprinting and Diagnostics,
Hyderabad, INDIA - 500 001e-mail: shubhada@cdfd.org.in
**Presenting author*

Motivation and Aim: The increased rate of tuberculosis infection and a fatality rate of ~23% have made it necessary to search for new ways to prevent the chronic disease. This work involves inferring genome-wide unctional linkages in *Mycobacterium tuberculosis*.

Methods and Algorithms: Support Vector Machine (SVM), a machine learning algorithm, was used for the prediction. The parameter values were obtained from genomic context methods such as phylogenetic profile, gene distance and operon method and gene expression correlations. The positive interacting pairs included homologous protein interactors of *E. coli* and the literature curated protein interactions of *M. tuberculosis*. The negative dataset was generated considering differential localization of proteins in the cell.

Results: We have predicted 32,546 binary interactions among 3,571 *M. tuberculosis* proteins with the accuracy of 84%. The protein interaction network has a degree exponent of 1.67 showing scale-free behavior. Markov Cluster Algorithm was used to derive clusters in the interaction network and pathway enrichment was studied for each cluster. The combined centrality values of the nodes were used to enlist key proteins in the network, which may be involved in carrying out essential functions in the organism. The examples includes ribosomal and replication proteins such as DnaA, DnaB, RpoA, RplB etc, lipid metabolism proteins such as MbtG, DesA1, KasA and intermediary metabolism proteins such as subunits of ATP synthase, HemD, PurF.

Conclusion: We propose that this resource will be helpful for the systems level analysis of *M. tuberculosis* and identification of potential drug targets.

MODELING AND SIMULATION OF CELL-CELL COMMUNICATION WITH PETRI NETS ?

R. Hofestädt
AG Bioinformatics and Medical Informatics, University Bielefeld, Bielefeld, Germany
Email: hofestae@techfak.uni-bielefeld.de

Keywords: *Modeling and simulation, cell communication, cell differentiation, Petri nets.*

Motivation: Until now the focus of Bioinformatics is the function of a cell. To understand the molecular mechanisms of life we also need to understand the cell-cell communication and cell differentiation processes. Checking the available literature, we can see that many different formal systems are published to model metabolic processes. During the last decade has been shown that the Petri net formalism is useful for this kind of modeling and simulation. The main advantages of Petri nets are: 1) powerful theory 2) and simulation environments are available. Furthermore, it is possible to start the modeling process with a simple discrete Petri net model, which can be extended anytime using kinetic parameters (functional Petri nets etc.). Contemplating modeling of cell-cell communication and cell differentiation processes we can see specific applications of the cellular automaton and/or the L-Systems. The cellular automaton was exactly defined to model simple communication processes. The L-Systems were defined by Lindenmayer only to model and simulate cell differentiation processes. Therefore, Lindenmayer defined parallel grammars which can exactly realize this kind of simulation processes.

Aim and method: This presentation will discuss the question if the Petri net can also be used for the modeling of cell-cell communication processes. Therefore we will discuss how we can extend Petri nets so that the communication between different Petri nets will be possible. We will develop and discuss different topologies of Petri net structures. Furthermore we will discuss the question if these kinds of Petri nets are useful to solve practical questions.

Application: In our case we will use the Quorum Sensing process of the bacteria to demonstrate how this kind of simulation can be used and can be useful for the solution of scientific questions.

References:

1. M. Chen, S. Hariharaputran, R. Hofestädt, B. Kormeier, S. Spangardt. (2009) Petri net models for the semi-automatic construction of large scale biological networks, Springer Science and Business. Journal 11047, Article 9151.

DNA BARCODING OF ZOOPLANKTON ORGANISMS FROM KAZAN LAKES

L.L. Frolova, B.I. Barabanchikov, A. Husainov*
Kazan State University, Kazan, Russia
e-mail: Shade0602@yandex.ru

Key words: *zooplankton, DNA barcoding, bioinformation analysis, freshwater lakes*

Motivation and Aim: DNA barcoding is a new technique that uses a short genetic marker in an organism's DNA for species-level identification in the frame of an international initiative of the Consortium for the Barcode of Life. At present, databases of CO1 sequences of different organisms are smallest in comparison with large represented sequences of other genes and they are not including the information about CO1 gene of zooplankton organisms of Kazan region lakes (Russia).

Our goal was the identification of zooplankton organisms from Kazan lakes by the mitochondrial cytochrome c oxidase subunit I gene (COI).

Methods and Algorithms: We used zooplankton organisms of freshwater Kaban lakes (Kazan, Russia). They were caught and identified in accordance with legacy methodology [1,2]. The "universal" DNA primers for the polymerase chain reaction (PCR) amplification of a 710-bp fragment of the mitochondrial cytochrome c oxidase

subunit I gene (COI) for invertebrates were used: LCO1490: 5'-ggtcaacaaatcataaagatattgg-3' and HC02198: 5'-taaacttcagggtgaccaaaaaatca-3'. The conditions of experiment are described by Folmer [3]. For the analysis of nucleotide sequences were used software from NCBI website (http://www.ncbi.nlm.nih.gov).

Results: As a result of the set of experiments we obtained five nucleotide sequences fragments of the mitochondrial cytochrome c oxidase subunit I gene (COI) for follow organisms from Kaban lakes of Kazan city: *Brachionus calyciflorus* with sequence length 661 pb, *Keratella cochlearis* - 705 bp, *Scapholebecis mucronata* - 658 bp, *Chydorus sphaeriaus* - 658 bp, *Mesocyclops leacrarta* - 658 bp. These sequences were analyzed with bioinformation methods and prepared to input to the database at NCBI website with Barcode Submission Tool application.

Conclusion: These positive results allow us to continue the identification of zooplankton organisms from Kazan lakes by the mitochondrial cytochrome c oxidase subunit I gene (COI) and become active participants of the international Barcode of Life projects.

Availability: Used software and databases available free of charge in on-line and off-line mode.

Acknowledgements: We thank to Albert Rizvanov for his kind assistance in experimental work.

References:

1. G.G.Vinberg, G.M.Lavrentieva. (1982) The Methodical recommendations on collection and processing material at hydrobiological study on freshwater reservoirs. Zooplankton and his product, L.: Gosniorh, ZIN Acad. of Sc. of USSR, 33 (in Russia).
2. Identification guide of freshwater invertebrates of Europeen part of USSR, L.: Gidrometeoizdat, 1977 (in Russian).
3. 0.Folmer et al. (1994) DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates, Molecular Marine Biology and Biotechnology, **3(5):** 294-299.

CUP1 5'-UTR IS ESSENTIAL FOR HSP104 MEDIATED REGULATION OF GENE EXPRESSION IN YEAST

V.V. Ignatova[1]*, A.A. Rubel[1,2], A.F. Saifitdinova[1,2]
[1]*Dept. of Genetics, St-Petersburg State University, St-Petersburg, Russia*
[2]*St-Petersburg Branch of Vavilov Institute of General Genetics, St-Petersburg, Russia*
e-mail: valentina.v.ignatova@gmail.com
**Corresponding author*

Motivation and Aim: Traditionally, heat-shock proteins are believed to be involved in protein folding trafficking and rescuing of improperly folded polypeptides. Nevertheless, a number of chaperones have been found out in interaction with macromolecules other then proteins. Some of the molecular chaperones can participate in regulation of mRNA translation. It was described for yeast chaperone Zuotin from Hsp40/DnaJ family [1] and plant chaperone Hsp101from Hsp100/ClpB [2]. We have shown that expression of the reporter gene under control of *CUP1* promoter is positively regulated by Hsp104 independently on the level of transcription [3]. It was suggested that the molecular chaperones regulate expression of some particular genes binding with untranslated region (UTR) of the corresponding mRNA.

Methods: Secondary structure of *CUP1* 5'-UTR was performed by mfold (version 3.2). Potential binding sites for transcription factors and control elements was found using MatInspector. Searching for translation regulation signals was done using BLASTN 2.2.10. Amino-acid sequence of chaperones that take part in regulation of gene expression were obtained from the GenBank. Sequence alignment and comparative analysis was performed using ClustalW and BioEdit. Protein structure was obtained from Pfam and PDB databases. Experiments were carried out with *Saccharomyces cerevisiae* strain BY4742 (Invitrogen). A set of reporter constructions was created using methods of molecular cloning and site-directed mutagenesis. Transcription rates were estimated using reverse transcription real time PCR technique with fluorescent TaqMan probes. Reporter protein translation was monitored using fluorescence microscopy. Amount of proteins was compared using immunoblotting followed with densitometry.

Results and Conclusion: Four tandem repeats (CAAU) in 5'-UTR from *CUP1* mRNA are essential for translation regulation. We have shown that *CUP1* 5'-UTR participates in Hsp104 mediated post-transcriptional regulation of gene expression. We have also found the potential domains that involved in Hsp104 – mRNA interaction and proposed the hypotheses of Hsp104 mediated regulation of gene expression in yeast *Saccharomyces cerevisiae*.

Acknowledgements: The work was supported by FTP "Academic and teaching staff of innovative Russia" of The Russian Ministry of Education and Science.

Reference:

1. Santanu Raychaudhuri S et al. (2006) Zuotin. A DnaJ molecular chaperone, stimulates cap-independent translation in yeast // Biochem. and Biophys. Res. Com. V.350, P.788-795.
2. Gallie DR et al. (1987) The 5'-leader sequence of tobacco mosaic virus RNA enhances the expression of foreign gene transcripts in vitro and in vivo // Nucl. Acids Res. V.15, P.3257-3273.
3. Rubel AA et al. (2008) Yeast chaperone Hspl04 regulates gene expression on the posttranscriptional level // Mol. Biol. T.42, №1 C.123-130.

QUANTIFYING, MODELING AND ANALYZING SELF-ORGANIZATION PATTERNS IN BACTERIAL SWARMS

H. Zhang, C. Xie and O.A. Igoshin*
Department of Bioengineering, Rice University, Houston, TX, US
** Corresponding Author: igoshin@rice.edu*

Key words: *pattern formation, biofilms, agent-based model*

Motivation and Aim: Myxococcus xanthus is a model bacteria famous for its coordinated cellular behavior resulting in formation of various multicellular structures. Examples of these structuresinclude fruiting body – aggregates in which bacteria sporulate under starvation conditions and ripples – dynamical bacterial density waves propagating through the colony during predation. Relating these complex self-organization patterns in *M. xanthus* swarms to motility of individual cells is a complex-reverse engineering problem that cannot be solved solely by traditional experimental research.

Methods and Algorithms: Our group addresses this problem with a complementary approach – a combination of agent-based modeling and biostatistical image quantification. To illustrate our approach we discuss our methods of quantifying emergent order in developmental aggregation under starvation conditions and discovering features that affect the aggregation dynamics.

Results: Quantification of the dynamical of changes in aggregate sizes and numbers revealed several stages of developmental process. With a variety of image processing techniques we extract various features of the aggregation patterns and show how this can be used to obtain new biological insights. A novel method to quantify randomness in the distribution of the aggregates on the plate showed that the distribution becomes progressively less random with increasing developmental time. This ordering of aggregates is under genetic control as some mutant strains lack the ordering phase. To further understand the mechanism of ordering we implemented an agent-based model to study the patterns emergent from intercellular interactions between *M. xanthus* cells. Our model based on reduction of cell movement in high density can reproduce several features of the experimentally observed aggregation dynamics. However, the model does not reproduce later stages of aggregates ordering.

Conclusion: These results indicate changes of intercellular interactions or cell motility must be associated with later aggregation stages. The developed model can also be tuned to reproduce other patterns observed during swarming, development and predation in *M. xanthus* swarms.

Acknowledgements: This work is in collaboration with labs of Profs. Roy Welch (Syracuse University) and Lawrence Shimkets (University of Georgia). The work is supported by NSF CAREER Award to OAI (MCB-0845919).

BISTABILITY AND LOW-PASS FILTERING IN THE NETWORK MODULE DETERMINING BLOOD STEM CELL FATE

J. Narula[1], A.M. Smith[2], B. Gottgens[2] and O.A. Igoshin[1*]
1 Department of Bioengineering, Rice University, 6100 Main St. Houston, TX, US
2 Cambridge Institute for Medical Research, University of Cambridge, Hills Road, Cambridge CB2 0XY, United Kingdom
*** Corresponding Author: igoshin@rice.edu**

Key words: *gene regulation, networks, hematopoeiesis, bistability*

Motivation and Aim: Combinatorial regulation of gene expression is ubiquitous in eukaryotes with multiple inputs converging on regulatory control elements. The dynamic properties of these elements determine the functionality of genetic networks regulating differentiation and development. Here we develop a method quantitatively characterize the regulatory output of distant enhancers and apply it for dynamical modeling of the network determining blood stem cell development.

Methods and Algorithms: The regulatory output of distant enhancers is derived from a biophysical approach that computes the occupancy of various promoter states from statistical thermodynamics. The parameters – interaction free energies of protein-protein and protein-DNA interactions – are recursively computed from experimental data on transcriptional reporter libraries.

Results: We apply this method to model the Scl-Gata2-Fli1 triad – a network module important for cell fate specification of hematopoietic stem cells. We show that this triad module is inherently bistable with irreversible transitions in response to physiologically relevant signals such as Notch, Bmp4 and Gata1 and we use the model to predict the sensitivity of the network to mutations. We also show that the transitions between steady states of the triad require signals to persist for longer than a minimum duration threshold thereby acting as a low-pass filter. We have found the autoregulation loop connecting the slow-degrading Scl to Gata2 and Fli1 to be crucial for this low-pass filtering property.

Conclusion: Taken together our analysis not only reveals new insights into hematopoietic stem cell regulatory network functionality but also provides a novel and widely applicable strategy to incorporate experimental measurements into dynamical models of transcriptional networks. Moreover proposed mechanism of gene regulation via distant enhancers shows surprising flexibility in generating logical input functions.

DROSOPHILA MELANOGASTER mtDNA HAPLOGROUPE DIVERSITY DEPENDS ON WOLBACHIA.

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

Key words: *mtDNA, Wolbachia, Drosophila melanogaster, population.*

Motivation and Aim: Wolbachia are intracellular bacteria found in many species of arthropods and nematodes. They are maternally inherited and capable manipulate the reproduction of their arthropod hosts in various ways, may play a role in host speciation and have potential applications in biological pest control. These endosymbionts may also affect the evolution of their hosts in more subtle ways, having specific effects on the population genetics and molecular evolution of host mitochondrial DNA. The mtDNA will be forced through a bottleneck of one host female, from which all mtDNA haplotypes in the population will be descended. Consequently, infected populations may have lower mtDNA diversity (1).

Methods and Algorithms: It was sequenced a 2757-bp region of the mtDNA genome, which contains *At6*, *At8*, three tRNA and partial *cox1* and *cox3* genes. Taking into account the strictly maternal transfer of the endosymbiont *Wolbachia* and the coinheritance with mitochondria, it was an attempt to describe mtDNA polymorphism as dependent on the infection status.

Results: In our study it has been demonstrated inheritance association in *D.melanogaster* between mitotypes and 4 genotypes - wMel, wMel2, wMelCS, wMelCS2 (2) of strain wMel *Wolbachia*.

Conclusion: In paper are discussed mtDNA diversity and some aspects evolution history of *D.melanogaster.*

References:

1. M. Turelliet al. (1992) Dynamics of cytoplasmic incompatibility and mtDNA variation in natural *Drosophila simulans* populations. Genetics **132:** 713–723.
2. M. Riegler et al. (2005) Evidence for a global Wolbachia replacement in *Drosophila melanogaster.* Curr Biol. **15:** 1428-1433.

MATHEMATICAL MODEL OF CELL VOLUME REGULATION IN RESPONSE TO HYPOTONIC SHOCK

A.V. Ilyaskin*[1], D.A. Medvedev[2], A.P. Ershov[2], G.S. Baturina[1], E.I. Solenov[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Lavrentyev Institute of Hydrodynamics SB RAS, Novosibirsk, Russia*
e-mail: ilyaskin.alex@yahoo.com
**Corresponding author*

Key words: *cell volume, membrane potential, ion fluxes, RVD, mathematical models*

Motivation and Aim: Renal collecting duct principal cells perform vasopressin regulated water reabsorption and form the final composition of egested urine. The osmotic pressure of the extracellular fluid varies significantly. To maintain viability and integrity in hypotonic conditions cells have mechanisms for regulatory volume decrease (RVD). Despite its importance, very little is known about the volume regulation of renal collecting duct cells.

The purpose of this study was to investigate the time course of cell volume changes in response to hypotonic shock and to create a mathematical model of renal collecting duct principal cells reaction to the change of extracellular osmolarity on the basis of the experimental data.

Methods: Changes of the cell volume were measured by fluorescent dye Calcein quenching method. Numerical integration of model differential equations was carried out by the Runge-Kutta fourth order method.

Results: The change of extracellular osmolarity in experiments leads to the rapid cell swelling which is followed by regulatory volume decrease (RVD) and recovery of the initial level of cell volume. After the second consequent hypotonic shock the cell behavior is close to the behavior of ideal osmometer.

On the basis of foregoing experimental data the mathematical model of a dynamic cell reaction to the hypotonic shock was created. This model calculates the time dependence of cell volume, transmembrane potential and intracellular amount of osmolytes such as Na^+, K^+, Cl^- and hypothetical organic anions. The quantitative estimation of the intracellular osmolyte loss during the RVD was made along with the calculation of the contribution of different membrane transport systems (channels, cotransporters, Na/K-pump) to this process.

Conclusion: Analysis of the model shown that there are several essential RVD mechanisms: 1) the increase of membrane K^+ and Cl^- permeability; 2) the activation of KCl-cotransporters; 3) the appearance of organic anions efflux; 4) the decrease of membrane osmotic water permeability. The simultaneous activation of all these mechanisms allows the cell to decrease the volume rapidly and saves it from the excessive swelling when the hypotonic shock is repeated.

ON THE GENERAL THEORY OF VARIABILITY

S.G.Inge-Vechtomov
Dept of Genetics and Breeding, St.-Petersburg State University, S.-Petersburg, Russia
e-mail: ingevechtomov@gmail.com

key words: *variability classification, interdependency of inherent and non-inherent variability, epigenetics, mutation, modification, template principle*

Variability is the most specific character of the living things. Nevertheless, there is no general theory of variability and the widely accepted contradictory phenomenological classification of variations is a common source of misunderstanding among biologists. We need clearer definition of mutation, which would shear it from recombination (homologous, non-homologous, illegitimate etc), genomic variations caused by cytoskeleton modifications, a wide range of epigenetic phenomena (which also have no satisfactory definition) etc. It is evident, that terms: transposon mutagenesis, chromosome and genome mutations contain obvious contradictions, the same as ontogenetic variations, which overlap with mutations, recombination and modifications through regulation of expression of genetic information during differentiation.

It is possible to avoid these problems basing upon the mechanisms of variability. Than we shall follow the template principle in biology, proposed by N.K.Koltsov (1928)[1] for chromosome replication and embodied by F.Crick (1958, 1970)[2] as the Central dogma of molecular biology. Some characteristics are common for all three template cell processes: replication, transcription, translation. All tree stages of the processes – initiation, elongation (or copying), termination are governed by Watson-Crick complementary interactions. It opens a possibility for overlap between functional domains of molecular machines responsible for stability and replication of genetic material and for the expression of genetic information. All three template processes possess of common characteristics: (1) polyvariancy (multiplicity, variability) of the copying molecular machines, which leads to ambiguity of the template processes[3], and (2) ability to repair or correct produced polymers. The balance of these two characteristics serves for evolutionary optimization of their variability levels. This idea permits to consider so-called inherent and non-inherent variability from the single point and to offer an alternative classification of variability types.

In addition to the first order or sequence templates mentioned above, we must consider the second order or conformational templates, responsible for prion (protein) inheritance in yeast

It becomes evident that connection and interdependency between different types of variability is not the matter of chance: there are common steps between mutations and some modifications through phenotypic expression of primary lesions of genetic material; genome duplications in evolution are dependent on modifications of cytoskeleton proteins; events such as SOS repair etc. Epigenetic variations and inheritance may be also considered from the same position as was shown by Tchuraev (2005)[4]. Conformational templates, must be discussed in the same context. Evolutionary significance of the variation types must be reconsidered.

In conclusion, the same mechanisms may underline both inherent and non-inherent variations depending on a systematic position and a life cycle stage of the organism under investigation. Absence of direct correlation between molecular mechanisms and phenomenology of variability suggests a biological indefinity principle.

1. N.K.Koltsov (1936) Inherent molecules. In: Organization of the Cell. Gos. Izd. Biol. Med. Moscow - Leningrad: 585-620. (Russ.)
2. F. Crick(1970) Central dogma of molecular biology. Nature. 227. 561-563.
3. S.G.Inge-Vechtomov (1976) Polyvariancy principle for template processes. In: “Researches in Genetics”. Leningrad State Univ. 7. 7-19. (Russ.)
4. R.N.Tchuraev (2005) The frame of noncanonical theory of heredity: from genes to epigenes. J. of General Biol. 66. 99-122. (Russ.)

GENETIC DISSECTION OF VARIOUS DNA REPAIR FUNCTIONS OF THE AP ENDONUCLEASES REVEALS THEIR RESPECTIVE BIOLOGICAL ROLES

A.A. Ishchenko[1*], A. Gelin[1], M.K. Saparbaev[1], N.A. Timofeyeva[2], O.S. Fedorova[2]

[1]*Groupe «Rüparation de l'ADN», UMR8200 CNRS, Institut Gustave Roussy, 94805 Villejuif Cedex, France*
[2]*Institute of Chemical Biology and Fundamental Medicine Lavrentyev Ave. 8, Novosibirsk 630090, Russia*
e-mail: ichtchen@igr.fr
[*]*Corresponding author*

Key words: *nucleotide incision repair, base excision repair, AP endonuclease APE1, 3′→5′-exonuclease*

Motivation and Aim:

Oxygen-derived species including free radicals can produce a variety of modifications in DNA including base and sugar lesions, strand breaks, abasic sites and clustered lesions. Oxidized DNA bases are substrates for two overlapping pathways: base excision repair (BER) and nucleotide incision repair (NIR). The major human AP endonuclease APE1 is multifunctional enzyme involved in both BER and NIR pathways. At present, the biological role of APE1-catalyzed NIR and 3'-5' exonuclease activities remains poorly understood. In this study, we have attempted to genetically separate AP endonuclease and nucleotide incision activities of APE1 to establish NIR as a distinct and separable function of AP endonucleases.

Results and Conclusions:

Biochemical characterization of the APE1 mutants carrying specific amino acid substitutions has demonstrated deficiencies in both NIR and 3'-5' exonuclease activities, suggesting that these two functions are governed by the same amino acid residues. Furthermore, expression of the APE1 mutant proteins in a repair-deficient strain of *E. coli* complemented its hypersensitivity to alkylation but not to oxidative DNA damage. The differential drug sensitivity of the mutants suggests that the NIR pathway removes lethal DNA lesions generated by oxidizing agents. We show that in living cells a major human AP endonuclease, Ape1, incises DNA containing alpha-anomeric 2'-deoxyadenosine indicating that the intracellular environment of human cells supports NIR activity. In vitro reconstitution of the human NIR pathway have established requirement of APE1, FEN1, DNA polymerase β and DNA ligase I to fully repair alpha-2'-deoxyadenosine residue in DNA. Finally, our data further substantiate that NIR is a distinct and separable function of the human AP endonuclease essential for handling lethal oxidative DNA lesions.

Acknowledgements: This work was supported by grants from "Centre National de la Recherche Scientifique" of France and Russian Ministry of Education and Science (State Contract 02.740.11.5012).

PECULIARITIES OF INTERACTION MIRNA WITH MRNA OF SOME ONCOGENES

O.A. Berillo, A.S. Isabekova, V.A. Khailenko, A.T. Ivachshenko*
Kazakh National University named al-Farabi, Almaty, Kazakhstan
e-mail: a_ivashchenko@mail.ru
**Corresponding author*

Key words: *2D mRNA, miRNA, oncogenes, exon, intron, 3′UTR*

Motivation and Aim: The majority of miRNAs regulate expression of genes in eukaryotic cells at mRNA translation stage. miRNAs promote splitting of mRNA on fragments, reduce speed of translation, or activate translation process considerably [1, 2]. Usually miRNAs co-operate with 3′UTR and 5′UTR. However their interaction with introns in pre-mRNA and with exons of mRNA is revealed. We studied peculiarities of miRNA interaction with two-dimensional (2D) structure of mRNA.

Methods and Algorithms: 2D structure of both pre-mRNA and mRNA was built under the program UNAFold.3.7 (http://dinamelt.bioinfo.rpi.edu). Interaction sites of 2D mRNA with miRNAs were extracted from microRNA.org (http://www.microrna.org). Interaction of miRNAs with mRNA of *BAX*, *FASLG*, *FTL*, *GAST*, *TNFRSF17* oncogenes has been studied.

Result: Principal causes of cancer development are modification of gene expression, also at level of translation processes [1]. It was revealed, that all miRNAs co-operate with mRNA in sites which have unpaired nucleotides. On the average the number of unpaired nucleotides in sites of 2D structures of mRNA, connecting miRNAs, makes 35-60% out of the number of pairs formed at of miRNAs with mRNA interaction. It provides stronger linkage of miRNAs with mRNA. In some cases miRNAs co-operate with both complement sites of 2D mRNA that increases significant change 2D structures of mRNA. In mRNA of *BAX*, *FASLG*, *FTL*, *GAST* and *TNFRSF17* oncogenes there are sites connected with 10, 23, 3, 3 and 5 miRNAs respectively. Hence, the probability of updating of gene expression increases accordingly. mRNA of *BAX* and *FTL* genes have miRNAs linkage sites in exons. Six miRNAs contact with exon-6 and two miRNAs contact with exon-7 of *BAX* mRNA. mRNA of *FTL* gene contains six sites of miRNA linkage in exon-3 and 11 sites in exon-4. In exon-3 these sites are located in the site adjoining exon-4. Such features of interaction of miRNAs with mRNA of these genes can cause translation of truncated proteins. It means that to proteins synthesised on mRNA and received as a result of alternative splicing, proteins are added as products of alternative translation. Some miRNAs can influence splicing as they contact intron near to splicing sites. So, 7 miRNAs contact with intron-3 at the site splicing exon-3/intron-3 and intron-3/exon-4 of *TNFRSF17* pre-mRNA. In some genes the sites of interaction mRNA with miRNAs revealed the increase of CUG triplets which are specific for the restriction RNA.

Conclusion: Translation of mRNA of *BAX*, *FASLG*, *FTL*, *GAST* and *TNFRSF17* oncogenes can be essentially modified by molecules of miRNAs, which interact with 3′UTR, exons and introns. The obtained data promote the creation of oncodiagnostic and oncotherapeutic methods.

References:

1. I. L.de Silanes et al. (2007) Aberrant regulation of messenger RNA 3'-untranslated region in human cancer, *Cellular Oncology*, **29**: 1–17.
2. T.H.Beilharz et al. (2009) microRNA-mediated messenger RNA deadenylation contributes to translational repression in mammalian cells, *PLoS One*, **4**: e6783.

INTRONS OPTIMISE NUCLEOTIDE CONTENT OF GENES

A. Ivachshenko*, A. Kabdullina, V. Khailenko, A. Asheulov, S. Atambayeva
Kazakh National University named al-Farabi, Almaty, Kazakhstan
e-mail: a_ivashchenko@mail.ru
**Corresponding author*

Key words: *DNA, RNA, genes, exon, intron, nucleotides*

Motivation and Aim: Nucleotide content of single strand DNA in virus, organelle, archaebacteria, eubacteria and eukaryote genomes is evaluated by the rule: fC/fG-fA/fT = *a*(fY-0,5) + *b*, where fT, fC, fA, fG are frequencies of nucleotides, *a* and *b* - regression parametres [1]. In single strand DNA with the length more than 50 Kbp the value of fC/fG-fA/fT is close to zero that corresponds to 2nd Chargaff's rule. In protein-coding intronless genes the value of fC/fG-fA/fT can differ from zero considerably. It reduces possibility of pre-mRNA and mRNA to form 2D structure with maximum A-U and G-C pairs. Introns change the value of fC/fG-fA/fT to zero and make 2D structure mRNA more stable. The purpose of the study is to reveal how introns reduce the value of fC/fG-fA/fT in eukaryote genes.

Methods and Algorithms: Nucleotide sequences genomes *Homo sapiens*, *Monodephis domestica*, *Plasmodium falciparum* and *Teileria parva* were extracted from GeneBank (http://www.ncbi.nlm.nih.gov). From each genome were formed samples of genes with 1, 2, 3, 4, 5, 6-9, 10-14, 15 and more introns. Calculations were performed by means of program Arabella Calculator-2007.

Result: In *H. sapiens* genes with 1-2 introns the value of fC/fG-fA/fT for the exon sum was equal -0.465 and was decreased by introns to -0.033, in genes with 6-9 introns the value of fC/fG-fA/fT decreased from -0.384 to 0.025. In genes with the higher intron number the value fC/fG-fA/fT was close to zero. It should be noted that the ratio of average intron lengths and exon lengths in *H. sapiens* genes was dependent on genes density in a chromosome and changed from 10 to 67 [2]. In *M. domestica* genome introns decrease the value of fC/fG-fA/fT in a gene to a less degree, than in *H. sapiens* genes. In genes with 1-2 introns the value of fC/fG-fA/fT decreased from -0.499 to -0.283, and in genes with 6-9 introns - from -0.243 to -0.038. The ratio of average intron lengths and exon lengths in *M. domestica* genes changed from 15 to 40. In unicellular eukaryote genomes introns also reduced the value of fC/fG-fA/fT. In one-intronic *P. falciparum* genes the value of fC/fG-fA/fT was -0.535, and in exons was -0.705. At 15 and more introns there in a gene they reduced the value of fC/fG-fA/fT from -0.711 to -0.302. In *T. parva* genomes the decrease of the value of fC/fG-fA/fT under introns influence is shown to less degree. In one-intronic genes the value of fC/fG-fA/fT in exons decreased from -0.373 to -0.286, and in genes with 10-14 introns from -0.332 to -0.196. The ratio of average intron lengths and exon lengths in *P. falciparum* and *T. parva* genes changed from 0.22 to 0.69 and from 0.22 to 0.40.

Conclusion: Introns in protein-coding genes of *H. sapiens*, *M. domestica*, *P. falciparum* and *T. parva* reduce the absolute value of fC/fG-fA/fT. This intron effect is shown more strongly in genes with the large intron number.

References:

1. Z.D.Kramskova et al. (2002) Four-nucleotide rule. Viral genomes. 3-d International conference in bioinformatics of genome regulation and structure. Novosibirsk, 167-169.
2. A.T.Ivashchenko et al. (2009) Variations of the length of exons and introns in human genome genes. *Russian Journal of Genetics*, **45**: 16-22.

ITMSYS: AN INTERACTIVE WEB-BASED TEXT-MINING SYSTEM FOR AUTOMATED ANALYSIS OF THE FULL-TEXT ARTICLES

Ivanisenko T.V.*[1], Demenkov P.S.[1], Ivanisenko N.V.[2], Ivanisenko V.A.[1]
[1]*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia.*
[2]*Novosibirsk State University, Novosibirsk, Russia.*
e-mail: slayer@bionet.nsc.ru
**Corresponding author*

Motivation and Aim: Reconstruction of the genetic networks requires an analysis of the huge number of scientific publications. It is necessary to extract various facts about molecular-genetic, genetic-genetic, metabolite-genetic and other types of interactions from such publications. Manual analysis of articles by experts has high degree of accuracy but is very time-consuming therefore it makes timely the task of development of interactive tools for analysis of a full-text articles which can be used for reconstruction of the genetic networks. The aim of this work was a development of a system that would allow expert to identify and display found names of different types of objects in text with various types of interactions between them.

Methods and Algorithms: Automated extraction of information about molecular-genetic, genetic-genetic, metabolite-genetic and other types of interactions in text was performed using the text-mining methods we developed. For text-mining, we also used the thesaurus we previously developed for the names of proteins, genes, metabolites, diseases, microRNAs, biological pathways, cells and organisms. A web-based interface allows text-mining module to work with different types of files (such as *.pdf, *.doc or *.txt) uploaded by experts and to demonstrate results of data extraction in user friendly way as well.

Results: An interactive web-based text-mining system for automated analysis of the full-text articles (ITMSys) was developed. The ITMSys software is equipped with tools for reconstruction and visualization of genetic networks and web-based interface. The ITMSys ensures analysis of full-text articles connected with genetic networks.

Conclusion: Developed system can be used by experts for the acceleration of genetic networks reconstruction process, as well as in the other fields of science related to automated analysis of the scientific texts.

Acknowledgements: Work was supported in part by state contract FASI №02.514.11.4123, interdisciplinary integration project SB RAS №111, №94 and program RAS 22. “Molecular and cellular biology”.

ANDCell AND ANDNanobiotech: ASSOCIATIVE NETWORK DISCOVERY SYSTEMS IN SYSTEMS BIOLOGY AND NANO-BIOTECHNOLOGY.

V.A. Ivanisenko*, P.S. Demenkov, T.V. Ivanisenko, N.L. Podkolodny
PBSoft ltd, Novosibirsk, Russia
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: salix@bionet.nsc.ru
**Corresponding author*

Computational systems (ANDCell and ANDNanobiotech) for automated extraction of knowledge from scientific texts (PubMed abstracts) and databases were developed. The ANDCell is targeted on the reconstruction of semantic associative networks in systems biology (molecular genetic interactions, gene regulation events, catalytic processes, proteolysis, polymorphic gene – disease associations, etc.). The ANDNanobiotech is targeted on nanobiotechnology (drug-delivery systems, labeling, diagnostics, treatment, microfluidics, implants, grafts, biosensors among others.). Information is extracted with the aid of original text-mining technology. The ANDCell knowledge database contains about 5 millions of facts, the ANDNanobiotech knowledge database contains about 1,5 millions of facts. The ANDVisio program provides the access to the knowledge databases and the representation of the results in a graphic form of reconstructed associative networks. The vertices of such networks are the molecular genetic objects (genes, proteins, microRNAs, metabolites, cellular components), diseases, processes, nanomaterials and nanobioconstructs while the edges between the vertices represent various types of associations. For molecular interactions and associations, data on the cell types and organisms are represented. The systems provide a user's friendly interface implemented links to the molecular-genetic databases and the articles from which information was extracted. The developed systems may be useful for resolving a wide range of tasks in biology, biomedicine and nanobiotechnology.

Availability: The ANDCell and ANDNanobiotech systems are available at request to the authors. Work was supported in part by Government contract FASI №02.514.11.4123, SB RAS interdisciplinary integration grant №111.

SCREENING OF NOVEL LIGANDS FOR HUMAN CYTOCHROME P450(51): INTEGRATION OF VIRTUAL AND SPR TECHNOLOGIES

A.S. Ivanov* [1, 2], O.V. Gnedenko [1], A.A. Molnar [1], A.V. Veselovsky [1], N.V. Adrianov [2], S.A. Usanov [3], A.I. Archakov [1, 2]

1 Institute of Biomedical Chemistry RAMS, Moscow, Russia
2 Russian State Medical University, Moscow, Russia
3 Institute of Bioorganic Chemistry NAS, Belarus
e-mail: alexei.ivanov@ibmc.msk.ru
**Corresponding author*

Key words: *human cytochrome P450(51), novel ligands, molecular docking, virtual screening, surface plasmon resonance, optical biosensor*

Motivation and Aim: The current trend in drug discovery is to integrate computational and experimental technologies in the frame of pipeline "From gene to drug" [1, 2]. Virtual and experimental screening are important components of pipeline and can be used as complementary to each other. The aim of our research was to apply the integration of these technologies in screening of novel ligands for human cytochrome P450(51). This enzyme catalyzes the oxidative removal of the 14α-methyl group (C-32) of the lanosterol, which is an essential reaction in the biosynthetic pathway for cholesterol [3], and is thus a useful target for anti-atherosclerosis drug design.

Methods and Algorithms: Virtual screening was carried out by using molecular docking program Dock 6.0, which docks ligands with flexible conformation into the rigid active site of enzyme. Crystal structure of human P450(51) (PDB code 3I3K) and the internal database containing structures of steroid and triterpenoid compounds available in our laboratory were used. Known P450 inhibitor ketokonazole was utilized as positive control. Experimental screening was carried out by using surface plasmon resonance (SPR) biosensor Biacore T-100.

Results: Virtual screening have resulted in obtaining the list of mostly perspective potential ligands. Some compounds from the top part of the list were included into the set of 22 compounds for experimental screening. SPR analysis of protein-ligand interaction yields 5 potential ligands, which bound with target protein. Among these five ligands there was one from the resulting top list of virtual screening.

Conclusion: Integration of virtual and biosensor screening technologies has allowed us to find some new potential ligands for target human cytochrome P450(51).

Acknowledgments: This work was supported in part by Russian Space Agency (development work "ISS-Science-IBMCh-2007-2009" in frame of contract 851-5383/07) and by Russian Foundation for Basic Research (grants 10-04-00796 and 10-04-90026).

References:

1. Veselovsky A.V., Ivanov A.S. (2003) Strategy of computer-aided drug design, *Current Drug Targets - Infectious Disorders*, **3:** 33–40.
2. Ivanov A.S. Et al. (2006) Bioinformatics platform development: from gene to lead compound. *Methods Mol. Biol.* **316:** 389-431.
3. Pikuleva I.A. (2006) Cytochrome P450s and cholesterol homeostasis. *Pharmacology & Therapeutics,* **112(3):** 761-773.

PROCESSING AND CRYOPRESERVATION OF HUMAN UMBILICAL CORD BLOOD HSC AT STEM CELLS BANK POKROVSKI.

A.B. Smolyaninov, D.A. Ivolgin*, K. Korovina, E. Kotelevskaya. O. Supilnicova
Stem Cells Bank Pokrovski, Saint-Petersburg, Russia
ida59m@mail.ru
**Corresponding author*

Key words: *cord blood hematopoetic stem cells, processing of cord blood, cryopreservation, cells separator "Sepax "*

Motivation and Aim: The study was aimed to establish a standard procedures for collection, separation, enumeration and cryopreservation of cord blood hematopoetic stem cells at the human umbilical cord blood bank.

Methods and Algorithms: 708 samples of cord blood were collected after delivery of infant prior to expulsion of placenta. The average cord blood volume collected was 102.6 ml. All cord blood units were processed by using the Sepax automated method (Biosafe, Switzerland). After finishing all tests and adding of 10% DMSO cord blood units were transferred to a controlled rate freezer (Planer, UK) that was pre-cooled to 4єC. The units were then cooled at 1єC/min to -12єC, cooled at 20єC/min to -60єC, followed by warming of the units at 10єC /min to -18єC, cooled at 1єC/min to -60єC and finally, 3єC/min to -100єC. Then cord blood units were stored in the liquid nitrogen for further application. Seven units of cord blood were thawed after one year of cryopreservation and recovery rate of nucleated cells and CD34(+) cells as well as viability of mononuclear cells were evaluated.

Results: The recovery rate of nucleated cells and mononuclear cells after the processing was 84.63±19,3% and 77,9±21,7% respectively, average number CD 34+ cells was 22,2 x 10^5 per unit, viability count of mononuclear cells was 99.31%. The recovery rate of nucleated cells and CD34(+) cells after thawing were 95.5±3 % and 85,2±23 % respectively and the viability count on revival of mononuclear cells was 85,9%.

Conclusion: It is suggested that the methods reported for processing and the algorithm of cryopreservation of human umbilical cord blood-derived hematopoetic stem/progenitor cells is effective.

TRANSGENIC MOUSE MODELS FOR RHEUMATOID ARTHRITIS

Yoichiro Iwakura
Center for Experimental Medicine and Systems Biology, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan
e-mail: iwakura@ims.u-tokyo.ac.jp

Motivation and Aim: Rheumatoid arthritis (RA) is a systemic, chronic inflammatory disorder that mainly affects joints. The disease is autoimmune in nature, and high levels of proinflammatory cytokines including IL-1, IL-6, IL-17, and TNF are detected in the affected joints.

Methods and Algorithms: We have generated two RA models, human T-cell leukemia virus type I (HTLV-I) transgenic mice and IL-1 receptor antagonist (Ra)-deficient (KO) mice, to elucidate the pathogenic mechanisms of the disease. Both models spontaneously developed arthritis with elevated autoantibodies.

Results: The histopathology showed marked synovial and periarticular inflammation with articular erosion, closely resembling that of RA in humans. We found that TNF-, but not IL-6-, deficiency suppressed development of arthritis in IL-1Ra KO mice, while IL-6 but not TNF was involved in the HTLV-I transgenic mouse model. IL-17 was important for both models.

We found that the expression of C-type lectin receptor (CLR) genes was augmented in the affected joints of these models using DNA microarrays. Dendritic cell immunoreceptor (DCIR) is one of such CLRs with a carbohydrate recognition domain in their extracellular carboxy terminus and an ITIM in its intracellular amino terminus. Because human shared syntenic locus containing the *Dcir* gene is linked to several autoimmune diseases including RA and SLE, we have generated *Dcir* KO mice to examine the roles of this gene in the immune system. We found that aged *Dcir* KO mice spontaneously developed sialadenitis and enthesitis associated with elevated serum autoantibodies. DCs were excessively expanded in *Dcir* KO mice after aging. *Dcir* KO mouse–derived bone marrow cells (BMCs) differentiated into DCs more efficiently than did wild-type BMCs upon treatment with GM-CSF, owing to enhanced STAT-5 phosphorylation. These findings indicate that DCIR is crucial for maintaining the homeostasis of the immune system, suggesting that Dcir is one of novel targets for the treatment of RA.

MODELING AND IN VIVO IMAGING OF MOLECULAR AND MECHANICAL REGULATIONS OF PLANT DEVELOPMENT

H. Jönsson*
Computational Biology & Biological Physics, Lund University, Lund, Sweden
e-mail: henrik@thep.lu.se
**Corresponding author*

Key words: *CLAVATA, WUSCHEL, PIN1, auxin, systems biology*

Motivation and Aim: The shoot apical meristem (SAM) is a stem cell niche acting as a main regulator of above-ground plant development. The maintenance of stem cells in the SAM, and the initiation of organs at the periphery are dependent on genetic regulation, hormone signaling, and mechanical anisotropies resulting in a complex dynamical system regulating organized differentiation and growth. Mathematical modeling has proven to be a useful tool to understand SAM development at a systems level, and the use of live microscopy has increased our knowledge of details in the protein dynamics associated with SAM development. This allow for creating computational models where microscopy data provides experimental templates for optimizing the models, and the models can introduce experimentally verifiable predictions.

Methods and Algorithms: We use confocal microscopy data as templates to our computational models that integrate Finite Element methods for the mechanical parts, and ordinary differential equation solvers for the molecular parts. We us simulated annealing algorithms to infer model parameters for the networks by comparing simulations with experimental data.

Results: New models for stem cell regulation and primordia formation in the SAM will be presented. Examples are models where hormone signaling and transport are combined with mechanical models for generating phyllotactic patterns [1,2] as well as models for stem cell regulation focusing on the CLAVATA WUSCHEL feedback[3]. Further, model predictions originating from parameter optimizations of the models will be presented, and our method of treating ensembles of parameter value sets will be discussed.

Acknowledgements: The work is in part supported by the Swedish Research Council.

References:

1. H. Jönsson, M. Heisler, B.E. Shapiro, E.M. Meyerowitz, and E. Mjolsness (2006)
An auxin-driven polarized transport model for phyllotaxis *Proceedings of the National Academy of Sciences USA* **103**, 1633-1638 (2006)
2. O. Hamant, M. Heisler, H. Jönsson, P. Krupinski, M. Uyttewaal, P. Bokov, F. Corson, P. Sahlin, A. Boudaoud, E. M. Meyerowitz, Y. Couder, and J. Traas (2008)
Developmental patterning by mechanical signals in Arabidopsis *Science* **322**, 1650-1655
3. H. Jönsson, M. Heisler, G.V. Reddy, V. Agrawal, V. Gor, B.E. Shapiro, E. Mjolsness, and E.M. Meyerowitz (2005)
Modeling the organization of the WUSCHEL expression domain in the shoot apical meristem *Bioinformatics* **21**(Suppl. 1), i232-i240

MODELING OF INTERPLAY OF MECHANICAL AND BIOCHEMICAL INTERACTIONS IN MORPHOGENESIS

P. Krupinski[3], M. Heisler[1,4], O. Hamant[2], M. Uyttewaal[2], C. Ohno[1,4], J. Traas[2], E.M. Meyerowitz[1], H. Jönsson[3*]

[1]Division of Biology, California Institute of Technology, Pasadena, California 91125, USA
[2] INRA, CNRS, ENS, Universitũ de Lyon, 46 Allũe d'Italie, 69364 Lyon Cedex 07, France
[3]Computational Biology and Biological Physics Group, Department of Theoretical Physics, Lund University, S-221 00 Lund, Sweden
[4]Present address: European Molecular Biology Laboratory, Heidelberg, Germany
e-mail:henrik@thep.lu.se
**Corresponding author*

Key words: *mechanical modeling, auxin transport, PIN1, morphogenesis, phyllotaxis*

Motivation and aim: Formation of specialized organs in plants requires, on the mechanical level, correlated regulation of the amount and direction of cellular growth. The former is influenced by polarized transport of the plant hormone auxin, while the latter is related to the directionality of the microtubule array. We use the combination of experimental and modeling techniques to elucidate how mechanical signals affect both microtubule alignment and auxin transport and thus contribute to emergence of phyllotactic patterns.

Methods and algorithms: We compare live imaging data of auxin patterning and microtubule cytoskeleton formation in Arabidopsis, with results of computer simulations involving a Finite Element Model of mechanical interactions and biochemical model of auxin transport in the Shoot Apical Meristem (SAM).

Results: The essential part of auxin-induced localized primordia formation is directional transport of the hormone directed by polarization of its efflux mediator, PIN1 membrane protein. Another factor contributing to morphogenesis of the SAM is anisotropy of the growth, which is influenced by alignment of the microtubules. We demonstrate that both PIN1 cellular redistribution and organization of microtubules respond to mechanical stress. We propose a model of the auxin transport in the SAM, where mechanical stresses in the cell walls convey information about auxin concentration in the neighboring cells. We show that such model is capable of forming phyllotactic patterns and is consistent with experimental results of cell ablations.

Conclusions: Plant organ morphogenesis involves coordination between biochemical and mechanical interactions in the SAM. Our results suggest that mechanical signals are not only passively influenced by auxin patterning, but also actively direct transport of auxin using mechanical stress as a common regulator of PIN1 localization and mechanical anisotropy.

Acknowledgments: We want to acknowledge the support from Swedish Research Council

UNRAVELING THE COMPLEXITY OF PRIMARY AND METASTATIC EWING'S SARCOMA USING HELICOS SINGLE MOLECULE SEQUENCING

P. Kapranov[1]*, R.J. Arceci[2], J. Buckley[3], G. Reaman[4], P. Reynolds[5], P. Sorensen[6], J. Thompson[1], P. Milos[1]*, T. Triche[3]*;

[1]Helicos BioSciences Corporation, Cambridge, MA, United States, 2John Hopkins University, Baltimore, MD, United States, [3]University of Southern California, Los Angeles, CA, United States, [4]Children's Oncology Group, Bethesda, MD, United States, [5]Texas Tech University, Lubbock, TX, United States, 6University of British Columbia, Vancouver, BC, Canada.

e-mails: pkapranov@helicosbio.com, pmilos@helicosbio.com, triche@usc.edu

[*]Corresponding authors

Motivation and Aim: Ewing sarcoma is a high-risk childhood cancer that represents a major treatment challenge, as survival has not improved significantly despite aggressive chemotherapy. Elucidation of molecular changes in progression of this cancer towards the metastatic and chemo-resistant states offers the potential to better understand the fundamentals of biological processes responsible for poor disease outcome.

Methods and Algorithms: We have applied Helicos single-molecule sequencing to build unbiased whole-genome profiles of the transcriptome, methylome and DNA of two cell lines derived from the primary tumor of a Ewing sarcoma patient and her chemo-resistant metastasis. This simple system is ideal for generation of hypotheses on relationships between coding and non-coding RNAs, the epigenome, primary DNA sequence, and the malignant phenotype. Unlike work on primary tumors, it also allows for biologic validation.

Results: Large differences between the primary and metastatic cells exist at both the genomic and transcriptome levels, including differential expression and methylation as well as copy number and sequence variants suggesting clonal evolution from the primary to the metastatic tumor. A significant fraction of the genome was found to represent RNAs expressed at different levels between the cell lines, of which ~60% corresponded to un-annotated transcripts. In some instances, the latter could span large (100's of kb) genomic regions that contain no annotated transcripts. Specific examples of the non-annotated and presumably non-coding RNAs differentially regulated during the metastatic progression will be discussed. About 25% of the transcripts upregulated in the metastatic clone were found to be associated with promoter demethylation when compared to the primary tumor, suggesting that promoter demethylation is specifically associated with up-regulation of transcripts in the metastatic/chemo-resistant tumor in a selective manner.

Conclusion: The fundamental changes occurring during the metastatic progression are very complex. Understanding of these processes at the basic levels requires a combination of approaches aimed at measuring different types of molecules and modifications of molecules in a cell. Availability of such data creates unprecedented opportunities to discover hitherto unknown relationships and yet provides significant challenges to our ability to integrate and analyze diverse data sets and extract biologically-relevant information from them.

INSIGHT INTO POPULATION HISTORY, EVOLUTION, AND DEMOGRAPHIC EVENTS IN NORTHWESTERN SIBERIA: IDENTITY AND INTERACTION.

T.M. Karafet[1,2]*, L.P. Osipova[1], Y.V. Kuzmin[3], and M.F. Hammer[2]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia;*
[2] *University of Arizona, Tucson, USA;*
[3]Institute of Geology & Mineralogy *SB RAS, Novosibirsk, Russia*
e-mail: tkarafet@email.arizona.edu
[*]*Corresponding author*

Motivation and Aim: Recent publications [1-3] have put forth highly controversial conclusions about the origin, the age and dispersal of Y-chromosome haplogroup N – the major lineage in Northwest Siberia. Our study examines Y chromosome genetic variation to reconstruct the historical events that led to the peopling of Northwest Siberia. Y chromosome results were compared with autosomal SNPs (Single Nucleotide Polymorphism) and mtDNA variation to understand the evolutionary forces at work in the history of Siberia.

Methods: Y-chromosome and autosomal SNPs and STRs (Short Tandem Repeats) were typed by both standard methods and by Sequenom. We sequenced 779bp of the mtDNA cytochrome oxidase 3 gene (CO3). The software package ARLEQUIN 3.0 was used to calculate population parameters and genetic distances. Dating estimates of haplogroup age were obtained using the program BATWING. The program STRUCTURE was used to investigate population structure.

Results: In order to reconstruct the history of Y-chromosome haplogroup N we genotyped samples in Siberian populations, as well as groups from Central and East Asia, Mongolia and European Russia with an extensive battery of Y chromosome SNPs and STRs. Included in our set of SNPs are two markers (P89.2 and P301) that have not been surveyed in this geographic region thus far, and provide increased phylogenetic resolution of one of the major lineages (haplogroup N) associated with the peopling of Northwest Siberia. Different patterns of population structure were detected using data from the autosomal, Y-chromosomes, and mitochondrial DNA.

Conclusions: We observed very different geographic distribution of Y-chromosome N haplogroups Taking into account haplotype variation, the age of mutations and archaeological evidence we put forward a colonization model that encompasses early Paleolithic settlements processes, as well as events in historic times. We demonstrated that distinctive marriage structure, sex-specific migration, and polygyny shaped genetic differentiation among populations in Northwest Siberia.

References:

1. S. Rootsi et al. (2007) A counterclockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. Eur J Hum Genet, 15: 211– 405.
2. M. Derenko et al. (2007) Y-chromosome haplogroup N dispersals from south Siberia to Europe. J Hum Genet, 52: 763– 770.
3. S. Mirabal et al. (2009) Y-chromosome distribution within the geo-linguistic landscape of northwestern Russia. Eur J Hum Genet, 17: 1260–1273.

BIOINFORMATIC SEARCH FOR PLANT HOMOLOGUES OF CHECKPOINT SERINE/THREONINE-PROTEIN KINASE BUB1

P.A. Karpov*, A.I. Yemets, Y.B. Blume
Institute of Food Biotechnology and Genomics, Natl. Acad. of Sci. of Ukraine, Osipovskogo str., 2a, *Kyiv* 04123
e-mail: karpov.p.a@gmail.com
**Corresponding author*

Key words: *BUB1, protein kinases, plants*

Motivation and Aim. Mitotic checkpoint serine/threonine-protein kinase BUB1 is essential for spindle-assembly checkpoint signaling and correct chromosome alignment. This kinase plays a key role in assembly of checkpoint proteins (CENPF, BUB1B, CENPE and MAD2L1) on kinetochore. BUB1 is required for kinetochore localization of PLK1 playing important role in defining SGOL1 localization and thereby affecting sister chromatids cohesion. BUB1 acts as a substrate for anaphase-promoting complex or cyclosome (APC/C) in complex with its activator CDH1 (APC/C-Cdh1). BUB1-BUB3 complex participates in inhibition of APC/C (by phosphorylating of CDC20), and phosphorylates MAD1L1. Also, BUB1 mediates cell death, essential during early and later stages of embryonic development, necessary for postimplantation embryogenesis and proliferation of primary embryonic fibroblasts and plays an important role in spermatogenesis and fertility. BUB1 was identified in *Mamalia*, *Fungi* and *Mycetozoa*, but the existence of plant homologs of BUB1 are still in doubt. So, the goal of the present work was the search of potential plant homologues of BUB1.

Methods and Algorithms. The sequences of "reviewed" BUB1 were obtained from UniProt. Domain architecture and functional region information was obtained from the Swiss-Prot and with SMART and Pfam tools. Sequence alignments were performed in ClustalX2. The search of plant homologues was carry out with SIB BLAST (Matrix: BLOSUM62) against the sequence of human BUB1 catalytic domain. Cladistic analysis was carried out using MEGA 4.0 package. Modeling and analysis of spatial structures was carried out in Swiss-PdbViewer V.4.0.1 and Discovery Studio Visualizer 2.5. Protein-protein interactions were analysed with STRING 8.2 tool.

Results. BLASTp scanning of UniProt and GenBank databases against sequence of BUB1 catalytic domain from *Homo sapiens* identified 7 potential plant homologues with 35-36% identity of catalytic (kinase) domains: A9T515 (*Physcomitrella patens*), NP_179656.4 (*Arabidopsis thaliana*), D1HL41 (*Vitis vinifera*), B9FXF7 (*Oryza sativa* subsp. japonica), B4FI36 (*Zea mays*), B9RVW6 (*Ricinus communis*) and C5X9D8 (*Sorghum bicolor*). At the same time, the similarity was 53-56% and 5-9% of gaps. The analysis in SMART and Pfam, confirmed the presence of Mad3_BUB1_I domain and typical for BUB1 domain organization for NP_179656.4 (*A. thaliana*), D1HL41 (*V. vinifera*), B9FXF7 (*O. sativa* subsp. japonica), B4FI36 (*Z. mays*) and B9RVW6 (*R. communis*). Neighbor-Joining (NJ) analysis (based on the profile alignment of catalytic domains) and results of sequence comparison identified D1HL41 (GSVIVT01020577001) from *V. vinifera* as closest plant homolog of human BUB1 (identities=36%, positives=54%). The reconstruction of catalytic domains spatial structures and subsequent fitting confirmed strong similarity of human BUB1 and *V. vinifera* D1HL41. The following search of protein-protein interactions partners for identified plant homologues in *H. sapiens*, *Drosophila melanogaster, Saccharomyces cerevisiae* and *Dictyostelium discoideum* confirmed their participation in spindle checkpoint mechanism and their functional homology to BAB1.

Conclusion. Thus it was confirmed the existence of mitotic checkpoint serine/threonine-protein kinases BUB1 in plants. BUB1 homologues were identified in *A. thaliana*, *V. vinifera*, *O. sativa* subsp. japonica, *Z. mays* and *R. communis*.

TESTING OF FUNCTIONAL ACTIVITY OF PUTATIVE DIOXIN RESPONSIVE ELEMENTS IN PROMOTER REGIONS OF GENES, ENCODING MACROPHAGEAL TRANSCRIPTION FACTORS AND CYTOKINES

E.V. Kashina* [1], E.V. Antontseva [1], A.V. Katokhin, E.A. Oshchepkova[1], D.Y. Oshchepkov[1], M.Y. Shamanina[1], A.Y. Grishanova[3], D.P. Furman[1,2], V.A. Mordvinov[1]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
[3]*Institute of Molecular Biology and Biophysics SB RAMS, Novosibirsk, Russia*
e-mail: alena_kashina@bionet.nsc.ru

Motivation and Aim: The environmental pollutant 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) is a byproduct in pesticides and a persistent chemical formed in various combustion processes that accumulates in the food chain [1]. TCDD is the prototype of aryl hydrocarbon receptor (AhR) ligands with a very high binding affinity for the AhR [1]. The complex ligand:AHR:ARNT functions as a transcription factor, binding to the dioxin responsive element (DRE) sequences in the regulatory regions of target genes. Considerable number of DREs were predicted using computer system SITECON in the regulatory regions of the genes encoding transcription factors and cytokines, expressed in macrophages [2]. The aim of this study was to test the functionality of the predicted DREs by experimental methods, i.e. electrophoretic mobility shift assay (EMSA) and real-time PCR.

Methods and Algorithms: Nuclear extract and total RNA were isolated from U937 macrophages, treated with 2nM TCDD for 24 hours. The binding of TCDD:AHR:ARNT transcription complex from the nuclear extract with double-stranded oligonucleotides, containing the putative DREs was studied by the EMSA. Isolated RNA was used for the study of the TCDD-mediated alteration of the gene expression levels using RT-PCR with SYBR Green I. The primers for each gene were designed using the Primer Express® Software v2.0 (Applied Biosystems).

Results: DRE-containing oligonucleotides, specific for *IRF1, REL, IL12* gene promoters interacted with nuclear extract. EMSA with competitors (oligonucleotides, containing proven DREs) and specific antibodies indicated that AhR was involved in the formation of DNA-probe in complexes. RT-PCR results demonstrated altered expression levels of genes harboring DREs under investigation.

Conclusion: Obtained data demonstrates functional activity of DREs in at least *IRF1, REL, IL12* gene promoters through AHR/ARNT transcription complex.

Acknowledgements: This work was supported by RFBR (No. 09-04-12209-ofi_m), projects of the RAS (No. 15 and 22.8) and SB RAS (Nos. 107 and 119 and Lavrentiev's project in biological sciences No. 3) Federal Agency for Science and Innovations (contract no. 02.512.11.2332); and grant NSh-2447.2008.4.

References:

1. C.F. Vogel et al. (2005) Induction of proinflammatory cytokines and C-reactive protein in human macrophage cell line U937 exposed to air pollution particulates. *Environ Health Perspect.* 113:1536-1541.
2. Furman DP, et al. (2009) Promoters of the genes encoding the transcription factors regulating the cytokine gene expression in macrophages contain putative binding sites for aryl hydrocarbon receptor. *Comput Biol Chem.,* 33(6):465-8.

AGE BIOINFORMATICS OF THE PERSON – THE NEW APPROACH TO STUDYING OF MECHANISMS OF AGEING OF THE PERSON

B.A. Kaurov
*Russian g*erontological *the scientifically-clinical centre, Moscow, Russia*
E-mail: bokar@mail.ru

Currently the scientific literature has accumulated an enormous amount of data on different aspects of human ageing, which are largely not yet been systematized and formalized. In this connection it is proposed to allocate from the bioinformatics section of biology as a separate subsection entitled "Age bioinformatics" for the individuals of different species. In particular, "Age bioinformatics of the person" can be defined as an interdisciplinary field, located at the intersection of bioinformatics, gerontology and geriatrics, studying age-related changes of the person of elderly and senile age with the help of computers based on the creation of specialized information databases and systems analysis using different methods applied statistics and mathematics, including methods of artificial intelligence [1]. The practical way out of this trend in gerontology should be to formulate recommendations for the creation of new geriatric means of prophylaxis and treatment of elderly and senile ages, and also to clarify the mechanisms of ageing of the person.

Special value in age bioinformatics of the person is played by preliminary structurization of the data and their formalized description. For this purpose it is necessary to create appropriate databases of age changes of the person, a specialized language for the formalized description of these changes, detailed schemes of ageing of the person on different levels of its organization. At present there are several such schemes [1,2] based on the graphical language Cmap Tools [3].

One of differences of the scheme of ageing offered by me [2] from existing schemes is its modular principle. The second difference is that in her age-related changes are seen not only from molecular-cellular level and then spread to a higher level, but also, on the contrary, with the organismal level and then on to the following levels.

At present in this scheme 45 specialized modules are allocated, including: factors damaging DNA; the molecular processes of destruction DNA; age changes the properties of nucleic acids, proteins; changes in functional activity of genes, proteins; factors synthesis of DNA, RNA, protein, energy-containing molecules, cellular membranes; factors of cell differentiation, mitosis, apoptosis, cell movement and migration; age-related changes biochemical, physiological, morphological, cellular, immunological; biomarkers of ageing and etc. In the future some of these modules must be linked with the corresponding genetic networks [4].

Except drawing up of schemes of ageing, now is actively working on filling created a database of age changes of the person on the basis of standard DBMS Access and Visual Fox Pro using the methods of artificial intelligence and developed a specialized language GERON [1,5].

References:

1. B.A. Kaurov. (2009) On the application of new information technologies in gerontology. *Science and Innovation. Academy of Sciences of Belarus,* N **8 (78)**: 15-17.
2. B.A. Kaurov. (2009) New scheme of ageing of the person. *VI Scientific and Practical Conference "Society, State and medicine for the elderly.*" Moscow: 31-32.
3. D.I. Muromtsev. (2009) *Conceptual modeling of knowledge in system Cmap Tools*. Petersburg. SPb SU ITMO: 83.
4. N.A. Kolchanov et al. (2005) The integration of gene networks that control the physiological functions of the organism. *Bulletin VOGiS.*, **v. 9, N 2:** 179-198.
5. B.A. Kaurov. (2009) About the formalized system model of ageing of the person. *Informatics and managerial systems*, **N 4 (22) :** 26-28.

MGSmodelsDB – THE DATABASE FOR STORING MATHEMATICAL MODELS OF MOLECULAR GENETIC SYSTEMS

F.V. Kazantsev[1,*], I.R. Akberdin[1], M.T. Ree[1], N.A. Ree[1], K.D. Bezmaternykh[1], V.A. Likhoshvai[1,2]

[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
e-mail: kazfdr@bionet.nsc.ru
[*]*Corresponding author*

Motivation and Aim:

Investigation of molecular-genetic systems mechanisms, patterns of relationship and structure is the system biology fundamental problem demanding the integration of modern experimental and theoretical approaches, in particularly mathematical modeling methods.

The natural hierarchy of gene networks structure allow us to reveal an elementary subsystems (promoters, enzymes etc.) and describe them separately. Thus we can use them for making models of random gene networks structure.

To implement the approach we developed the mathematical models database of elementary cell subsystems that we call MGSmodelsDB.

Methods and Algorithms:

The MGSmodelsDB is developed under the ORACLE database system and uses XML support functionality for storing and retrieving data.

Results:

The database structure was developed for storing mathematical models of two types of elementary cell subsystems: the enzymatic reactions and gene expression regulatory sites. The system has WEB user interface and allows to request mathematical models by setting name of interested reactant which takes part in the elementary cell subsystem. The MGSmodelsDB supports set of reactant synonyms. You can export the selected mathematical model in SBML standard for further analysis by other programs that support the standard.

Conclusion

The database, at the moment, contains 100 elementary mathematical models describing 35 promoters, 60 enzyme reactions in which about 50 metabolites concerning processes of synthesis/utilization of nucleotides and processes of respiration in *E.coli* cell participate.

Each reactant in the database has link at external sources like KEGG and EcoCyc and the synonyms list that compatible with the external sources too.

The database available at: http://samurai.bionet.nsc.ru/MGSmodelsDB

Acknowledgements:

This work has been partially supported by the RFBR №08-04-01008-a and the Grant NSh-2447.2008.4, and by the SB RAS Interdisciplinary project Nos. 107, 119, and also by the RAS projects 21.26, 22.8, 23.29.

FROM OMICS TO DRUGS. COMBINATORIAL TARGETING KEY NODES IN APOPTOSIS NETWORK.

A. Kel[1,2,*], A. Gluch[1], V. Poroikov[3], O. Koborova[3], A. Zakharov[3] and G. Selivanova[4]
[1] *BIOBASE GmbH, Wolfenbuettel, Germany*
[2] *Institute of Chemical Biology and Fundamental Medicine Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia*
[3] *Institute of Biomedical Chemistry of Russian Academy of Medical Sciences, Moscow, Russia*
[4] *Microbiology and Tumor Biology Center (MTC), Karolinska Institutet, Sweden*
e-mail: ake@biobase.de
[*] *Corresponding author*

Motivation and Aim: In this study, we analyzed a large scale gene expression study of treatment of breast cancer cell line by antitumor drugs – RITA and Nutlin, whose direct targets are p53 and Mdm2.

Methods and Algorithms: We have applied ExPlain™ computer system [1], which allows to analyze composite structure of promoters, identify transcription factors involved in regulation of these processes, and perform topological modeling of signal transduction processes leading to stopping of cell proliferation and apoptosis.

Results: We found that several transcription factors, such as E2F1, Fox and Egr factors, and c-myc play an important role at the early stages of these processes switching specific regulatory program of entrance into cell quiescence phase (G0 phase) or programmed cell death. We also found that transcription factors as ETF, c-myb, ZEB1, Crx, Oct, and Fox families are key factors contributing to the downregulation of pro-survival genes upon p53 reactivation by RITA antitumor drug leading to apoptosis of cancer cells. Topological modeling of the signal transduction network upstream of these transcription factors using ExPlain™ tools allows us to reveal key-nodes of such network, which might master-regulate the whole program of cell survival which is balancing versus the program of entrance into cell quiescence phase and apoptosis. Among them we identified subunits of PI3K kinase. Experimental inhibition of this kinase in combination with RITA triggers tumor cells to apoptosis. We consider such key-nodes as the most perspective targets for novel anticancer drugs.

Conclusion: Finally, we applied the powerful chemo-informatics approach based on the computer toll PASS [2], which allowed us to search for prospective leads among libraries of small molecular compounds targeting in combined manner the key nodes found by the ExPlain™ system. We identified several drug candidate whose application on the cancer cells in the combination with RITA low concentration treatment shifts balance of survival mechanisms in tumor cells towards apoptosis.

Acknowledgements: This work was partially supported by EU grant Net2Drug (LSHB-CT-2007-037590).

References:

1. Kel A et al. ExPlain: finding upstream drug targets in disease gene regulatory networks. SAR QSAR Environ Res. 2008;19(5-6):481-94.
2. Alexey Lagunin et al. PASS: prediction of activity spectra for biologically active substances. Bioinformatics, 2000, 16(8): 747-748.

REGULATORY-TARGET GENE RELATIONSHIPS IN TRITICEAE ALLOPOLYPLOID AND HYBRIDE GENOMES

E.K. Khlestkina*, O.Yu. Tereshchenko, E.A. Salina
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: khlest@bionet.nsc.ru
**Corresponding author*

Key words: *allopolyploid wheat, wide hybrids, transcription, flavonoid biosynthesis*

Motivation and Aim. Alien germplasm keeps stirring interest as a source of genes useful for crop plant species. Interspecific hybrids display a range of novel phenotypes, from which favorable ones can be introduced into the breeding processes. With this in mind, it is timely to know how genetic material derived from different species or genera manages to coexist in a common nucleus. The questions are, how genes function in a foreign background, are the homoeologous genes co-expressed, or are some suppressed in the hybrid genome? In the current study we investigated conservation of expression networks in the flavonoid biosynthesis pathway (FBP) in allopolyploid wheat (*Triticum aestivum* L.) and wheat-alien wide hybrids.

Methods and Algorithms. Analysis of FBP structural (target) genes transcription by qRT-PCR was performed in allopolyploid wheat (genome AABBDD, *2n=6x*=42) genotypes carrying different homoeoalleles of FBP regulatory genes (*Rc* – red coleoptile, *Rg* – red glume), wheat-rye chromosome substitution lines carrying rye target gene *F3h-R1* (encoding for the key FBP enzyme flavanone-3-hydroxylase) instead of one of the three wheat *F3h-1* homoeologues, wheat-rye, wheat-Aegilops and wheat-barley hybrids carrying alien regulatory genes (*Rc-R1, Rc-S1* or *Rc-H1*) at wheat genetic background.

Results. Gene silencing was detected for none of the *F3h-1* homoeologues in allopolyploid wheat {*F3h-A1*, *F3h-B1*, *F3h-D1*} or wheat-rye substitution line {*F3h-A1*, *F3h-B1*, *F3h-R1*} coleoptiles where *F3h-1* expression is governed by the wheat *Rc* genes. Activity of *F3h-A1*, *F3h-B1*, and *F3h-D1* was equal, while that of *Rc-A1*, *Rc-B1*, and *Rc-D1* was significantly different. The lack of any genome-specific relationship between the regulatory (*Rc* or *Rg*) and the target (*F3h* or *Chi* - chalcone-flavanone-isomerase) genes in allopolyploid wheat was observed, suggesting high conservation of FBP networks among the three diploid genomes of allohexaploid wheat. However, in the wheat-rye chromosome substitution line we observed unequal expression of wheat and rye *F3h* homoeologues. In wheat-alien hybrids, transcription of wheat target genes (*F3h-1*) under regulation of rye, Aegilops or barley regulatory factors was reduced depending on genetic distance between wheat and a donor species.

Conclusion and Availability. (1) FBP genes provide an effective tool for studies into gene regulation in plants having complex genomes; (2) FBP gene regulation cuts across genomes of allopolyploid wheat; (3) regulatory FBP genes contribute more to the functional divergence between the diploid genomes of allopolyploid wheat than do the structural genes; (4) phenomenon of transcriptional dominance takes place in the wheat-rye chromosome substitution line; (5) regulatory *Rc* genes of different Triticeae species are able to activate wheat target genes *F3h-1*, demonstrating good cooperation of the wheat and alien FBP gene systems within the hybrid genomes; however (6) the bigger genetic distance between wheat and a donor species, the lower transcriptional level of wheat *F3h-1* genes is, suggesting that successful insertion of alien gene into recipient genome still does not guarantee the desirable level of its transcription, because of the donor-recipient gene expression networks divergence.

VARIABLE PART OF GENE EXPRESSION PROFILES IN LIVER AND KIDNEY OF PIGS

*N.S. Khlopova, T.T. Glazko
Russian State Agrarian University – MTAA named after K.A.Timiryazev, Moscow, Russia
e-mail: hns86@mail.ru
**Corresponding author*

Key words: *microarrays, gene expression profiles, key genes, variable expression.*

Motivation and Aim: Estimation of organ-specific gene expression profiles using DNA microarrays is used for searching the genes which transcription can bring in the key contribution to organ metabolome. Usually such analysis excludes the genes with individual differences in expression between the investigated animals. In order to find out the possible reasons of such variability we carried out the analysis of the interrelationships between genes, expression of which showed the essential differences in transcription level in a liver and kidney between six Landrace pigs.

Methods and Algorithms: The analysis of hybridization intensity of cDNA, synthesized on the mRNA of the liver and kidney of six Landrace pigs was carried out on DNA microarrays consisting of 19980 of 70-mer oligonucleotide probes[1]. The arrays were scanned on GSI Lumonics ScanArray 5000 laser scanner. Mathematical processing was carried out with Statistica.

Results: The analysis of 600 genes, distinguished by hybridization intensity between liver and kidney on microarray probes for more than on 20000 standard units of the luminescence in investigated pigs allowed forming two groups of the genes. The first group included genes with the same contribution in gene expression profiles in different animals ("constitutive" genes) and the second group joined the genes with varied expression levels in different animals ("variable" genes). A total of 24 genes with the significant individual variability in expression were revealed. The individual animal variability in gene expression correlated between liver and kidney in 11 of these 24 genes ($r > 0,96$; $P < 0,05$). That was allowed to assume the presence of common regulatory factors for both organs for these 11 genes. Subdividing of 24 variable genes into groups with inner statistically significant correlations in individual variability of gene expression was revealed. As a rule, genes in individual groups with the interconnected expression belonged to the general metabolic way. Totally 8 groups were allocated - 6 genes which products participated in formation of a blood clot; 6 genes - in transport and a lipid metabolism; 5 genes represented markers of blood cells and lysosomes; products of 3 genes participated in programmed cell death and four groups in 2 genes which products participated in Ca^{2+} transport, in intercellular matrix creation, reflected the mitochondrion functional activity and the hormone depended genes.

Conclusion: The obtained data testified that individual variability in gene expression between animals could be caused by the variability of regulatory factors external for organs and different blood fullness of organ samples.

References:

1. Glazko T.T., Khlopova N.S., Fahrenkrug S., Glazko V.I. (2009) Gene expression profiles in liver and kidney of pigs, Izvestia of Timiryazev Academy, Special Issue: p. 55-60

PREDICTION OF REGULATORY REGIONS OF EUKARYOTIC GENES BY EXPERTDISCOVERY SYSTEM

I.V. Khomicheva*[1,2], E.E. Vityaev[2], T.I. Shipilov[2]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,*
[2]*Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*
e-mail: khomicheva@bionet.nsc.ru
**corresponding author*

Key words: *integrated system, hierarchical analysis*, *prediction, regulatory regions of genes*

Motivation and Aim: Discovery of transcriptional regulatory regions is the most important and simultaneously the most difficult task of modern bioinformatics. To approach the decision it is important to consider the variety of biological signals, sufficient to recognize the entire biological situation (tissue, stage specifity, etc.).

Methods and Algorithms: 'ExpertDiscovery' system discovers the hierarchical complex signals, based on elementary signals revealed by state-of-art programs. It was successfully applied to the prediction of transcription factor binding sites and discovering the regularities over them [1].

Results: In current research we integrated by 'ExpertDiscovery' system the statistically overrepresented motifs found by YMF [2]. We performed the analysis of large-scaled promoter sequences (-500 to +200 bp) of eukaryotic genes (using lipid-metabolism system as an example). For different characteristics of motifs (length, degenerate symbols and spacers) the accuracy comparison of 'ExpertDiscovery' performance was made according to bootstrap procedure. The best were prefect hexamers. More then 18% of control data were predicted with false negative rate lower then 10^{-7} (Fig 1).

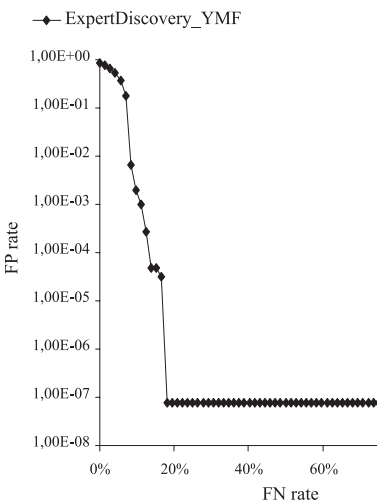


Figure 1. False positives vs false negatives.

Conclusion: Each biological situation is unique and requires its specific solution. The methodology of 'ExpertDiscovery' implies the flexible search of decision starting from the prior knowledge about the problem (elementary signals and interpretable operations over them).

Availability: http://www.math.nsc.ru/AP/ScientificDiscovery/pages/projects.html

Acknowledgements The work is supported by the Russian Foundation for Basic Research 08-07-00272-a and Integration projects no. 111, 119. Authors are grateful to Dr. Ignatieva E.V. for the provided data for analysis.

References:

1. I.V. Khomicheva et al. (2008) ExpertDiscovery system application for the hierarchical analysis of eukaryotic transcription regulatory regions based on DNA codes of transcription, *Intelligent Data Analysis*, **12(5):** 481-494.
2. S. Sinha, M. Tompa (2002) Discovery of novel transcription factor binding sites by statistical overrepresentation, *Nucleic Acids Research*, **30(24):** 5549-5560.

MICROARRAY DATA ANALYSIS PLUGIN FOR BIOUML

I.N. Kiselev [2,*], A.A. Shadrin [2], Y.V. Kondrakhin[1, 2], F.A. Kolpakov[1, 2]
[1]Institute of Systems Biology, Novosibirsk, Russia; [2]Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia;
e-mail: axec@developmentontheedge.com
**Corresponding author*

Key words: *microarray, statistical analysis, BioUML, up and down regulated genes, gene regulation models.*

Motivation and Aim. Huge amount of microarray data has become available during the past decade. However, dealing with large tables of information and extracting useful data from them is complicated. The aim of this work was to develop software for automatization and simplification solving such problems as finding up and down regulated genes, discovering gene expression differences between microarrays data sets, analysis of gene expression patterns and finding putative regulators for target genes by matching their patterns (building gene regulation models).

Methods and Algorithms. BioUML is open source Java framework for formal description and modeling of biological systems. It has plugin-based architecture so adding new plugins to it is natural and enhances abilities for analysis. Besides using BeanExplorer technology simplifies method parameters visualizations for user. To accomplish initial aims, the following methods were added to plugin:

For statistical analysis:

- Student, Wilkokson, Lehman-Rosenblatt and Kolmogorov-Smirnov tests;
- hypergeometric analysis and meta-analysis [1];
- multiple and polynomial regressions, correlation analysis.

For gene regulation modeling:

- Linear, nonlinear [2] and stochastic nonlinear models [3].

Results. Java-based plugin was developed and integrated into BioUML workbench. Program has user-friendly interface, allows quick processing of large amounts of data and provides different analysis methods both already known methods and novel (such as hypergeometric test). It was successfully tested and now is used for data analysis within EU grants FP6 “Net2Drug” and FP7 “LipidomicNet”.

Conclusions. Developed software provides researchers with comprehensive tool for microarray analyzing. Furthermore, its architecture allows simple extending plugins with new methods of analysis.

Availability: Software is freely available as a part of BioUML on website: http://www.biouml.org/

References:

1. Y.V. Kondrakhin, R.N. Sharipov, A.E. Kel, F.A. Kolpakov. (2008) Identification of Differentially Expressed Genes by Meta-Analysis of Microarray Data on Breast Cancer, In Silico Biology, 8: 383-411.
2. K. Chen et al. (2005) A stochastic differential equation model for quantifying transcriptional regulatory network in Saccharomyces cerevisiae. Bioinformatics, 21: 2883-1890.
3. J. Vohradsky, T.T. Vu (2007), Nonlinear differential equation model for quantification of transcriptional regulation applied to microarray data of Saccharomyces cerevisiae. Nucleic Acids Research, 35: 279-287.

INTRANUCLEAR ACTIN AND ITS FUNCTION

E. Kiseleva
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: elka@bionet.nsc.ru

Key words: *actin, myosin, preribosomal subunits, nuclear pore, rRNA nuclear export*

Motivation and Aim: Several recent studies have demonstrated that the intranuclear actin has important function in transcription by all three RNA polymerases. It is a component of chromatin remodeling complexes and ribonucleoprotein particles. It was also shown that actin cooperates with nuclear myosin 1 (NM1) in the recruitment of the chromatin remodeling complex, a mechanism that is required after the initial phases to promote pre-rRNA elongation. It has been suggested that NM1 together with actin may also be implicated in post-transcriptional control of rRNA biogenesis. Actin and myosin has been found in pre-60S ribosomal subunits, and in link with the NE however it was not known whether a fraction of NM1 associates with rRNA transcripts and binds to the nuclear pore. Aim of our studies was to analyse this question using different approaches.

Methods and Algorithms: Molecular biology methods, immunofluorescence and light microscopy, as well as the transmission, field emission in lenz scanning electron microscopy and immuno-electron microscopy have been used. Objects: Hela cells and early Xenopus oocytes.

Results: Ultracentrifugation on HeLa nucleolar extracts showed RNA-dependent NM1 coelution with preribosomal subunits. In RNA immunoprecipitations (RIPs), NM1 coprecipitated with pre-rRNAs and 18S, 5.8S, and 28S rRNAs, but failed to precipitate 5S rRNA and 7SL RNA. In isolated nuclei and living HeLa cells, NM1 or actin inhibition and selective alterations in actin polymerization impaired 36S pre-rRNA processing. Immunoelectron microscopy (IEM) on sections of manually isolated *Xenopus* oocyte nuclei showed NM1 localization at the NPC basket. Field emission scanning IEM on isolated nuclear envelopes and intranuclear content confirmed basket localization and showed that NM1 decorates actin-rich pore-linked filaments. NM1 facilitates maturation and together with actin accompanies export-competent preribosomal subunits to the NPC, thus modulating export. Finally, RNA immunoprecipitations on cross-linked HeLa cells demonstrated that NM1, CRM1, and Nup153 precipitate same 18S and 28S rRNAs but not 5S rRNA.

Conclusion: We present evidence that in the nucleolus, NM1 associates with rRNA; it appears to be loaded onto nascent preribosomal subunits and together with actin accompanies export competent ribosomal subunits to the nuclear pore basket. These results suggest a novel role for NM1 in modulating maturation and together with actin the export of preribosomal subunits to the cytoplasm through the NPCs.

Acknowledgements Work was supported by RFFI and MKB program of RAN.

References:

1. Obrdlik et al. (2010) Nuclear myosin 1 is in complex with mature rRNA transcripts and associates with the nuclear pore basket, FASEB J. 24:146-57.

PERSONAL REFERENCE DASHBOARD: THE SOFTWARE FOR COGNITIVE BIOINFORMATICS

Evdokimov P., Kistanova V., Kudryavtsev A., Ilgisonis E.
1 Knowledge Technologies Ltd., Moscow, Russia
2 KuB Ltd., Moscow, Russia

Motivation and Aim: The notion of cognitive bioinformatics was introduced by Kuchar and co-authors [1] as problem-solving system, which takes into account the scientists and their research environment. The ever-growing compendium of molecular-biology resources is processed through "thinking engine", where individual research sessions are abstracted to the so-called domains of knowledge. However, the proposed cognitive model lacks the programming implementation. In this work we present the Personal Reference Dashboard as an example of cognitive bioinformatics software.

Methods and Algorithms: The client of the Dashboard is distributed as a Web-browser plug-in. This plug-in records the user access to the Web-resources, listed in PubMed as URLs of respective scientific journals. Each article is translated to the PubMed identifier, which is further used to retrieve the associated MeSH terms. Therefore, users of the Dashboard are assigned with the profiles of MeSH terms, extracted from the papers they read.

Results: Client software was developed for Internet Explorer and Mozilla Firefox browsers. The server part of the Dashboard provides information on the personal access log, indicating the title of accessed papers. Alternatively the most frequent MeSH terms can be viewed. Several users (*e.g.* from one lab) can be aggregated to get an averaged list of shared MeSH terms. The time-dependent changes in MeSH-profiles can be displayed to analyze the evolution of the concepts.

Conclusion: Personal Reference Dashboard provides a new paradigm to tackle the data integration debacle, which currently threatens the postgenomic science.

Availability: http://ws.bioknowledgecenter.ru

References:

1. O. A. Kuchar, J. F. Reyes-Spindola, M. Benaroch (2004) Cognitive Bioinformatics: Computational Cognitive Model for Dynamic Problem Solving. *Proceedings of the Third IEEE International Conference on Cognitive Informatics*, 84 – 92.

2n/4n MOUSE CHIMAERAS AS AN AVAILABLE NEW ANIMAL MODEL

E.A. Kizilova*, A.N. Golubitsa, A.I. Zhelezova, A.A. Kruglova, N.M. Matveeva, P.Yu. Povarnitcyna, D.O. Belokrylova, N.S. Nazarko, N.N. Khodeneva, O.L. Serov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: pinus@bionet.nsc.ru
**Corresponding author*

Key words: *ES-cells, tetraploidy, embryo, chimaeras, tissue, GFP, mouse*

Motivation and Aim: Many previous studies have shown that tetraploid (4n) cells rarely contribute to the derivatives of the epiblast lineage. Mouse 4n/2n chimeras and 4n/2n mosaics usually show a restricted tissue distribution. The aim of the present study was to determine either the phenomenon depends on the features of 4n partner, particularly on 4n ES-cell origination.

Methods and Algorithms: Two basic methods – injection of ES-cells and "sandwich"-aggregation were applied to produce 2n/4n and 4n/4n chimaeric embryo with hybrid ES-cells of tetraploid and near-tetraploid caryotypes. Hybrid ES-cell lines were produced earlier by fusion of GFP-marked *E14Tg2aSc4TP6.3* and fetal fibroblasts (*tef*) [1] or by fusion GFP-marked *E14Tg2aSc4TP6.3* and ES-cells *D3* (*D3T*) [2]. Tissue samples taken from embryos, fetuses and adult (6-9 moths) chimaeras were analyzed using disposable microscopic facilities (http://www.bionet.nsc.ru/microscopy/index.html).

Results: Some chimaeric blastocysts were examined *in vitro*, but most of them were transferred into recipient females and then successfully developed up to live-burn chimaeric pups. *In vitro* culturing of chimaeric embryos produced by "sandwich"-aggregation of 2n/4n partners in different combinations revealed that despite of ploidy and origination ES-cells always colonized ICM. It was produced 19 adult chimaeras with *tef* and 21 with *D3T,* respectively. Microscopic analysis shown the presence of hybrid cells derivates in most types of tissues come from ectoderm, endoderm and mesoderm. Gonads were colonized just once in case of abnormal juvenile male.

Conclusion: In contrast with 4n blastomeres, ES-cells (at least hybrid ones) of tetraploid or near-tetraploid caryotypes are capable to colonize most of organ and tissues not only in himaeric embryos but in adult himaeras. The developmental potency of the 4n ES-cell lines have been studied does not significantly depend on origination of cells have been fused with.

Availability: Such type of 2n/4n chimaerisms can be discussed as a new available animal model for biological and biomedical investigations both in fundamental and applied aspects.

Acknowledgements: This work was supported by the Russian Foundation for Basic Reseach (09-04-01369-a) and Carl Zeiss Group Grant (2009).

References:

1. A. A. Kruglova et al. (2008) Embryonic stem cell/fibroblast hybrid cells with near-tetraploid karyotype provide high yield of chimeras, Cell Tissue Res. **334**:371-380.
2. Matveeva N. M. et al. Unequal segregation of parental chromosomes in embryonic stem cell hybrids, Mol. Reprod. Dev, **71**: 305-314.

SYMBOLIC ALGORITHMS IN RESEARCH OF GENE NETWORKS FROM ONE CLASS

A.A. Evdokimov, S.E. Kochemazov*, A.A. Semenov
Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia,
Institute for System Dynamics and Control Theory SB RAS, Irkutsk, Russia
e-mail: veinamond@gmail.com
**Corresponding author*

Key words: *discrete models of gene networks, fixed points, functional graph, synthesis of gene network graph adjacency matrix, threshold function, SAT.*

Motivation and Aim. A gene network is a group of interacting genes, providing the formation of a particular trait of an organism. Different methods are used in mathematical modeling of functioning of gene networks: from logical-dynamic models to systems of differential equations on graphs. We consider one simple model [1] of a gene network and propose a research method based on symbolic algorithms for it.

Methods and Algorithms. A regulatory loop is a connected digraph $G=G(U,D)$ with n vertices, identified with the genetic elements (RNA, proteins), without multiple arcs, and a set of arcs, associated with regulatory relations. Each vertex $i \in U$ has a weight with a threshold p, representing a concentration of a product, identified with the given vertex. Functioning of a gene network is defined by automaton mapping A_G that changes states (n-vectors in the alphabet <0,1,...,p-1>, vectors of weights of vertices in G). The fixed point is such n-vector that is mapped into itself with mapping A_G.

Results. We propose fixed points search algorithms for this class of gene networks. These are symbolic algorithms based on SAT-approach to the problem of inversion of discrete functions [2]. In the course of the work there was proved a theorem, a consequence of which was the fact that it is possible to recover the adjacency matrix of a graph of a gene network based on constraints on a structure of the graph (for example, specified fixed points of automaton mapping can be such constraints). The algorithms proposed are highly efficient on a number of numerical experiments.

Conclusion. Numerical experiments proved this approach to be highly perspective for research of gene network discrete models. It is interesting to apply these methods to real gene networks.

Availability. The program for the fixed point search and adjacency matrix synthesis was also developed. This program finds all the fixed points of the automaton mapping on gene network graph specified by an adjacency matrix. The program provided also can find elements of an adjacency matrix using some constraints.

Acknowledgements. The work was supported by the Russian Foundation for Basic Research (08-01-00871, 09-01-00070) and the interdisciplinary grant 119 of SB RAS.

References:

1. E.D. Grigorenko, A.A. Evdokimov, V.A. Likhoshvai, I.A. Lobareva. (2005) Fixed points and cycles of the automata mapping for the gene network simulation (Russian). *Herald of Tomsk State University,* **14**: 206 - 212.
2. A.A.Semenov (2009) Decomposition Representations of Logical Equations in Problems of Inversion of Discrete Functions, *Journal of Computer and Systems Sciences International,* Vol. 48, **5**: 718–731.

ROLE OF N-END AMINO ACIDS IN TRANSLATION OF HUMAN PROTEINS

O.A. Volkova, A.V. Kochetov*
Institute of Cytology and Genetics SB RAS & Novosibirsk State University, Novosibirsk, Russia,
**Corresponding author ak@bionet.nsc.ru*

Key words: *translation initiation, AUG codon context*

Motivation and Aim

Accurate prediction of the efficiency of eukaryotic translation initiation signal is of importance for evaluation of both the mRNA coding potential and translation rate. However, this task was not solved yet. The recognition of the AUG triplet as translation initiation site depends on its nucleotide context. The relative importance of -3 and +4 context positions was experimentally shown. The roles of other context positions are still unknown. This especially concerns the 3'-end part located at the beginning of CDS.

Material and Methods

To analyze the role of nucleotides in positions +4,+5,+6 in translation initiation we performed a comparative analysis of human mRNA sample (24154 nucl. sequences) characterized with either "optimal" (RnnAUG, R=purine) or "suboptimal" (YnnAUG (Y=pyrimidine) nucleotides in the key position -3. We hypothesized that if the nucleotides pos. +4, +5, +6 participate in start codon recognition, these samples will be statistically different.

Results

1. The context variant considered to be "optimal" (RnnAUG) was found to be heterogeneous in respect with the nucleotide preferences in pos. +4,+5,+6: it is likely that G in pos. +4 is important only for functioning of GnnAUG context variant.

2. The second position of amino acid sequences of proteins encoded by mRNA samples with different start codon contexts was characterized by different preferences. Notably, the proteins encoded by mRNAs with AnnAUG context were characterized by specific and significant over-representation of serine, whereas the presence of GnnAUG context correlated with a higher occurrence of alanine and glycine.

Conclusion and Availability

The observed significant correlations probably resulted from the interplay between the nucleotides in positions -3, +4, +5, +6 and the second amino acid of the encoded protein. It may be assumed that the formation of the first peptide bond between the initiator Met and Ser allows ribosome to avoid some steric constrains resulting from the presence of A in position -3. In turn, over-representation of Ala and Glu in the 2nd position of proteins encoded by GnnAUG mRNA subsample is likely to be less dependent on the selection at the level of amino acids and can reflect the noticeable selection at the level of nucleotides in positions +4,+5,+6. This (after experimental verification) can be used to improve a prediction accuracy of start codon position(s) and mRNA initiation efficiency

Acknowledgements

This work was supported by RFBR (08-04-00525), grants from Ministry of Science & Education (2.1.1/6382; 02.740.11.0277; P128) and the Program of RAS (Molecular and Cellular Biology). We also thank SD RAS Complex Integration Program for partial support.

"BLUEPRINT" MODELLING OF THE NUCLEAR RECEPTOR NETWORK.

A. Kolodkin[a]*, N. Plant[b], F.J. Bruggeman[c], E. Wiedemann[b], M.J. Mone[a], B. Siebers[d], B.G. Olivier[a], M. Swat[a], J.L. Snoep[e], P. Goldfarb[b] and H.V. Westerhoff[f,a]

[a] Molecular Cell Physiology, VU University, Amsterdam, The Netherlands; [b]Centre for Toxicology, University of Surrey, Guildford, UK; [c] Regulatory networks group, NISB, the Netherlands; [d]Molecular Enzyme Technoloye und Biochemistry, University of Duisburg-Essen, Germany; [e]Department of Biochemistry, Stellenbosch University, South Africa; [f]Manchester Centre for Integrative Systems Biology, the University of Manchester, UK

e-mail: alexey.kolodkin@falw.vu.nl

[*]Corresponding author

Key words: *systems biology; nuclear receptor (NR); virtual human; blueprint modeling*

Motivation and Aim: "Blueprint" modeling means adjusting one master-model to different systems with similar network design that may be useful for building integrative computer models and eventually for making a "virtual human" – whole body model highly anticipated for a comprehensive understanding of body functioning and for intensification of drug discovery [1,2]. We have studied the potency of "blueprint" approach using nuclear receptor (NR) network as an example. NRs are transcription factors whose activity is modulated by different intra- and extracellular signals [3]. There are 48 different NRs: some NRs (e.g. PXR) wait for the signal (hormone ligand) in the nucleus, others (e.g. GR) are predominantly localized in the cytoplasm and shift to the nucleus upon the addition of ligand. Biologically it looks very different. But could all NR models be considered as instantiations of the same network design?

Methods and Algorithms: kinetic modeling in COPASI and Mathematica-6 software.

Results: Canonical "blueprint" kinetic model of NR signaling has been built and, by varying only several kinetic parameters, successfully parameterized for different types of nuclear receptors such as GR, PXR and VDR.

Conclusion: We have demonstrated that different NR systems can be considered as instantiations of the same reaction network. Our example demonstrates the power of "blueprint" approach in the perspectives of using this strategy whilst building the virtual human.

Availability: available on request from the authors

Acknowledgements: NucSys, BioSim, EC-MOAN, YSBN, UNICELLSYS, SulfoSYS, MOSES, NWO-FALW, BBSRC (www.systembiology.net/support)

References:

1. 1. Westerhoff, H.V., et al. (2008) Integrating systems approaches into pharmaceutical sciences, European Journal of Pharmaceutical Sciences, **35**(1-2): 1-4.
2. 2. Westerhoff, H.V., et al. (2009), Systems biology towards life in silico: mathematics of the control of living cells. J Math Biol, 2009. **58(1-2):** 7-34.
3. 3. Carlberg, C. and T.W. Dunlop (2006) An integrated biological approach to nuclear receptor signaling in physiological control and disease, Critical Reviews in Eukaryotic Gene Expression, 2006. **16(1)**: 1-22.

THE TRANSCRIPTIONAL PROFILE OF RETINAL PIGMENT EPITHELIUM/CHOROID OF OXYS RAT AS A BACKGROUND FOR THE RETINOPATHY DEVELOPMENT

O.S. Kozhevnikova, V.M. Efimov, A.M. Markovets, N.G. Kolosova*
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
*E-mail: kolosova@bionet.nsc.ru *Corresponding author*

Motivation and Aim: The etiology of age-related macular degeneration (AMD), the leading cause of blindness in the developed world. A major focus of AMD research is the retinal pigment epithelium (RPE)/choroid, which remains poorly understood. Recently, we showed that senescence-accelerated OXYS rats are a useful model for the study of pathogenesis and design of new therapeutic targets of AMD. The aim of this study was to characterize the changes in the RPE/choroid with age that may contribute to the development of AMD.

Methods: The transcriptional profiles of the RPE/choroid of OXYS rats and controls (Wistar rats) were compared using microarray analyses at the age of 20 days (without symptoms), 3 mo (with manifestation of the first signs), and 16 mo (a stage of pronounced AMD). Samples were hybridized to a custom cDNA microarray containing 12,000 genes and ESTs, and the data from the quantified scanned images were analyzed using Bioconductor and SAM. Preferential expression of some genes was confirmed by real-time PCR.

Results: There were 132, 151, and 168 genes differentially expressed (up- or down-regulated at least two-fold, $p<0.05$) in RPE/choroid of OXYS rats at the age of 20 days, 3, and 16 mo, respectively. Major functional categories and pathways where expression of genes was changed in RPE/choroid of OXYS rats at the age of 20 days included: “energy metabolism” – 9 genes, “hypoxia response” – 7 genes, and “response to stress” – 5 genes. Expression of septin SEPT3, which had not previously been reported to be expressed in the retina, was increased in OXYS rats of all ages. SEPT3 was exclusively expressed in neurons and was enriched in presynaptic nerve terminals, which suggests that it might have unique functions in synapse formation or maturation or, perhaps, play a role in exocytosis/endocytosis. The expression of membrane palmitoylated protein-4 (MPP4) was downregulated at age 20 days and increased two-fold at the age of 3 months. We observed similar changes in the expression of HSP70 gene. The results of microarray analyses of MPP4 and HSP70 expression were confirmed by real-time PCR. MPP4 is a retina-specific scaffolding protein of the membrane-associated guanylate kinase family that has been implicated in organizing presynaptic protein complexes in the photoreceptor ribbon synapse. Recently, Stцhr et al (2009) described a presynaptic protein complex in photoreceptors consisting of the key adaptor proteins MPP4, PSD95, and VELI3, which also include PMCAs and the Ca^{2+}-binding protein recoverin. HSP70 was associated with proteins of the MPP4 multiprotein complex, which supports a role of MPP4 in Ca^{2+} regulation in the synaptic terminals. Our work appears to be the first report of alterations of MPP4 expression in the retina of rats, especially, in the context of AMD development. Additionally, we show early changes in the retinal HSP70 expression that are associated with aging and AMD. *Conclusion:* Changes in expression of certain genes precede detectable signs of AMD in OXYS rats; however, any deleterious impact of these gene-expression changes on RPE/choroid presumably depends on cumulative effects that occur between 20 days and 3 mo of age.

Acknowledgements: This work was supported by grants from RFBR 08-04-00722.

ANALYSIS OF P53 BINDING SITES BY USING CHIP-SEQ DATA

I.S. Yevshin[1,2], Yu.V. Kondrakhin[1,3], M. Turunen[4], T. Kivioja[4], F. Nikulenkov[5], R.N. Sharipov[1,6], J. Taipale[4], G. Selivanova[5], F.A. Kolpakov[1,3,*]

[1]Institute of Systems Biology, Ltd, Novosibirsk, Russia; [2]Novosibirsk State University, Novosibirsk, Russia; [3]Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia; [4]National Public Health Institute, Helsinki, Finland; [5]Karolinska Institutet, Stockholm, Sweden; [6]Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

e-mail: fedor@biouml.org

**Corresponding author*

Key words: *p53, binding site recognition, ChIP-seq, position weight matrix*

Motivation and Aim: Transcription factor p53 is a well-known tumor suppressor. Mutations of p53 binding sites are basic hallmarks of many types of cancer. The canonical structure of p53 binding sites is described by two decameric sequences PuPuPuC(A/T)(T/A)GPyPyPy separated by a spacer of 0–13 bp. Known models for recognition of p53 sites were built earlier on the base of separate small training sets obtained by less comprehensive methods than ChIP-seq. Recently, five sets of ChIP-seq data were obtained in the frameworks of "Net2Drug" project that allowed to perform: 1) comparison of methods for identification of transcription factor (TF)-binding fragments; 2) construction of more accurate method for p53 binding site prediction.

Methods and Algorithms: ChIP-seq data were obtained in the experiment of treatment of breast cancer MCF7 cell line by activators of p53 10 uM Nutlin3a, 0.1 and 1 uM RITA (Reactivation of p53 and Induction of Tumor cell Apoptosis) and 100 uM 5-FluoroUracil. Three methods SISSRs (Site Identification from Short Sequence Reads), MACS (Model-based Analysis of ChIP-Seq) and KOLI were applied to identification of p53-binding fragments. For recognition of potential binding sites the extended position weight matrix (PWM) method [1] was used. Our new alignment method and clusterization method adapted to ChIP-seq data (unpublished) were applied to construct a set of more optimal PWMs.

Results and Conclusion: Different methods for TF-binding fragments identification generate quite distinct sets of fragments. Thus, in case of 1uM RITA the methods SISSRs, MACS and KOLI demonstrated overlapping of 48.13% of fragments only. On the base of analyzed ChIP-seq data more effective PWMs for p53 were built. The sensitivity and False Discovery Rate were selected as a measure of accuracy of recognition procedures. According to our PWMs, the most typical motif of p53-binding sites is (A/C/T)NN(A/G)(G/a)(A/G)**CATG**(C/T)CCA(G/a)(A/g)**CATG**(C/t)(C/t)(C/t)NN. Analysis of p53-binding sites allowed to conclude that spacers of 6, 7 and 8bp are more considerable. Analysis of human 6th chromosome demonstrated that SINE/Alu and LINE/L1 repeats are also significantly enriched by p53-motifs. To avoid this effect we constructed additional PWMs specific to these repeats that allowed us to decrease essentially the false positives of p53-binding sites prediction by our method.

Acknowledgements: This work was supported by EU grant №037590 "Net2Drug".

References.

1. E.A.Ananko et al. (2007) Recognition of interferon-inducible sites, promoters, and enhancers, *BMC Bioinformatics*, **8:56:** 1-14.

ANALYSIS OF GLUCOCORTICOID RECEPTOR BINDING SITES BY USING CHIP-SEQ DATA

V.M. Merkulov[1], T.I. Merkulova[1], I.S. Yevshin[2,3], Yu.V. Kondrakhin[2,4], R.N. Sharipov[1,2], F.A. Kolpakov[2,4,*]

[1]Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia; [2]Institute of Systems Biology, Ltd, Novosibirsk, Russia; [3]Novosibirsk State University, Novosibirsk, Russia; [4]Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia

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

Key words: *glucocorticoid receptor, ChIP-seq, binding site recognition*

Motivation and Aim. Glucocorticoids are important regulators of wide spectra of human physiological processes. Glucocorticoid receptor (GR) is a ligand-inducible transcription factor belonging to the nuclear receptor super-family. The structure of GR binding sites is described by a hexanucleotide TGTTCT or a palindrome AGAACANNNTGTTCT. The first step to systematical analysis of GR binding sites was made in [1]. This analysis was performed on the set of 160 GR binding sites from TRRD database. Recently, the results of ChIP-Seq experiment for GR have also been published [2]. Grounding on data from [1] and [2] we aimed at the two following tasks: 1) construction of more accurate method for GR binding site recognition; 2) analysis of GR-binding regions to reveal the transcription factors that potentially interact with GR.

Methods and Algorithms. For identification of the set of GR-binding fragments we used two published methods: SISSRs (Site Identification from Short Sequence Reads) and MACS (Model-based Analysis of ChIP-Seq). For recognition of potential binding sites and construction of new position weight matrices (PWM) we used methods from [3].

Results and Conclusion. Analyzing the set of GR-binding fragments we constructed several PWMs that recognized potential GR binding sites more effectively in comparison to known matrices. The sensitivity and False Discovery Rate were used as a measure of accuracy of recognition methods. The composite element GR/GR was detected on the base of analysis of distance between predicted GR-sites. The significant effect of multiplicity (i.e. occurrence of tandem repeated hexanucleotides) was also revealed. Application of chi-squared independence test allowed us to reveal the set of transcription factors, whose binding sites significantly (p-value<0.05) cooccurred in the set of GR-binding fragments. The PWMs for potential coregulators were extracted from the TRANSFAC and JASPAR databases (http://jaspar.genereg.net). Obtained results allow to conclude that NF-kappaB, c-Myb, STAT5A, AP-1, FOXO3, c-Ets-2 and RSRFC4 are the most significant candidates for the role of GR co-regulators.

Availability: by request to the corresponding author.

Acknowledgements. This work was supported by EU grant №037590 “Net2Drug”.

References.

1. V.M.Merkulov and T.I. Merkulova (2009) Structural variants of glucocorticoid receptor binding sites and different versions of positive glucocorticoid responsive elements: analysis of GR-TRRD database, *J* Steroid *Biochem Mol Biol*, **115**: 1-8.
2. T.E.Reddy et al. (2009). Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation, *Genome Res*, **19**: 2163-2171.
3. E.A.Ananko et al. (2007) Recognition of interferon-inducible sites, promoters, and enhancers, *BMC Bioinformatics*, **8:56:** 1-14.

BIOUML – INTEGRATED PLATFORM FOR BUILDING VIRTUAL CELL AND VIRTUAL PHYSIOLOGICAL HUMAN

F.A. Kolpakov[1,2*], N. Tolstyh[1], E.O. Kutumova[1,2], I.N. Kiselev [1,2], A.A. Shadrin[1,2], T.F. Valeev[1,3], A. Ryabova[1], R.N. Sharipov[1,4], A.E. Kel[1]
[1]Institute of System Biology, Novosibirsk, Russia;
[2]Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia; [3]Institute of Informatics Systems SB RAS, Novosibirsk, Russia;
[4]Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: fedor@biouml.org
**Corresponding author*

Key words: *BioUML, virtual cell, virtual physiological human*

Motivation and Aim: Building of virtual cell and virtual physiological human are main challenges for systems biology in this decade. Integrated and extensible informational platform is essential part for solving these tasks.

Methods and Algorithms. BioUML is an open source integrated platform for systems biology that spans the comprehensive range of capabilities including access to databases with experimental data, tools for formalized description and visual modeling of complex biological systems. Due to scripts support (R, JavaScript) it provides powerful possibilities for analyses of high-throughput data. Plug-in based architecture (Eclipse run time from IBM is used) allows to add new functionality using plug-ins.

Results and Conclusion: BioUML platform consists from 3 parts:

- BioUML server – provides access to biological databases;
- BioUML web edition – web interface based on AJAX technology;
- BioUML workbench – standalone application.

Brief description of BioUML platform functionality:

- supports main standards used in systems biology: SBML, SBGN, CellML, BioPAX, OBO, PSI-MI;
- supports access to main biological databases:
  - catalolgs: Ensembl, UniProt, ChEBI, GO and other;
  - pathways: KEGG, Reactome, EHMN, BioModels, SABIO-RK, TRANSPATH, EndoNet, BMOND and other;
- database search - full text search using Lucene engine;
- graph search, graph layout engine;
- visual modeling:
  - support for hierarchical models;
  - simulation engine supports (ODE, DAE, hybrid, stochastic, 1D PDE);
  - parameters fitting;
- genome browser (supports DAS protocol);
- data analyses and workflows – specialized plug-ins for microarray analysis, integration with R/Bioconductor, JavaScript support, interactive script console.

BioUML platform is successfully used for building components of virtual cell (integrated model of apoptosis, lipid related pathways) and virtual physiological human (models of haemodynamics, heart and kidney).

Availability: http://www.biouml.org.

Acknowledgements: This work was supported by EU grants №037590 “Net2Drug” № 202272 «LipidomicNet», integration and interdisciplinary grants №16, 91 of SB RAS.

ENTROPY ASPECT OF HYDROPHOBICITY IN QSAR RESEARCH

G.A. Isaeva, S.P. Konovalenko*, P.P. Isaev

Taganrog State Pedagogical Institute, Taganrog, Russia
e-mail: svetlana_s12@mail.ru
**Corresponding author*

Key words: *hydrophobicity, lipophilicity, QSAR, absolute entropy*

Motivation and Aim: Lipophilicity parameter is crucial for judging the traffic potential of substances and biological objects and therefore widely used in QSAR. Hydrophobicity depends on the absolute entropy, so you need to find the optimal method for calculating the absolute entropy. At present there are two approaches to determining the values of absolute entropy ¬ thermo chemical (calorimetric) and statistical (spectroscopic). A class of adamantane derivatives is considered. This class possesses brightly expressed biological activity.

Methods and Algorithms: The absolute entropy is calculated using three methods: 1) The method of Anderson, Bayer, Watson [1], basing on which the absolute entropy was calculated proceeding from the entropy of formation of compounds; 2) An additive method for calculating the entropy ($S = \Sigma n_i S_i$); 3) Approximate methods for evaluation of absolute entropy in solidity and liquidity ($S^s = 1.1 * C_p$, $S^l = 1.4 * C_p$).

Results: Function of state has the property of additivity, hence the entropy of the entire system can be folded from the entropies of its individual parts. Additive method of calculating the absolute entropy of adamantane derivatives is proposed. Additive calculation method and the method of Anderson, Bayer, Watson give almost equal values of absolute entropy. When correlating one method to another, we obtain high statistics (correlation coefficient $r^2 = 0,99$; Fisher's criterion $F = 6620,2$), and the schedule is presented by a straight line almost passing through the center of coordinates at an angle of 45^0. When calculating the absolute entropy according to an additive scheme using the absolute values of entropy, presented in the method of Anderson, Bayer, Watson, we get statistical data not high enough and observe spread of points from a straight line (correlation coefficient $r^2 = 0,92$; Fisher's criterion $F = 226,42$). Approximate calculation methods for predicting the values of absolute entropy in this class of compounds are not appropriate, because the values obtained for the absolute entropy deviate greatly from the real data.

Conclusion: Existing approximate methods of calculation do not allow to r estimate the absolute entropy of adamantane derivatives reliably. The method of Anderson, Bayer, Watson takes into account the entropy of formation of compounds, so it is most advisable to be used for calculating the absolute entropy. Additive scheme of calculating the absolute entropy has the right to existence in this class of compounds and has good correlation characteristics. All considered methods are not universal. It is necessary to create a mathematical model satisfying the experimental data for calculating the entropy.

Availability: This method can be used by program ChemOffice and system of computer mathematic Maple.

References:

1. R.Reid, T. Sherwood. (1971) Ideal gas thermal capacity, heat energy and free energy of formation. In: *Properties of gases and liquids,* 265-268 (Chemistry Leningrad).

STUDY OF POLYMORPHISMS IN GLUTATHIONE S-TRANSFERASE M1 AND T1 (*GSTM1* AND *GSTT1*) GENES IN SELKUPS AND TUNDRA NENETS

R.P. Korchagina*[1], L.P. Osipova[1], N.A. Senkova[1], N.A. Kostrykina[2], M.L. Philipenko[2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Institute of Chemical biology and Fundamental medicine SB RAS, Russia*
e-mail: Kruosana@mail.ru
**Corresponding author*

Key words: *Selkups, Tundra Nenets, biotransformation of xenobiotics, GSTM1 0/0, GSTT1 0/0*

Motivation and Aim: study of polymorphisms in *GSTM1* and *GSTT1* genes is important in aspect of medicine and population genetics because products of these genes – enzymes of xenobiotics biotransformation phase II – participate in detoxification of endotoxins, exotoxins, carcinogens and other like substances. We intended to study the distribution of polymorphisms in glutation-S-transpherase M1 and T1 genes (*GSTM1*0*, *GSTT1*0*) in native small peoples of Northern Siberia (Selkups and tundra Nenets), who have conserved traditional way of life in extreme conditions of Northern Siberia and adapted to this environment; to find out the presence of ethnic differences in the frequency of "null" genotypes and to evaluate the risk for development of diseases which are connected with deletions in *GSTM1* and *GSTT,* in these populations.

Methods and Algorithms: DNA samples were formed from a pool of ethnically pure persons of Selkups and Tundra Nenets. The presence of deletions in *GSTM1* and *GSTT1* genes was detected by "Real-time PCR" using intercalating fluorescent dye SYBR Green and analysis of melting curve.

Results: the polymorphism of *GSTM1* gene was studied in 147 Selkups and 139 Tundra Nenets, and *GSTT1* was studied in 146 and 140 peoples, respectively. The frequency of *GSTM1 0/0* genotype appeared to be higher in Tundra Nenets than in Selkups (28,8 % versus 19%). Contrariwise, frequency of *GSTT1 0/0* genotype is reliably lower in Tundra Nenets than in Selkups (6,4% versus 17,8%, X^2=6,75; P<0,01).

Conclusion: our study revealed differences in distribution polymorphisms of *GSTM1* and *GSTT1* genes among these ethnical groups. Both "null" variants (*GSTM1 0/0* and *GSTT1 0/0*) have nearly equal frequency (19%) in Selkups, but the frequency of *GSTM1 0/0* is 1,5 times higher, and the frequency of *GSTT1*0/0* is 3 times lower in the population of Tundra Nenets as compared with Selkups. As it is known, *GSTM1*0/0* and *GSTT1*0/0* are associated with risk for different diseases, we can suppose that nearly 20% of individuals from our study are in risk group.

FUNCTIONAL ANALYSIS OF PROMOTER REGION OF THE *Xist* GENE IN MOUSE (*Mus musculus*).

A.M. Korotkova
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: korotkovaanna@yahoo.com

Key words: *Xist, X chromosome inactivation, promoter, transcription factor.*

Motivation and Aim: Genes dosage compensation on sex chromosomes in eutherian mammals is realized by a random inactivation of one of the X chromosome in females. This process is regulated by the X chromosome inactivation center (Xic) located on the X chromosome and consists of a number of elements. The *Xist* gene is a crucial factor of the inactivation process. It produces a long non-coding RNA that covers all future inactive X chromosome resulting in a heterochromatinisation and transcription silencing of majority of the genes on the inactive X. We are studying the promoter region of *Xist* to investigate the mechanism of its function in the inactivation process.

Methods and Algorithms: To study promoter functions we have constituted reporter constructions on the base of a luciferase gene supported vector under the control of different parts of the promoter region of the *Xist* gene, and their activity has been measured. The promoter DNA sequence was also subjected to a computer analysis in order to find potential binding sites for transcription factors.

Results and Conclusion: The reporter constructions activity was studied on different types of female mouse cells: differentiated and undifferentiated. During these experiments we obtained a similar expression pattern. The test results were analyzed and compared with data derived from the computer analysis (we located potential binding sites for transcription factors ER, RAR, SRY etc). It was found that six regions likely to possess either enhancer or silencer activity. The conservative region among all mammals contains two ER transcription factors. ER is a nuclear receptor for a steroid hormone estrogen. It was suggested that estrogen might take part in the inactivation process during the embryo development and cells differentiation. To examine this hypothesis the reporter construction activity was tested as well in estradiol subjected cells. As a result of the experiment activity of one reporter construction was increased almost by a factor of two. It is possible therefore that estrogen effects the site directly or through a mediator. The made experiments argue for estrogen effect on the inactivation process in the period of embryo development.

ANALYSIS OF DNA HYPERMETHYLATION IN BILE SAMPLES OF PATЭENTS WITH CHOLANGIOCELLULAR CARCINOMA

A. Atay[1], T. Semerci[1], S. Cuhadar[1], E. Alper[2], M.H. Koseoglu[1*]
[1]*Department of Biochemistry and Clinical Biochemistry,*
[2] *Department of Gastroenterology*
Ataturk Training and Research Hospital, Basin Sitesi, Izmir Turkey
e-mail: mkoseoglu@yahoo.com
** Corresponding author*

Motivation and Aim: Genetic alterations play crucial role in carcinogenesis. A number of genes may be hypermethylated in some tumors and this methylation often occurs very early during carcinogenesis. Analysis of methylation of DNA in bile samples might provide a novel strategy for detection of cholangiocellular carcinoma (CCC) which are malignant tumors of the intra or extrahepatic biliary tract.. To investigate this approach, CCC related gene promotors (SFRP2, p16, DAPK1, HIC1, MGMT) were assessed in bile samples from patients who had prediagnosed for CCC.

Methods and Algorithms: Twenty-one patients undergoing therapeutic endoscopic retrograde cholangiopancreatography (ERCP) for routine clinical examination of CCC prediagnosis were enrolled in this study. The patients (n=21) ranged in age from 48 to 82 years. Brush cytology and bile samples were taken from patients. DNA was isolated from bile samples by use of the Spin DNA Extraction Kit. Methyl Spin Kit was used to obtain bisulfite-treated DNA which was the second step of DNA hypermethylation analysis. Onco Strip Assay Colon was used to identify the DNA hypermethylation based on polymerase chain reaction (PCR) and hybridization. The assay covers methylation patterns in the promotor region of 5 genes.

Results: 11 (52%) of 21 patients were positive for malignancy. The most frequent methylated genes were SFRP2 and HIC1 genes in CCC patients (90.9%, 81.8% respectively). It was also found that SFRP2 gene had 58.8% positive predictive value and 75% negative predictive value for detecting CCC.

Conclusion: This is the first and preliminary study that reports DNA hypermethylation in bile samples of patients with CCC. These results may be important in designing of genetic screening programs, determination of prognosis, early detection and treatment for patients with CCC.

COMPUTER SIMULATION OF INTERACTION OF PROTEIN PLASTOCYANIN WITH TRANSMEMBRANE PROTEIN COMPLEXES PHOTOSYSTEM I AND CYTOCHROME BF

I.B. Kovalenko*, A.M. Abaturova, O.S. Knyazeva, G.Yu.Riznichenko, A.B.Rubin
Biological faculty of Moscow State University, Moscow, Russia
e-mail: kovalenko78@mail.ru
**Corresponding author*

Key words: *computer simulation, photosynthesis, protein-protein interaction*

Motivation and Aim: Mobile electron carrier plastocyanin is a small copper protein which transfers electrons from the cytochrome bf complex to the reaction center of photosystem 1. There is a conceptual difficulty to explain the mechanism of plastocyanin diffusion in lumen. It is not clear if plastocyanin can rapidly diffuse over a distance of hundreds of nanometers shuttling electrons between granal and stromal areas. It is widely recognized that electrostatic interactions play a crucial role in the binding of plastocyanin to its reaction partners, photosystem 1 and cytochrome bf. Computer simulation can give information on plastocyanin diffusion and interaction in chloroplast thylakoid lumen.

Methods and Algorithms: We designed a computer 3D model of diffusion and interaction of proteins [1-3]. The model is multiparticle, it can describe the interaction of several hundreds of proteins in a cell environment. In our model the interacting proteins are represented as rigid bodies with spatial fixed charges. Translational and rotational motion of proteins is the result of the effect of stochastic Brownian force and electrostatic force. Poisson-Boltzmann formalism is used to determine the electrostatic potential field generated by the proteins and photosynthetic membrane.

Results: Using the computer model we studied kinetic characteristics of plastocyanin-cytochrome bf and plastocyanin-photosystem I complex formation at a variety of ionic strength values. The computer models demonstrated non-monotonic dependences of complex formation rates on the ionic strength as the result of electrostatic interactions. We simulated the electron transfer from cytochrome bf to photosystem I by plastocyanin and calculated the rate of this process at different conditions.

Conclusion: The simulation method presented in this paper can be applied for the description of diffusion and functioning of many macromolecules that interact in the heterogeneous interior of subcellular systems.

Availability: available on request from the authors.

References:

1. I.B.Kovalenko et al. (2006) Direct simulation of plastocyanin and cytochrome f interactions in solution. *Phys. Biol.*, **3**: 121–129.
2. I.B.Kovalenko et al. (2008) Computer simulation of plastocyanin–cytochrome f complex formation in the thylakoid lumen. *Biophysics,* **53**(2): 140-146.
3. I.B.Kovalenko et al. (2009) A novel approach to computer simulation of protein–protein complex formation, *Doklady Biochemistry and Biophysics,* **427**: 215–217.

A SYSTEM FOR PROCESSING OF DIGITAL IMAGES ACQUIRED WITH MODERN MICROSCOPY TECHNIQUES

K.N. Kozlov*, A.S. Pisarev and M.G. Samsonova
St.Petersburg State Polytechnical University, St.Petersburg 195251, Russia
e-mail: kozlov@spbcas.ru
**Corresponding author*

Key words: *Confocal microscopy, image processing and analysis, Drosophila*

Motivation and Aim: Recent advances in microscopy techniques in combination with easily available high-performance computer hardware are making the image-based nanoscale characterization of objects feasible. The important prerequisite for success of such studies is the ability to automatically detect and quantitatively characterize objects under study. Automation also means a possibility to construct complex scenarios for image processing within an integrated problem-solving environment, that enables a user without strong programming skills to perform a variety of tasks related to scientific computing.

Methods and Algorithms: We have developed a new software package ProStack (*Pro*cessing of *Stack*s) that is designed to automate the analysis 2D and 3D digital images of biological objects acquired with modern microscopy techniques. ProStack provides flexible and convenient platform for processing of experimental images. It implements geometrical, morphological, histogram, segmentation and other domain specific and domain-independent methods as separate modules. Several modules can be joined in a complex image processing scenario through a graphical user interface and intermediate results can be visualized during workflow enactment. The processing operations are highly accurate and afford tuning to ensure customization and flexibility without loss of efficiency. The system can be easily integrated with open source and commercial packages, e.g. Matlab. The designed workflow can be saved as a complex program module and re-used in other workflows.

Results: The ProSrack package was successfully applied to automatically identify and count objects in images of the *Drosophila* Kc167 cells, correctly orient, rotate and crop the confocal images of expression patterns of the Drosophila segmentation genes, remove background from confocal images, as well as to extract quantitative data from the images of expression patterns of genes controlling development in *Drosophila melanogaster* and *Nematostella vecentis*.

Availability: http://urchin.spbcas.ru/downloads/ProStack/ProStack.htm

Acknowledgements: We are very thankful to J. Reinitz, M. Kreitman and A. Shtein for many valuable discussions. The support of the study by the NIH Grant RR07801, the CRDF GAP Awards RUB-1578-ST-05, the RFBR Grants 09-04-01590-a, 08-01-00712-a is gratefully acknowledged.

SYSMO –DB: DATA MANAGEMENT FOR SYSTEMS BIOLOGY PROJECTS

Olga Krebs[*,3], Katy Wolstencroft[1], Stuart Owen[1], Finn Bacall, Jacky Snoep[2], Wolfgang Mueller[3] and Carole Goble[1]
[1]*School of Computer Science and*
[2]*Manchester Centre for Integrative Systems Biology, University of Manchester, UK*
[3]*Heidelberg Institute for Theoretical Studies, Germany*
e-mail: olga.krebs@h-its.org
[*]*Corresponding author*

Key words: *model, SysMO-DB, Systems Biology data*

Abstract

SysMO-DB is a web-based exchange environment for scientists to share and exchange Systems Biology data and models. It was designed to hold the research outcomes from the SysMO consortium (Systems Biology of Microorganisms), but the principles and methods employed could be generally applicable to Systems Biology research.

SysMO-DB is composed of the SysMO SEEK and the SysMO JERM. The SEEK is a Yellow Pages and assets catalogue, describing who holds what resources and where they can be accessed. It is the main user interface to the system and links out to data, models, protocols and analysis methods hosted by individual projects. The JERM ("Just Enough Results Model") is the underlying model for understanding the structure and content of assets, and extracting them from their source. A JERM for any one type of data (i.e. microarray data, or metabolomic data) is the minimum data schema that the SysMO projects agree to share. SysMO-DB is a light-touch approach to integration, allowing scientists to associate models and data in the context of experiments and the investigations they are a part of. In this talk I will describe SysMO-DB and its support of the interaction between data and models in Systems Biology.

CHIPMUNK: DISCOVERY OF TRANSCRIPTION FACTOR BINDING MOTIFS IN CHIP-SEQ DATA

I.V. Kulakovskiy*[1], V.A. Boeva[1,2,3,4,5], A.V. Favorov[1,6], V.J. Makeev[1]
[1]*Research Institute for Genetics and Selection of Industrial Microorganisms, Moscow, Russia*
[2]*Institut Curie, 26 rue d'Ulm, Paris, F-75248 France*
[3]*INSERM, U900, Paris, F-75248 France;* [4]*INSERM, U830, Paris, F-75248 France*
[5]*Mines ParisTech, Fontainebleau, F-77300 France*
[6]*Johns Hopkins University, Baltimore MD, USA*
e-mail: ivan.kulakovskiy@gmail.com
[*]*Corresponding author*

Key words: *motif discovery, ChIP-Seq, transcription factor binding sites*

Background: The task of identification of transcription factor binding motifs in a limited number of short DNA sequences has a long history. Recently upcoming ChIP-Seq data provided a new challenge for motif discovery. Such data consist of thousands of sequences where a 'short' overrepresented motif is to be found. Fortunately, in the case of ChIP-Seq data one has additional information, which helps to select the correct signal. This information is coverage profile constructed for DNA fragments obtained from ChIP-Seq experiments.

Methods and Algorithms: Here we present a ChIP-Seq extension of our motif discovery algorithm ChIPMunk [1]. It is an expectation-maximization-with-bootstrapping algorithm which uses coverage profile information by setting a specific weight profile over each sequence to represent a prior probability for the sequence position to overlap with a binding site.

Results: We have tested our algorithm on several ChIP-Seq data sets, including those for NRSF, GABP and for the oncogenic protein EWS-FLI1. Using coverage profile information as the motif positional preference prior, we successfully identified the correct motifs without traditional strict truncation of large enriched regions. We show that ChIPMunk motif recognition quality is the same or better than that of the traditional (MEME [2]) or ChIP-Seq-oriented (HMS [3]) tools while the speed is dramatically better.

Conclusion: ChIPMunk can be effectively used to analyze large sequence sets and therefore is a helpful tool for prediction of transcription factor binding motifs in ChIP-Seq data.

Availability: ChIPMunk is freely available within the ru_genetika Java package. Web-based version is also available: http://line.imb.ac.ru/ChIPMunk

Acknowledgements: This work was supported by RFBR#07-04-01623 and #07-04-01584 and RSASI#02.740.11.5008.

References:

1. T.L. Bailey et al. (2009) MEME Suite: tools for motif discovery and searching, *Nucleic Acids Res,* **37:** W202-W208.
2. Ming Hu et al. (2010) On the detection and refinement of transcription factor binding sites using ChIP-Seq data, *Nucleic Acids Res,* doi:10.1093/nar/gkp1180.
3. I.V. Kulakovskiy, V.J. Makeev (2009) Integration of data obtained by different experimental methods to determine the motifs in DNA sequences recognized by transcription-regulating factors [in Russian], *Biofizika*, **54(6)**: 965-74.

MOLECULAR MECHANISMS UNDERLYING FIDELITY OF RNA SYNTHESIS BY BACTERIAL RNA POLYMERASE

A. Kulbachinskiy[1]*, D. Pupov[1,2], N. Miropolskaya[1], D. Esyunina[1,2]
[1]*Institute of Molecular Genetics, Russian Academy of Sciences, Moscow, 123182*
[2]*Moscow State University, Biological Faculty, Moscow, 119991*
e-mail: akulb@img.ras.ru
[*]*Corresponding author*

Key words: *RNA polymerase, transcription fidelity, RNA cleavage, proofreading*

High fidelity of transcription is required for correct expression of genetic material. Transcription errors may lead to accumulation of nonfunctional RNA molecules and mutant variants of proteins and affect cell viability. The fidelity of RNA synthesis by RNA polymerase depends on two processes: (1) efficient discrimination of correct and incorrect substrates during nucleotide addition to the RNA 3'-end in the RNA polymerase active center, and (2) removal of incorrectly incorporated nucleotides by RNA cleavage that occurs in the same active center as RNA synthesis. Substrate discrimination depends on interactions of the incoming nucleotide with the complementary template DNA nucleotide and with several parts of the active center that contact the base, ribose and triphosphate groups of the substrate. These contacts promote closing of the substrate in the active center; this is required for efficient catalysis and supports more efficient incorporation of correct substrates. The reaction of RNA cleavage occurs in backtracked elongation complexes and is regulated by specialized protein factors (Gre factors in bacteria). RNA cleavage likely depends on RNA polymerase elements that contact the RNA 3'-end and promote backtracking of the elongation complex. RNA cleavage can be greatly stimulated by incorporation of incorrect nucleotides in the RNA 3'-end resulting in efficient correction of transcription errors. Processive and precise RNA synthesis depends on coordinated regulation of polymerizing and cleavage activities of RNA polymerase in the transcription cycle, including regulation by various transcription factors that change catalytic properties of the RNA polymerase active center.

NEW MOUSE STRAINS FOR BEHAVIORAL AND PSYCHO-PHARMACOLOGICAL GENETICS

Alexander Kulikov*, Daria Bazovkina, Daria Osipova, Maria Tikhonova, Elena Kondaurova, Nina Popova
Institute of Cytology and Genetics, Russian Academy of Sciences, 630090 Novosibirsk, Russia, e-mail: v_kulikov@bionet.nsc.ru
*Corresponding author

Key words: *Inbred mice, congenic mice, behavioral disorders, serotonine, cytokines*

Motivation and Aims: The involvement of brain neurotransmitters, cytokines and trophic factors in the molecular mechanisms of behavior and psychotropic drug action is commonly accepted. Behavior correction with psychotropic drugs is usually studied using pharmacological approach and knockout of the genes involved in brain neurotransmission. Here new mouse lines for study the role of serotonin and cytokines in regulation of various kinds of normal and pathological behavior were presented.

Methods and Algorithms: In the laboratory of behavioral phenogenetics of the Institute of Cytology and Genetica (Novosibirsk, Russia) three new mouse lines were created: 1) the ASC/Icg (Antidepressant Sensitive Catalepsy) mouse line selectively bred from a backcross population between CBA/Lac and AKR/J strains for high predisposition to catalepsy; 2) the AKR.CBA-D13Mit76 congenic mouse line with the 61-70 cM CBA-derived fragment of chromosome 13 transferred to the AKR genome (Kulikov et al., Genes, Brain Behav., 2008, 7:506-512); and 3) the B6-1473G congenic mouse line with the 1473G allele of the tph2 gene decreasing activity of the rate-limiting enzyme of serotonin synthesis in the brain, tryptophan hydroxylase-2, transferred to the C57BL/6J genome (Osipova et al., J. Neurosci. Res., 2009, 87:1168).

Results: The major gene defining predisposition to catalepsy was mapped on the 61-70 fragment of mouse chromosome 13 and linked to the Il6st gene coding the gp130 protein associated with cytokine receptors. The ASC mice showed numerous depressive-like traits and altered serotonin neurotransmission compared with the parental CBA and AKR strains. Chronic antidepressant treatment decreased catalepsy and normalized immunity in ASC, but did not affect these traits in CBA mice. The transfer of the CBA-derived 61-70 cM fragment of chromosome 13 to the AKR genome increased predisposition to catalepsy, aggression and sensitivity to lipopolysaccharide in the AKR.CBA-D13Mit76 congenic mice. The transfer of the 1473G allele of the tph2 gene to the C57BL/6 genome significantly affected the aggression and depressive-like immobility in the forced swim test.

Conclusion: 1. The ASC mouse line meets face, predictive and construct validity criteria of animal model of depression and antidepressant drugs screening.

2. The AKR.CBA-D13Mit76 congenic mouse line with altered gp130 protein is a promising model to study the role of cytokines in behavior regulation.

3. The B6-1473G congenic mouse line with altered tph2 gene is a valuable model of the serotonergic mechanisms of behavior regulation.

THE DISCRETE MODEL OF THE GENE NETWORKS REGULATORY LOOPS WITH THE THRESHOLD FUNCTIONS

A.A. Evdokimov, E.O. Kutumova*
Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia
e-mail: e.o.kutumova@gmail.com
**Corresponding author*

Key words: *gene network, discrete model, regulatory loop, functional graph, working vector, threshold function, cycles and fixed points, dynamics of states.*

Motivation and Aim. Gene networks have an important role in the living systems functioning. A characteristic feature of its organization is their ability to regulate itself through the regulatory loops with positive and negative feedbacks. These two types of loops make it possible to maintain a certain functional state or slide to another state of a gene network [1]. We consider one of the methods for description and modeling of gene networks – in terms of the discrete models of the regulatory loops functioning.

Methods and Algorithms. The regulatory loop is a connected digraph with n vertices, identified with the genetic elements (RNA, proteins), and a set of arcs, associated with the regulatory relations. Circulant digraphs $G_{n,k}$, where (k–1) is a number of inputs, are considered [2]. Variables taking integer values with the threshold p are corresponded to all vertices. Each value represents a concentration of a product, identified with the given vertex in a moment of time. The functioning is characterized by the stepwise changing of states - n-vectors in the alphabet <0,1,...,p-1>. Working vectors are the vectors which have inputs in the functional graph of the regulatory loops [3].

Results. We propose an approach to the complexity analysis for the discrete models of gene networks on the basis information on their functioning. We present the theoretical and computer analysis for various parametric data of the regulatory loops. The theorems characterizing structural properties, fixed points and cycles of the functional graphs were received. In particular, the recurrent relation for the number of working vectors in the case $p = 2$ was founded, and the asymptotic behavior of this number was described. Also the explicit formula was obtained for the arbitrary values of n, k, p.

Conclusion. Investigation of the regulatory loops functioning represented here provides an opportunity to understand the regulatory mechanisms of the processes under the control of gene networks and possibility for a directional impact on them.

Availability. The program for the functional calculation of the regulatory loops was also developed. This program finds the number of fixed points, cycles and other parameters of the functional graphs and allows exploring the dynamics of the states changing.

Acknowledgements. The work was supported by the Russian Foundation for Basic Research (08-01-00671, 09-01-00070) and the interdisciplinary grant 119 of SB RAS.

References:

1. V.A. Likhoshvai, V.P. Golubyatnikov, G.V. Demidenko, A.A. Evdokimov, S.I. Fadeev. (2005) Theory of the gene networks (Russian). In: *System computerized biology*, Eds. N.A. Kolchanov and S.S. Goncharov, 430–576, Novosibirsk, SB RAS.
2. E.D. Grigorenko, A.A. Evdokimov, V.A. Likhoshvai, I.A. Lobareva. (2005) Fixed points and cycles of the automata mapping for the gene network simulation (Russian). *Herald of Tomsk State University,* **14**: 206 - 212.
3. A.A. Evdokimov, E.O. Likhovidova. (2008) The discrete model of the circulant gene network with the threshold functions. *Herald of Tomsk State University,* **2(3)**:18-21.

MODULAR MODELING OF THE APOPTOSIS MACHINERY

E.O. Kutumova[1,2,*], R.N. Sharipov[1,3], I.N. Lavrik[4], F.A. Kolpakov[1,2]
[1]*Institute of Systems Biology, Novosibirsk, Russia*
[2]*Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia*
[3]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[4]*German Cancer Research Center (DKFZ), Heidelberg, Germany*
e-mail: e.o.kutumova@gmail.com
**Corresponding author*

Key words: *apoptosis, modular modeling, comprehensive map*

Motivation and Aim. Apoptosis is a highly regulated and evolutionary conserved process of cell death that plays a critical role in development and maintenance of tissue homeostasis. Formal description of pro- and anti-apoptotic machinery was enriched significantly for the last years and a range of mathematical models was created. However these models mainly describe different segments of implicated pathways and integrated model of apoptosis regulation has not been developed yet.

Methods and Algorithms. Existing models from literature and the Biomodels database were used as the source of data for modeling. BioUML workbench (http://www.biouml.org) was used for creation and parameters fitting of the integrated model of apoptosis. The BMOND database (http://bmond.biouml.org) was used as a repository of diagrams, mathematical models and results of simulation.

Results. We developed the integrated model of apoptosis consisting of 13 functional modules: TRAIL, CD95, TNF-α, EGF, NF-kB, p53, the mitochondria, cytochrome c, SMAC, type I cells, caspase-12, PARP-1 and apoptotic execution phase modules. Each module implements the independent pathway (or its part) for transmission of the intracellular signals. Input and output sets of the biological components are uniquely determined and each module can be processed and enhanced on the base of experimental data independently from others. Totally the model comprises 286 species in five different compartments and 684 reactions. It contains about 700 unknown parameters (initial values for protein concentrations and constants for reaction rates). To evaluate those parameters and fit them we used 24 experimental datasets. First, we fitted parameters for each module individually. Then obtained values were used as initial approximation for fitting of parameters for combinations of modules and, finally, for the whole model.

Conclusion. Here we have demonstrated how the model of the complex biological system with the huge number of parameters can be constructed by the modular way and fitted to the experimental data.

Acknowledgements: This work was supported by EU grant №037590 “Net2Drug”.

Availability. The integrated model is available in the BMOND database (http://bmond.biouml.org).

References.

1. G.Joshi-Tope et al. (2005) Reactome: a knowledgebase of biological pathways, *Nucleic Acids Research,* **33**: D428-D432.
2. M.Krull et al. (2003) TRANSPATH: an integrated database on signal transduction and a tool for array analysis, *Nucleic Acids Research,* **31(1)**: 97–100.

DWARF FORM OF *MALUS BACCATA* (L.) BORKH: INITIAL STAGE OF THE PARAPATRIC SPECIATION?

E.V. Kuznetsova*[1], T.E. Peretolchina[2], A.V. Rudikovskiy[1], D.Yu. Sherbakov[2,3]
[1]*Siberian Institute of Physiology and Biochemistry of Plants SB RAS, Irkutsk, Russia*
[2]*Limnological Institute SB RAS, Irkutsk, Russia*
[3]*Irkutsk State University, Faculty of Biology and Soil Science*
e-mail: elenakuznetsova01@gmail.com
[*]*Corresponding author*

Key words: *Apple-trees, M. baccata, dwarf forms, microsatellites, speciation.*

Motivation and Aim: Many authors consider the parapatric speciation as important way of speciation among animals, however, it is not so common in plants. One rare example of the possible parapatric speciation in plants is Siberian apple tree *Malus baccata* (L.) Borkh and its rare form dwarf apple tree which is found on the territory of the Republic of Buryatia in the contact zone of forest and steppe. Taxonomic status of the dwarf apple tree remains undecided. So the aim of our investigation was determining of the phylogenetic relationships between different forms of Siberian apple trees and clarifying of the origin of the dwarf apple tree.

Methods and Algorithms: The samples were collected on the territory of the Republic of Buryatia. We chose four geographic locations of plants: №1 is mixed, it consists of dwarf and tall forms of the apple-trees and grows in a valley of Zagustaj river in Republic Buryatia. The second group is represented by the dwarf forms growing near Yagodnoe village near Gusinoozersk along a stream. The group №3 is natural typical tall forms of *M. baccata* that occupy the territory around Kabansk village. Group №4 is man-made planting of tall forms in the Yagodnoe village. Genomic DNA was extracted by the modified methods of Doyle and Dicson from leaves.

Six pairs of specific micosatellite primers were used for analysis of genetic diversity within the studied groups of plants: 01ab, 02b1, 04H11, 05G8, 23G4, 28F4. Genetic structure of studied apples-trees was determined using software STRUCTURE 2.3.1. (number of populations K from 2 to 5, burnin period – 50 000, number of MCMC reps after burn – 1 000 000). Phylogenetic tree based on micosatellite loci was reconstructed using distance matrix DAS.

Results: Microsatellite phylogenetic tree has shown that the all studied groups of dwarf and tall forms of Siberian apple tree form independent clades and also dwarf forms is separated from tall plants. Dwarf forms look polyphyletic on the tree.

Conclusion: At the edge of the areal the differences between the specific habitats are quite high and may play a role in a selection, which is reflected in the genetic differentiation of various growth forms of apples. On the other hand, only extreme biotopes (for example, highly arid places) where there is a dwarf form, can favor to the survival of such marginal forms which, apparently, are more adapted for these conditions. The question of the heritability of dwarfism remains open, due to the long duration of generations within apple trees. The phylogenetic analysis has shown that the dwarf form has descended from tall *M. baccata*. This fact suggests that dwarf apples are ecological forms of the Siberian apple which, probably, represents the initial stage of the parapatric speciation.

GENE LOSS AND ACQUIRING IN EVOLUTION OF PROKARYOTIC COMMUNITIES – MODELING WITH EVOLUTIONARY CONSTRUCTOR PROGRAM

S.A. Lashin*, V.V. Suslov, Yu.G. Matushkin
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
e-mail: lashin@bionet.nsc.ru
**Corresponding author*

Key words: *complexity evolution, modeling, prokaryotic community, gene loss.*

Motivation and Aim: Prokaryotes are – the world of tiny genomes and huge size communities. Consuming external nonspecific substrate (NS) populations of various species divide functions in community and act as compartments in associated metabolic reactions. By virtue of exchange of metabolites and other specific substrates (SS) various species form trophic networks (graphs) [1]. Reduction of genome size is considered to be one of the major trends of prokaryotic evolution, as the replication determines reproduction rate (i), and specialization of species in community facilitates loss of functions and, consequently, genes (ii). In the present study the role of both ecocenotic and genetic factors in evolution of prokaryotic genomes size was investigated *in silico*.

Methods and Algorithms: The modeling tool "Evolutionary constructor" – "EC" [2] was used for computer simulations of evolution.

Results: We have modeled evolution of trophic networks of haploid organisms' populations, providing each other with SS and consuming the common NS. The time of evolution simulation was up to 30000 generations. Growth rate of a population depended on number of consumable SS, efficiency of SS/NS consumption/utilization (both factors are genetically determined), and total genome size of population cells (bigger genomes get bigger penalties). Deficiency of NS in environment could be partially compensated with an excess of SS. Against the backdrop of NS concentration oscillations (from sublethal to excess) both gene loss and horizontal gene transfer (HGT) could occur. It has been shown that trend of gene reduction is adaptive only in conditions of NS excess. In this case it involves all members of community. In conditions of NS deficiency the adaptive trend is increase of genomes size owing to HGT. In normal conditions events of gene loss and transfer redistributed genes in populations-members of community which changed proportions of their sizes.

Conclusion: In highly-integrated community genome size reduction becomes a major trend in comfortable conditions, and occasionally occurs in sub-comfortable ones. However in non-comfortable conditions integrated metabolism conduces community members to acquire genes.

Acknowledgements: The work is supported by the RFBR grant 10-04-01310-a, RAS program 22 (subprogram 8), RAS program "Biosphere origin and evolution".

References:

1. G.A.Zavarzin (2003) Natural microbiology. Lectures. M.: Nauka, 348 p.
2. S.A.Lashin et al. (2010) Comparative modeling of coevolution in communities of unicellular organisms: adaptability and biodiversity // JBCB (in press).

POLY(ADP-RIBOSE) POLYMERASE 1 IS A KEY REGULATOR OF DAMAGE PROCESSING IN BASE EXCISION REPAIR

O.I. Lavrik*, S.N. Khodyreva, E.S. Ilina, M.V. Sukhanova, M.M. Kutuzov
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
e-mail: lavrik@niboch.nsc.ru
**Corresponding author*

Key words: *PARP1, base excision repair, affinity modification, chemically reactive DNA*

Motivation and Aim: Poly(ADP-ribose) polymerase 1 (PARP1), functioning as DNA nick-sensor, interacts with base excision repair (BER) DNA intermediates containing single-strand breaks. Bound to DNA breaks, PARP1 catalyzes synthesis of poly(ADP-ribose) covalently attached to some nuclear proteins and itself. Autopoly(ADP-ribosyl)ation of PARP1 facilitates its dissociation from DNA breaks and is considered as a factor regulating DNA repair. The aim is to study the role of PARP1 and its autopoly(ADP-ribosyl)ation in regulation of damage processing by the BER enzymes.

Methods and Algorithms: Affinity modification, MALDI-TOF-MS, enzyme functional tests.

Results: We identified PARP1 among the BER proteins cross-linked to the photoreactive branch-point BER DNA intermediate. Apurinic/apyrimidinic endonuclease 1 (APE1), DNA polymerase β (Pol β), flap endonuclease 1 (FEN1) were found to crosslink to the same BER intermediate. By functional assays in reconstituted systems, as well as in cell extracts, we demonstrated that PARP1 and its poly(ADP-ribosyl)ation is involved in regulation of activity of the base excision repair enzymes – APE1 (3'-5' exonuclease activity), Pol β, FEN1. PARP1 was shown to more efficiently influence DNA synthesis in long patch BER and its poly(ADP-ribosyl)ation pathways. PARP1's ability to interact with intact AP sites and AP sites processed by APE1 via covalent via Schiff base intermediate was demonstrated in cell extracts and with pure PARP1 protein. The identity of PARP1 as the target for cross-linking to AP sites in cell extracts was proved by peptide mapping on the data of MALDI-TOF-MS. PARP1 is unable to cleave AP sites, but instead forms with AP sites a stable intermediate. PARP1 is unable to be autopoly(ADP-ribosyl)ated upon binding with AP sites until processing of AP site by APE1.

Conclusion: Thus, in addition to well-known role of PARP1 as nick-sensor, we demonstrated potential role of this protein in regulation of earlier stage of the BER process preceding incision of sugarphosphate backbone of DNA. Interaction of PARP1 with AP sites along with the previously detected interaction with DNA breaks demonstrates that PARP1 is a key sensor of lesions appeared in BER. PARP1 can regulate initial stage of BER persisting on AP site until APE1 could come and initiate AP site cleavage. The PARP1 activation and resultant automodification appears to facilitate the BER factor recruitment to stimulate the repair process. In the absence or deficiency of APE1, PARP1 can provide temporal protection of AP sites. PARP1 is able to interact also with AP sites cleaved by APE1. However, in the absence of Pol β, when the removal of 5' deoxyribose is failed, interaction of PARP1 with this intermediate stimulates synthesis of poly(ADP-ribose) polymer, which is known as a death signal.

Acknowledgements: This work was supported by RFBR, project № 10-04-01083, Program of RAS "Molecular and Cellular Biology" and State contract № 02.740.11.0079.

TOWARDS UNDERSTANDING LIFE/DEATH DECISIONS AT CD95

I.N. Lavrik*, C. Pforr, N. Fricker, R. Eils, P.H. Krammer
German Cancer Research Center, DKFZ, Heidelberg
i.lavrik@dkfz.de
**Corresponding author*

Key words: *apoptosis, modeling, ODEs, CD95, caspases*

Motivation and Aim: Apoptosis is a complex but highly defined cellular programme of cell demolition. Recently, substantial progress in understanding apoptotic signaling networks has been achieved using systems biology. Here we show how systems biology allows to understand life/death decisions made in the cell and to establish new ways of treating the diseases associated with defects in apoptosis on the basis of our work on modeling CD95-induced apoptosis.

Methods and Algorithms: In this study we implemented modeling using Ordinary Differential Equations (ODEs) and all data were generated biochemically using quantitative western blots. The data generated were used for building a systems biology model of CD95 signaling, which combines the apoptotic and the non-apoptotic signaling network.

Results: CD95 (APO-1/Fas) is a member of the death receptor (DR) family, a subfamily of the TNF-R (tumor necrosis factor receptor) superfamily. Engagement of CD95 leads to the induction of the apoptotic and non-apoptotic signaling pathways. First step in signaling is the formation of the DISC, death-inducing signaling complex. We have shown that two proteins of the DISC: procaspase-8 and c-FLIP regulate both apoptotic and non-apoptotic pathways in a dynamic way.

Conclusion: Using systems biology approach we have established the switch between apoptotic and non-apoptotic signaling. The predictions and implications of this model of life and death will be discussed.

SIMULATING PULSE WAVE IN 1D HEMODYNAMIC MODEL

T.I. Leonova[1,2,3,*], E.A. Biberdorf[3,4], F.A. Kolpakov[1,2].
[1]Institute of Systems Biology, Novosibirsk, Russia; [2]Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia; [3]Novosibirsk State University, Novosibirsk, Russia; [4]Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia
**e-mail: zlove.t@gmail.com*

Key words: *blood flow circulation, 1D model of hemodynamics, pressure pulse wave*

Motivation and Aim. Pulse wave velocity (PWV) is a measure of arterial stiffness and could indicate a state of arterial wall in case of different diseases including essential hypertension. Simulation of pulse wave in human arterial system can be used for validation of one-dimensional (1D) model of hemodynamics [1].

Methods and Algorithms. BioUML technologies (http://www.biouml.org) and "Hemodynamics" plug-in were used for simulation PWV in 1D model of hemodynamics. Blood dynamics has been simulated for hypothetical vessel of the length 1.5m as well for model of human arterial system consisting of 55 arteries. The results of simulation has been compared with Moens-Korteweg equation: $V_{PW} = \sqrt{\frac{Eh}{2\rho_b r}}$, where E is the elastic modulus, h is the thickness of the arterial wall, r is the internal radius of the artery and ρ_b r $_{b\ b}$– blood density [2].

Results. Dependence on the thickness of the arterial wall and the internal radius of the artery has been analyzed. Figure 1 demonstrates PWV obtained by numerical simulation and calculated by Moens-Korteweg equation. Carotid-femoral PWV has been obtained from the model of 55 arteries and its value 8.2m/s in a normal case.

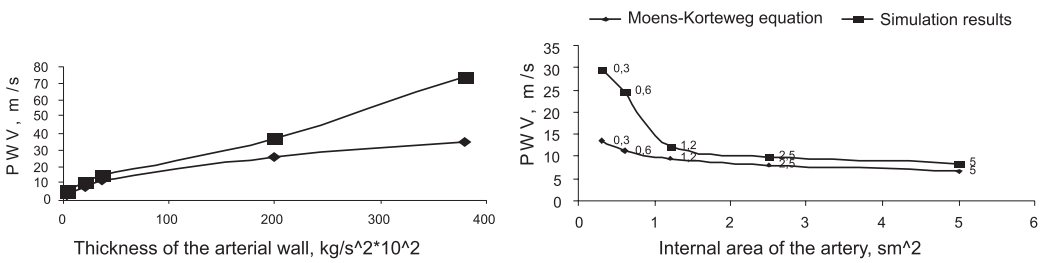


Figure 1. Comparison of PWV obtained by numerical simulation and calculated by Moens-Korteweg equation. A) Area=2.5cm2; B) Thickness=2*103kg/s2.

Conclusion. Developed model of hemodynamics can reproduce adequately phenomenon of pulse wave. In physiological range from 5 to 11 m/s PWV obtained by numerical simulation and calculated by Moens-Korteweg equation are in good correspondence. PWV value for model from 55 arteries corresponds well to empirical values (5–10m/s).

Acknowledgements. This work was supported by integration and interdisciplinary grant №91 of Siberian Branch of Russian Academy of Sciences.

References:

1. Blood circulation system and arterial hypertension: biophysical and genetic-physiological mechanisms, mathematical and computer modeling. Eds. Ivanova L.I., Blokhin A.M., Markel A.L. Novosibirsk: Siberian Branch of Russian Academy of Sciences Press, 2008, 252pp. (Integration projects SB RAS; Issue 17)
2. Biophysics. Eds. Antonov V.F. Moscow, Vlados, 2000, 287pp.

NUMERICAL ANALYSIS OF COMPLEX MODEL OF HUMAN BLOOD FLOW CIRCULATION USING 1D HEMODYNAMIC MODEL

T.I. Leonova[1,2,5,*], E.A. Biberdorf[3,5], F.A. Kolpakov[1,2], A.M. Blokhin[3,5], A.L. Markel[4]
[1]Institute of Systems Biology, Novosibirsk, Russia; [2]Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia; [3]Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia; [4]Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia; [5]Novosibirsk State University, Novosibirsk, Russia
**e-mail: zlove.t@gmail.com*

Key words: *blood flow circulation, essential hypertension, 1D model of hemodynamics, orthogonal sweep method, BioUML*

Motivation and Aim. A complex mathematical model of blood flow circulation has been designed for the purpose of scientific research of human essential hypertension progress. Aims of the work were verification and validation of one-dimensional (1D) model of hemodynamics and development of a closed model of the human blood flow circulation including a filtration block and a model of the heart.

Methods and Algorithms. The model of the human arterial system was created as a graph of 55 arteries (edges). The blood flow in an artery is described by 1D model. Methods of lines and orthogonal sweep were used for calculations. BioUML (http://www.biouml.org) technologies were used for development of "Hemodynamics" plug-in that provides user friendly interface for creating model of arterial tree as a graph. The plug-in automatically generates Java or Matlab code for hemodynamics simulation based on this graph.

Results. The model of the human arterial system was created on the base of one-dimensional model of blood flow circulation in 55 main arteries using methods of lines and orthogonal sweep. Numerical data for blood flow dynamics were obtained for all arteries and cross-sections examined. Simulation results can be visualized as a plot as well as saved as a text file [1]. The model was verified and validated on the basis of researching a velocity of the pressure pulse wave. Numerical modeling of an inflatable cuff effect was based on linear and nonlinear functions and has been designed for purposes of validation of the model and researching Korotkoff sounds. The virtual blood circulatory system was closed by coupling 1D hemodynamics, two-chambered heart and blood filtration models. The complex model of human blood circulation will be used in investigations of essential hypertension development.

Conclusion. Concluding the obtained results we assume that the model is able to simulate basic physiological processes that take place in the human arterial system.

Availability. Software is available as a "Hemodynamics" plug-in for BioUML on website: http://www.biouml.org/download.shtml?0.8.6/workbench

Acknowledgements. This work was supported by integration and interdisciplinary grant №91 of Siberian Branch of Russian Academy of Sciences.

References:

1. Blood circulation system and arterial hypertension: biophysical and genetic-physiological mechanisms, mathematical and computer modeling. Eds. Ivanova L.I., Blokhin A.M., Markel A.L. Novosibirsk: Siberian Branch of Russian Academy of Sciences Press, 2008, 252pp. (Integration projects SB RAS; Issue 17)

DETECTION OF RELATION BETWEEN PRESENCE OF CONTEXT/REGULATORY SIGNALS IN PROMOTERS OF GENES AND THEIR DIFFERENTIAL EXPRESSION

Levitsky[1,2] V.G., Oshepkov[1,2] D.Y., N.I. Ershov[1], L.O. Bryzgalov[1], T.I. Merkulova[1,2]
[1]*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
e-mail: levitsky@bionet.nsc.ru
[*]*Corresponding author*

Key words: *microarray data, transcription factor binding sites, tandem repeats*

Motivation and Aim: Microarray data gave tremendous promise for understanding of regulation of gene expression. Since detection of significant differences in gene expression depends on multiple biological and technical variables, extracting biological knowledge from microarray data requires adequate statistical treatment.

Methods and Algorithms: We used rat liver as the model to study mechanisms of toxic action of different cancerogens. We analysed the microarray data of hepatic transcriptome changes of rats treated by aminoazo dyes – hepatocarcinogenic 3′-MeDAB or non-carcinogenic OAT. The data of microarray were referred to 5934 unique accession numbers of RefSeq database (http://www.ncbi.nlm.nih.gov/RefSeq/). They were used to extract curated nucleotide sequences of upstream regions of rat genes ([-2000;+1] relative to annotated transcription starts) from http://hgdownload.cse.ucsc.edu/downloads.html (UCSC Genome Browser Downloads site). χ^2 tests for 2x2 contingency tables were used to evaluate the significance of enrichment/depletion of potential binding sites (PBS) of liver-specific transcription factor (TF) FoxA and not specific for liver TF SREBP and SF-1. PBS were recognized by SITECON method [1] in upstream regions of genes; for these sequences we also computed frequencies of tandem repeats of types $(N_1N_2N_3N_4)_3$ (N – any nucleotide).

Results: The log ratios reflecting the compound-specific change of expression were used to rank all genes. 20% of genes showing the smallest change were used as the control group (ranked as [40%;60%]). We compared the enrichment/depletion of PBS and $(N_1N_2N_3N_4)_3$ repeats in upstream regions of genes ranked either as [X%; (X+20)%] (repression) or [(80-X)%; (100-X)%] (activation) against those for the control group, where X denotes 0, 1, ...19%. χ^2 test was used to verify reliability of observed difference. Since we did multiple comparisons at once Bonferoni correction for 20 tests was taking into account.

Conclusion: Upstream regions of genes that were activated by both compounds were enriched by PBS of liver-specific TF FoxA, but this was not the case for TF SREBP and SF-1. These upstream regions also have shown the enrichment of $(TTTG)_3$ repeats which earlier showed strong binding affinity to FoxA proteins in vitro [2].

Acknowledgments

The work was supported by state contact no. 02 740 11 0705.

References:

1. Oshchepkov D.Y., et al., (2004) SITECON: a tool for detecting conservative conformational and physicochemical properties in transcription factor binding site alignments and for site recognition. *Nucl. Acids Res,* **32:**208-212.
2. Bryzgalov L.O., et al., (2008) Detection of target genes of FOXA transcription factors involved in proliferation control. *Biochemistry (Mosc),* **73**:70-75.

RECOGNITION OF POTENTIAL BINDING SITES IN CHIP-SEQ DATA

Levitsky[1,2] V.G., Oshepkov[1,2] D.Y., N.I. Ershov[1], Vasiliev G.V., T.I. Merkulova[1,2]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
e-mail: levitsky@bionet.nsc.ru
**Corresponding author*

Key words: *ChIP-Seq experiments, transcription factor binding sites, weight matrix.*

Motivation and Aim: Extracting biological information from ChIP-Seq data requires application of appropriate statistical analysis. Transcription factor (TF) FoxA (old name HNF3) is critical to the development and function of the liver [4].

Methods and Algorithms: We applied optimized position weight matrices (oPWM) [1] to recognize potential binding sites (BS) of TF FoxA in a ChIP-Seq derived data.

Results: Nucleotide sequences of 37 functional FoxA BS were retrieved from TRRD [2]. Three partially overlapping training sets of BSs were constructed on the basis of degenerated motives YRTTTRYBYWDD, BNTSTTTKBHBW and HTVTTTGBDBH [3]. On the basis of these training sets three oPWM were constructed for recognition of potential BS of FoxA. 11475 mouse genomic sequences, that contained ten or more reads of DNA sequences derived from FoxA ChIP-Seq experiment [4], were extracted from whole chromosomes. Since the lengths of reads were in the range from 100 to 300 nt [4] we considered any position in a sequence as approved by PWM if it was spaced from closest hit of PWM not farther than 200 nt. Simultaneous application of three PWM methods allowed to predict FoxA BS in substantially larger fraction of sequences derived from FoxA ChIP-Seq experiment than those for any single method. This conclusion was confirmed for a wide range of permissible stringencies for three types of PWM and ChIP-Seq (number of overlapping reads).

Conclusion: Detailed analysis of training BS and discrimination of several types of BS may be the critical step for reliable verification of genome-scale ChIP-Seq data by TFBS recognition programs.

Acknowledgments
The work was supported by state contact no. 02 740 11 0705.

References:

1. V.G. Levitsky, et al., (2007) Effective transcription factor binding site prediction using a combination of optimization, a genetic algorithm and discriminant analysis to capture distant interactions. *BMC Bioinformatics,* **8:**481.
2. N.A. Kolchanov, et al., (2002) Transcription Regulatory Regions Database (TRRD): its status in 2002. *Nucl Acid Res*, **30:**312-317.
3. V.G. Levitsky, (this issue) Application of motif discovery tool for FoxA binding sites analysis.
4. E.D. Wederell, et al., (2008) Global analysis of in vivo Foxa2-binding sites in mouse adult liver using massively parallel sequencing. *Nucl Acid Res,* **36:**4549-64.

APPLICATION OF MOTIF DISCOVERY TOOL FOR FOXA BINDING SITES ANALYSIS

Levitsky V.G.
Institute of Cytology & Genetics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
e-mail: levitsky@bionet.nsc.ru

Key words: *transcription factor binding sites, motif discovery, genetic algorithm*

Motivation and Aim: Motif discovery is still is an important step in annotation of sequenced genomes [1]. This task consists in detection of overrepresented motives in a dataset of nucleotide sequences. It is supposed that sequences contain multiple occurrences of binding sites (BS) of known or unknown transcription factor (TF). IUPAC 15-letter code is widely used for representation of ambiguous pattern of nucleotides (motif) in a given DNA sequence, where any character may represent more than one nucleotide. Since typical length of BS is 8-12 nt, exhaustive search in 15-letter alphabet requires calculation in the spaces of respectively 15^8 - 15^{12} dimensions (order of magnitude from 9 to 14). Stochastic method developed here may substantially decrease this search space.

Methods and Algorithms: A genetic algorithm (GA) based approach is employed to search for motives in 15-letter degenerate code. Markov model was used to measure the overrepresentation of motives.

Results: Developed approach was applied to construct alternate alignments of BS of TF FoxA. Nucleotide sequences of 37 functional FoxA BS were retrieved from Sample database [2]. GA scored motives based on: (a) T, portion of training sequences with at least one occurrence of a motif ($0 < T \leq 1$) b) F, estimated expected length of a background sequence that contain a motif. Background sequences were modeled by 0^{th} order markov model. For whole training dataset the value T = 0.55 was chosen since the next growth of this parameter caused the fall of F score below reasonable threshold. Among motives with highest F scores YRTTTRYBYWDD, BNTSTTTKBHBW and HTVTTTGBDBH were selected since for them the lowest mutual overlapping of respective locations in training sequences were observed. These motives were found respectively in 21, 22 and 21 sequences of training dataset; pairs of 1^{st} & 2^{nd}, 1^{st} & 3^{rd}, and 2^{nd} & 3^{rd} motives were found concurrently in 14, 12 and 15 sequences, respectively. Thus, 33 sequences out of total 37 contained at least one occurrence of any motives. Analysis of ChIP-Seq data [3] approved that developed approach is very promising for prediction of potential BS of FoxA.

Conclusion: Developed method may be used for the fast and effective search of degenerate motives.

References

1. T. Marschall, S. Rahmann (2009) Efficient exact motif discovery. *Bioinformatics,* **25:**i356-364.
2. N.A. Kolchanov, et al., (2002) Transcription Regulatory Regions Database (TRRD): its status in 2002. *Nucl Acid Res*, **30:**312-317.
3. V.G. Levitsky, et al., (this issue) Recognition of potential binding sites in ChIP-Seq data.

DETAILED CHARACTERIZATION OF SMALL SUPERNUMERARY MARKER CHROMOSOMES REVEALS BREAKPOINT HOT SPOTS

T. Liehr[1]*, N. Kosyakova[1], K. Mrasek[1], E. Ewers[1], S.W. Cheung[2], A. Weise[1]
[1]*Jena University Hospital, Institute of Human Genetics and Anthropology, Kollegiengasse 10, D-07743 Jena, Germany*
[2]*Baylor College of Medicine, Houston, Texas, USA*
e-mail: i8lith@mti.uni-jena.de
**Corresponding author*

Key words: a*rray-comparative genomic hybridization (aCGH), chromsoome microdissection, small supernumerary marker chromosomes (sSMC)*

Motivation and Aim: Breakpoint characteristics of small supernumerary marker chromosomes (sSMC) are not well studied by now.

Methods and Algorithms: Array-comparative genomic hybridization (aCGH) was done in 64 sSMC. The studied sSMC-specific DNA was derived from glass-needle based microdissection, avoiding by that the problem of mosaicism as present in ~40% of corresponding patients. Furthermore, a detailed analysis of overall 128 characterized breakpoints in non-heterochromatic chromosomal regions of sSMC was done.

Results: It turned out that 82.8% and 53.1% of the breakpoints are located within copy number variant regions and regions with segmental duplications, respectively. 6.3% of the breakpoints locate within sequence gaps, making an overall of 90.6% of association of non heterochromatic sSMC-related breakpoints with 'critical genome structure'. Moreover, approximately three quarters of the breakpoints were concordant with fragile sites. Still, there was a 7.0% overlap of the observed breakpoints and interspersed telomeric sequences (ITS), but only two out of 128 breaks were within an olfactory receptor gene family region.

Conclusion: Overall, we present the largest ever done aCGH study in sSMC and provide evidence for hotspots involved in sSMC formation. Deduced from that data it was also possible to characterize regions causing clinical problems if present additionally, for all human chromosomes except for chromosomes 6, 13, X and Y.

Availability: http://www.med.uni-jena.de/fish/sSMC/00START.htm

Acknowledgements: Supported in parts by the Dr. Robert Pfleger-Stiftung, DLR/BMBF ARM 08/001 and BLR 08/004, the DFG (LI820/17-1, LI 820/15-1, WE 3617/2-1), the Boehringer Ingelheim Fonds, the Erwin-Riesch Stiftung, the Evangelische Studienwerk e.V. Villigst, DAAD (D07/00070 and A0896742) and Prochance 2008 of the Friedrich Schiller University Jena 21007091.

IgE-MEDIATED IL-4 PRODUCTION BY MAST CELLS IS SPECIFICALLY REGULATED BY MKK3

Tong-Jun Lin and Adam J. MacNeil
Department of Microbiology and Immunology, Dalhousie University, Halifax, NS, Canada B3K 6R8

Motivation and Aim: Mast cells play a central role in allergic inflammation and are activated through crosslinking of receptor-bound IgE, initiating a signaling cascade resulting in production of biologically potent mediators. Signaling pathways in the regulation of specific mediators remain incompletely defined. Here, we examined the role of mitogen-activated protein kinase (MAPK) kinase 3 (MKK3) in IgE-dependent mast cell activation *in vitro* and *in vivo*.

Methods and Algorithms: In an *in vivo* model of passive cutaneous anaphylaxis, MKK3 knockout (KO) mice showed a deficit in IgE-dependent inflammation. To characterize the mechanism of this deficiency, we cultured bone marrow-derived mast cells (BMMCs) from wild-type (WT) and MKK3 KO mice.

Results and Conclusion: Mast cells from MKK3 KO mice matured normally *in vitro*, suggesting that MKK3 has no effect on mast cell development. We found that IgE-mediated mast cell activation induced rapid MKK3 phosphorylation by 5 minutes, diminishing slowly over 6 hours. In MKK3 KO BMMCs, phosphorylation of p38 was significantly reduced at both early and later time points, while other MAPKs were unaltered. Among 40 cytokines and chemokines tested using a protein array assay, IL-4 was the only cytokine specifically down-regulated in MKK3 KO cells. This finding was confirmed by ELISA. Thus, MKK3 is a specific regulator for IgE-induced IL-4 production in mast cells. We next examined the activity of candidate IL-4 promotor-binding transcription factors and found Early Growth Response factor (Egr) activity was diminished from 1-6 hours suggesting that MKK3 plays a role in the activation of Egr-1 which may in turn drive IL-4 expression.

COGNITIVE BIOINFORMATICS: USING MESH TERMS TO CREATE KNOWLEDGEBASES

A.V. Lisitsa*[1], Ponomarenko E.A. [1], Ilgisonis E.V. [1], Gusev S.A. [2], Archakov A.I. [1]
[1]*Institute of Biomedical Chemistry RAMS, Moscow, Russia*
[2]*Crytome Research, Moscow, Russia*
e-mail: lisitsa.av@gmail.com
[*]*Corresponding author*

Key words: *cognitive bioinformatics, text-mining, data-mining, web-usage mining*

Motivation and Aim: Cognitive bioinformatics deals with solving complex problems of heterogeneous data integration and presentation [1]. The major challenge is to produce such output, which is intelligible to the individual researcher.

Methods and Algorithms: The individual research profile is collected by monitoring the selective access of the user to the scientific papers. As most of the biomedical papers are associated to the list of MeSH terms, these are extracted to produce the user-specific research profile.

Results: We present results of 4-month experiment in monitoring the scientific Web-usage activity in 6 laboratories of the Institute of Biomedical Chemistry and in few labs from other research institutions. From the profiles of approx. 50 participants it was possible to visualize major research domains, where several distinct clusters were observed: proteomic methods, DNA-arrays, drug delivery, nano-safety and others. It was shown, that depending on the user profile, the specific relationships between the research directions could be elucidated.

Conclusion: Web-usage activity of biomedical research comprises the knowledgebase, which facilitates the coordination of multi-lab projects and initiate the sprouts of collaboration.

Availability: http://ws.bioknowledgecenter.ru

References:

1. O. A. Kuchar, J. F. Reyes-Spindola, M. Benaroch (2004) Cognitive Bioinformatics: Computational Cognitive Model for Dynamic Problem Solving. *Proceedings of the Third IEEE International Conference on Cognitive Informatics*, 84 – 92.

INTERACTION BETWEEN NUCLEOME AND PLASTOME: HEAT SHOCK RESPONSE REGULATION IN PLASTIDS OF PLANTS

V.A. Lyubetsky*, O.A. Zverkov, L.I. Rubanov, A.V. Seliverstov
Institute for Information Transmission Problems RAS, Moscow, Russia
e-mail: lyubetsk@iitp.ru
**Corresponding author*

Key words: *competition of RNA polymerases, RNA polymerases of bacterial and phage types, promoter binding efficiency, gene transcription level*

Motivation and Aim: Understanding interactions between nuclear and plastid genomes is important. Here we consider one example: a suggested mechanism of the heat shock response in chloroplasts. The mechanism well explains known experimental data for chloroplasts.

Methods and Algorithms: To study a competition of RNA polymerases of different types in plastids we developed a model based on a sophisticated system of interacting stochastic and deterministic processes. It is implemented in two versions: a program for multiprocessor clusters (tested with the 1024 parallel CPU on MBC-100K supercomputer of the Joint Supercomputer Center RAS) and a standalone PC program with real-time computing progress graphic monitor. Both programs are available at http://lab6.iitp.ru/rivals/, along with detailed algorithm descriptions and usage examples.

Results: Let's consider an example. *Hordeum vulgare* contains two copies of the following set of genes: *trnI*–*rpl*23–*rpl*2–(*trnH*)–*rps*19. One set competes with neighboring gene *psbA*: P1-*trnI-rpl23-rpl2*-(*trnH*-P2)-*rps19*-(*psbA*-P3), and the other set adjoins the next operon on the same strand: P1-*trnI-rpl23-rpl2*-(*trnH*-P2)-*rps19-rpl22-rps3-rpl16-rpl14-rps8-infA-rpl36-rps11-rpoA*. The transcription level ratios were measured experimentally for these sets at the temperatures of 21°C and 40°C [1].

Conclusion: Our model predictions conform within experimental error with the *in vitro* measurements for values of the promoter binding efficiency P1=1.4, P2=0.7, P3=0.3s^{-1}, and the RNA polymerase elongation rates R_{21}=9.2 and R_{40}=36.8bp/s at lower and higher temperatures, respectively, which also agrees with independent observations [2]. Thus, the modeled competition of RNA polymerases can explain the heat shock response mechanism at least in isolated chloroplasts. Noteworthy, our quantitative predictions are in good agreement with the sigma subunit knockout experiments in other loci, where we also predict the elongation termination sites verified with the multiple alignment and biological data. Such is locus P1–*psbB*–*psbT*–(*psbN*–P2)–*psbH*–*petB*–*petD*–(*rpoA*–*rps*11–P3–*rpl*36–*trnI*–N) in *Arabidopsis thaliana,* where the sites are conserved 44bp palindromes (presumably DNA cross-hairpins), e.g. TTAACGTAATCAGCCTCCAAATATTTGGA GGCTGATTACGTTAA, downstream the *psbT* gene.

References:

1. 1. Ya.O. Zubo, E.A. Lysenko, A.Yu. Aleinikova, V.V. Kusnetsov, N.L. Pshibytko. (2008) Changes in the transcriptional activity of barley plastome gene under heat shock, *Russian Journal of Plant Physiology*, 55(3): 293-300.
2. 2. E.A. Abbondanzieri, J.W. Shaevitz, S.M. Block. (2005) Picocalorimetry of transcription by RNA polymerase, *Biophysical. J.: Biophys. Letters*, doi: 10.1529/biophysj.105.074195.

LACK OF CONSERVATION OF BACTERIAL TYPE PROMOTERS IN PLASTIDS OF STREPTOPHYTA

A.V. Seliverstov, L.I. Rubanov, V.A. Lyubetsky*
Institute for Information Transmission Problems RAS, Moscow, Russia
e-mail: lyubetsk@iitp.ru
**Corresponding author*

Key words: *bacterial type promoters, plastids in Streptophyta, conservation*

Motivation and Aim: We report a study of PEP-promoters of plastome genes in representatives of the green line (Viridiplantae, etc.) and the red line (Rhodophyta, stramenopiles, including Bacillariophyta, Pelagophyceae, Raphidophyceae, etc.). Plastid genes and their promoters are believed to be evolutionarily conserved across large taxonomic lineages, although the authors are unaware of such systematic studies.

Results: The below table contains the species predicted to possess at least one *widely* conserved promoter in the plastome. Predictions are identical for their close relatives with a corresponding orthologous gene. Within flowering plants the promoter sequences are similar and well aligned, therefore we illustrate results on *Arabidopsis thaliana* and *Spinacia oleracea* only. Our analyses suggest that *widely* conserved promoters are absent elsewhere in streptophyte plastomes. We believe that the scarcity of the widely conserved PEP-promoters upstream coding plastome genes may be related to the evolutionary changes of sigma subunit paralogs and phage type RNA polymerases that lead to rapid replacements of the transcription initiation PEP-promoter. Coordinates are relative to the start codon. The "Ex" means the presence of the TG -extension in "-10" box, "Ps" marks a pseudogene and a negative prediction, "=" is a negative prediction when there is the gene.

Species	*psaA*	*psbA*	*psbB*	*psbE*	*rbcL*
Arabidopsis thaliana	Ex -188	Ex -77	-170	-125	-177
Spinacia oleracea	Ex -179	Ex -82	-175	-150	-176
Cycas taitungensis	Ex -156	-60	Ex -170	-141	-156
Cryptomeria japonica	Ex -142	-58	Ex -142	-137	-161
Pinus koraiensis	Ex -158	-52	-193	-148	-136
Pinus thunbergii	Ex -158	-52	-180	-145	-127
Welwitschia mirabilis	Ex -156	Ex -51	-271	-31	-136
Adiantum capillus-veneris	Ex -163	Ex -55	Ex -291	-191	-157
Angiopteris evecta	Ex -152	Ex -69	Ex -181	-142	-148
Psilotum nudum	Ex -147	Ex -53	Ex -178	-127	-140
Huperzia lucidula	Ex -153	Ex -55	Ex -187	-134	-150
Anthoceros formosae	Ex -155	=	=	-143	-160
Aneura mirabilis	Pseudo	Ex -54	Pseudo	Pseudo	-148
Marchantia polymorpha	Ex -149	Ex -53	=	-132	-124
Physcomitrella patens	Ex -161	Ex -53	=	-145	-143
Chara vulgaris	Ex -199	-121	Ex -179	=	-154
Chaetosphaeridium globosum	Ex -154	Ex -57	Ex -161	-119	-102
Staurastrum punctulatum	Ex -235	Ex -59	-190	-154	-219
Zygnema circumcarinatum	Ex -157	Ex -58	-159	-122	-168
Chlorokybus atmophyticus	=	=	-266	=	=
Mesostigma viride	=	Ex -53	-89	=	=
Bigelowiella natans	=	-136	=	=	=

THE FUNCTIONAL ANALYSIS OF POLYMORPHIC VARIANTS OF HUMAN DNA-POLYMERASE IOTA

A.V. Makarova[1*], A.V. Kulbachinskiy[1], Y.I. Pavlov[2]
[1] *Institute of Molecular Genetics, Russian Academy of Science, Moscow 123182, Russia*[1]
[2] *Eppley Institute for Research in Cancer, University of Nebraska Medical Center, Omaha, NE, USA*
e-mail: amakarova-img@yandex.ru
[*] *Corresponding author*

Key words: *DNA polymerase iota, active site, fidelity of DNA synthesis*

Motivation and Aim. DNA polymerase iota (Pol ι) is a member of recently discovered Y-family of DNA polymerases which are specialized in translesion DNA synthesis, participate in DNA repair and somatic hypermutagenesis. Due to special organization of the active site these enzymes are able to bypass various DNA lesions and are characterized by very low fidelity of DNA synthesis (10^{-1} – 10^{-3}) on undamaged templates. Human Pol ι shows the lowest accuracy of synthesis among eukaryotic DNA polymerases and preferentially misincorporates dGTP opposite template T (Tissier, 2000; Zhang, 2000). The mechanisms underlying low fidelity of DNA synthesis of Pol ι are still unknown. To evaluate the role of evolutionary polymorphic and conserved residues of the Pol ι active site in catalysis and fidelity we created Pol ι mutant variants with substitutions of amino acids that, as revealed by recent crystallographic analysis, contact the template DNA, incoming nucleotides or coordinates metal ions: Q59A, L62I, L62A, Y61A, L78G, D126A and E127A.

Methods and Algorithms: Mutant genes encoding for different human Pol ι variants were obtained by site-directed mutagenesis. GST-tagged Pol ι and Pol ι mutant variants were produced in baker yeast *S. cerevisiae* and purified by affinity chromatography. Biochemical properties of wild-type and mutant Pol ι variants were studied by primer extension and kinetic analysis.

Results: The functional analysis of mutant variants of human Pol ι showed that a double change of metal-coordinating amino acids D126A, E127A dramatically decreased enzyme activity and point substitutions Q59A and L62A led to a significant reduction of the Pol ι activity. Pol ιvariants with L62I, L78G and Y61A substitutions did not differ significantly from the wild-type enzyme in the efficiency of DNA synthesis. For all mutant variants of Pol ι, we analyzed fidelity of DNA synthesis and misinsertion specificity on DNA templates of various sequence context.

Conclusion: The results of the work demonstrate an important role for several conserved and polymorphic amino acid residues in the active center of Pol ι in DNA synthesis. Analysis of available three-dimensional structures of Pol ι (Nair, 2004; Kirouac, 2009) allows to propose the functional roles of these amino acids in catalysis and substrate selection.

References:

1. A. Tissier et al. (2000) Pol iota, a remarkably error-prone human DNA polymerase, *Genes Dev*, **14**: 1642-1650.
2. D.T. Nair et al. (2004) Replication by human DNA polymerase-iota occurs by Hoogsteen base-pairing, *Nature*, **430**: 377-380.
3. K.N. Kirouac, H. Ling. (2009) Structural basis of error-prone replication and stalling at a thymine base by human DNA polymerase ι, *EMBO J*, **28**: 1644-1654.
4. Y. Zhang et al. (2000) Preferential incorporation of G opposite template T by the low-fidelity human DNA polymerase iota, *Mol Cell Biol*, **20**: 7099-7108

HIGHLY INACCURATE MODE OF DNA REPLICATION TRIGGERED BY A DOUBLE-STRAND BREAK.

Malkova A[1], Chabes A[2], Deem A[1], Keszthelyi A[2]
[1]*Department of Biology, School of Science, IUPUI, Indianapolis, IN, 46202-5132*
[2]*Department of Medical Biochemistry and Biophysics Umee University, Umee Sweden*

Key words: *double-strand breaks, yeast, recombination, frameshift mutations, dNTP pools*

Motivation and Aim: New DNA must be synthesized both for purposes of genome duplication and DNA repair. While the former is a highly accurate process, short-patch synthesis associated with repair of DNA damage is error-prone. Break-induced replication (BIR) is a unique cellular process that mimics normal DNA replication in terms of its processivity, rate, and capacity to duplicate hundreds of kilobases, but which is initiated at double-strand breaks (DSBs) instead of at replication origins. This study was designed to determine the mutation rate associated with BIR.

Methods and Algorithms: To measure frameshift mutagenesis associated with BIR, we used an experimental system in yeast *Saccharomyces cerevisiae*, wherein a galactose-inducible DSB was initiated at the *MAT***a** locus of the truncated, recipient copy of chromosome III, while the donor copy of chromosome III contained an uncleavable *MAT*α-inc allele and served as the template for DSB repair. Elimination of all but 46 bp of homology on one side of the break on the recipient molecule via replacement by *LEU2* and telomeric sequences results in repair of approximately 80% of DSBs in this wild type strain through BIR. To study mutagenesis during BIR, frameshift reporters were inserted at several positions on the donor chromosome. We used three frameshift reporters: A_4, A_7, and A_{14}, which are all alleles of the *LYS2* gene with an insertion of approximately 60 bp. Insertion of any of the three alleles results in a "+1" shift in the reading frame and a Lys^- phenotype, while a Lys^+ phenotype can be restored by a frameshift mutation that occurs in an approximately 71-bp region of the allele that includes the inserted sequence and restores the reading frame.

Results: Here we demonstrate that the rate of frameshift mutagenesis during BIR is up to 3000-fold higher than during normal replication. Importantly, this high rate of mutagenesis was observed not only close to the DSB where BIR is less stable, but also far away, suggesting BIR is highly mutagenic even after it has become fast and processive. Further, we established that polymerase proofreading and mismatch repair operate during BIR, and that BIR-associated mutagenesis is largely independent of translesion DNA polymerases ζ or η. In addition, dNTP levels were elevated during BIR and this contributed to BIR-related mutagenesis.

Conclusion: We found that BIR is a highly mutagenic pathway of DSB repair, despite participation of highly accurate replication polymerases, and active proofreading and MMR error-correction pathways. We propose that the primary reason for this mutagenesis is polymerase errors, with Pol δ producing many of these errors. Elevated dNTP levels during BIR repair were partially responsible for increased mutagenesis during BIR. Hypermutability during BIR could play an important role in evolution and human diseases, especially during tumorigenesis where cells may be especially prone to BIR repair of broken chromosomes.

REVISITING THE RIBOSOMAL DATABASE PROJECT CLASSIFIER

T.S. Ghosh, H.M. Monzoorul, and S.S. Mande*
Innovation Labs, Tata Consultancy Services Limited, Hyderabad, India
e-mail: sharmila@atc.tcs.com
**Corresponding author*

Key words: *16S rRNA, metagenomics, RDP classifier, taxonomic diversity.*

Motivation and Aim: One of the primary objectives of any metagenomic analysis is to accurately estimate the relative abundance of 'all' taxonomic groups present in a given environment. Given that most metagenomic studies use 16S rRNA profiling for estimating taxonomic diversity, the accuracy/reliability of obtained taxonomic estimates is intricately dependent on the efficiency of available 16S rRNA classification algorithms. It is thus extremely important to validate the performance of 16S classification algorithms across all individual taxonomic groups. Furthermore, the validation should be done using typical read sequences with errors which are obtained using current sequencing technologies. Given the metagenomics context wherein data sets contain sequences originating from entirely new clades, it is also essential to validate the 16S rDNA classification algorithm using a 'leave one clade out' strategy rather than a 'leave one species out' strategy. With this motivation, we have re-evaluated the performance of the RDP classifier [1] using data sets/database scenarios which simulate the metagenomics context.

Methods: Validation was performed using four different sets of test sequences simulating reads with errors obtained from 454 and Sanger sequencing technologies. To simulate a typical metagenomic context, evaluation was done using four different database variants which simulated scenarios wherein query 16S rDNA fragments originated from a new genus or family or order or class respectively.

Results and Conclusion: Our validation results have indicated that the accuracy rate of the RDP classifier is not uniform across various taxonomic groups. For example, while the RDP classifier demonstrates high classification accuracy (>90%) for certain phyla (Proteobactia, Firmicutes etc), the accuracy falls below 60% for certain other phyla (Tenericutes, Bacteroidetes, Cyanobacteria etc). Given that the proportions of certain phyla, like Bacteroidetes have been used for drawing inferences with respect to physiological disorders such as obesity [2], the implication of our findings are expected to be of high significance for the metagenomics community.

References

1. Cole *et al*., (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucl. Acids Res, 37 (Database issue), 141-145. doi:10.1093/nar/gkn879
2. Turnbaugh *et al*., (2006) An obesity associated gut microbiome with increased capacity for energy harvest. Nature, 444(7122):1027–1031

INDUS: AN ALIGNMENT-FREE ALGORITHM FOR RAPIDLY ESTIMATING THE TAXONOMIC DIVERSITY OF METAGENOMIC SAMPLES.

H.M. Monzoorul, N.K. Singh, T.S. Ghosh, C.V.S.K. Reddy, and S.S. Mande*
Innovation Labs, Tata Consultancy Services Limited, Hyderabad, India
e-mail: sharmila@atc.tcs.com
**Corresponding author*

Key words: *Metagenomics, binning, Mixed Memory Markov Models, alignment-free*

Motivation and Aim: Obtaining estimates of taxonomic diversity using binning algorithms is one of the first steps in metagenomic analysis. Earlier studies have indicated that out of the two binning approaches, namely similarity-based and composition-based, similarity-based methods have relatively high binning accuracy and specificity. However, since similarity-based approaches involve alignments of sequences, performance of these algorithms are very slow as compared to composition-based approaches. The objective of the current study was to develop an alignment-free binning approach, that is not only rapid in execution, but also has binning accuracy and specificity comparable to existing similarity-based binning approaches. We have developed a novel composition-based binning algorithm called INDUS which achieves high binning accuracy and specificity by using a unique work-flow that incorporates two layers of composition-based filters.

Methods: For a given query sequence, INDUS first identifies a set of organisms, whose genome fragments have an oligonucleotide composition similar to the composition of the query sequence. Subsequently, the query is scored against precomputed 'Mixed Memory Markov Models' [1] generated for the identified set of organisms. The query sequence is finally assigned to the organism having the highest score. In cases where comparable scores are obtained with models corresponding to multiple organisms, the query is assigned to a taxonomic clade using a 'most common phylogeny' approach.

Results and Conclusion: The INDUS approach has been validated using data sets/database variants that simulate typical metagenomic scenarios. Results indicate that the accuracy and specificity obtained using the INDUS algorithm is comparable to existing alignment-based approaches. In addition, the binning time by the INDUS algorithm is observed to be an order of magnitude lower than existing alignment-based approaches. These observations coupled with the fact that this algorithm can be executed on a simple desktop with modest hardware specifications highlight the immense utility of the INDUS algorithm for metagenomic research groups having limited computational resources.

References

1. Vidyasagar, M. et al. (2008) The 4M (Mixed Memory Markov Model) Algorithm for Finding Genes in Prokaryotic Genomes,. IEEE Trans, (53), 26-37

ProVIDE – PROGRAM FOR VIRAL DIVERSITY ESTIMATION

T.S. Ghosh, H.M. Monzoorul, K. Dinakar, S.S. Mande*
Innovation Labs, Tata Consultancy Services Limited, Hyderabad, INDIA
e-mail ; sharmila@atc.tcs.com
**Corresponding author*

Key words: *viral metagenomics, binning algorithm, alignment parameters, taxonomy*

Motivation and Aim: Given the absence of a universal marker gene (such as 16S rRNA) in the viral kingdom, researchers typically use similarity-based approaches like BLAST (with stringent E-values) for taxonomic classification of viral metagenomic sequences. However, since a majority of sequences in typical metagenomes originate from hitherto unknown viral groups, the use of such stringent thresholds will result in a large fraction of sequences remaining unclassified. Furthermore, using less stringent E-values (observed for BLAST hits with poor alignment quality) will result in a high number of incorrect taxonomic assignments. Our recently published SOrt-ITEMS algorithm [1] provides an approach by which the above issues can be addressed. Based on alignment parameters, an elaborate work-flow is followed by SOrt-ITEMS for assigning reads originating from genomes of hitherto unknown archaeal/bacterial organisms. Thresholds of these alignment parameters were generated by observing the pattern of sequence divergence within and across various taxonomic groups belonging to bacterial and archaeal kingdoms. However, many taxonomic groups within the viral kingdom are characterized by the absence of a typical Linnean-like taxonomic hierarchy (phylum, class, order, family, genus and species). This motivated us to develop ProViDE (Program for Viral Diversity Estimation) - a novel algorithm that uses a customized set of alignment parameter thresholds/ranges, specifically suited for the accurate taxonomic labelling of viral metagenomic sequences. These thresholds take into the account the pattern of sequence divergence and the non-uniform taxonomic hierarchy observed within/across various taxonomic groups of the viral kingdom.

Methods: In order to determine thresholds of alignment parameters customised for viral metagenomic sequences, simulated data sets were generated from diverse viral genomes. BLASTx outputs were obtained by querying these sequences against the nr database. Subsequently, correlations between the quality of alignment (reflected in the values of various alignment parameters) and the extent of taxonomic divergence between the organisms corresponding to the query and the hit sequence were inferred. Based on these inferences, steps were devised (for various query lengths) to identify an appropriate taxonomic level of assignment for a given sequence.

Results, Conclusion and Availability: Performance evaluation with data sets/databases variants simulating typical metagenomic scenarios indicates that ProViDE has significantly high specificity and accuracy. To the best of our knowledge, ProViDE is the first ever similarity-based binning algorithm that provides an accurate and specific taxonomic label to all reads constituting viral metagenomic data sets. Details of the algorithms and the results will be discussed during the conference.

References:

1. Monzoorul HM, Tarini S, Dinakar K, Sharmila SM (2009) SOrt-ITEMS: sequence orthology based approach for improved taxonomic estimation of metagenomic sequences. *Bioinformatics* 25(14):1722–1730 2009

IDENTIFICATION OF CONDITIONAL ENRICHMENT OF MOTIF STRUCTURES IN THE COMBINED PROTEIN AND GENE REGULATORY NETWORKS OF *ESCHERICHIA COLI*

Khushbu Pal, Shubhada R Hegde and Shekhar C Mande*
Centre for DNA Fingerprinting and Diagnostics, Hyderabad, INDIA - 500 001e-mail: shekhar@cdfd.org.in
**Presenting author*

Motivation and Aim: Cellular networks are composed of small structural elements termed motifs. Very often these motifs imply structural and functional characteristics of the underlying network. Of all possible motif patters, a few are selected in biological networks. Since biological interactions vary with varying environmental conditions, this dynamics is expected to be represented by the motif structures as well. To test this hypothesis we studied the conditional enrichment of all possible three and four node motifs in the condition specific networks of *E. coli*.

Methods and Algorithms: The network is constructed by merging both protein functional linkages and transcription regulatory interactions. Considering this as a parent network, the gene expression data was employed to build conditional networks. All possible three-node and four-node motifs were searched in the 466 conditional networks using FANMOD. The motif, if present in significantly more number (p-value < 0.05), is considered enriched.

Results: Of all possible 26 three-node motifs, we find 9 motifs to be significant in all the conditions tested. One of the examples is feed forward loop. However, the motif such as feedback loop was highly under represented. Similarly, out of 328 possible four-node motifs, only 58 were significant and 90 were insignificant in all. The motif structure termed bi-fan is an example for highly represented class. This analysis, we believe, will enable us to understand the differential selection of motif structures and their implications in biological systems.

RAT MODELS OF HUMAN HYPERTENSION

A.L. Markel
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e mail: markel@bionet.nsc.ru

Key words: *arterial hypertension, models, genes*

Motivation and Aim: Numerous experimental and genetic rat models for cardiovascular diseases including arterial hypertension have been developed over some past decades. The all the models of hypertensive disease may be divided into two main groups: pathophysiological and genetic.

The first one includes the following hypertensive state.

(i) Arterial hypertension produced by renal ischemia via renal artery stenosis. This model has first been developed by Harry Goldblatt in the 1930s.

(ii) Hypertension induced by salt-water retention. It may be produced by mineralocorticoids (DOCA) infusions and high salt diet.

(iii) Chronic hypertension may be caused by blockade of nitric oxide (NO) generation. L-NAME (NG-nitro-l-arginine methyl ester) given intraperitoneally or orally may produce this type of hypertension.

(iv) Chronic infusion of low doses of the vasoconstrictor peptide angiotensin II (50–200 ng/kg min) by osmotic minipump is another classical model for hypertension in the rat.

Methods and Algorithms: The group of genetic models of arterial hypertension includes:

(i) Spontaneously hypertensive rats (SHR) were developed by Okamoto K, Aoki K (1963) at Kyoto University School of Medicine, Kyoto, Japan. These rats are the most widely used rat model of hypertensive disease.

(ii) Bianchi G, Ferrari P, Barber BR (1984) from Istituto di Scienze Mediche, Milano published a Chapter in Handbook of Hypertension (de Jong W ed., Elsevier Science, Oxford) on a new genetically hypertensive rat strain. It was demonstrated crucial role of alpha-adducin gene in hypertension development in this model.

(iii) Another hypertensive strain of rats was bred by Vincent M, Dupont J, and Sassard J (1979) in France (UA CNRS 606, Faculty of Pharmacy, Lyon).

(iv) Genetic hypertensive (GH) rat strain was generated by Smirk FH, Hall WH (1958) at the University of Otago.

(v) Two rat strains with the salt-sensitive hypertension were developed by Dahl LK, Heine M, and Tassinari L (1962) at Brookhaven National Laboratory, Upton, New York, and by Ben Ishay D, Saliternik R, and Welner A (1972) at The Hebrew University, Jerusalem.

(vi) Finally, Heller J, Hellerova S, Dobesova Z, Kunes J, and Zicha J (1993) reported on a new model of genetic hypertension, the Prague hypertensive rat (PHR).

Results and Conclusion: Our experimental model of arterial hypertension ISIAH (Inherited stress induced arterial hypertension) rat strain was developed in the Institute of Cytology and Genetics (Novosibirsk, Russia) by Markel AL (1985) and was characterized firstly in the Handbook of Hypertension, Amsterdam, Elsevier, 1999. Now this strain subjected to extensive genetic and molecular study with aim to find genes responsible for stress sensitive hypertension development. Some results of this work will be considered in the presentation.

PIGMENT EPITHELIUM-DERIVED FACTOR (PEDF) AND VASCULAR ENDOTHELIAL GROWTH FACTOR (VEGF) EXPRESSION AND MORPHOLOGICAL CHANGES DURING NORMAL AGING AND DEVELOPMENT OF RETINOPATHY IN WISTAR AND OXYS RAT`S RETINA

A.M. Markovets*, A.A. Zhdankina
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia;
Siberian State Medical University, Tomsk, Russia
e-mail: markovets@bionet.nsc.ru
** Corresponding author*

Key words: *age-related macular degeneration, OXYS rats, molecular mechanism of pathogenesis, vascular endothelial growth factor, pigment epithelium-derived factor.*

Motivation and Aim: OXYS rat is a unique animal model of complex socially significant disease – age-related macular degeneration (AMD), which became the first in developed country as a reason of irreversible blindness. Exact mechanism of AMD pathogenesis remains poorly understood. Recently, we have shown that disease`s development in OXYS rats associated with VEGF gene expression, as it occurs in humans. VEGF and PEDF are produced by retinal pigmented epithelial (RPE) cells and play critical role in vascular homeostasis in retina as antagonist. Their imbalance is a key contributor to development of exudative (dry) form of AMD, but role of these genes in development early stages is unclear. The aim of this work is comparison age- and disease-related changes of rat`s retina and expression of key genes.

Methods and Algorithms: Real-time PCR with specific primers to VEGF and PEDF gene in retina and morphological analysis of retina`s semi thin section had been performed using 20 days, 3, 12, 17, 24 months-aged OXYS and Wistar (control) rats.

Results: The OXYS rats of 20 days did not have signs of retinopathy and we did not find differences in gene expression between OXYS and Wistar rats. Expression of VEGF gene decreased with age in 5-10 times ($p<0,000$), PEDF – in 5 times ($p<0,000$), but in OXYS rats reduction occurred faster than in Wistar. Already at 3 months-aged OXYS rats VEGF expression had decreased almost in 10 times ($p<0,000$) and PEDF gene – in 3 times ($p<0,009$) in comparison with 20 days animals and 3 months-aged Wistar. At the age 24 months expression both of those genes in Wistar rats reached half of expression level at 20 days ($p<0,0003$), while in OXYS rats such effect wasn't observed. Morphological analysis showed deterioration with age state of Wistar`s retina: decrease average square of RPE cells, nuclear layer, open capillary's amount and increase of vessels with thrombosis, amount of pyknotic nucleus in neuron. And in this case deterioration in OXYS rat`s retina state occur faster: parameters of 3 months-aged OXYS correspond to 12-17 months-age Wistar.

Conclusion: Age-related deterioration of retinal status in OXYS rats is critical for disease development and occurs faster than in Wistar. First AMD signs in OXYS rats appeared at 3 months actively against the background of significant reduced expression level of VEGF and PEDF genes associated with decline of RPE cells. Early age-related decline of gene expression in RPE cells results in atrophy of choroid, developing of ischemia with consequence neuroretina deaths. Compensatory reaction on the pathological changes in retina is absent in OXYS rats for no obvious today reasons. And it cause further retinopathy progression. By the nature OXYS rat`s retina changes correspond to the atrophic (dry) form of humans AMD.

Acknowledgements: This work was supported by grants from RFBR 08-04-00722.

CORRELATION BETWEEN NUCLEOSOME FORMATION POTENTIAL OF 5'-UTR AND ELONGATION EFFICIENCY INDEX OF CODING SEQUENCES IN S.CEREVISIAE AND *S. POMBE GENOMES.*

Yu.G. Matushkun*, V.A. Likhoshvai, A.V. Orlenko, V.G. Levitsky
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: mat@bionet.nsc.ru
**Corresponding author*

Key words: *evolution, modeling, elongation efficiency, transcription initiation, correlation.*

Motivation and Aim: Optimization of gene expression requires coordinated optimization of both transcriptional and translational processes. High rates of translation elongation and transcription initiation are necessary for high-expressing genes. As for eukaryotes, the high rate of transcription initiation may require non-condensed DNA packaging in nucleosomes. We have analyzed correlation between high rate of translation elongation and non-condensed DNA packaging in nucleosomes.

Methods and Algorithms: In order to estimate gene expression efficiency as a function of its nucleotide content the Elongation Efficiency Index (EEI) [1] was used. Nucleosome Formation Potential (NFP) values were estimated by Recon method [2].

Results: We have analyzed relations between NFP and EEI in respect to the sample of all yeast ORF (6301 sequences – *S.cerevisiae*, 4681 - *S.pombe,* from GenBank). The profiles of correlation coefficient between NFP of regions upstream translation start and EEI of coding sequences were built. We considered (-600; +600) regions towards translation starts to build these profiles. Reliable negative correlation between NFP in promoter region of a gene and EEI is shown for all genes and the sample of high-expressing genes (10% EEI-highest genes). Similarly, reliable positive correlation between NFP and EEI is shown for low-expressing genes (10% EEI-lowest genes). We consider this correlation to be explained by expression optimization. Transcription initiation should be facilitated for high-expressing genes (high EEI) – to maximize mRNA number (i.e. nucleosome packaging should not be condensed) and vice versa for low-expressing genes. In order to explain found relations between EEI and NFP we have analyzed nucleotide sequences of promoter regions of yeasts. Proximal promoter region (-300;+1) of *S.cerevisiae* was found to be polyA-tract-enriched (tract length ≥ 5 and ≤ 9). There is reliable positive correlation ($p<0.5$) between EEI and presence of polyA-tracts in promoter regions, in particular $r=0.25$ for polytract AAAAA. Similar results are also obtained for *S.pombe.*

Conclusion and Availability: The concordance of rates of transcription initiation and translation elongation is shown in the context of expression optimization.

Acknowledgements: RAS Program 22.8, RFBR 10-04-01310, RAS Program 15.2.

References:

1. N.V.Vladimirov et al. (2007), Correlation of Codon Biases and Potential Secondary Structures with mRNA Translation Efficiency in Unicellular Organisms, Molekulyarnaya Biologiya. 41, No. 5: 926–933 [in English].
2. Levitsky V.G. (2004), RECON: a program for prediction of nucleosome formation potential, Nucl. Acids. Res. 32: 346-349.

STABILITY OF SOLUTIONS OF DELAY DIFFERENTIAL EQUATIONS WITH PERIODIC COEFFICIENTS

G.V. Demidenko[1,2], V.A. Likhoshvai[2,3], I.I. Matveeva[1,2]*
[1]*Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia;*
[2]*Novosibirsk State University, Novosibirsk, Russia;*
[3]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[*]*Corresponding author: e-mail: matveeva@math.nsc.ru*

Key words: *population dynamics, gene networks, blood cell production, mathematical models, delay differential equations, asymptotic stability, attraction domain*

Motivation and Aim: Various biological processes are modeled by delay differential equations. Time delays have been incorporated into biological models to represent resource regeneration times, maturation periods, reaction times, etc. For example, such equations arise when modeling of population dynamics, gene networks, the blood cell production, etc. Therefore development of methods for study of qualitative properties of solutions of such equations has theoretical as well as practical importance.

Methods and Algorithms: In the present paper we use methods developed by the authors when studying stability of solutions of ordinary and delay differential equations with periodic coefficients (for example, see [1, 2]).

Results: We consider the systems of delay differential equations of the form

$$\frac{d}{dt}y(t) = \mu A(t)y(t) + B(t)y(t-\tau) + F(t, y(t), y(t-\tau)), \quad t > \tau > 0. \tag{1}$$

where $A(t)$, $B(t)$ are matrices with periodic entries, $\mu > 0$ is a parameter, $F(t,u,v)$ is a vector-function. We establish conditions under which an equilibrium of (1) is asymptotically stable, obtain estimates characterizing the decay rate of solutions of (1) as $t \to \infty$, and find attraction domains of the equilibrium.

Conclusion: Equations of the form (1) arise in various biological problems (Hutchinson's model, Nicholson's blowflies model, Allee effect models, Mackey-Glass models, etc.) The obtained results give ample opportunities to conduct both theoretical and computational investigations and obtain qualitative characteristics for biological processes.

Acknowledgements: This research was supported by the Russian Foundation for Basic Research (project no. 10-01-00035), by the Federal Target Grant "Scientific and Educational Personnel of Innovation Russia" for 2009-2013 (government contract no. 02.740.11.0429), and by the Siberian Branch of the Russian Academy of Sciences (interdisciplinary project no. 107).

References:

1. G.V. Demidenko, I.I. Matveeva. (2004) On stability of solutions to quasilinear periodic systems of differential equations, *Sib. Math. J.*, **45:** 1041-1052.
2. G.V. Demidenko, I.I. Matveeva. (2007) Stability of solutions to delay differential equations with periodic equations of linear terms, *Sib. Math. J.*, **48:** 824-836.

WHY RecA/RAD51, A KEY PROTEIN OF HOMOLOGOUS RECOMBINATION, HYDROLYZES ATP

A.V. Mazin
Department of Biochemistry and Molecular Biology, Drexel University College of Medicine, Philadelphia, PA 19102-1192, USA
e-mail: amazin@drexelmed.edu

Key words: *Homologous recombination, Rad51, RecA, Branch migration, DNA strand exchange, Holliday junctions, ATP hydrolysis*

Motivation and Aim: The Holliday junction, the crossover point at which the two DNA helices are joined together, is a key intermediate in recombination, DNA repair, and replication. Several proteins are known to bind Holliday junctions and promote their branch migration (BM) by translocating along DNA at the expense of ATP hydrolysis. Surprisingly, the bacterial recombinase RecA and its eukaryotic homologue Rad51 also promote BM of HJs despite the fact that they do not bind Holliday junctions preferentially or translocate along DNA. RecA/Rad51 plays a key role in DNA double-stranded break repair and chromosome segregation during meiosis. RecA/Rad51 binds to ssDNA and forms contiguous filaments that promote the search for homologous DNA sequences and DNA strand exchange. The mechanism of BM promoted by RecA/RAD51 is unknown.

Results and conclusion: Here, we demonstrate that RecA/Rad51 drives the BM of Holliday junctions through a unique mechanism that relies on the RecA/RAD51 polar polymerization on DNA. We show that polar polymerization of RecA and Rad51 proteins requires ATP hydrolysis.

Acknowledgements. This work was supported by the NIH Grant CA100839, and the Leukemia and Lymphoma Society Scholar Award 1054-09

COMPUTER SYSTEM SITEX FOR ANALYZING PROTEIN FUNCTIONAL SITES IN EUKARYOTIC GENE STRUCTURE

I.V. Medvedeva*, P.S. Demenkov, V.A. Ivanisenko
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: brukaro@bionet.nsc.ru
**Corresponding author*

Key words: *protein functional sites, database, eukaryotic gene structure*

Motivation and Aim: Analyzing protein structure projection on exon-intron structure of corresponding gene through years led to several fundamental conclusions about structural and functional organization of the protein [1]. According to these results we decided to map the protein functional sites and analyze the specificity of the coding positions. Although there are some databases (SEDB, ExDom, XdomView) that store the information about protein structure mapping, there is no information about protein functional sites with one-to-one scheme of projection.

Methods and Algorithms: We have integrated the resources from PDB and Ensembl to match protein tertiary structure and corresponding gene as one-to-one projection. Also we added the descriptive information about domains (from Pfam) and protein classification (from SCOP) and map their borders. To evaluate the protein functional site discontinuity we counted two coefficients for each of them: one coefficient measures the discontinuity in protein sequence and the other between exons, encoding the site. So we collected for about 10 000 unique sites from 2 500 unique sequences from 27 organisms (75% are represented by human proteins). Among this we included the BLAST search and 3D similar structure search using PDB3DScan for the polypeptide encoded by one exon, participating in organizing the functional site. *Results:* We applied correspondence analysis (CA) for the codon usage bias in different positions (encoding protein functional sites, donor and acceptor splicing sites) from the SitEx sample [2]. So the results for donor and acceptor splicing sites matched the splicing consensus (A/CAG|G/A) [3]. For the codons that code the functional sites the maximum relative inertia was for CAC, CAT (H), TGC, TGT (C).

Conclusion: We created the database SitEx that keep the information about this mapping and included the BLAST search and 3D similar structure search using PDB3DScan for the polypeptide encoded by one exon, participating in organizing the functional site. This will help: 1) to study the positions of the functional sites in exon structure; 2) to make the complex analysis of the protein function; 3) to exposure the exons that took part in exon shuffling and came from bacterial genomes; 4) to study the peculiarities of coding the polypeptide structures.

Availability: http://samurai.bionet.nsc.ru/~demps/brukaro/

Acknowledgements: This work was supported by State contract FASI №02.514.11.4123, Interdisciplinary integration project SB RAS №111, №94, Program RAS 22. Molecular and cellular biology.

References:

1. H. Kaessmann, S. Zöllner, A. Nekrutenko, W. H. Li. (2002) Signatures of domain shuffling in the human genome. Genome Res. 2002. 12: 1642-1650.
2. Perrière G., Thioulouse J. (2002) Use and misuse of correspondence analysis in codon usage studies. Nucleic Acids Res. 30:4548–4555.
3. Abelson J., Trotta C.R., Li H. (1998) tRNA splicing. J Biol Chem 273*:*12685–12688

PRION-ASSOCIATED PROTEINS IN YEAST: COMPARATIVE ANALYSIS OF YEAST STRAINS, DISTINGUISHED BY THEIR PRION CONTENT.

E.V. Mikhailova*, O.V. Nevzglyadova, A.V. Artyomov, T.R. Soidla
Institute of Cytology RAS, Saint-Petersburg, Russia
e-mail: mikhailovaev@list.ru
**Corresponding author*

Key words: *prion, amyloid, heterokaryon, cytoduction, Saccharomyces cerevisiae*

Motivation and Aim: systematic detection and identification of yeast proteins, that can associate with prions, forming the so-called prion aggregates.

Methods and Algorithms: our approach to the identification of prion-associated proteins includes isolation of the pellet fraction, enriched by prion aggregates of yeast cells crude lysates and comparison of isogenic yeast strains, differing only by their prion composition. These strains were obtained using the cytoduction and various prionotropic treatments. To get [*PSI*$^+$] strains we applied overexpression of the *SUP35* gene located on a plasmid. To obtain [*psi*$^-$] strains we used GuHCl treatment or overexpression of the plasmid-born *HSP104* gene. 2D electrophoresis followed by MALDI analysis of the pellet proteins of [*PSI*$^+$] and *[psi*$^-$] strains permitted identification of the prion-associated proteins.

Results: More than 30 proteins whose presence in the pellet fraction correlate with the change of prion(s) content were identified. Approximately a half of these proteins belong to chaperons and to enzymes of glucose metabolism. Chaperons are known to be involved in prion metabolism and are expected to be present in prion-containing aggregates. Nevertheless, several recent data suggest that the presence of glucose metabolism enzymes is not accidental too (3). We also detected six proteins involved in oxidative stress response, eight – in translation, and several proteins involved in proteolytic degradation.

Conclusion: We would like to conclude that our method seems to reveal the same core prion-associated proteins, as other approaches (1, 2, 3). At the same time our approach is less restrictive and has enabled us to identify some additional proteins – all of them likely to be a respond to signals of protein damage and misfolding. Most of identified proteins seem to be prion-associated, but we can't exclude the possibility that several proteins may propagate as prions.

References:

1. Bagriantsev S.N., Gracheva E.O., Richmond J.E., Liebman S.W. 2008. Variant-specific [*PSI*$^+$] infection is transmitted by Sup35 polymers within [*PSI*$^+$] aggregates with heterogeneous protein composition. Mol. Biol. Cell. 19 : 2433–2443
2. Wang Y., Meriin A.B., Costello C.E., Sherman M.Y. (2007) Characterization of proteins associated with polyglutamine aggregates: A novel approach towards isolation of aggregates from protein conformation disorders. Prion, 1 : 128–135.
3. Wang Y., Meriin A.B., Zaarur N., Romanova N.V., Chernoff Y.O., Costello C.E., Sherman M.Y. (2008) Abnormal proteins can form aggresome in yeast: aggresome-targeting signals and components of the machinery. FASEB J., 23 : 451–463.

IN SILICO ANALYSIS OF AUXIN-REGULATED ROOT APICAL MERISTEM PATTERNING

V.V. Mironova*[,1], N.A. Omelyanchuk[1], E.S. Novoselova[2], V.A. Likhoshvai[1,2]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
e-mail: kviki@bionet.nsc.ru
**Corresponding author*

Background: In plant roots, auxin is critical for patterning and morphogenesis. In root apical meristems (RAM) of different plants the auxin maximum formation was described that functions in maintaining of the stem cell niche in RAM. Auxin concentration maxima in plant tissues are mainly formed due to active auxin transport between cells that is mediated by the PIN family carriers [1].

Aims and Motivation: It has been previously shown that the auxin maximum in RAM forms by the "reverse fountain" mechanism in the cell layout with the pre-assigned positioning and levels of PINs [1,2]. According to this model, the acropetal and basipetal auxin flows arranged in the RAM structure are coordinated to generate and maintain an auxin distribution in the root tip. However, in a number of work it has been shown that the formation of an auxin gradient precedes tissue patterning so the self-organizing mechanism for auxin patterning in RAM have to exist.

Methods: We created a 2D mathematical model of auxin distribution in the root initial taking into account PIN protein dynamics that is regulated by auxin. The cell layout in the model consists only of the two types of layers: protoxylem and epidermis. For the protoxylem layers the following processes were defined (1) auxin flow from the shoot, which is a sole source of auxin; (2) irreversible losses of auxin; (3) auxin diffusion; (4) active auxin transport in the acropetal direction, which is regulated by the PIN1 protein; (5) synthesis and degradation of PIN1 protein within cells, depending on the local auxin concentration. For the epidermal layers only (2) and (3) processes were described.

Results: We demonstrated *is silico* that the auxin-regulated PIN dynamics in RAM accounts for (1) the generation of the auxin distribution pattern in the root tip that can further govern the RAM patterning (2) the changes in the auxin distribution pattern under different growth conditions (increase in auxin flow from the shoot, cut of the root tip, treatment with exogenous auxin or auxin transport inhibitors). In the 2D model the auxin concentration maximum in the root tip arises from the cell layout consisting of the two types of cell, where only the acropetal flow was defined. Comparing to the previous model [2] where five types of cell, which arrange acropetal, basipetal and lateral flows as well as lateral redistribution in the root cap, are necessary for generation of the same pattern.

Conclusion: The model allows to make some predictions about auxin-regulated RAM patterning under different growth conditions and in different plant species.

Acknowledgements: This research was supported in part by the US NSF (PHY05-51164 and EF-0330786), RFBR (№ 08-04-01214-a), Nsh-2447.2008.4, Goscontract PIN RAS №10104-37/P-15/110-327/150409-001, the SB RAS Integration Projects No. 107, 119 and RAS Programms №№ 21(26), 22(8), 23(29).

References:

1. 1. Blilou I. et al. (2005) The PIN auxin efflux facilitator network controls growth and patterning in Arabidopsis roots. Nature, 433: 39-44.
2. 2. Grieneisen VA. et al. (2007) Auxin transport is sufficient to generate a maximum and gradient guiding root growth. Nature, 449: 1008-1013.

PROXIMAL PROMOTERS ARE ENRICHED WITH AUXIN RESPONSIVE ELEMENTS IN EARLY AUXIN INDUCED GENES

V.V. Mironova*, N.A. Omelyanchuk, D.Yu. Oschepkov, V.G. Levitsky
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: kviki@bionet.nsc.ru
**Corresponding author*

Aims and Motivation: Auxin regulates many plant developmental and physiological processes. Auxin response transcription factors (ARFs) mediate auxin-induced gene expression. ARFs specifically bind to TGTCnC-containing auxin response elements (AuxREs) [1]. Since TGTCnC motifs are widely spread throughout the *A. thaliana* genome, a more precise approach for recognition of AuxRE is required.

Methods and Algorithms: A sample (1) of 11 experimentally proven TGTCnC-containing AuxREs from arabidopsis, rice, pea and soybean and a sample (2) with promoters of 53 auxin-regulated genes were created using published data. The optimized PWM method [2] was used for prediction of AuxREs. To verify predicted targets we used the microarray data on gene expression after auxin treatment [3]. We calculated the ratios reflecting the auxin-induced changes in gene expression. A group of genes showing the smallest changes were used as the control group. We compared the enrichment of AuxREs in the upstream regions of *A. thaliana* genes that respected to either repression or activation. χ^2 test was used to verify reliability of observed difference.

Results: First, the set of *A. thaliana* promoters with proved transcription start sites (TSS) was created. For that purpose, we superposed the data on the upstream regions of *A. thaliana* genes [3] to the experimentally proved TSS distribution [4], by this way we obtained 8688 promoters. Analysis of the sample (2) showed, that the density of predicted AuxREs in proximal promoters (-300 to +1 region relative to an annotated TSS) nearly twofold exceeded that for the (-2000; -300) regions. By the optimized PWM method using the sample (1) and with the threshold estimated for the sample (2) we performed a genome-wide prediction of the AuxREs in the sample of proximal promoters for genes with proved TSS, which have a unique reference to the microarray data. The analysis of the presence of putative AuxREs in the proximal promoters of these genes, showed a significant enrichment of potential AuxREs for the genes displaying early auxin-induced repression or activation of expression.

Conclusion: A novel effective approach for AuxREs recognition was created. We detected that proximal promoters of early auxin induced genes are enriched by AuxREs.

Acknowledgements: This research was supported in part by the RFBR (№ 08-04-01214-a), Nsh-2447.2008.4, the SB RAS Integration Project No. 119 and RAS Programs №№ 22(8), 23(29).

Reference:

1. T.J. Guilfoyle and G. Hagen (2007) Auxin response factors, *Curr Opin Plant Biol* 2007, **10:**453–460
2. V.G. Levitsky, et al., (2007) Effective transcription factor binding site prediction using a combination of optimization, a genetic algorithm and discriminant analysis to capture distant interactions. *BMC Bioinformatics,* **8:**481.
3. http://arabidopsis.org/
4. http://ppdb.gene.nagoya-u.ac.jp/

A POSSIBLE MARKER FOR INDETERMINATE PLANT GROWTH HABIT

V.V. Mironova*, N.A. Omelyanchuk, K.V. Gunbin
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: kviki@bionet.nsc.ru
**Corresponding author*

Background: Auxin influences functioning of apical meristems in plant growth. TIR1 is an auxin receptor that acts as a part of the ubiquitin protein ligase SCF^{TIR1} complex [1]. TIR1 mediates rapid degradation of Aux/IAA proteins and consequent changes in expression of auxin-regulated genes.

Aims and Motivation: TIR1/AFB family of F-box proteins are conserved across land plant lineages and fall into four clades [2]. The evolutionary history suggests that the members of each clade may have distinct functions.

Methods: The phylogenetic tree of TIR1/AFB family was reconstructed by PHYML 3.0 (CAT substitution model) using amino-acids sequences of TIR1/AFB proteins from 12 plants species. Reconstruction of the 3D structure of AFB6 proteins was performed by Swiss-model server.

Results: The TIR1/AFB family has a clade named as AFB6 [2]. In the our reconstructed phylogenetic tree of the TIR1/AFB family, proteins from this clade are present in *P. sylvestris, P. trichocarpa, P. dactylifera, V. vinifera, C. papaya, R. communis, S. lycopersicum* and *P. sativa* (AFB6-plus plants). However, these proteins were not found in *A. thaliana, O. sativa, P. patent, S. bicolor*, *Z. mays* and *B. rapa* (AFB6-minus plants). Analysis of AFB6 homologs distribution in plants showed that AFB6-minus and AFB6-plus plant species differ in at least one phenotypic character. All AFB6-minus plants are characterized by determinate growth habit where all AFB6-plus species have a possibility for indeterminate habit. Plants with indeterminate growth maintain an active meristem in the fully developed organ whereas determinate plants do not [3]. This character is an agriculturally important trait: determinate tomatoes varieties as well as indeterminate tobacco and maize varieties are goals of many plant breeding programs. To investigate the possible functions of the AFB6 homologs, we reconstructed the 3D structure of these proteins. We found that AFB6 homologs have the conservative domains for auxin and Aux/IAA binding, whereas the domain for binding to ubiquitin protein ligase SCF complex significantly differs from TIR1. The analysis allows to conclude that the AFB6 homologs may regulate development and maintaining of apical meristems in indeterminate plants by mediating their sensitivity to auxin via previously unknown mechanism.

Conclusion: We found the possible marker for plants with indeterminate growth and propose its function. Further experimental investigation of the AFB6 proteins may contribute to creation of new agricultural plant varieties.

Acknowledgements: This research was supported in part by the RFBR (№ 08-04-01214-a, №09-04-01641-a), Nsh-2447.2008.4.

References:

1. Dharmaziri N. et al. (2005) The F-box protein TIR1 is an auxin receptor. *Nature*. **435**: 441-445.
2. Parry G et al., (2009) Complex regulation of the TIR1/AFB family of auxin receptors. *PNAS*. **106**:22540–22545.
3. Shishkova S. et al. (2008) Determinate Root Growth and Meristem Maintenance in Angiosperms. *Annals of Botany*. **101**: 319–340.

INVESTIGATION OF AUXIN RESPONSE FACTORS (ARFS) GENE FAMILY EVOLUTION

K.V. Gunbin, V.V. Mironova*, M.P. Ponomarenko
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: kviki@bionet.nsc.ru
**Corresponding author*

Background: Auxin is a key regulator of plant development. Auxin response factors (ARFs) are the main transcription factors, which mediate auxin-dependent changes in gene expression. In plants, ARFs are encoded by relatively large gene families comprising 23 members in arabidopsis, 25 in rice and 38 in populous, between them there are activators as well as inhibitors of gene expression [1].

Aims and Motivation: ARF transcription factors regulate a plenty of processes in plant development, however due to the redundancy in their functions, it is very difficult to differentiate them between different ARFs. Investigation of ARFs evolution could clarify the diversification in ARFs functions.

Methods: Analysis of the sequences of ARFs core promoters performed by the original method [2] that allows to estimate an affinity of TBP/DNA binding and predict the presence ($TATA^-$) or absence ($TATA^+$) of TATA-box in a gene core promoter.

We developed a novel method (SAMEM) which implements the permutational test for comparing the molecular evolution model of the ARF protein families with the experimental data. The SAMEM was used for finding rare amino acid substitutions on the ARFs phylogenetic trees branches. We performed the correlation analysis of the evolutionary changes of the 531 amino acid physicochemical properties (PCP) with the protein domain co-occurrence (PDC) (the SUPERFAMILY database 1.73).

Results: First, we analyzed the core promoter regions of ARFs in arabidopsis and rice. We showed that the activator and inhibitor statuses of ARFs transcription factor are correlated with the $TATA^+/TATA^-$ status of a gene. We also showed that the TATA+/TATA- statuses of AtARFs are correlated with the gene expression shown by microarray [3]. Second, the phylogenetic trees were reconstructed for ARFs activators and inhibitors separately on 8 plants species. We cannot found any bursts of rare amino acid substitutions for the tree branches on which of ARFs paralogs diversified. It is consistent with experimental data on ARF paralogs redundancy [1]. Nevertheless, we found several individual rare amino acid substitutions on the inner tree branches on which Rosids species evolved (3 substitutions in activators and 7 substitutions in inhibitors). The correlation analysis between the separated ARF proteins evolution and the PDC allows us to discover the most significant (97.6% PCP, $p<0.05$) correlation for ARF3 and ARF4 genes. Also some correlations were found for the ARF6, 8 paralogs (3.6% PCP, $p<0.05$) and ARF7,19 paralogs (0.9% PCP, $p<0.05$).

Conclusion: The relations between ARFs functions and the primary gene and protein sequences evolution were found.

Acknowledgements: This research was supported in part by the RFBR (№ 08-04-01214-a, №09-04-01641-a), Nsh-2447.2008.4.

References:

1. Paponov IA et al. (2009) The evolution of nuclear auxin signalling. BMC Evolutionary Biology, 9:126
2. Ponomarenko et al. (2008) A step-by-step model of TBP/TATA box binding… DAN, 419:828-832.
3. http://arabidopsis.org/

MODELING OF THE SUPPRESSIVE EFFECT OF HCV NS3 PROTEASE INHIBITOR ON HCV SUBGENOMIC REPLICON REPLICATION IN HUH CELLS

K.D. Bezmaternykh,[1] E.L. Mishchenko,[1*] V.A. Likhoshvai,[1] V.A. Ivanisenko,[1] N.A. Kolchanov[1,2]

[1]Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia; [2]Novosibirsk State University, Novosibirsk, Russia.

e-mail: elmish@bionet.nsc.ru

**Corresponding author*

Key words: *HCV, simulation, mathematical modeling, gene network*

Motivation and Aim: The goal of this study was to develop a mathematical model of suppression of hepatitis C virus (HCV) subgenomic replicon replication in Huh-7 cells by NS3 protease inhibitor. This model would allow assessment of NS3 protease-inhibiting antiviral drugs in an infected cell and search for best treatment procedures.

Methods and Algorithms: The MGSModeller program (Kazantsev et al. 2008) was used for the model development and calculations.

Results: The new model was developed on the base of the earlier model (Mishchenko et al. 2007, Mishchenko et al. 2009) by including the descriptions of some processes: formation and operation of endoplasmic reticulum vesicles, which are compartments of viral RNA replication; suppression of intracellular antiviral defense by the viral NS3 protein; and reactivation of intracellular antiviral defense with the presence of NS3 protease inhibitor. The model correctly reproduces: (1) steady-state concentrations of viral proteins, (+) and (–) RNA strands, the proportions of virus components in membrane vesicles and cytoplasm of an infected cell: (2) the time profile of the achievement of steady-state concentrations of components after replicon transfection; and (3) kinetic and concentration dependences of viral RNA suppression by the inhibitor.

Conclusion: Analysis of up-to-date information allowed reconstruction of our earlier model by including the simulations of the induction of intracellular antiviral defense with the presence of NS3 protease inhibitor and the maturation and operation of vesicles, the compartments of viral RNA replication. The model allows simulation of the course of viral RNA and protein suppression with the presence of NS3-inhibiting candidate drugs in a broad K_i range. It can be also applied to the search for optimal procedures for their use.

References:

1. F.V. Kazantsev et al. (2008). MGSmodeller – a computer system for reconstruction, calculation and analysis mathematical models of molecular genetic system // Proceedings of the sixth international conference on bioinformatics of genome regulation and structure, p.113.
2. E.L. Mishchenko et al. (2007). Mathematical model for suppression of subgenomic hepatitis C virus RNA replication in cell culture. *J. Bioinform. Comput. Biol*. **5:** 593-609.
3. E.L. Mishchenko et al. (2009). Mathematical modeling of effects of potential antivirals on hepatitis C virus replicon replication in a cell. *VOGIS Herald*, 2009, **1**(13): 208-217.

INVESTIGATION OF *OPISTHORCHIS FELINEUS TRANSCRIPTION PROFILE BY DIRECT SEQUENCING CDNA LIBRARY'S CLONES*

S.I. Tatkov[1], I.I. Brusentsov[1], M.Yu. Pomaznoy[1], O.V. Nosareva[1], A.M. Nayakshin[2], S.V. Guselnikov[2], E.V. Brenner[2], G.V. Vasiliev[1], A.V. Katokhin[1], V.A. Mordvinov[1*]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*
e-mail: mordvin@bionet.nsc.ru
**Corresponding author*

Key words: *Opisthorchis felineus*, cDNA library, transcriptome

Motivation and Aim: An Investigation of transcription patterns of different organisms or different tissues in them can lead to discovering of basic laws in according with that biological systems are organized and functioned. We have selected as an object for study *Opisthorchis felineus*, the trematodes belonging to the family Opisthorchiidae. This parasite, being liver flukes, is widely distributed in Siberia regions and the huge part of regional population is infected by it. Since any vaccines for prevention of parasitoses are yet unavailable, chemotherapy plays the main role in treatment of invasions. The detailed transcription patterns study of this liver fluke can reveal new drug targets.

Methods: from total polyA mRNAs of *Opisthorchis felineus* cDNAs copies were obtained and cloned into phagemid pBluescript II(+) (Agilent Technologies, Inc., USA). Ligated DNAs were transformed into *E.coli* XL1 Blue MRF' by electroporating. There were obtained 53,000 colonies on LB agar (50 mkg/ml ampicillin). From those 11,000 clones were isolated, cultivated in 96-deep well plates, filled by LB with ampicillin, and then to each well glycerol was added up to final 15%, and plates were frozen at -70^{0}C and kept at this temperature for further investigations. For preliminary assessment the obtained cDNA-library was analyzed by 1,5% agarose electrophoresis, and size of cDNA-inserts were estimated after PCR-amplification with forward and backward primers (T3 and T7 respectively). To obtain biomass for purifying plasmids, the clones were picked up into 96-deep well plates and cultivated under intensive aeration in thermostats GFL 3033 (GFL, Germany). The plasmid DNAs for sequencing were purified by routine protocols on QiaCube (QIAGEN, Germany) and "Tecan Freedom Evo" (Tecan, Switzerland).

Results: From 11,000 clones selected randomly 2,500 ones were amplified by PCR with T3 and T7 primers. The size of cDNA-inserts varied from 400 till 1500 bp, portion of clones without inserts was about 5%. The sequencing of the first 400 clones of cDNA-library of O. felineus has revealed the dominate role of housekeeping proteins (ribosomal proteins S3A, S18, actin, different proteases, aldolase, succinate dehydrogenase, different glutation-S-transferases, myoglobin, subunits of ATP-sintase, that are encoded in chromosomes). Some of the sequenced mRNA has no significant homologs in the existing database of NCBI.

MODELING CYP51 INTERACTIONS WITH SUBSTRATES AND DERIVATIVES

D.V. Mukha*, S.A. Usanov
Institute of Bioorganic Chemistry of NAS, Minsk, Belarus
e-mail: d.v.mukha@gmail.com
**Corresponding author*

Key words: *cytochrome P450, CYP51, 14α-demethylase, lanosterol, docking, molecular dynamics, MMPB-SA, binding energy*

Motivation and Aim: CYP51A1 (sterol 14α-demethylase) is an enzyme which catalyzes an essential reaction in the cholesterol biosynthesis path. It is shown that the reaction product of lanosterol 14α-demethylation is involved in the regulation of germ cells maturation, affecting the reproductive system functioning [1]. Inhibition of CYP51 is widely used to treat fungal infections. The aim of our work was to explain the mechanism of substrate specificity of the human enzyme, which is important for the development of new harmless drugs.

Methods and Algorithms: Modeling of cytochrome P450 51A1 complexes with substrates and their derivatives was carried out in a software package Amber 10 [2]. UCSF Dock v6.3 was used for the docking of substrates in the active site of CYP51. The molecular complexes were equilibrated by the 1 ns long molecular dynamics. Equilibrated systems were used to calculate the enthalpy of binding by MMPB(GB)SA algorithm [3]. Modeling of the interaction of steroids with β-cyclodextrin was carried out using the GAFF and GLYCAM_06 force fields. Contribution of residues in the free energy change upon binding was determined by using alanine scanning. The calculations necessary for docking and molecular dynamics were carried out on cluster supercomputer SKIF in United Institute of Informatics Problems of NAS of Belarus.

Results: The GB algorithm gave much higher enthalpy magnitudes compared with the PB. The enthalpy of complex formation with hydrophobic substrates was overestimated by both algorithms, since water was used as a solvent. When taking into account in calculations an enthalpy of substrate binding to β-cyclodextrin, used as a model of hydrophobic environment, as well as changes in entropy, values of free energy yielded by PB algorithm become comparable to the experimental (± 1 kcal/mol). Intermediate aldehyde-derivatives of lanosterol show higher affinity to the active center than the original substrates, and the latter, in turn, bind stronger than the final products of the reaction. This may indicate on the mechanism presuming three successive hydroxylation reactions, in which the intermediate products do not leave the active site of the enzyme until the end of the reaction. Removal of the methyl group at the 14th carbon atom and the formation of the double bond greatly reduces the affinity of the enzyme to the ligand, which leads to the release of the product. Essential for substrate binding amino acids were identified.

Conclusion: The created model for interaction of sterol 14α-demethylase with ligands proved to be useful for assessing the affinity of inhibitors to the enzyme. It will facilitate to determine the side effects of drugs aimed at suppressing the activity of host cytochrome P450s. On the other hand, the development of highly specific inhibitors of CYP51A1 leads to the creation of new drugs against hypercholesterolemia.

References:

1. Stromstedt M. et al. (1998) Elevated expression of lanosterol 14alpha-demethylase (CYP51) and the synthesis of oocyte meiosis-activating sterols in postmeiotic germ cells of male rats, Endocrinology, 139 (5): 2314-21.
2. Case D. A. et al. (2008) AMBER 10, 304 pp. (University of California, San Francisco)
3. Kollman P. A. et al. (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models, Acc Chem Res, 33 (12): 889-97.

REGRESSION SYSTEM FOR PREDICTION OF ERRORS IN THE DATA ON GENE EXPRESSION IN SITU OBTAINED FROM CONFOCAL IMAGES

E.M. Myasnikova*, S.Yu. Surkova and M.G. Samsonova
St.Petersburg State Polytechnical University, St.Petersburg, Russia
e-mail: myasnikova@spbcas.ru
**Corresponding author*

Key words: *confocal microscopy, image clipping, linear regression prediction.*

Motivation and Aim: In our previous work [1] we have presented an algorithm for estimation of errors in the data on gene expression in situ extracted from confocal images. The application of the method is limited by the requirement of additional information that is not available from the standard procedure of data acquisition, and the special design of experiments is needed. The aim of the work is to create a learning system for the prediction of error size in the data obtained from a confocal image based on the information about the parameters of the microscope.

Methods and Algorithms: The common way to reduce the noise in confocal images is averaging of multiple frames that however leads to the biased data in case of clipped single frames. The size of data error caused by clipping depends in the first place on the values of microscope parameters that are used to set brightness and contrast of an image. In [1] we published a method based on censoring technique for the estimation and correction of this kind of errors, but the method implementation required the availability of all the confocal scans along with the averaged image.

To predict error size in the data extracted from the averaged image we developed a regression system. The linear regression function is constructed from the learning sample composed of images obtained at different combinations of microscope parameters, and for each member of the sample all the scans are saved as separate images. The values of microscope parameters are used as independent variables. Dependent variable is the estimated error value for the given mean intensity level.

Results and Conclusions: The regression system was applied to the data on segmentation gene expression in Drosophila stored in the FlyEx database (http://urchin.spbcas.ru/flyex/). The predicted errors proved to be of small size not exceeding 5-7% of the mean intensity level in the embryo nucleus

High values of data errors are usually caused by improper choice of the microscope settings used in order to increase the image brightness and contrast. An important application of the current work is the possibility to accurately correct this kind of errors thereby allowing to obtain images of the higher dynamical range and thus to extract more detailed quantitative information from them.

Acknowledgements This work is supported by NIH grant RR07801, GAP award RUB1-1578 and RFBR grants 08-04-00712-a, 08-01-00315-a and 09-04-01590-a.

References:

1. E. Myasnikova, S. Surkova, L. Panok, M. Samsonova, and J. Reinitz. (2009) Estimation of errors introduced by confocal imaging into the data on segmentation gene expression in *Drosophila*, *Bioinformatics*, **25**:346–352.

VETERINARY WELL-BEING OF LABORATORY ANIMALS

V.A. Naprimerov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
E-mail: nva61@bionet.nsc.ru

Key words: *Laboratory animals, SPF status, GLP standard.*

Motivation and Aim: to inform the specialists related to laboratory animals that to maintain specific pathogen free (SPF) status in the newly constructed facility of the Animal Center of the Institute of Cytology and Genetics SB RAS is rather uneasy, but the extremely necessary task.

Methods and Algorithms: The rules of the keeping and breeding of laboratory animals are defined by FELASA and GLP standards. Following these standards will allow the SPF-Animal Center in Novosibirsk to be integrated into the Federation of International Mouse Resources (FIMRe).

Results: The list of pathogens to be excluded from the SPF animals (mice and rats) consists of 40 specified pathogens. The main vectors of transmitting infection are wild mice and rats, people and administrated substances.

The conclusion: Only joint efforts of all interested persons will provide the health and well-being of animals in modern SPF facility.

References:

1. Good Laboratory Practice Standards (US Environmental Protection Agency, 40 CFR, Part 792, 1989)
2. GOST ISO 14644-1-2002 Pure premises and the controllable environments connected with them.
3. Sidorchuk A. A (2009) Infectious diseases of laboratory animals. A.A.Sidorchuk, A.A.Glushkov. Publishing house: the Fallow deer.

SEQUENCE ANALYSIS OF COG3868 AND COG2342 FAMILIES

O.O. Stepuschenko[1], D.G. Naumoff*[2]
[1] *Kazan State University, Department of Genetics, Kazan, Russia*
[2] *Institute for Genetics and Selection of Industrial Microorganisms, Moscow, Russia*
e-mail: daniil_naumoff@yahoo.com
[*] *Corresponding author*

Motivation and Aim: The endo-α-1,4-polygalactosaminidase (EC 3.2.1.109) is a very rare enzyme, which has been found only in two strains of bacteria (*Streptomyces griseus* C-10 and *Pseudomonas* sp. 881). It belongs to the GH114 family of glycoside hydrolases (or COG3868). According to the CAZy database (http://www.cazy.org/), this family includes 56 proteins. COG3868 is closely related to functionally uncharacterized COG2342. The latter includes proteins with TIM-barrel type of the three-dimensional structure (PDB, 2AAM). This type of 3D structure has been shown for the catalytic domains of four clans (a hierarchical group higher than family) of glycoside hydrolases. Relationships within the GH114 family are still unclear and became the purpose of the work, as well as its evolutionary connections with other protein families.

Methods and Algorithms: Protein sequences were retrieved from the NCBI database. Multiple sequence alignment of 117 GH114 domains was made in BioEdit program (very similar and partial sequences were omitted). The phylogenetic trees were built using programs of PHYLIP package. 34 members of COG2342 family were used as outgroup. Interfamily relationships were established using PSI Protein Classifier. This program analyzes results of PSI-BLAST searches. Several most divergent representatives from the GH114 and COG2342 families were used as a query.

Results and Conclusion: We have revealed 128 non-identical protein sequences of GH114 domains using the blast algorithm. They include representatives of several phyla of Bacteria (Actinobacteria, Aquificae, Chloroflexi, Deferribacteres, Deinococcus, and Proteobacteria), as well as some Eukaryota (Alveolata, Ascomycota, Basidiomycota, Chlorophyta, and Oomycetes). The majority of the proteins have similar length (251–375 amino acid residues) and contain only one domain (GH114). Protein from *Hahella chejuensis* (GenPept, ABC28688.1) consists of two GH114-domains, protein from *Endoriftia persephone* (ZP_ 02533448.1), in addition to GH114, has a GH9-domain.

Three main clusters can be recognized on the GH114 phylogenetic trees (both Neighbor-Joining and Maximum Parsimony). Two of them have high bootstrap support and include mainly representatives of Actinobacteria. The third cluster is formed by proteins from Ascomycota, but they are not very well separated from some bacterial proteins. The tree topology supports the important role of horizontal transfer in the evolution of GH114 proteins. Two very conserved residues (Asp and Glu) most probably are the nucleophile and proton donor in the active center, respectively.

Iterative screening of the protein database allowed us to reveal relationship of GH114 and COG2342 with GH5 (clan GH-A), GH13 (clan GH-H), GH18 (clan GH-K), GH20 (clan GH-K), GH27 (clan GH-D), GH29, GH31 (clan GH-D), GH35 (clan GH-A), GH36 (clan GH-D), GH42 (clan GH-A), GH66, GH97, GH101, COG1306, COG1649, GHL3, and GHL4 families. These data support the common evolution origin of all TIM-barrel type glycoside hydrolase catalytic domains, as we suggested earlier (D.G. Naumoff, BGRS'2006).

SEQUENCE ANALYSIS OF YEAST GLYCOSIDE HYDROLASES

D.G. Naumoff
State Institute for Genetics and Selection of Industrial Microorganisms, Moscow, Russia
e-mail: daniil_naumoff@yahoo.com

Motivation and Aim: On the basis of sequence similarity, catalytic domains of glycoside hydrolases [EC 3.2.1.-] have been grouped into more than 100 families (GH1–GH115, except GH21, GH40, GH41, GH60, GH69, and GH91) in the CAZy classification (http://www.cazy.org/). Various ascomycetous yeasts have about 40–50 glycosidase genes, in total representing only 32 CAZy families. The Gĭnolevures database (http://www.genolevures.org/) provides classification of all proteins encoded in 9 species of hemiascomycete yeasts into about 8000 families. The purpose of the work is to clarify the correspondence between CAZy and Gĭnolevures families and to analyze phylogeny of yeast glycoside hydrolases.

Methods and Algorithms: Multiple sequence alignments of glycoside hydrolases from 9 hemiascomycete yeast species were retrieved from the Gĭnolevures site. The alignments were manually edited in BioEdit and representatives (if any) from *Schizosaccharomyces pombe* were added as outgroup. The phylogenetic trees were built using programs of PHYLIP package.

Results and Conclusion: Among the 32 CAZy families analyzed, eight are not presented in Gĭnolevures genomes (GH10, GH45, GH71, and GH115) or represented by a single protein, all in *Debaryomyces hansenii* (GH20, GH51, GH105, and GH114). Among the remaining 24 families, seven do not include *S. pombe* members. For 18 families we found one-to-one family correspondence. CAZy families GH15, GH17, GH27, GH28, GH32, GH37, GH38, GH47, GH63, GH65, GH72, GH78, GH81, GH85, and GH92 correspond to GL3R1486, GL3C0057, GL3R4232, GL3R3998, GL3R0923, GL3R0613, GL3R2870, GL3C0154, GL3R1048, GL3R2013, GL3R0042, GL3R4175, GL3C0407, GL3R4002, and GL3R3768, respectively, in the Gĭnolevures database. Families GH2, GH31, and GH88 correspond to GL3R4262, GL3C0670, and GL3R3758, but each of the former three also includes a singleton (unclassified) protein from *D. hansenii*. GH5 includes members of GL3C0106 and GL3R0891 families, as well as a singleton from *Lachancea* (*Saccharomyces*) *kluyveri*. GH3, GH16, GH18, and GH76 correspond to pairs of Gĭnolevures families: GL3R0575 & GL3R4191, GL3C0130 & GL3R0461, GL3C0583 & GL3C0738, and GL3R0174 & GL3R3972, respectively. Families GL3C0220, GL3R2291, and GL3R2899 compose GH13.

Phylogenetic analysis allowed us to distinguish five subclusters of hemiascomycete proteins in GH72 (GL3R0042) family. Two *S. pombe* proteins compose a sister group for one of the clusters, but the other two *S. pombe* proteins lie clearly out of the clusters. The data obtained suggest that the common ancestor of all hemiascomycete yeasts had five GH72 proteins and the corresponding duplications had happen at different stages of the evolution. We found several clusters in seven other families: GL3C0057 (at least 4 clusters), GL3C0106 (at least 3 clusters), GL3C0130 (3 clusters), GL3C0154 (3 clusters), GL3R0174 (at least 2 clusters), GL3C0407 (2 clusters), and GL3R0461 (2 clusters). Our results demonstrate a high level of evolutionary plasticity of the carbohydrate utilization system in yeasts. Gene loss as well as multiple gene duplications have happen during the evolution of several yeast glycoside hydrolase families. These events can be only partly explained by total genome duplications.

Acknowledgements: This work has been supported by Russian Federal Agency for Science and Innovation (contract 02.740.11.5008).

INVESTIGATION OF B-TYPE PHOSPHOGLYCERATE MUTASE AND NEURON-SPECIFIC ENOLASE INTERACTIONS USING MOLECULAR DYNAMICS SIMULATIONS

D. Hakobyan[1], K. Nazaryan[1,2*]
[1]*Laboratory of Computer Modeling, Institute of Molecular Biology, Yerevan, Armenia*
[2]*Department of Bioinformatics, Biomedical Faculty of Russian-Armenian (Slavonic) University, Yerevan, Armenia*
e-mail: karen.nazaryan@gmail.com
**Corresponding author*

Key words: *NSE, enolase, mutase, 2PG, functional interaction, channeling, EEF1*

Motivation and Aim: Interaction specificities and possibility of substrate channelling phenomenon between most of glycolytic enzymes remain poorly investigated due to weak forces and short interaction periods between them. The interaction as well as substrate channeling phenomenon between human neuron-specific enolase and B-type phosphoglycerate mutase (dPGM-B) have been investigated in details with CHARMM [1] – Molecular Dynamics (MD) simulation program.

Methods and Algorithms: Interactions of 31 orientations of these enzymes have been investigated with effective energy function (EEF1) implicit solvation method [2]. The initial 3D structures of dimeric NSE (1TE6) and dPGM-B (1YFK) were obtained from Protein Data Bank (PDB). The model of 2-phosphoglycerate product was added to each of the two active pockets of dPGM-B subunits to investigate the cleavage site and possibility of channeling. Procedure of periodically increased temperature dynamics was used to obtain close conformation of dPGM-B as well as to study the details of 2PG cleavage from the enzyme.

Results: Interactions of 31 orientations of these enzymes were investigated and grouped into 5 distinctive interactions. Interactions between active regions of the enzymes occurred preferentially as in three of the five groups the enzymes interacted with their active regions. Cleavage of 2PG from dPGM-B through the residue loop Trp^{16}–Gly^{24} was observed for a separate subunit of dPGM-B. Channeling of 2PG between the enzymes was observed for one of the 31 orientations studied, when NSE with the residues of its active site covered the Trp^{16}–Gly^{24} region of dPGM-B [3].

Conclusion: Interaction between the active regions of the enzymes seems to be favorable. The dPGM-B Trp^{16}–Gly^{24} regions may appear not only the cleavage site of 2PG but also another entrance site for a substrate. The details of channeling suggest that substrate transfer between enolase and phosphoglycerate mutase may favor from weak interaction of the enzymes.

References:

1. B. R. Brooks et al. (1983) CHARMM: A program for macromolecular energy minimization and dynamics calculations, J. Comp. Chem. 4: 187-217.
2. T. Lazaridis, M. Karplus. (1999) Effective energy function for proteins in solution, Proteins 35: 133-152.
3. D. Hakobyan, K. Nazaryan. (2010) Molecular dynamics study of interaction and substrate channeling between neuron-specific enolase and B-type phosphoglycerate mutase, Protein: Structure, Function, and Bioinformatics *in press*.

DIVERSITY OF SPECIES AVAILABLE FOR BIOMEDICAL RESEARCH

Timo Nevalainen
Laboratory Animal Centre, University of Eastern Finland, Kuopio, Finland
e-mail: Timo.Nevalainen@uef.fi

Motivation and Aim: Within the species used in biomedical studies mice are overly dominating. Within mammals out of every three animals used, two are mice. When coldblooded animals and birds are added to the statistics, more than half of 12 million vertebrate animals used annually in the European Union (EU) are mice. In countries where pharmaceutical industry is missing and fundamental research is the main user of laboratory animals, mice are even more dominating.

Methods and Algorithms: During the 80's numbers of laboratory animals used were going down, but since the creation of transgenic animals, the trend has been upwards. This is tightly associated with mice as the target species; in fact creation GM-strains in other species is not currently that common. Moreover, mice are economical to breed, house and use.

Results and Conclusion: When going through the species variety at hand, it is clear that there are not many species in the same weight category with humans. Size is a decisive factor when designing and testing artificial organs and materials. Large dogs may qualify for the purpose, but because of social pressure their numbers are going down, and indeed the most common breed, beagle, is a quite small dog. Pigs do resemble humans, but size of farm pigs is an obstacle to their use. Miniature swine is an animal with reasonable size when sexually mature. Swine is a good example of spontaneous model for coronary circulation and stenosis displaying abundant collaterals and associated high frequency of arrhythmias dues to myocardial ischemia. Over the last years, numbers of pigs and sheep have increased in EU at the expense of dogs.

In Europe there is a strong trend to implement the Three R strategies, *i.e.* the alternatives to the use of animals. In short, these are replacing sentient animals with systems incapable of feeling pain, refining harmful procedures and reducing the numbers of animals used to absolute minimum. As to species, this implies moving away from mammals, and indeed we see increase in numbers of birds and fish to be used in research. Production of polyclonal antibodies serves as an example. There is a trend to move away from the use of rabbits associated with aggressive adjuvant and bleeding to chicken, and to look for possibilities of oral immunisation.

The new EU Directive will introduce stringent constraints on the use of nonhuman primates (NHP). The use of great apes will in essence be banned, which is not a difficult decision because they are not used in Europe at all. For other NHP species, the directive brings requirement for F2 animals, but there will be a transition period of several years to allow sufficient breeding capacity to be established. The acceptable aims for use of NHP are studies on devastating or debilitating diseases or basic research in general.

LABORATORY ANIMAL SCIENCE – FROM PHILOSOPHY TO EXPERIMENTAL DESIGN

Timo Nevalainen
Laboratory Animal Centre, University of Eastern Finland, Kuopio, Finland
e-mail: Timo.Nevalainen@uef.fi

Motivation and Aim: Laboratory animal science (LAS) is a small, but unique discipline with more diverse stakeholders than most traditional ones. As such it can be seen as a well demarcated entity cutting across many basic biological sciences. Experimental design is one of key topics taught in competence courses for scientists and LAS specialists. There is ample evidence that we are encountering problems in design features, which clearly hamper creditability and repeatability of animal studies.

Methods and Algorithms: On top of the statistical tests themselves, a variety of non-statistical issues of concern can be identified. The most common ones appear to be lack of randomisation, bias and incorrect choice of experimental unit. The most recent discussion has been raised on whether purposely inbuilt increased heterogeneity of *e.g.* environments would lead to less variation. A closer look at those articles shows that there can be important treatment and/or strain x environment (*e.g.* laboratory) interactions, such that some treatment effects are only statistically significant under some environmental conditions. Apart from the new terminology used, there is hardly nothing new; this has been shown by statisticians for many decades ago and relates to applicability of the results.

Results: It is one thing to do something in the experiment, and other what will be written into the article. Unfortunately we seem to be in deep trouble here as well. The guidelines on study description published over 30 years ago, has not led to the desired outcome. Crucial elements of animal, housing and procedure descriptions are left out from articles making it impossible for others to repeat the study; this must have resulted in unnecessary use of animals. Apparently, the editors of scientific journals aim at saving printing costs, and while a few journals practice having supplemental information *e.g.* on internet, this has not become common practise. Supplemental information would also allow presenting all numerical data, which is often impossible to extract from illustrations thus rendering subsequent meta-analysis impossible.

Conclusion: It is clear that improvements in these non-statistical items and study description are long overdue. The key persons in this process are scientific journal editors and their referees. It is clear that many of them know the situation, but lack the incentive to do so. However, the obvious scientific and animal welfare motives should be compelling enough to make the change.

THE CELL GROWTH AND DIVISION CAN DESTROY STEM CELL NICHE IN A REACTION-DIFFUSION MODEL

Sergey Nikolaev[1]*, Ulyana Zubairova[1], Stanislav Fadeev[2], Eric Mjolsness[3], Nikolay Kolchanov[1]

[1.] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[2.] *Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*

[3.] *Computer Science Department, Center for Computational Morphodynamics, University of California, Irvine, USA*

e-mail: nikolaev@bionet.nsc.ru

**Corresponding author*

Key words: *shoot apical meristem, stem cell niche, mathematical model, DL-system formalism.*

Motivation and Aim: A minimal 1D-model of stem cell niche structure regulation along vertical axis of the SAM was developed on the basis of a qualitative hypothesis of interplay between the CLV and WUS genes. Previously it was shown that there is a set of parameters supplying a stationary solution in qualitative correspondence with experimental observations [1]. But the question arises what will be the model dynamics under cell growth and division.

Methods and Algorithms: Using DL-system formalism [2] we developed a mathematical model of stem cell niche structure regulation on 1D-array of growing and dividing cells. A number of computer simulations were performed to study the model dynamics.

Results: In the issue the dependence of probability of the stem cell niche destruction on cell cycle duration relative to diffusion time scale was obtained. Increase of the specific cell growth rate results in monotonic increase of system destruction probability and in decrease of its mean lifetime.

Conclusion: Cell divisions account for relevant perturbation in the SAM structure and may result in destruction of it. The stem cell niche survivability depends on relations between model parameters.

Acknowledgements: This work was supported by the Russian Foundation for Basic Research (08-04-01214-a), SB RAS Integration project 119, RAS program №22 (project 8).

References:

1. Nikolaev S.V., Kolchanov N.A., Fadeev S.I., Kogai V.V., Mjolsness E. (2006) A study of a one-dimensional model for control of renewing zone size in biological tissue. *Computational technologies*, **11**. № 2: 67-81.
2. P. Prusinkiewicz, A. Lindenmayer (1990) The algorithmic beauty of plants. *N.Y.:Springer-Verlag*.

A MODEL OF SHOOT APICAL MERISTEM COMPARTMENTALIZATION BASED ON CLV/WUS INTERPLAY

Sergey Nikolaev[1]*, Ulyana Zubairova[1], Eric Mjolsness[2], Pavel Smal[1], Bruce Shapiro[3], Nikolay Kolchanov[1]

[1.] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[2.] Computer Science Department, Center for Computational Morphodynamics, University of California, Irvine, USA

[3.] Biological Network Modeling Center, Division of Biology, California Institute of Technology, Pasadena, USA

e-mail: nikolaev@bionet.nsc.ru

[*]*Corresponding author*

Key words: *shoot apical meristem, stem cell niche, mathematical model.*

Motivation and Aim: There is a qualitative hypothesis of the interplay between CLV and WUS genes as a mechanism for the SAM spatial differentiation into central zone (CZ – stem cells), organizing center (OC), and peripheral zone (PZ). The following is an important summary of the hypothesis: CLV3 expression occurs in the central cells of 3 upper layers (CZ), while WUS expression occurs in the cells of OC, just below CZ; and CLV3 by means of binding with putative receptor CLV1/CLV2 inhibits WUS expression, while WUS activates CLV3 expression. This interplay is believed to be able to regulate stem cell niche structure in the SAM.

Methods and Algorithms: Previously we developed a mathematical model of a spatially distributed molecular-genetic mechanism for such a spatial differentiation of the SAM to test the above hypothesis [1]. We added a hypothetical gene expressing in the uppermost cells. And we supposed regulatory molecules propagate across the SAM by diffusion. Now the model has been realized in the Cellzilla package [2], and numerically solved to obtain a stationary solution on a 2D domain representing vertical cut of the SAM.

Results: We obtained model parameters that provide a stationary solution for the spatial gene expression domains of the modeled genes in qualitative accordance with experimentally observed data on vertical cut of the SAM.

Conclusion: The hypothesized mechanism for stem cell niche structure control in the SAM captures central features of the experimentally observed interactions between compartments.

Availability: http://www.computableplant.org –> models.

References:

1. S.V. Nikolaev, A.V. Penenko, V.V. Lavreha, E.D. Mjolsness, N.A. Kolchanov. (2007) A Model Study of the Role of Proteins CLV1, CLV2, CLV3, and WUS in Regulation of the Structure of the Shoot Apical Meristem, *Russian Journal of Developmental Biology*, **38:** 383–388.
2. B.E. Shapiro, J. Lu, M. Hucka, E.D. Mjolsness (2007) Platforms for Modeling in Systems Biology: Recent Developments in MathSBML and Cellerator, Poster and Proceedings extended abstract, ICSB 2007, Long Beach California, Oct 2-4 2007. URL: http://www.icsb-2007.org/proceedings/abstracts/G24.pdf

CALCULATION OF THE PROPERTIES OF STRUCTURE OF FORMYL- METHIONINE -tRNA BY QUANTUM MECHANIC

M. Noei [a]*, F. Mollaamin[b] , A. Nouri[c] , M. Monajjemi[d]
[a] *Department of chemistry , Mahshahr Branch , Islamic Azad university , Mahshahr , Iran*
[b] *Department of chemistry , Qom Braneh , Islamic Azad university , Qom , Iran .*
[c] *Science and Research Branch , Islamic Azad university , Tehran , Iran .*
[d] *Science and Research Branch , Islamic Azad university , Tehran , Iran .*
E-mail: HUmaziar.noie@yahoo.comU

Key words: *formyl – methionine (fMet); Hartree – Fock (HF) ; density functional theory (DFT)*

Motivation and Aim: In all these studies it has been assumed that binding of fMet-tRNA to ribosomes is the polynucleotide. in bacteria the start codon AUG is recognized by fMet-tRNA. Understanding properties of structure of fMet-tRNA is very important because this structure is start of the proteins biosynthesis.

Methods and Algorithms: In our current study, extensive quantum mechanical calculations of structure of fMet- tRNA and solvent effects on structure of fMet-tRNA and calculations of NMR parameters and IR spectra have been Performed on a Pentium-4 based system using GAUSSIAN 98 program .

Results: The HF and DFT energies are of particular interest because they provide results for interactions appearing in solvent medium considered in this letter, which are in accord with biological behavior of Adenine + fMet of fMet-tRNA. Furthermore, recent papers often tend to ask about the role of water solvent effect on the stability of Adenine + fMet of fMet-tRNA structure.

Conclusion: The geometries of Adenine + fMet of fMet-tRNA structure at a constant pressure of 1 atmosphere and temperature of 310 K were optimized using the 3-21G basis set, and vibrational frequencies were obtained at the same level. Also in connection Adenine + fMet of fMet-tRNA, carbon bond from adenine organic base and Nitrogen from agent NH_2 and (C-N) have the most intensities vibration and also is the most instable bond in this connection. By using results of the IR and NMR spectra, we understood the properties of the structure of formyl-methionine-tRNA.

Availability: For more information about these calculation you can see www.gaussian.com

REVEALING OF EVOLUTIONARY RELATIONSHIPS AMONG non-LTR RETROTRANSPOSONS BY NON-METRIC MULTIDIMENSIONAL SCALING.

A.S. Novikov*, V.M. Efimov, O.S. Novikova, A.G. Blinov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: anov@bionet.nsc.ru
**Corresponding author*

Key words: *Non-LTR retrotransposons, non-metric multidimensional scaling, phylogenetic trees*

Motivation and Aim: The field of phylogeny has the goals of working out the relationships among species, populations, individuals, genes. It has been shown that evolutionary trees reconstructed by different methods and based on the same data set can contain inconsistencies to each other. Multidimensional scaling (MDS) is a method for visualizing proximity data, that is, data where objects are characterized by dissimilarity values for all pairs of objects. An efficient algorithm for non-metric multidimensional scaling analysis (nMDS) has been developed recently as a maximally unsupervised data mining method for very large data sets. In this study phylogenetic investigation of more than 100 non-LTR retrotransposons was performed. Since nMDS is a very natural tool to visualize the salient features in the data in general, we have compared information, which was obtained using the distance based method of phylogeny reconstruction and results of nMDS analysis. Particularly, we were interested in relationships between ancient and novel clades of non-LTR retrotransposons.

Methods and Algorithms: The input for nMDS is the rank order of dissimilarities among the objects (mobile elements (MEs) in the present study). Dissimilarity matrices were obtained in MEGA 4.0 program. All obtained matrices had more than 0.99 rank correlation. Therefore, the choice of amino acid substitution model can not affect results. We used two types of evalution of the goodness of the configuration that we obtain by the nMDS algorithm. The first one was a pointwise criterion to evaluate individual objects, and the second one was based on estimation of the number of correctly embedded objects.

Results: Since more than 90% of the objects are correctly embedded, we may conclude that nMDS has succeeded in visually capturing the information in the data. 3D embedding is essentially similar to this. This solution have only one worse-than-0.01%-level ME according to the pointwise embedding criterion. Almost all clades recognized by the phylogenetic analysis can be recognized as well-defined clusters. It is clear that MEs from both the ancient and novel clades can be easily separated by line. nMDS seems to capture extra relations. For example, retrotransposon PgtNLR7 from the Deceiver clade embedded separately from two other retrotransposons from the same and evolutionary close (RTE, L1) clades from run to run. Thus, we can suppose that this ME groups on the tree with two other elements by chance and does't truly belong to the Deceiver clade. More data are needed to prove it.

Conclusion: nMDS can reproduce initial structure without any metric input. This fact is well known, but the larger dataset has been used in our investigation. nMDS cannot directly infer phylogeny, but it seems to capture extra relations that cannot be captured by tree diagrams. We should have more experiences with large-scale datasets.

INFLUENCE OF HYDROPHOBICITY OF BETA-BLOCKERS ON THEIR BINDING AFFINITY

V.N. Novoseletsky*, R.G. Efremov
Institute of Bioorganic Chemistry RAS, Moscow, Russia
e-mail: valeryns@gmail.com
**Corresponding author*

Key words: *beta-blockers, hydrophobicity, binding affinity*

Motivation and Aim: G-protein-coupled receptors (GPCR) play a key role in signal transduction in living cells, participate in regulation of numerous biological functions and, thus, are important therapeutic targets in a variety of disease states. Beta-adrenoreceptor antagonists are among the most widely used drugs in clinical practice. It has been shown earlier that antagonist binding is accompanied by a hydrophobic interaction between receptor and ligand [1]. To further investigate this issue and arrive to numerical models, we used theoretical modeling approach to the estimation of hydrophobic interactions between the beta2-adrenoreceptor and a set of its antagonists with experimentally measured binding constants.

Methods and Algorithms: To calculate hydrophobic interactions we used web-server PLATINUM (http://model.nmr.ru/platinum) [2]. It is an easy-to-use and customizable tool for estimation of the hydrophobic/hydrophilic match or mismatch at the interface of two molecules based on the concept of the Molecular Hydrophobicity Potential (MHP). It utilizes empirical atomic hydrophobicity constants derived from the water-octanol partition coefficients for organic compounds. Recently determined spatial structure of adrenoreceptor [3] was used to perform molecular docking of antagonist molecules. Resulting complexes were then optimized to satisfy experimental spatial restraints. Original approach was proposed for estimation of hydrophobic match of beta-blockers in the binding site of beta2-adrenergic receptor.

Results: Detailed analysis of hydrophobic match/mismatch of the beta-blockers reveals some properties of substituents which are crucial for high binding affinity. Hydrophobicity of particular fragments is used for construction of scoring function for binding affinity.

Conclusion: Strong correlation ($R^2 = 0.8$) between pKd and hydrophobic match was found. Therefore our method of estimating hydrophobic match can be further used to discover new potent inhibitors to beta2-adrenoreceptor.

Acknowledgements This work was supported by the grant of Russian Foundation for Basic Research № 07-04-01514-a, by the Programme RAS MCB and by the grant SS-4728.2006.

References:

1. M.L.Contreras et al. (1986) Thermodynamic properties of agonist interactions with the beta adrenergic receptor-coupled adenylate cyclase system. II. Agonist binding to soluble beta adrenergic receptors, *J Pharmacol Exp Ther,* **237:** 165-172.
2. T.V.Pyrkov et al. (2009) PLATINUM: a web tool for analysis of hydrophobic/ hydrophilic organization of biomolecular complexes, *Bioinformatics,* **25:** 1201-1202.
3. V.Cherezov et al. (2007) High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor, *Science*, **318:** 1258-1265.

LIMITED COMPONENTS IN DIET INCREASE LIFE SPAN IN FRUIT FLIES: SISTEM ANALYSIS

V.N. Novoseltsev*, J.A. Novoseltseva
Institute of Control Sciences RAS, Moscow Russia
e-mail: novoselc@yandex.ru
**Corresponding author*

Key words: *gerontology, caloric restriction, life span, nutrients content, reproduction.*

Motivation and Aim: It is posed in modern gerontology that life span increase is related to caloric restriction. They thought that reproduction stoppage due to the caloric restriction liberates energy, which is used for somatic maintenance thus enlarging life span. Nonetheless this statement is prejudiced during the last years. Firstly, it is shown that caloric restriction affects life span also in sterile fly females, in which energy economy is principally impossible (Mair et al. 2005). Secondly, experiments were made in which the diet components (yeast and sugar), not calories, were restricted (Scorupa et al. 2008). It was shown that life span is regulated not by caloric contents but the nutrients in the diet. Still it was not clear what mechanisms underlay the phenomenon.

Methods and Algorithms: We analyze systematically the processes in a fruit fly organism and simulate them to find out the underlying mechanisms. Our study is based on Partridge's and Pletcher's groups results (Mair et al. 2005, Scorupa et al. 2008). The first ones demonstrate increase of life span in sterile *Drosophila* females, the second, the dependence of reproduction and life span as well as obesity and protein accumulation on the yeast and sugar concentration in the diet. We studied the following processes in a fly organism: energy consumption and balance, eggs production, protein accumulation, adiposity, starvation and ageing. We proposed a mathematical model describing these processes and elaborated the corresponding Matlab schemes to simulate them.

Results: Firstly, we have stressed that caloric restriction breaks the energy balance. The disturbance provokes reproduction stopping. If it is not enough to restore the energy balance, metabolic processes are to be remodeled to diminish energy departure and to increase energy supply. We show that energy economy due to cessation of the egg production is enough to produce the experimentally observed increase in life span.

Secondly, simulation of the processes in component-restricted experiments widely confirmed the findings of Pletcher's group. Changes in life span are normally follow egg production and obesity, and protein and sugar shortage drastically restricts it.

Particularly we have found that increase in protein supply under starvation conditions shorten the time lag preceding start of egg laying.

Conclusion: System analysis of component restriction in diet made it possible to find out the underlying mechanisms of life span increase. From one side, the main factors affecting the increase are stopping of egg production and obesity beginning. From the other, starvation both in protein and sugar restricts life span. The opinion that "lifespan is optimized on a balanced diet as the interaction between cues from each dietary component maintain energy balance allowing sufficient resources for reproduction, energy storage, and somatic maintenance" (Scorupa et al. 2008) looks as premature. Life span is formed by normal interference of the processes of egg production, obesity, and starvation.

References:

1. W. Mair et al. (2005) Calories do not explain extension of life span by dietary restriction in *Drosophila. PloS Biol.* **3**:1305-1311.
2. D.A. Scorupa et al. (2008) Dietary composition specifies consumption, obesity, and lifespan in *Drosophila melanogaster. Aging Cell.* **7**:478-490.

COMPARATIVE STUDY OF PROTEIN MOLECULAR DYNAMICS TRAJECTORIES OBTAINED WITH DIFFERENT COMPUTATIONAL PARAMETERS

A.Yu. Nyporko
Taras Shevchenko Kiev National University, Kiev, Ukraine
e-mail: dfnalex@gmail.com

Key words: *molecular dynamics, calculation parameters, reliable trajectories.*

Motivation and Aim. Molecular dynamics computation is very useful method for investigation of protein structure [1]. This approach gives a possibility to discriminate and evaluate the nuances of protein structure and behavior which are indiscernible for other methods (in particular, distant correlative motions [2] and distant effects of amino acid replacements [3,4]) as well as allow to observe perturbations of protein structure for some of time. However, despite usability and wide facilities, results obtained via molecular dynamics sometimes are ambiguous. Main claim to molecular dynamics is the slow but stable gradual increase of protein oscillation amplitude and, as a result, lack of relaxation of molecular system that is observed in some cases. Thus, the aim of this research is selection of optimal criteria for molecular dynamics calculations (force fields, calculation parameters, etc.) to generate time-stable protein motions trajectories which correlate with NMR-derived motions' patterns.

Methods and Algorithms. Protein Data Bank was scanned to retrieve the spatial structures of water-soluble proteins, participating in cytoskeleton functioning, which are resolved by NMR spectroscopy and have no less than 20 deposed conformers. Among them the four proteins were selected in a random way for 50 ns molecular dynamics calculation with various criteria/conditions (PDB access codes are 1UNC, 1AJ3, 1GM1 and 1T0Y). Calculations were carried out with GROMACS 4.0.3 software using force fields Gromos 96, Gromacs and OPLS. The next parameters were varied: *nstlist* – frequency to update the neighbor list (0, 1, 10, 100), *comm_mode* – center of mass motion removal function (none, linear, angular), *T-coupling* – thermostat function (berendsen, V-rescale), *rcoulomb* – computation method for electrostatic interactions (cut-off, PME). Structural stability was estimated by conformational energy dynamics (using *g_energy* module), and levels of molecular oscillations (using *g_rms* module). Motion patterns for each studied protein were obtained from 20 molecular dynamics' conformers with lowest potential energies and compared with appropriated patterns calculated from NMR-derived conformers.

Results and Conclusion. Among studied parameters of molecular dynamics calculations PME method of electrostatic interactions evaluations shown the most stabilize effect on protein dynamics. Application of this method in combination with any others resulted in stable horizontal plateau of molecular oscillations and decrease of average oscillation level and oscillation amplitude. Using V-rescale thermostat resulted in accelerated stabilization of motions level. Variation of *nstlist* hasn't produced significant consequences (except nstlist=1 resulted in accelerated stabilization of system similar to V-rescale action). Effects of center of mass motion removal function were ambiguous. Among studied force fields OPLS appeared to be the most optimal for molecular dynamics calculations. Profiles of oscillations of individual amino acid residues obtained from trajectories calculated with OPLS are similar to NMR motion patterns for all studied protein. Using Gromacs force field revealed good correlation with NMR data for 1GM1 and 1T0Y proteins, using Gromos 96 force field – with NMR data for 1T0Y only. Thus, application of OPLS force field with PME electrostatic calculation and V-rescale thermostat seems to be optimal for molecular dynamics calculation and gives a possibility to obtain the protein dynamics trajectories comparable by a quality with NMR data.

References

1. M. Meli, G. Colombo (2009) Molecular simulations of peptides: a useful tool for the development of new drugs and for the study of molecular recognition. *Methods Mol Biol,* **570**:77-153.
2. O. F. Lange, H. Grubmaller (2008) Full correlation analysis of conformational protein dynamics. *Proteins,***70***:*1294-312.
3. P. Lazar et al. (2009) Molecular modeling study on the effect of residues distant from the nucleotide-binding portion on RNA binding in *Staphylococcus aureus* Hfq. *J Mol Graph Model.* **28**:253-260.
4. A.Yu. Nyporko, Ya.B. Blume (2008) Influence of amino acid replacements associated with multidrug resistance on в-tubulin molecular dynamics. (The Sixth Intl. Conference on Bioinformatics of Genome Regulation and Structure (Novosibirsk, Russia, June 22-28, 2008). Book of Abstracts, Novosibirsk, 2008, p. 177.

AN EXPERIMENTAL AND COMPUTATION APPROACH TO SEARCH FOR THE TRANSCRIPTION FACTOR GAGA BINDING SITES

E.S. Omelina*, D.Y. Oshchepkov, E.M. Baricheva, T.I. Merkulova
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: kuoma@ngs.ru
**Corresponding author*

Key words: *transcription factor GAGA, binding sites, Drosophila melanogaster*

Motivation and Aim: Development of computer assisted methods to predict transcription factor binding sites (TFBSs) is very important for eukaryotic genomes annotation. But the high false positive rate is a serious issue in attempts to reliably predict TFBSs. The most efficient way for quality estimation of the TFBS prediction methods is based on the direct experimental verification. The aim of this study was to develop the accurate methods to search for transcription factor GAGA binding sites in the *Drosophila* genes and experimental verification of the predicted sites.

Methods and Algorithms: The SITECON method [Oshchepkov *et al*., 2004] was used for recognizing the GAGA binding sites. Electrophoretic mobility shift assay (EMSA) was used to verify GAGA binding sites predicted in the regulatory regions of individual genes.

Results: To develop the recognition method we collected a sample of the nucleotide sequences of 128 experimentally identified GAGA binding sites from 15 genes previously described in the literature. Analysis of the sample has shown that the structure of only 25% of GAGA sites corresponds to GAGAGAG consensus sequence [Wilkins and Lis, 1998]. This type of sites may be included to microsatellite category $(GA)_{3\text{-}9}$ that combines 31% of all GAGA sites. 11% and 20% of GAGA sites are organized as direct repeat of trinucleotide GAG separated by one or three base pairs correspondently, 16% of sites contain only half-site GAG. On the basis of samples of GAGnGAG and GAGnnnGAG types two variants of SITECON method were developed and a number of the potential GAGA sites were found in the *Drosophila* genes. To estimate the recognition quality, we tested the capability of double-stranded oligonucleotides corresponding to the predicted sites to interact with the purified recombinant GAGA protein by EMSA. The experimental tests demonstrated that more than 85% of the predicted sites, indeed, interact with the GAGA-factor.

Conclusion: Thus, the perspective approach for recognizing different types of GAGA binding sites was created basing on program package SITECON.

Availability: http://wwwmgs.bionet.nsc.ru/mgs/programs/sitecon/

References:

1. D.Y.Oshchepkov et al. (2004) SITECON: a tool for detecting conservative conformational and physicochemical properties in transcription factor binding site alignments and for site recognition, *NAR*, W208-12.
2. R.C.Wilkins, J.T.Lis. (1998) GAGA factor binding to DNA via a single trinucleotide sequence element, *NAR*, 26(11):2672-8.

THE TIME OF TUMOR SUPPRESSOR PROTEIN FAMILIES EMERGENCE AND ITS FUNCTION IN GERM-LINE.

L.V. Omelyanchuk*, J.A. Pertseva**
*Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia.
**Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia.*
e-mail: ome@bionet.nsc.ru
**Corresponding author*

Key words: *tumor suppressor, multicellarity, meiosis*

Motivation and Aim: Earlier we showed that Merlin tumor suppressor protein family was formed at the time of first metazoa emergence [Golovnina et al. 2005]. This rise the question if this is also the case for the other tumor suppressor protein families. Our experimental studies document the function Merlin in *Drosophila melanogaster* male germ line [Dorogova et al. 2008], so we asked when this additional function emerged.

Methods and Algorithms: Phylogenetic trees were constructed by Mega-4 or Clustal-X.

Results: LGL protein is a classical tumor suppressor in *Drosophila.* This protein family has evident homology to Sro7p and Sro77p of *S. cerevisiae.* The phylogenetic tree constructed for LGL protein family (Tom1 of *C.elegans* was used as an out group) shows the protein family was emerged before *Trichoplax adhaerens* (considered as the most primitive metazoa) LGL derivate from the LGL of metazoan common ancestor [Omelyanchuk, Pertseva, 2010]. Phylogenetic tree constructed for the Hyd protein family shows that the formation of protein family takes place before *Trichoplax adhaerens* protein diverge from the common ancestor [Omelyanchuk et al. 2009a]. HRS protein was shown to be a binding partner of tumor suppressor Merlin. The phylogenetic tree constructed for this protein shows the protein family was formed at the time when C. elegans protein diverged from common ancestor [Omelyanchuk et al. 2009b]. The study of spermatogenesis in those cases documents meiotic abnormalities. So the tumor suppressors my have primary target in meiosis in unicellular eukaryotes and were integrated in the proliferation control when the metazoa were formed. Since LGL family has clear progenitors in unicellular *S. cerevisiae* (Sro7p and Sro77p) we tested this point considering available microarray data for genes involved in sporulation in this organism. It occurs that those proteins did not change its expression during sporulation. This means that unicellular progenitors of tumor suppressors unlikely to have the meiotic functions.

Conclusion: We showed that the tumor suppressor protein families Merlin, LGL, Hyd and HRS emerged simultaneously with multicellarity formation for the purpose of proliferation control in the primitive cell aggregates.

References:

1. K.Golovnina, A.Blinov E.M.Akhmametyeva, L.V.Omelyanchuk, L.S.Chang. (2005)Evolution and origin of Merlin, the product of the Neurofibromatosis type 2 (NF2) tumor suppressor gene, BMC Evol. Biol. 5: 69.
2. N.V.Dorogova, E.M.Akhmametyeva, S.A.Kopyl, N.V.Gubanova, O.S.Yudina, L.V.Omelyanchuk, L.S.Chang. (2008) The role of Drosophila Merlin in spermatogenesis. BMC Cell Biol., 9(1):1.
3. L.V.Omelyanchuk, Yu.A.Pertseva (2010) Phylogenetic Study of Formation of the Lethal (2) Giant Larvae Tumor Suppressor Protein Family, Doklady Biochemistry and Biophysics, 2010, 430:50–52.
4. L.V.Omelyanchuk, J.A.Pertseva., J.M.Saul (2009a) Evolutionary Origin of the Tumor Suppressor Hyperplastic Discs Protein, In Silico Biology.9:203–207.
5. L.V.Omelyanchuk, J.A.Pertseva, S.S.Burns, L.S.Chang (2009b) Evolution and Origin of HRS, a Protein Interacting with Merlin, the Neurofibromatosis 2 Gene Product, Gene Regulation and Systems Biology. 3:143–157.

PROTEIN FUNCTION INFORMATION FROM FUNCTIONAL SITE PREDICTION

M.J. Ondrechen*
Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA 02115 USA
e-mail: M.Ondrechen@neu.edu

Key words: *Protein function prediction; THEMATICS; POOL; functional sites*

Motivation and Aim: The development of computational tools for the prediction of protein function from the 3D structure is a very important problem in the post-genomic era. To date there are over 8700 structural genomics protein structures in the Protein Data Bank, most of which are of unknown or uncertain function. We and others have developed computational methods for the prediction of functionally important residues in proteins. The utilization of those predictions for functional annotation is now addressed.

Methods and Algorithms: Previously we have reported on THEMATICS, a method for the prediction of functionally important residues in protein 3D structures [1]. More recently, Partial Order Optimum Likelihood (POOL) [2] utilizes THEMATICS and other input features in a new, monotonicity-constrained maximum likelihood machine learning method, for enhanced performance in prediction of catalytic and binding residues.

Results: To carry functional residue prediction to the next step of function prediction, we have placed predicted functional residues onto a structural alignment of superfamily members, in order to sort the superfamily into functional subclasses [3]. This technique is now applied to structural genomics proteins of unknown function. In one example of a structural genomics protein from a domain of unknown function from *Bifidobacterium adolescentis*, the closest 3D structure match is to DNA polymerase III and the phylogenetic tree also predicts DNA replication activity, albeit inferred from electronic annotation. However, our best local structure match at the predicted active site is to a phosphatase. Experiments have confirmed this functional prediction. Direct biochemical assay shows phosphatase activity and primer extension analysis shows no DNA replication or DNA repair activity. Results from some other selected superfamilies are shown.

Conclusion: Functional annotation by a local structural match at the predicted active site can be a valuable tool for the discovery of protein function from 3D structure.

Availability: THEMATICS server: http://pfweb.chem.neu.edu/thematics/submit.html

Acknowledgements: US National Science Foundation grant MCB-0843603

References:

1. Y. Wei, J. Ko, L.F. Murga, and M.J. Ondrechen (2007) Selective Prediction of Interaction Sites in Protein Structures with THEMATICS. *BMC Bioinformatics*. **8**: 119.
2. W. Tong, Y. Wei, L.F. Murga, M.J. Ondrechen, and R.J. Williams (2009) Partial Order Optimum Likelihood (POOL): Maximum Likelihood Prediction of Protein Active Site Residues Using 3D Structure and Sequence Properties. *PLoS Comp Biol* **5**(1): e1000266.
3. Y. Wei, D. Ringe, M.A. Wilson, and M.J. Ondrechen (2007) Identification of Functional Subclasses in the DJ-1 Superfamily Proteins. *PLoS Comp Biol* **3**(e10): 120-126.

COMPARATIVE ANALYSIS OF CPG AND CNG CLUSTERS IN THE VICINITY OF MAMMALIAN ORTHOLOGOUS GENES.

Oparina Nina[1,2*], Fridman Marina[2], Makeev Vsevolod[2]
[1]Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
oparina@gmail.com
[2]Institute of genetics and selection of industrial microorganisms, GosNIIgenetika, Moscow, Russia
e-mail: *oparina@gmail.com*
*corresponding author

Motivation and Aim: Mammalian genomes contain less CpG dinucleotides due to demethylation. By the way, multiple clusters of CpGs traditionally called CpG-islands are frequent in these genomes. It was shown that CpGs remain mutable even in CpG islands besides they are considered to be unmethylated there. The recent data demonstrated that CpGs remain highly mutable even in CpG islands and in several phyla CpG-islands are depleted and shortened. Nevertheless, even in these genomes there're still a lot of CpG islands. Multiple algorythms were used for CpG-islands detection in the genome. Nevertheless there still no good ideas for detection of functional CpG-islands – we still don't know there exact fucntions.

Methods and Algorithms: We used UCSC GenomeBrowser and Galaxy, RepeatMasker Web Server, Bioconductor and R for clusterization tasks, CpGcluster, CpGProD? Evola and RoundUp orthologs databases and some others.

Results: We have found that despite overall mutability numerous CpG-enriched clusters were found in the vicinity of orthologous mammalian genes, including there promotor regions, 3'-ends, UTRs. In several cases orthologous introns containing both CpG-clusters and mammalian-conserved elements were detected. Neary half of such sequences were highly conserved in mammals but the others were not similar but were surrounded by conserved elements. We have classified CpG-clusters including: non-conserved, located at the same orthologous loci (with no sequence conservation of CpG island), and highly conserved. Conserved CpG islands were detected mostly in gene-overlapping loci, while a subset of intergenic conserved CpG was described. Surprizingly, the most conserved CpG-islands were enriched with CNG motifs. We have compared CNG content in various CpG-clusters and found that they are also surprizingly enriched in germline methylated islands.

Conclusion: CpG-enriched sequences, both detected by modern CpG-island detection methods and small or GC-poor undetected are frequently conserved and/or located in the vicinity of the same orthologs in mammals. We have constructed classification of these sequences available for further funtional study of CpG-islands. We have shown that conserved CpG-clusters are frequently methylated in germline and also are enriched with CNG motifs. The understanding of evolutionary stability of such islands needs further investigations.

References:

1. Bird AP: CpG islands as gene markers in the vertebrate nucleus. Trends Genet 1987, 3:342-347.
2. Gardiner-Garden M, Frommer M: CpG islands in vertebrate genomes. J Mol Biol 1987, 196:261-282.
3. Takai D, Jones PA: Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci USA 2002, 99:3740-3745.
4. Hackenberg M, Previti C, Luque-Escamilla PL, Carpena P, Martinez-Aroza J, Oliver JL: CpGcluster: a distance-based algorithm for CpG-island detection. BMC Bioinformatics 2006, 7:446.
5. Han L, Su B, Li WH, Zhao Z: CpG island density and its correlations with genomic features in mammalian genomes. Genome Biol 2008, 9:R79.
6. Jiang C, Han L, Su B, Li WH, Zhao Z: Features and trend of loss of promoter-associated CpG islands in the human and mouse genomes. Mol Biol Evol 2007, 24:1991-2000.

GENOME WIDE NUCLEOSOME OCCUPANCY AND TRANSCRIPTION FACTORS BINDING IN YEAST GENOME

Y.L. Orlov*, W.S. Goh, J. Li, J.-Q. Run, X. Xue, M. Huss, N.D. Clarke
[1]*Genome Institute of Singapore, 60 Biopolis Street, Genome, 138672 Singapore.*
e-mail: orlovy@gis.a-star.edu.sg
[*]*Corresponding author*

Key words: *next generation sequencing, yeast genome, nucleosome, TF binding*

Motivation and Aim: Study of high-resolution nucleosome position data that has recently become available in yeast together with transcription factor binding data provides a model for global regulation of gene expression. Chromatin immunoprecipitation microarray experiments (ChIP-chip) years ago revealed an under-representation of nucleosomes in promoter regions, relative to transcribed regions. In contrast, TFs are underrepresented in transcribed regions and enriched in promoters.

Methods and Algorithms: We revisit the data of Kaplan et al. (2009) as well as new, more quantitative ChIP-qPCR experiments (Goh et al., 2010). The maps have been obtained from both crosslinked and uncrosslinked chromatin in vivo, and from chromatin assembled from genomic DNA and nucleosomes in vitro. Additionally we used new in vivo nucleosome sequencing data from yeast in different environmental conditions. We confirm a role for intrinsic nucleosome binding preferences in the binding of transcription factors. To assess quantitatively the correlation between low nucleosome occupancy and TF binding, we asked how well nucleosome tag counts correctly distinguish TF-bound sites from random sites selected from yeast promoters. We use the area under the ROC curve as a measure of this association.

Results: The analysis was done to the 41 yeast TFs for which there are at least 50 binding motifs bound according to the ChIP-chip data of Harbison et al. (2005) ($p<1e-3$) and the subsequent motif re-analysis by the same group. Bootstrap analysis of ROC curve areas shows a significant association between TF occupancy and nucleosome depletion for most of the TFs under analysis. To assess more directly how much of an effect on TF binding is encoded by the intrinsic DNA specificity of nucleosomes, we determined the binding occupancies of 107 perfect consensus binding sites in the genome using ChIP-qPCR. ChIP enrichment values are inversely correlated with nucleosome occupancy.

Conclusion: Two different methods in their nucleosome mapping experiments, one involving formaldehyde crosslinking and the other a more traditional non-crosslinking protocol (Kaplan et al., 2009). We have shown that TF binding sites tend to be associated with excess nucleosome counts in crosslinked chromatin vs. uncrosslinked (nucleosome occupancy difference). In conclusion, we argue that regional nucleosomal density in yeast promoters is generally more relevant to TF binding than precise nucleosome position.

References:

1. Goh W.S., Orlov Y., Li J., Clarke N.D. (2010) *PLoS Comput Biol.* 6(1):e1000649.
2. Harbison C.T., et al. (2004) *Nature* 431: 99-104.
3. Kaplan N., et al. (2009) *Nature* 458:362-6.

EXTENDING MAPS OF TF BINDING IN EMBRYONIC STEM CELLS BY CHIP-SEQ

Yuriy L. Orlov*, Mikael Huss, Bing Lim, Huck-Hui Ng
Genome Institute of Singapore, 138672, Singapore
e-mail: orlovy@gis.a-star.edu.sg
**Corresponding author*

Motivation and Aim: Success of high-throughout sequencing combined with chromatin immunoprecipitation (ChIP-seq) allows study TF landscape in genome scale. Identification of the DNA binding specificities of sequence-specific transcription factors (TFs) is important for understanding transcriptional regulatory networks. We used self-renewal mouse embryonic stem cell as a model. Induced pluripotent stem (iPS) cells can be obtained through the introduction of defined factors into somatic cells1. The combination of Oct4, Sox2 and Klf4 (OSK) constitutes the minimal requirement for generating iPS cells from mouse embryonic fibroblasts.

Methods and Algorithms: We have applied ChIP sequencing to build unbiased whole-genome map of transcription factor binding in mouse ESC (embryonic stem cells). The transcriptional reprogramming of a genome in response to signals involves the coordinated regulation of hundreds of genes, mediated by transcription factors, changes in chromatin structure, and the action of regulatory RNAs. We are using genomic and molecular approaches to understand genome-wide regulatory networks involving transcription factors and their downstream targets, some of which can themselves be regulators. In addition to Nanog, Oct4, and other 12 TFs we consecutively used ChIP-seq to construct genome binding maps for Eset (Yuan et al., 2009), Nr5a2 (Heng et al., 2010), and Tbx3 (Han et al., 2010). Binding sites peak calling was done by MACS program using specific chromatin IP and control input sequencing data from the same cells as described in (Chen et al., 2008). Co-localization count of ChIP-seq TFs binding sites in mouse genome was done based on overlaps of genomic coordinates.

Results: We developed methods for ChIP-seq peak detection and downsteam analysis including chromatin modifications and co-motif enrichment. We have shown two main groups of TFs in networks: promoter-specific Myc associated TFs and distal regulators associated with Oct4. Through the genomic analyses of ESC genes that have roles in pluripotency and fusion-mediated somatic cell reprogramming, we identified Tbx3 as a transcription factor that significantly improves the quality of iPS cells. Induced-PS cells generated with OSK + Tbx3 (OSKT) are superior in both germ cell contribution to the gonads and germ-line transmission frequency. However, global gene expression profiling could not distinguish between OSK and OSKT iPS cells. Genome-wide ChIP-sequencing analysis of Tbx3 binding sites in ESCs suggests that Tbx3 regulates pluripotency-associated and reprogramming factors, in addition to sharing many common downstream regulatory targets with Oct4, Sox2, Nanog and Smad1.

Conclusion: The fundamental changes occurring during the metastatic progression are very complex. Understanding of these processes at the basic levels requires a combination of approaches aimed at measuring different types of molecules and modifications of molecules in a cell. Availability of such data creates unprecedented opportunities to discover hitherto unknown relationships and yet provides significant challenges to our ability to integrate and analyze diverse data sets and extract biologically-relevant information from them.

References:

1. Chen X. et al. (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133(6):1106-17.
2. Yuan P. et al. (2009) Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. *Genes Dev*. 23(21):2507-20.
3. Heng J.C. et al. (2010) The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. *Cell Stem Cell* 6(2):167-74.
4. Han J. et al. (2010) Tbx3 improves the germ-line competency of induced pluripotent stem cells. *Nature* 463(7284):1096-100.

PROMOTERS OF THE GENES ENCODING THE MACROPHAGE TRANSCRIPTION FACTORS CONTAIN BINDING SITES FOR ARYL HYDROCARBON RECEPTOR

E.A. Oshchepkova[1]*, D.Y. Oshchepkov[1], E.V. Kashina [1],
E.V. Antontseva [1], M.Y. Shamanina[1], V.A. Mordvinov[1], D.P. Furman[1,2]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
e-mail: nzhenia@bionet.nsc.ru
[*]*Corresponding author*

Key words: *macrophage, dioxin, TCDD, transcription factor, promoters, binding sites.*

Motivation and Aim: 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) is the most toxic among the dioxin xenobiotics [1]. TCDD mediates gene expression via AhrR/Arnt transcription complex activation, which binds to dioxin responsive elements (DRE) in the gene regulatory regions. The role of dioxin in activity regulation of the genes involved in development of the immune response can be regulated both directly, by the transcription complex containing dioxin as a ligand, and indirectly, via intrinsic transcription factors, containing fumctional DREs in it's regulatory regions.

Methods and Algorithms: The regulatory regions of the genes, encoding key TFs and it's subunits were searched for DREs using the SITECON software package, designed for searching for the transcription factors binding sites in genomic sequences [2]. EMSA and Rt-PCR experiments were performed to test sites functionality.

Results: DRE sites have been detected in the promoter regions of 10 of the 12 genes encoding the transcription factors functioning in macrophages, namely, NFkB3 and STAT3 (three in each gene); CEBPβ and RelB (two in each); and NFkB1, NFkB2, JUN, IRF1, and IRF4 (one in each). Several DREs' functional activity was experimentally proven.

Conclusion: The double regulation, via DRE and the binding sites for the corresponding transcription factors in the promoter regions of macrophage genes, and the interregulation of the genes providing for the immune response allow the system to rapidly respond to a provocative agent (xenobiotic) and finely tune its function.

Acknowledgements: This work was supported by RFBR (No. 09-04-12209-ofi_m), projects of the RAS (No. 15 and 22.8) and SB RAS (Nos. 107 and 119 and Lavrentiev's project in biological sciences No. 3) Federal Agency for Science and Innovations (contract no. 02.512.11.2332); and grant NSh-2447.2008.4.

References:

1. P.K. Mandal. (2005) Dioxin: a review of its environmental effects and its aryl hydrocarbon receptor biology. *J Comp Physiol B.* 175: 221-230.
2. D.Yu. Oshchepkov et al. (2004) SITECON: a tool for detecting conservative conformational and physicochemical properties in transcription factor binding site alignments and for site recognition. *Nucl. Acids Res.* 32: W208–212.

DETECTION OF DIOXIN RESPONSIVE ELEMENTS IN THE PROMOTER REGIONS OF THE GENES ENCODING MACROPHAGE CYTOKINES

E.A. Oshchepkova[1]*, D.Y. Oshchepkov[1], E.V. Kashina [1], E.V. Antontseva [1], D.P. Furman[1,2], V.A. Mordvinov[1]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
e-mail: nzhenia@bionet.nsc.ru
[*]*Corresponding author*

Key words: *macrophage, dioxin, TCDD, DRE, cytokines, promoters, transcription.*

Motivation and Aim: Dioxin (and its prototype TCDD) is a one of the most potent toxicants. TCDD induces a broad spectrum of biological responses, including disruption of normal hormone signaling pathways, reproductive and developmental defects, immunotoxicity, liver damage, wasting syndrome, and cancer [1]. On the cell level, the dioxin mediates gene expression via AhR/ARNT transcription complex activation, which binds to dioxin responsive elements (DRE) in the regulatory regions of the inducible genes. Previously the TCDD–mediated modulation of the expression of genes encoding cytokines was shown experimentally [2]. For better understanding whether AhR/ARNT action is direct, or indirect through the immanent transcription factors, we have performed the search of the putative DREs in the regulatory regions of the genes, encoding cytokines, expressed in activated macrophages.

Methods and Algorithms: The regulatory regions of the genes, encoding cytokines were searched for DREs using the SITECON software package, designed for searching for the transcription factors binding sites in genomic sequences [3]. First series of EMSA and Rt-PCR experiments were performed to test sites functionality.

Results: DRE sites have been detected in number of macrophageal cytokine gene promoters, with *IL12a, IL12b, IFNA, IL24* among others. A few DREs' functional activity was experimentally proven.

Conclusion: Obtained results sound in favor of the possibility that TCDD can directly mediate the gene expression of the genes encoding macrophageal cytokines, containing DREs in their regulatory regions, thus affecting the immune response.

Acknowledgements: This work was supported by RFBR (No. 09-04-12209-ofi_m), projects of the RAS (No. 15 and 22.8) and SB RAS (Nos. 107 and 119 and Lavrentiev's project in biological sciences No. 3) Federal Agency for Science and Innovations (contract no. 02.512.11.2332); and grant NSh-2447.2008.4.

References:

1. P.K. Mandal. (2005) Dioxin: a review of its environmental effects and its aryl hydrocarbon receptor biology. *J Comp Physiol B.* **175**: 221-230.
2. C.F. Vogel et al. (2007) Modulation of the chemokines KC and MCP-1 by 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in mice. *Arch Biochem Biophys.* 461:169-175.
3. D.Yu. Oshchepkov et al. (2004) SITECON: a tool for detecting conservative conformational and physicochemical properties in transcription factor binding site alignments and for site recognition. *Nucl. Acids Res.* **32:** W208–212.

SIMILARITIES AND DIFFERENCES IN THE STRUCTURES OF ARTIFICIAL AND NATURAL TRANSCRIPTION FACTOR BINDING SITES: IMPACT OF FLANKING SEQUENCES TO RECOGNITION PERFORMANCE

Oshepkov D.Y. *, Levitsky V.G., and Khlebodarova T.M.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
e-mail: diman@bionet.nsc.ru
** Corresponding author*

Key words: *transcription factor, binding site, in vitro selected site*

Motivation and Aim: Recent development of new methods has provided ample information on the structures of transcription factor (TF) binding sites selected *in vitro*. However, it is still questionable whether these data reflect the actual structure of natural sites and how they can be used in methods of seeking sites in various genomes. In this study, the similarity between natural and artificial eukaryotic TF binding sites (BSs) is analyzed and their optimal length for recognition in the genome is assessed.

Methods and Algorithms: Nucleotide frequency matrices constructed from samples of natural and in vitro selected TF sequences were retrieved from the ArtSite database [1]. The matrices were compared by using the Kullback-Leiber divergence measure [2]. The accuracy of TFBS recognition with regard to sequence length was assessed by the oPWM method [3].

Results: The degrees of similarity between natural and artificial binding sites for 34 TFs from various species were compared. High degrees of similarity between core sequences of such BSs (7–14 bp) were found for about 80% of the TFs analyzed. However, analysis of BS recognition performance for these TFs showed that the best accuracy was achieved by using much longer sequences, 25–50 bp, which was not typical of artificial sites. Elongation of BS flanks increased the recognition performance two to tenfold for half of the BS. For 12 TFBSs this increase was more than tenfold: MEF2, SOX5, EGR1, etc. Exceptions were CEBPβ and HMG, where the increase in recognition performance with longer flanks was less than twofold, and AhR/ARNT with no such increase.

Conclusion: In spite of the fact that the core sequences of natural and *in vitro* selected BSs for 80% of TFs analyzed are highly similar, use of the latter for probing candidate sites in a genome can produce a great number of false positive results. For the overwhelming majority of TFs under study (>90%), flanking regions add much to the specificity of corresponding BSs.

Acknowledgments: This work was supported by RFBR (No. 08-04-01008), projects of the RAS (No. 22.8) and SB RAS (No. 119 and Lavrentiev's project in biological sciences No.3) and grant NSh-2447.2008.4.

References:

1. Khlebodarova T.M. et al. (2006) In: *Bioinformatics of Genome Regulation and Structure II.* N. Kolchanov and R. Hofestaedt (Eds), 55-65 (Springer Science+Business Media, Inc.)
2. Aerts S. et al. (2003), *Bioinformatics*, **19** (Suppl 2): ii5-ii14.
3. Levitsky V.G. et al. (2007), *BMC Bioinformatics*, **8:**481.

DEPPDB – DNA ELECTROSTATIC AND OTHER PHYSICAL PROPERTIES DATABASE

A.A. Osypov*, G.G. Krutinin, E.A. Krutinina, P.M. Beskaravayny, S.G. Kamzolova
Laboratory of Cell Genome Functioning, Institute of Cell Biophysics of RAS, 142290, Pushchino MR, Russia
e-mail: aosypov@gmail.com
**Corresponding author*

Key words: *DNA electrostatic potential, physical properties, data integration, genomics*

Motivation and Aim: Electrostatic and other physical properties of genome DNA influence its interactions with different proteins, in particular the regulation of transcription by RNA-polymerases. DEPPDB – DNA Electrostatic and other Physical Properties Database – 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, which are organized on a taxonomical basis.

Methods and Algorithms: The electrostatic potential around the double-helical DNA molecule was calculated by the original method [1, 2] using a new program package. It has an advantage comparing with the previously used calculation system as provides more accurate and free of the moire artifacts data handling, and calculations on demand on the fly. Calculations of other physical properties are based on the dinucleotide content. Different smoothing and cross-correlation algorithms are applied.

Results: Currently, the database contains all the completely sequenced bacterial, viral, mitochondrial and plastids genomes according to the NCBI RefSeq [3] as well as some model eucariotic genomes. Data for promoters, regulation sites, binding proteins *etc.* are incorporated from BioCyc [4] and manual annotation. All the data are fully integrated and several tools are provided to support different forms of analysis. Calculation on the fly of the user-provided sequences is available.

Availability: The database [5] is available for academic use via the web interface at http://deppdb.psn.ru (http://electrodna.psn.ru can be used as an alternative address).

Acknowledgements: The authors are grateful to Saveljeva E. G. for technical support and the Institute of Mathematical Problems of Biology of RAS for hosting the Database.

References:

1. R. V. Polozov et al. (1999) Electrostatic potentials of DNA. Comparative analysis of promoter and nonpromoter nucleotide sequences, J. Biomol. Struct. Dyn., 16(6), 1135-43.
2. A. A. Sorokin et al. (2006) Electrostatic properties of promoter recognized by E. coli RNA polymerase Esigma70, J Bioinform Comput Biol, 4(2), 455-467.
3. D. K. Pruitt, T. Tatusova, and D. R. Maglott. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins, NAR, 35, D61–D65.
4. P.D. Karp et al. (2005) Expansion of the BioCyc collection of pathway/genome databases to 160 genomes, NAR, 19:6083-89
5. A. A. Osypov, G.G. Krutinin, S. G. Kamzolova. (2010) DEPPDB – DNA Electrostatic Potential Properties Database. Electrostatic Properties of Genome DNA, J Bioinform Comput Biol, *in press*

CAMP-CRP BINDING SITES IN E.COLI GENOME. NEW INSIGHTS INTO PROTEIN-DNA ELECTROSTATIC INTERACTIONS

E.A. Krutinina, G.G. Krutinin, S.G. Kamzolova, A.A. Osypov*
Laboratory of Cell Genome Functioning, Institute of Cell Biophysics of RAS, 142290, Pushchino MR, Russia
e-mail: aosypov@gmail.com
**Corresponding author*

Key words: *DNA electrostatic potential properties, E.coli, promoters, cAMP-CRP binding sites*

Motivation and Aim: Electrostatic properties of genome DNA are well recognized to influence its interactions with different proteins, in particular the primary recognition and regulation of transcription by RNA-polymerases. To reveal the role of electrostatic properties of DNA in the recognition of the transcription regulation proteins binding sites we studied the binding sites of the cAMP-CRP complex in the genome DNA in *E.coli* K12 MG1655.

Methods and Algorithms: DEPPDB – DNA Electrostatic and other Physical Properties Database and its tools [1] were used to carry out the analysis.

Results: The averaged profiles of the DNA electrostatic potential aligned around the CRP dimer binding sites centres exhibit the pronounced rise in the negative potential value with the characteristic W-like profile in the concensus area of 16 bp (TGTGA-N6-TCACA palindrome). The extensive (around 150-300 bp long), symmetrical overall potential rise can not be explained by the influence of the concensus alone and reflects the sequence organization of the flanking regions, contributing to the high potential area formation. Apparently this sequnce organization was selected evolutionary to support the binding site recognition by the regulation protein molecula and its retention.

It is worth noting that this high potential area is relatively AT-enriched though doesn't possess any textual consensus properties. Such enrichment is commonly accepted as facilitating the DNA melting in the promoter regions. However, it is clearly not the case in the present system, as the promoter core (and the initially DNA melting point) lies downstream of the CRP binding sites, sometimes more than 100 bp apart. So the formation of the high electrostatic potential trap for the CRP complex is rather the function of this enrichment, especially given the fact of the symmetry of this electrostatic valley.

Conclusion: The data obtained reveal the role of electrostatic properties of DNA in the recognition of the transcription regulation proteins binding sites, further confirming their universal importance in the protein-DNA interactions beyond the classical promoter-RNA polymerase recognition and regulation. They demonstrate the necessity of the studies of the electrostatic properties of genome DNA in addition to the traditional textual analysis of its sequence.

Acknowledgements: The authors are grateful to Saveljeva E. G. for technical support.

References:

1. A. A. Osypov, G.G. Krutinin, S. G. Kamzolova. (2010) DEPPDB – DNA Electrostatic Potential Properties Database. Electrostatic Properties of Genome DNA, J Bioinform Comput Biol, *in press*

COMPARING TWO CONTIG ASEMBLY PROGRAM CAP3 AND PHAP WITH OLIVE EST'S COLLECTION

N. Ozdemir*[1], S. Vural Korkut[1], F. Oruç[1], U. Sezerman[2], C.Un[3]
1 Yildiz Technical University, Department of Biology, Istanbul,Turkey
2 Sabanci University, Faculty of Engineering and Natural Sciences, Istanbul, Turkey
3 Ege University, Department of Biology, Izmir, Turkey
e-mail: nehirozdemir@yahoo.com
**Corresponding author*

Key words: *CAP3, Phred/Phrap, EST collection, olive*

Motivation and Aim: Established EST's collection from olive leaf and fruit (*Olea europaea L.*) need to be assembled. Two programs CAP3 and Phrap which are using very common were applied to the fragment assembly. The results from both program was compared to evaluate the results.

Methods and Algorithms: 3734 EST's were generated from leaf and fruit cDNA libraries. The vector fragment was removed with Phred / Phrap, VecScreen programmes. The different computer programs were generated to compare and to show overlap of the ESTs. CAP3 and Phrap software were selected to compare the assembly results and used to alignment (Ewing and Green 1998; Ewing et al. 1998, Huang et al., 1999).

Results: While CAP3 program was applying with default value, Phrap program was tested with 10, 20, 30 bp as overlapping value. Overlapping value was accepted 20 bp for Phrap. From 3734 ESTs 924 contig and 1302 singlet were obtained with the Phrap software. On the other hand CAP3 software designated quite different result than Phrap. While 299 contigs were generating, 2368 singlets were remain with CAP3. Both results were evaluated with the program Consed / Autofinish which was developed for editing and representing of alignment. With control of Consed program we have noticed some inaccurate results with the Phrap. Although program represented some contigs consisted of two EST, with the program Consed we have discovered this contigs included just one EST. The data obtained from the CAP3 program were submitted to the GenBank database of NCBI using.

Conclusion: According comparison two program, CAP3 was accepted as a reliable source and their result was accepted to continue further analysis.

Availability:

Acknowledgements The Scientific and Technological Research Council of Turkey

References:

1. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment, *Genome Res*, **8(3):**175-185.
2. Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res*, **8(3):**186-194.
3. Huang X, Madan A. (1999) CAP3: a DNA sequence assembly program, GenomeRes, **9:** 868–77.

PREPARING AND ANALYSING EST's FOR THE MAMMARY TISSUE OF SHEEP IN PRENATAL AND POSTNATAL PERIOD

N. Ozdemir*[1], Z.Omeroglu[1], F. Oruç[1], K. Oztabak[2], C.Un[3]
1 Yildiz Technical University, Department of Biology, Istanbul,Turkey
2 Istanbul University, Faculty of Veterinary Medicine, Istanbul, Turkey
3 Ege University, Department of Biology, Izmir, Turkey
e-mail: nehirozdemir@yahoo.com
**Corresponding author*

Key words: *CAP3, Phred/Phrap, EST collection, sheep, colostrum, cDNA library*

Motivation and Aim: Because the colostrum is essential nutrient for new born lamb, the colostrum secretion, which secrets during first 48 h after the parturition, play very critical role on sheep productivity and sheep husbandry. With a EST collection we want to figure out of the gene expression profiles of colostrum secretion for prenatal and postnatal period in Kivircik sheep which is an important local Turkish sheep according to their meet quality and milk productivity.

Methods and Algorithms: In this study Kivircik sheep in farm of the Faculty of Veterinary science at University of Istanbul was used. The mammary tissues from the same sheep were taken by biopsy twice. End of the prenatal period just before 1 week of parturition first tissue was taken. 18-36 h during high period of the colostrum secretion after the parturition second mammary gland tissue was taken. From two cDNA libraries, two different EST collection was obtained and analayzed with Phred/Phrap, CAP3, BLAST (Ewing and Green 1998; Ewing et al. 1998, Huang et al., 1999).

Results: After construction of two cDNA libraries, total 3072 colonies which are randomly selected from the two libraries were sequenced to establish two ESTs collection for mammary gland tissue in pre- and postnatal period. For analysis of the raw EST data, the low-quality sequences and the vector fragment was removed with Phred / Phrap computer programmes. Fragment assembly was done with the CAP3 software. Putative functions of all unique sequences were designated by gene homology based on BLAST. Total 429 ESTs which show over % 80 homolgy to known sequences of other organism in NCBI have been determined. Also according to BLAST result we have compared two EST collections and listed differences of between prenatal and postnatal gen expression profiles.

Conclusion: This EST data are very valuable resource for functional genome studies of sheep.

Availability: All EST's data process in submitting NCBI

Acknowledgements Yildiz Technical University, Scientific Research Project Coordination

References:

1. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment, *Genome Res*, **8(3):**175-185.
2. Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res*, **8(3):**186-194.
3. Huang X, Madan A. (1999) CAP3: a DNA sequence assembly program, GenomeRes, **9:** 868–77.

A BIOPHYSICAL MODEL FOR GENOME-WIDE NUCLEOSOME AND TRANSCRIPTION FACTOR BINDING

Evgeniy A. Ozonov*, Erik van Nimwegen
Biozentrum, University of Basel, Basel, Switzerland
e-mail: evgeniy.ozonov@unibas.ch
**Corresponding author*

Key words: *nucleosome, transcription factors, biophysical model*

Motivation and Aim: The genomes of all eukaryotic organisms are packaged into nucleosomes - fundamental structural units of chromatin which consist of approximately 147 base pairs of DNA wrapped around a histone octamer. It has been well established that nucleosomes play an important role in transcription regulation [1]. The recent development of ChIP-Chip and ChIP-seq technologies have enabled the construction of genome-wide maps of nucleosome positioning and modifications at high resolution across various conditions. Such experimental data have already uncovered a number of general patterns, such as the fact that there generally are well-positioned nucleosomes at the 5' and 3' ends of genes and that a nucleosome-free region (NFR) occurs immediately upstream of active genes [2]. However, a clear mechanistic understanding of these patterns is currently lacking. Therefore, the aim of this work is theoretical analysis of transcription factors (TF) and nucleosome binding processes using thermodynamical model.

Methods and Algorithms: Starting from the binding specificities of 163 yeast transcription factors we model the biophysics of the binding of TFs and nucleosomes to transcription factor binding sites (TFBSs) genome-wide.

Results: Our analysis shows that predicted positions of NFRs at the 5' and 3' ends of genes are in agreement with experiment [2]. We explain that positions of NFRs are determined by binding of TFs.

Conclusion: We show that, through competition between nucleosomes and TFs for binding DNA, several of the observed genome-wide patterns of nucleosome positioning are reproduced, even in the absence of any sequence-specificity in nucleosome-DNA interaction. In particular, our analysis suggests that competitive binding of nucleosomes and TFs provokes effect of statistical positioning of nucleosomes around 5' and 3' ends of genes [3].

References:

1. Shivaswamy A, Bhinge A, Zhao Y, Jones S, Hirst M, Iyer V. Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. Plos Biology 2008.
2. Lee W, Tillo D, Bray N, Morse R, Davis R, Hughes T, Nislow C. A high-resolution atlas of nucleosome occupancy in yeast. Nature Genetics 2007.
3. Kornberg R, Stryer L. Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. Nucleic Acids Research 1988.

USE OF HIGH THROUGHPUT SEQUENCING TO OBSERVE GENOME DYNAMICS AT A SINGLE CELL LEVEL

Parkhomchuk[1]* D., Amstislavskiy[1] V.S., Soldatov[1] A., Ogryzko[2] V,
[1]*Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany.*
[2]*CNRS UMR 8126, Institut Gustave Roussy, Villejuif, France*
e-mail: parkhomc@molgen.mpg.de
[*]*Corresponding author*

Key words: *mutations, stochasticity*

Motivation and Aim: With the development of high throughput sequencing technology, it becomes possible to directly analyze mutation distribution in a genome-wide fashion, dissociating mutation rate measurements from the traditional underlying assumptions.

Results: We sequenced several genomes of *Escherichia coli* from colonies obtained after chemical mutagenesis and observed a strikingly nonrandom distribution of the induced mutations. These include long stretches of exclusively G to A or C to T transitions along the genome and orders of magnitude intra- and intergenomic differences in mutation density. Whereas most of these observations can be explained by the known features of enzymatic processes, the others could reflect stochasticity in the molecular processes at the single-cell level.

Conclusion: Our results demonstrate how analysis of the molecular records left in the genomes of the descendants of an individual mutagenized cell allows for genome-scale observations of fixation and segregation of mutations, as well as recombination events, in the single genome of their progenitor.

CONTAINING EDITING DEAMINASES

Y.I. Pavlov[1*] and I.B. Rogozin[2]
[1]*Eppley Institute for Research in Cancer and Department of Biochemistry and Molecular Biology, and Department of Microbiology and Pathology, University of Nebraska Medical Center, Omaha, NE, U.S.A.*
e-mail: ypavlov@unmc.edu
**Corresponding author*
[2]*National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, U.S.A.*

Key words: *editing deaminases, hypermutagenesis, mutation hotspots*

Motivation and Aim. DNA/RNA editing cytosine deaminases of the AID/APOBEC family participate in various cellular processes. The founder of the superfamily, APOBEC1, edits mRNA. AID is required for immunoglobulin gene diversification; APOBEC3G and F (A3G/F) destroy the HIV virus. Functions of other members of the family, APOBEC2, APOBEC4 and APOBEC5, are unknown. AID/APOBEC proteins operate on ssDNA substrates, converting cytosine to uracil, which could trigger downstream DNA repair events leading to death, hypermutagenesis or recombination. Why do they not destroy the genome? Despite extensive studies, little is known about the mechanisms of regulation of action of different AID/APOBEC proteins. We do not know what allows them to act on specific targets only.

Results and Conclusions. Editing of the cognate targets by AID and A3G leads to multiple mutations per Ig locus or HIV virus. Similarly, experiments *in vitro* indicate that deaminases operate processively on ssDNA and perform multiple deaminations. In contrast, when AID/APOBECs are overexpressed ectopically in model organisms, the resulting mutations are predominantly a single base change per reporter gene. We propose that genes, which are the natural targets for AID/A3G in vivo, possess more target DNA motifs and are found in cells in a special environment where certain components protecting the rest of the genome from deaminases are missing. It is plausible that various proteins of DNA metabolism that are associated with ssDNA interfere with deamination *in vivo.* We are studying the factors that can limit the effects of deaminases in yeast.

USE OF MATHEMATICAL MODELLING OF GENETIC NETWORKS FOR THE ANALYSIS OF INFLORESCENCE STRUCTURE EVOLUTION.

A.A. Penin[1]*, P.A. Prudkovskii[2]
[1]Department of Genetics, [2]Department of Quantum Radiophysics, Lomonosov Moscow State University, Moscow, Russia
E-mail: alekseypenin@gmail.com
**Corresponding author*

Key words: *inflorescence evolution, terminal flower, gene net modeling*

Motivation and Aim: Flowering plants are very polymorphic in terms of the structure of flowers and inflorescences. However the genetic mechanisms that determine this diversity are not well studied by now. One of the methods to study them is the mathematical modeling. This approach is based on the construction of simulation models inferred from the data on the genetic control of development. These models are used to conduct computational experiments aimed at revealing the alterations in the genetic network functioning that may cause observed alterations in morphology. In the present work we studied the genetic network controlling flower and inflorescence development and its alterations that underlie the presence or absence of terminal flowers.

Methods and Algorithms: For the purposes of model construction the genes of interest were grouped into several functional classes: those responsible for the transition to flowering, for the maintenance of meristem activity etc. The interaction of these classes of genes was described by six differential equations. Numerical solution of these equations allowed determining conditions leading to the termination of proliferating activity of inflorescence meristem and formation of terminal flower and time required for this termination, depending on the state of the genetic network. Exact values of empirical coefficients in equation were chosen according to experimental data on time of inflorescence terminalization in Arabidopsis wild type and mutants under long and short day.

Results and Conclusion: Analysis of the model demonstrates that the transition between different types of inflorescences can be caused not only by the alterations of functioning of genes controlling transition to flowering but also by the alterations of genes controlling (directly or indirectly) termination of meristem proliferation. It was also shown that the second mechanism is mediated by the unevenness in the activation of genes repressing meristem activity and their negative regulators. These data suggest that the transition between determinate and indeterminate inflorescences may be caused by alterations in the functioning of the genetic network controlling inflorescence development that do not necessarily affect orthologous genes.

Acknowledgements: This work supported by RFBR (№ 09-04-01363-a and 10-04-00859-a) and Federal agency of science and innovations (МК-159.2009.4 and Program for support of leading scientific schools), Federal target program «Scientific and scientific-pedagogic personnel of innovative Russia».

IN SILICO COMPARISON OF THE CONTRIBUTIONS OF TRANSPOSONS AND RETROTRANSPOSONS TO DUPLICATION FORMATION IN THE *y cn bw sp Drosophila melanogaster* GENOME

M.P. Perepelkina* and L.P. Zakharenko
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: zakharlp@bionet.nsc.ru
**Corresponding author*

Key words: *Drosophila melanogaster*, transposable element, duplication, recombination.

Motivation and Aim: Duplications of genome fragments are currently considered to be one of the major evolution factors. They cause polymorphism and genetic instability. They can also cause hereditary diseases in humans via gene dosage changes (Koszul et al.2009). One of the ways by which duplications arise is unequal crossingover of repetitive sequences. Low copy repeats (LCRs) constitute 5% of the human genome. They are detected at the ends of repeats that change gene dosage and cause diseases.

Genomes of practically all eukaryotic species are rich in transposable elements (TEs), which can also be involved in recombinations. It is known, for example, that locus-specific instability in *Drosophila melanogaster* can be caused by recombinations between TE copies. Transposable elements are also present at ends of some species-specific inversions.

Methods and Algorithms: The involvement of various TE classes to the formation of repeats in the genome of *Drosophila melanogaster* stock *y cn bw sp* was studied *in silico*. Nucleotide sequences adjacent to TE copies were compared. Insertions of TEs to the *D. melanogaster* genome were retrieved from www.flybase.org, and sequences adjacent to TEs, from Download sequence region NCBI (*Drosophila melanogaster, Release 5.10)*. Sequence alignment was carried out with Blast 2 Sequences NCBI software.

Results: We considered expressing (EST-positive) TEs, which formed direct repeats at the insertion site. With the assumption that the presence of direct repeats at the insertion site correlates with later TE insertion event, we suggest that the majority of full-length copies of virtually all TEs can migrate in principle. For all TEs under consideration, examples of their involvement in repeat formation were detected. Depending on TE type, 2 to 17% of copies occur within repeats. The sizes of repeats formed by TEs vary from 2 to 32 kb. Both full-length and truncated TE copies can take part in repeat formation. Most TEs occur in repeats as passive elements. In such cases, the repeats are formed by recombination between sequences adjacent to the TEs.

Conclusion: No significant difference was found between transposons (*hopper*, *hobo*, *FB*, *HB*, *1360*, etc.) and retrotransposons (*1731*, *412*, *copia*, *gypsy*, *roo*, *Doc*, *jockey*, etc.) in contribution to repeat formation.

Acknowledgments: We are grateful to Yu. A. Tkachev for designing a program for automated sequence analysis. This work was supported by the Russian Foundation for Basic Research, project 09-04-0213a, the RAS program *Biodiversity*, projects 23.29 and 23.30 and Grant of Carl Zeiss NSU2/11KC.

COMPARATIVE STUDY OF THE PHYLETIC EVOLUTION IN THE SPECIES FLOCK OF BAICALIAN ENDEMIC MOLLUSCS.

T.E. Peretolchina*[1], A.V. Novikov[1,2], T.Ya. Sitnikova[1], D.Yu. Sherbakov[1,2]
[1]*Limnological Institute SB RAS, Irkutsk, Russia*
[2] *Irkutsk State University, Faculty of Biology and Soil Science*
e-mail: tanya@lin.irk.ru
[*]*Corresponding author*

Key words: *Lake Baikal, Baicalia, mtCO1, ITS1, speciation, introgression.*

Motivation and Aim: Patterns of genetic variety of species belonging to species flocks contain evidences of many evolutionary events other then speciation. One of the most spectacular effects of phyletic evolution is the mitochondrial introgression causing dramatic discrepancy between phylogenies inferred from nuclear and mitochondrial sequences. It is generally believed to be the result of secondary gene flows between sister species differing dramatically in population numbers. In this study we examine the relationship between climatic variation, geographic isolation, and speciation in genus *Baicalia*.

Methods and Algorithms: Mollusks were collected by the dredge from the Lake Baikal at a depth of 10-40 m. Genomic DNA was extracted by modified methods of Sokolov from muscle tissue. Sequences were obtained by DNA analysis system CEQ 8800 (Beckman Coulter Inc) and in sequencing center of Novosibirsk. Phylogenetic reconstructions were made using MrBayes v. 3.1.2 and PHYML. The isolation-with-migration model, as implemented in ima2, was applied to estimate the prevalent direction of gene flow between the species and/or populations of *Baicalia*.

Results: In this study we have used nucleotide sequences of nuclear (internal transcribed spacer 1) and mitochondrial (cytochrome *c* oxidase subunit 1) markers in order to elucidate geographical and ecological details of speciation and phyletic processes in five species currently comprising genus *Baicalia*: *B. carinata, B. turriformis, B. carinatocostata*, *B. dybowskiana* and *B. rugosa,* and correlate them to major transformations of Baikalian ecosystem. We have shown that these changes could have caused speciation events and mitochondrial transgression, where the main mechanism was the change of reproductive behavior of the snails and sometimes shifts in substrate preferences. Different substrate preferences were caused by major climate and tectonic changes at the lake resulting in different availability of niches.

Conclusion: During the evolution of the genus, there were three cases of complete mitochondrial transgressions from *B. carinata* to other species. The transgressions were not synchronous and could due to secondary contacts between the more abundant *B. carinata* with her rarer sister species. For different pairs of species this have happened several times during the period between 1,8 - 0,2 MYA when rapid tectonic changes could likely reshuffle dramatically the biotopes and change relative selective advantages of reproductive strategies among stone and sand dwelling species. Only one speciation event (*B. rugosa*) happened during this time span.

Acknowledgements: RFBR 09-04-00972 “The Study of Evolutionary Mechanisms Generating Organisms’ Communities”

SEQUENCE OPTIMIZATION OF PROTEIN α-HELICES - A NEW METHOD FOR BIOENGINEERING OF THERMOSTABLE ENZYMES

M.A. Surzhik[1], A.E. Schmidt[2], A.V. Shvetsov[1], A.P. Yakimov[1], T.N. Kozhina[1], D.L. Firsov[3], E.A. Glazunov[1], M.G. Petukhov[1,2*]

[1] Petersburg Nuclear Physics Institute RAS, Gatchina, Leningrad district, Russia
[2] St.-Petersburg State Polytechnic University, St.-Petersburg, Russia
[3] University of Lausanne, Lausanne, Switzerland
e-mail: pmg@omrb.pnpi.spb.ru
** Corresponding author*

Key words: α-*helix, sequence optimization, tunneling algorithm, protein thermostability*

Motivation and Aim: Rational design of peptide and protein helices is not only of practical importance but also may provide a better understanding of protein folding. Modern theoretical models of helix-coil transitions in monomeric peptides allow accurate predictions of its helix stability. This provides a good basis for design of peptide and proteins with elevated stability. Here we report development of a new method of global sequence optimization of protein α-helices (SEQOPT) [1] and its first application for thermostabilization of fungal Glucoamylase, the one of most widely used industrial enzyme [2].

Methods and Algorithms: SEQOPT is based on AGADIR, the statistical mechanical model for helix-coil transitions in peptides and the Tunneling algorithm of global optimization. We designed short synthetic peptides with optimized sequences and investigated its α-helix stability by far UV CD-spectroscopy. Also we designed amino acid replacements in α-helix D of Glucoamylase and biochemically investigated its effects on the protein thermostability.

Results: The most stable peptide helices were found to consist of mainly Leu, Met, Trp, Tyr, Glu and Arg residues with good intrinsic helical propensities and high potential for other stabilizing interactions. The insoluble poly-L and poly-W motifs are among the most stable sequences indicating that multitude of hydrophobic interactions fully compensates loss of other helix stabilizing factors. There are also many soluble hetero-polymer sequences that are just a bit less stable than the homopolymer sequences. CD measurements showed that maximum achievable α-helix content in 10-13 residue peptides with fully optimized sequences at 5°C is expected to be ~70-75%. The experiments also showed a complex pattern of alternate influence of point mutations in α-helix D of Glucoamylase on thermoinactivation rate constants of the enzyme.

Conclusion: The results show that the helical potential of amino acids is enough to allow stable α-helices as short as of 10 residues in length. This effect also provides a valuable reserve of conformational stability in globular proteins.

Availability: SEQOPT is accessible at http://biod.pnpi.spb.ru/mm/seqopt/

References:

1. M. Petukhov et al. (2009) Design of stable α-helices using global sequence optimization, *J. Pept. Sci.*, **15**: 359-365.
2. М.А. Суржик et al. (2010) Влияние точечных аминокислотных замен во внутренней α-спирали на термостабильность глюкоамилазы из Aspergillus awamori x100, *Прикладная биохимия и микробиология*, **46**: 1-6.

POLYMORPHISM OF MICROSATELLITE REPEATS AND THEIR RELATION WITH THE MECHANISMS OF GENE TRANSCRIPTION REGULATION

A.V. Pheophilov, V.I. Glazko
Russian State Agrarian University – Moscow Agricultural Academy named after K.A. Timiryazev, Moscow, Russia
E-mail: foton87@yahoo.com, vglazko@yahoo.com

Keywords: *inverted repeats, ISSR, purine/pyrimidine tracks, transcription regulation*

Motivation and Aim: Control and management over genetic resources of wild and domestic animals and plants is determined by availability of testing systems which allow reliable identification of specific traits of genetic structures of species, breed and varieties. One direction of such systems' creation is an estimation of polymorphism of DNA inverted repeats' positioning (ISSR-PCR).

Methods and Algorithms: In purposes of estimation of their informational value the comparative analysis of DNA of several cattle, sheep breeds, bisons and aurochs by using of micro satellite loci fragments $(GA)_9C$, $(AG)_9C$, $(AC)_9T$ $(ACC)_6G$, $(GAG)_6C$, $(AGC)_6C$, $(AGC)_6G$, $(CTC)_6C$ as primers in PCR was carried out. Data about species and breed specific features of combinations of DNA fragments of various length (amplicons) were obtained.

Results: Most of them were revealed by using as the primers in PCR of purine/pyrimidine tracks which could take part in triplex DNA structure formation and gene expression regulation [1]. Amplicon spectra received by using two of such tracks $(AG)_9C$ and $(GA)_9C$ on DNA of cattle and sheep breeds showed more polymorphous DNA fragments for $(AG)_9C$ in comparison with $(GA)_9C$ ones. Genbank search with using BLAST algorithms allowed to localize these tracks in genes of II class MHC, moreover $(GA)_9C$ was observed both in sheep and cattle exons, that could be lead to relative higher conservatism of amplicon length in $(GA)_9C$ spectra. At the same time GAGAG sequence is the consensus one for the GAF transcription regulation factor [2, 3], and $(GA)_n$ repeats are the targets of CTCF binding which plays the key role in chromatin folding, changes of which related with the replacing of the gene expression programs [4]. Data obtaining testified the dependence between length polymorphism of DNA fragments flanked by inverted repeats of microsatellite loci and nucleotide sequence of such locus and higher levels of polymorphism of those amplicon spectra which flanks belong to purine/pyrimidine tracks, among which it's lower for GA containing sequences. Conservatism of the latter may be caused by involving of GA repeats into binding with the proteins, which take part in regulation of gene transcription programs, in particular by means of influence on chromatin folding.

Conclusion: Using of such fragments for polyloci genotyping of animal genome DNA allows reveal the combinations of DNA fragments with the different lengths which may be considered as gene pool "barcode" for species, breeds and individual animal.

References:

1. F.A.Rogers, J.A.Lloyd, P.M.Glazer. (2005) Triplex-forming oligonucleotides as potential tools for modulation of gene expression, *Curr Med Chem Anticancer Agents,* **Jul 5(4):319**-26.
2. N.L.Adkins, T.A.Hagerman, P.Georgel. (2006) GAGA protein: a multi-faceted transcription factor, *Biochem. Cell Biol,* **84**: 559-567;
3. 3. E.B.Tchoubrieva, J.B.Gibson. (2004) Conserved $(CT)_n$·$(GA)_n$ repeats in the non-coding regions at the Gpdh locus are binding sites for the GAGA factor in Drosophila melanogaster and its sibling species, *Genetica,* **121:** 55-63;
4. 4. A.Ottaviani, C.Schluth-Bolard, S.Rival-Gervier, A.Boussouar, D.Rondier, A.M.Foerster, J.Morere, S.Bauwens, S.Gazzo, E.Callet-Bauchu, E.Gilson, F.Magdinier. (2009) Identification of a perinuclear positioning element in human subtelomeres that requires A-type lamins and CTCF, *EMBO J.* **28, 16**: 2428-2436.

THE SUBSTITUTIONS G245C AND G245D IN THE ZINC BINDING POCKET OF THE p53 PROTEIN RESULT IN DIFFERENCES OF CONFORMATIONAL FLEXIBILITY OF THE DNA BINDING DOMAIN

S.S. Pintus*, V.A. Ivanisenko
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: pintus@bionet.nsc.ru
**Corresponding author*

Key words: *p53 protein, molecular dynamics, conformational flexibility, DNA binding domain (DBD), binding site*

Motivation and Aim:

P53 is a tumour suppressor protein, that activates transcription of a number of its target genes to inhibit the cell growth or to induce apoptosis. Binding of the p53 subunit to the target DNA depends on the zinc binding by the allosteric functional site. Some mutations of p53 result in decrease of range of its target genes. Two of such mutations, G245C and G245D reside in the zinc binding pocket of the DNA binding domain. Our aim was to test whether these mutations affect conformational flexibility of the DNA binding domain.

Methods and Algorithms:

Our method is based on multiple molecular dynamics simulations of complexes of tertiary structures of the mutant forms of the DNA binding domain using package Gromacs [1]. The initial coordinates for complexes are obtained from homology based modeling and search of zinc binding sites as in earlier work of Ivanisenko's group [2]. The resulting trajectories are truncated at the equilibrium state and subject to further statistical analysis. The Z statistics criterion is used to find C_α atoms of the mutant forms, which fluctuations are significantly different from the wild type.

Results:

The G245C and G245D substitutions differently affect conformational flexibility of the DNA binding domain[3]. The C-terminal α-helix that is known to bind the major groove of the target DNA [3] is mostly affected by G245C substitution, while the L3 loop that binds the minor groove of the target DNA is mostly affected by G245D substitution.

Conclusion: (text)

The substitutions in the zinc binding pocket of p53 have allosteric effect on the DNA binding site that may explain their effect on the protein function.

Acknowledgements

The work is supported by integration projects of SB RAS nn. 1, 26, 94, 111, 113 and 119, State contract FASI №02.514.11.4123 and Program RAS n. 22 (Molecular and cellular biology.)

References:

1. D. Van Der Spoel et al (2005) GROMACS: fast, flexible, and free, *J Comput Chem*, **26**: 1701-1718.
2. V. Ivanisenko et al (2005) PDBSite: a database of the 3D structure of protein functional sites, *Nucleic Acids Res,* **33:** W99-W104.
3. Y. Cho et al (1994) Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations, *Science*, **265:** 346-355.

INVESTIGATING THE TRANSCRIPTION OF LPS-STIMULATED MACROPHAGES AND ITS CORRELATION WITH CHROMATIN STRUCTURE

Mohammad Pirmoradian*, Olof Emanuelsson
Biotechnology, The Royal Institute of Technology, Sweden
**Corresponding author*

Key words: *Macrophage inflammatory response, microarray, ChIP-Seq profiling, LPS*

Motivation and Aim: Macrophage inflammatory response requires an orchestrated expression of specific genes to initiate inflammation and eliminate pathogens. Our aim is to correlate the macrophage-specific transcription in response to lipopolysaccharide (LPS) with chromatin structure.

Methods and Algorithms: We use published genome- level expression pattern data (microarray), together with knowledge of the transcription factors encoded by the mammalian genome, and the promoter regions. To this we add the results from ChIP-Seq profiling related to the chromatin structure (histone acetylation) of the human macrophage cell line THP-1 (with and without LPS).

Results: We perform functional analyses of the up and down regulated genes for relevant pathways and gene networks, followed by molecular functions and biological processes.

THE NANOBIOTECHNOLOGY DATABASE

V.A. Ivanisenko[1], N.L. Podkolodnyy*[1,2], P.S. Demenkov[1], T.V. Ivanisenko[1], N.N. Podkolodnaya[1], T.M. Khlebodarova[1], E.V. Ignatieva[1], O.A. Podkolodnaya[1], E.A. Ananko[1], and N.A. Kolchanov[1,3]

[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*
[3]*Novosibirsk State University, Novosibirsk, Russia*
e-mail: pnl@bionet.nsc.ru
**Corresponding author*

Key words: *nanobiotechnology, database, nanomaterials, nanomedicine, nanotoxicity*

Motivation and Aim. Nanomaterials are commonly defined as materials designed and produced to have structural features with at least one dimension of 100 nanometers (nm) or less. The nanosize-specificity bioactivity of nanomaterials was produced by changes of chemical and physical properties of nanoobjects. The huge number of publications on nanobiotechnology contains information important for future research efficiency but hard-to-reach. Development of database on nanobiotechnology with formalized information for searching and using in research is bioinformatics challenge.

Methods and Algorithms. Nanomaterials knowledge via the MEDLINE/PubMed and Inspec databases was extracted by text-mining technology and used for databases on nanobiotechnology development.

Results.

The controlled vocabularies for description of objects and different relations in the field of nanobiotechnologies were developed with text mining technology. About 20 000 terms using in scientific papers in the field of nanobiotechnology were found. The patents database on nanobiotechnology (more 50 000 records), database on nanomaterials using in nanobiotechnology and biomedicine (more 10 000 records), database on methods and technologies important for nanomaterials engineering, measuring and applications in the field of nanobiotechnology, biomedicine and nanobioengineering (more 1000 records) and database on nanoproducts (more 2000 records) were developed. Web interface for information search were created.

Conclusions. The databases on nanobiotechnology were developed. The databases contains information on nanomaterials using in nanobiotechnology, the nanomaterials properties important for biological activity, nanotechnologies using for nanomaterials engineering and nanoproducts using in biomedicine, pharmacology and bioengineering.

Acknowledgements: This research was supported by the Russian Ministry of Education and Science (contract № 01.647.11.2013).

COMPUTER SYSTEM FOR ANALYSIS OF MECHANISMS OF TRANSCRIPTION REGULATION IN EUKARYOTES

N.L. Podkolodnyy[1,2]*, E.V. Ignatieva[1], D.A. Rasskazov[1], E.A. Ananko[1], O.A. Podkolodnaya[1], N.N. Podkolodnaya[1], E.M. Zalevsky[1]

1Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

[2]Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia

e-mail: pnl@bionet.nsc.ru

[*]Corresponding author

Key words: *transcription regulation, database, data integration*

Motivation and Aim: In multicellular eukaryotic organisms, the transcriptional activity of a particular gene is dependent of the type of cell, organ, or tissue, developmental stage of the organism, cell cycle, or cell differentiation, also on numerous inducers or repressors, etc. This fine and complex regulation is provided by the involvement of a great variety of regulatory proteins (general transcription factors, other transcription factors, mediator proteins and co-regulators which include chromatin-remodeling and chromatin-modifying activities) and mechanisms underlying their functioning. The aim of this work is to develop the RETRA system for analysis the mechanisms of tissue-specific transcription regulation in three eukaryotic species: human, mouse and rat.

Methods and Algorithms: RETRA system was developed through the integration of data from TRRD, EntrezGene, UCSC genome browser, RefSeq, GNF SymAtlas v1.2.4., Swiss-Prot, Gene Ontology. Knowledge on chromatin modifications effecting gene transcription activity includes description of DNA or proteins modification types, their positions, and their effect on gene activity.

Results: RETRA system stores data supporting analysis of mechanisms of transcription regulation: i) the localization of human, mouse and rat genes in genomic sequences (~$25*10^3$, ~$27*10^3$, and ~$27*10^3$ genes respectively); ii) The expression levels of genes in different cells, tissues, organs (79 human, 61 mouse, 49 rat), estimated on the basis of analysis of microarray data (more than $6*10^7$ entries). RETRA system provides the possibility to find genes with similar expression patterns and to get nucleotide sequences of the promoter regions.

Conclusion: The RETRA system for resolve the following issues was developed: i) search of sets of co-expressed genes; ii) search of organizational modules of the promoters of co-expressed genes; iii) reconstruction transcription regulatory networks, and hereafter iv) reconstruction of hypothetic mechanisms of gene transcription regulation taking into account the knowledge on structures of protein regulators expressing in the specific cells situations, the various roles of protein regulators in transcription regulation, structures of co-expressed gene promoters, and etc.

Acknowledgements: This research was supported by the Russian Ministry of Education and Science (contract № P721), Interdisciplinary integration project of SB RAS № 119, Scientific School 2447.2008.4.

DE-NOVO PREDICTION OF TRANSMEMBRANE HELICAL DIMERS FOR BITOPIC PROTEINS

*A.A. Polyansky, P.E. Volynsky, R.G. Efremov.
M.M. Shemyakin and Yu.A. Ovchinnikov Institute of Bioorganic Chemistry Russian Academy of Sciences, Moscow, 117997, Russia
e-mail: newant@gmail.com
** Corresponding author*

Key words: *membrane proteins, protein oligomerization, molecular hydrophobicity potential, analysis of molecular surfaces, drug design*

Motivation and Aim: Oligomerization of membrane proteins (MB) represents a basis of various functions in the living cell (e.g. reception of extracellular signals). Transmembrane (TM) fragments of MB, usually presented as a bundle of α-helices (polytopic proteins) or just as a single helix (bitopic proteins), play an important role in such a process. Single point mutations in the TM region of bitopic proteins (e.g. receptor tyrosine kinases) can modulate an efficiency of the dimerization and thus lead to dramatic changes of their biological function, including development of oncologic disorders. An efficient prediction of the spatial structure of TM helical dimers provide this way unique opportunities to targeted modification of MB activity and design potential therapeutic agents [1]. Here we present a novel approach for *de-novo* reconstructions of 3D models for dimers of TM helices from their primary structures.

Methods and Algorithms: The proposed algorithm includes several consequence stages: (i) building "ideal" helices from TM-sequence, (ii) calculation of their surface and mapping hydrophobic / hydrophilic properties using conception of the molecular hydrophobicity potential (MHP), (iv) identification of the dimerization interface evaluating complementary of helical landscapes and MHP properties, (v) reconstruction of 3D model of the dimer and estimation of its quality according to scoring functions.

Results: The approach was tested using database of dimeric structures of bitopic proteins. Most of these 3D structures were experimentally obtained in Laboratory of biomolecular NMR, IBCH RAS. Results of prediction display high efficiency and accuracy. RMSD between predicted and experimental structures in all case do not exceed 2 E by backbone atoms. The effect of single point mutations for several proteins (e.g. glycophorin A) was also investigated. Thus, silent mutations slightly affect the dimer structure, whereas critical ones result in formation of structures with weak packing properties.

Conclusion: The algorithm for *de-novo* predictions TM helical dimers was elaborated and successfully tested using a set of know structures and mutagenesis data. Principles underlying the prediction approach can be proposed as fundamental basis of oligomerization of protein helices in biological membranes.

Availability: The prediction algorithm is implemented as a package of utilities and currently employed to design new peptides with predefined selectivity to TM fragments of several bitopic proteins (e.g. ephrin receptors).

References:

1. H. Yin et al. (2007) Computational design of peptides that target transmembrane helices. Science, **315:**1817-22.

OLIGOMERIZATION OF THE SEROTONIN RECEPTORS: BIOLOGICAL PROOF OF COMPUTATIONAL MODELS

Ponimaskin Evgeni*[1], Niv Masha[2], Kowalsman Noga[2], Renner Ute[3]
*Corresponding author
1. Cellular Neurophysiology, Medical School Hannover, Germany
2. Hebrew University of Jerusalem, Israel
3. DFG Research Centre Molecular Physiology of Brain, Goettingen, Germany

Key words: *G-protein coupled receptor, oligomerization*

Motivation and Aim: Serotonin (5-hydroxytryptamine or 5-HT) is an important neurotransmitter involved in regulation of anger, aggression, body temperature, mood and sleep. Serotonin receptors belong to the G-protein coupled receptors (GPCRs) family. These seven transmembrane proteins are involved in many signal transduction pathways and are important targets for drug development. In the recent years it became clear that many GPCRs act as oligomers, but the role of the oligomerizastion and its mechanisms are still not clear. We have recently demonstrated oligomerization of 5-HT1A receptors. To elucidate the structural and functional role of the 5-HT1A receptor oligomerization we aimed to design a computational model of receptor oligomers. This model was used to construct receptor mutants with impaired ability to interact.

Methods and Results: We modelled the structure of the 5HT1A serotonin receptor monomer and used it to predict amino acid residues that may participate in the formation of the 5HT1A homo- and hetero-dimers. Based on our prediction of the interaction interface, we performed the site-directed mutagenesis of defined amino acid residues within the transmembrane helices 1, 4 and 5. The mutated receptors were then tested for their ability to disrupt the dimer formation. For that, we applied different biochemical methods (co-immunoprecipitation and native protein gel electrophoresis) in combination with biophysical approaches, including Foerster resonance energy transfer (FRET).

Conclusion: Our results demonstrate that computational modeling of the interaction interface represents the powerful methods allowing the analysis of the molecular mechanisms involved in formation of the receptor-receptor dimers. In addition, site-directed mutagenesis of residues predicted to be involved in interaction can be used to study the functional consequences of receptor oligomerization.

Acknowledgments: Funding support: DFG grants: Po732 and CMPB. Volkswagen Stiftung.

TRANSCRIPTIONAL NETWORKS IN BRAINS OF ALCOHOLIC AND NONALCOHOLIC INDIVIDUALS.

I. Ponomarev*, S. Wang, L. Zhang, A. Harris, D. Mayfield
piatut@mail.utexas.edu
University of Texas at Austin, Austin, Texas, USA
** Corresponding author*

Key words: *Microarrays, Gene coexpression analysis, Human brain, Alcoholism*

Motivation and Aim: Transcriptional network analysis can be used to understand molecular and cellular mechanisms underlying normal function and pathology in complex tissues. To identify key genes and molecular pathways contributing to alcoholism, we profiled gene expression levels of 33 postmortem human brains (3 regions of the amygdala and superior frontal cortex; 16 controls and 17 alcoholic cases) using Illumina microarrays and a weighted gene coexpression network analysis (WGCNA: Zhang and Horvath, 2005).

Methods and Algorithms: We first explored coexpression relationships among all genes that passed the detection threshold, using WGCNA and identified modules of genes that were highly correlated (coexpressed) across all samples. We then identified alcohol-related modules, i.e., modules significantly enriched for genes differentially expressed between alcoholics and controls. Finally, we used DAVID database (http://david.abcc.ncifcrf.gov) and functional over-representation analysis to determine organizing principles of alcohol-related modules.

Results: In general, genes were clustered into modules according to their involvement in different physiological processes, functional groups, chromosomal locations and cell types. Several alcohol-responsive modules were similar across the brain regions, while others were unique for a particular region, indicating both common and brain region-specific alcohol-related changes. Modules enriched with cell type-specific genes were mainly region-specific, while examples of common changes included genes located on chromosome 19 and genes involved in immune/inflammatory responses.

Conclusions: Our transcriptional network analysis integrated multiple levels of data and connected molecular pathways to nervous system function and brain pathology. This approach provides contextual information about biological function that may not be available when examining single genes and assists in formulation of new refined hypotheses.

Acknowledgments: Funding support: NIH, NIAAA grants: AA012404 to AH, INIA grants (AA013518 to AH, AA016648 to DM, AA013517 Pilot Projects to DM and IP, AA013476 Subcontract to IP).

Reference:

1. B. Zhang, S. Horvath (2005), A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol,* 4:Article17, PMID: 16646834 [PubMed].

MOLECULAR DYNAMICS MODELLING BECOMES USER-FRIENDLY WITH GUI-BIOPASED

A.V. Popov, Y.N. Vorobjev*
SB RAS Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, Russia
**Corresponding author ynvorob@niboch.nsc.ru*

Key words: *molecular modeling, molecular dynamics, graphical interface.*

Motivation and Aim: Work with software packages of the molecular dynamics method is quite complicated, especially for non-experts in computational structural biology. One of conditions of program usage by experimentalists is possibility of fast access and formation of a task at intuitively clear level using the graphic user interface (GUI). Error-free files necessary for the computational experiment can be formed on the graphic user interface that performs analysis, error check and correction. Web version entitled GUI-BioPASED is developed for BioPASED, the program of a molecular dynamics method.

Methods and Algorithms: The program is designed to work with three-dimensional structures of proteins and nucleic acids for the purpose of the structure analysis, optimization of conformational energy, statistical analysis of structure dynamics and free energy calculation of conformation of the biopolymer. The classical molecular mechanics method with AMBER force field is used. The concept of the cross-platform web application is a basis for GUI-BioPASED program.

Results: GUI-BioPASED allows one to create control files for the computational module of BioPASED at intuitively clear level, download them and execute on the client side computer. Automatic control file creation allows one to avoid errors and thus to simplify the process of adjustment of parameters for computations. PDB-file validation and correction allows one to create a valid file that can be successfully read by the BioPASED program. GUI-BioPASED is supported by all common web browsers.

Conclusion: The developed program automates a creation of correct input information for molecular modeling by molecular dynamics method. New user-friendly interface performs data check, reports errors and hints at possible ways to correct them.

Availability: GUI-BioPASED program is available by the Internet [1]. The interface of the program supports Russian and English languages.

Acknowledgements: This research was supported by RFBR (09-04-00136, 08-04-00596), SB RAS integration projects (No. 16, No. 119, 2009-2011) and Russian Ministry of Education and Science (02.740.11.0079/NS-652.2008.4).

References:

1. GUI-BioPASED program. http://biopased.niboch.nsc.ru/

COMPUTER-AIDED APPROACHES TO VIRTUAL SCREENING AND RATIONAL DESIGN OF MULTITARGETED DRUGS

V.V. Poroikov*, O.A. Filz, O.N. Koborova, A.A. Lagunin, T.A. Gloriozova, D.A. Filimonov

Institute of Biomedical Chemistry of Rus. Acad. Med. Sci., Moscow, Russia
e-mail: vladimir.poroikov@ibmc.msk.ru
**Corresponding author*

Key words: *multitargeted drugs, synergistic/additive/antagonistic action, computer-aided prediction, virtual screening, rational design, natural products evaluation*

Motivation and Aim: Many diseases have a complex etiology, which treatment often requires multiple actions on several pharmacological targets. On the contrary, the majority of current drugs were designed to interact with a single target, which sometimes leads to activation/blockade of other elements in the appropriate signal regulatory pathway. As a consequence of negative feedbacks, expected pharmacotherapeutic effect may be significantly decreased or even completely suppressed. Therefore, the multitargeted drugs, due to their additive, synergistic or antagonistic action, might have some advantages comparing to the monotargeted medicines. The purpose of our study was to develop computer-assisted methods for identification of the most promising targets; finding and rational design of multitargeted agents with the required biological activity profiles.

Methods and Algorithms: NetFlowEx – software for simulation of behavior of signal regulatory pathways and identification of the most promising targets and their combinations. PASS (Prediction of Activity Spectra for Substances) - software, which predicts about 4000 kinds of biological activity on the basis of structural formula with mean accuracy about 95%. PharmaExpert – software for analysis of PASS predicted biological activity spectra and selection of compounds with the required biological activity profiles. PASS Constructor – software for design of drug-like molecules with the required profiles of biological activity.

Results: By analysis of signal regulatory pathways the promising targets for treatment of breast cancers were identified. Based on computer prediction of biological activity for 24 mln chemical compounds 64 molecules were selected for experimental testing. 26 samples were purchased and antineoplastic activity was confirmed experimentally in some molecules. Application of PASS & PharmaExpert to the analysis of biological activity of phytoconstituents demonstrated that computer prediction provides the scientific explanations for traditional usage of these preparations and reveals the hidden unexploited potential of medicinal plants.

Availability: PASS predictions are freely available at http://www.ibmc.msk.ru/PASS

Conclusion: Computer-aided methods are rather useful in design and discovery of the most prospective pharmacological targets and their ligands.

Acknowledgements. This work was partially supported by European Commission FP6 grant LSHB-CT-2007-037590 (Net2Drug), FP7 grant 200787 (OpenTox), ISTC grants 3197 and 3777, RFBR grants 05-07-90123 and 06-03-08077.

MODELING AND ANALYSIS OF PROTEIN CONFORMATIONS OF TICK-BORNE ENCEPHALITIS VIRUS WITH DIFFERENT PATHOGENICITY FOR HUMANS

U.V. Potapova, E.V. Romanova, S.I. Belikov
Limnological Institute SB RAS, Irkutsk, Russia
e-mail: shuana1983@yandex.ru

Key words: *tick-borne encephalitis virus, protein, mutation, 3D structure*

Motivation and Aim: Tick borne encephalitis (TBE) is the most important transmissible natural foci infection of the forest zone of the Eurasian continent. The TBE incidence is recorded in more than 25 countries of Europe and 7countries of Asia. In the Russian Federation (RF) the disease is recorded in 46 administrative territories, including 18 territories of Siberia and the Far East [1]. The research is based on previously obtained data on the mutations in complete genomes of tick borne encephalitis virus strains (TBEV), isolated from patients with varying severity of disease. Three mutations previously identified led to substantial amino acid changes, which probably could affect the infection process severity. It is suggested that two associated mutations, deletion of amino acid 111 in the capsid protein C and substitution (Ser1534 - Phe) in the NS3 protein influence strictly coordinated polyprotein processing, disturbing correct arrangement of viral particles. This process can result in the development of defect viral particles, containing no RNA. Mutation (Ser917 - Gly) in nonstructural protein NS1 results in the substitution of hydrophilic amino acid, specific to highly virulent strains, by the hydrophobic one. This could influence the effectiveness of viral replication complex, thereby affecting the infectivity of tick borne encephalitis virus strains. To confirm this hypothesis, we need to construct for strains with different virulence of a series of 3D structures of proteins TBEV.

Methods and Algorithms: To reach this aim, we constructed 3D structures of NS1, NS2B-NS3 protease, NS5 in which mutations are important, by used a computer simulation method. We constructed 3D models for proteins NS1, NS5 by using I-TASSER web-service [2]. Homology modeling was performed using the JACKAL. JACKAL is a package for protein structure modeling [3]. The homology model of NS2B-NS3 protease was constructed by using the template structure NS2B and NS3 protease West Nile virus. A comparative analysis was performed of conformational changes in proteins of various strains with the influence of major mutations.

Results: We have constructed 3D models for proteins NS1, NS2B-NS3 protease, NS5 TBEV strains with different virulence and pathogenicity for humans.

Conclusion: Analysis of the 3D models showed significant differences in the conformations of proteins for unapparent strains and the high virulence strains TBEV. In addition, analysis of the protein complex NS2B-NS3 protease reveals a strong influence substitution Ser-Phe on the formation of protein complex and compactness of the protein globule of the NS2B and NS3 protease.

Availability: The practical value of work is the possibility of early prediction of severity of illness among the population during seasonal outbreaks of infection, as well as mapping of the distribution centers of high-risk variants of the virus to different regions of the country. In the future, finding the relationship between nucleotide sequence and pathogenicity of tick-borne encephalitis virus makes it possible to develop a mechanism of molecular blocking the virus and as a consequence of molecular drug disign.

Acknowledgements (I.G. Kondratov, E.V. Pavlenko, G.N. Leonova)

References:

1. G.G.Onishchenko. (2003) Epidemiological Situation and the Strategy of Struggle against Tick Borne Encephalitis at the Present Stage, in Kleshchevoi i drugie virusnye entsefality (Tick Borne and Other Virus Encephalitises), *Ross. Akad. Med. Nauk*, 5- 8.
2. Y.Zhang. (2009) Protein structure prediction: when is it useful?, *Structural Biology*, **19:** 145-155
3. D.Petrey, Z.Xiang, B.Honig et al. (2003) Using multiple structure alignments, fast homology building, and energetic analysis in fold recognition and homology modeling, *Proteins: Struc., Func. and Genet.* **53:** 430-435.

VSDOCKER TOOL: USING AUTODOCK 4 ON WINDOWS-BASED COMPUTER CLUSTERS FOR VIRTUAL SCREENING

N.D. Prakhov[1,2]*, A.L. Chernorudskiy[2], M.R. Gainullin[2,3]
[1]Faculty of Biology, Nizhny Novgorod State University, Nizhny Novgorod, Russia; [2]Postgenomic Technology Group, Institute of Applied and Fundamental Medicine, Nizhny Novgorod, Russia; [3]IBILI, Faculty of Medicine, University of Coimbra, Azinhaga de Sta Comba, Coimbra, Portugal.
e-mail: nikita.prakhov@gmail.com
**Corresponding author*

Key words: *docking, virtual screening, software*

Motivation and Aim: Virtual screening is based on molecular docking which currently plays an essential role in drug design. AutoDock4 is one of the most widely used instruments for molecular docking [1], but it is not adapted for parallel working. An example of AutoDock 4 adaptation for Linux system is DOVIS. However, there was no similar utility for MS Windows systems to date. To address this problem we have developed a new software tool termed VSDocker 2.0 [2].

Methods and Algorithms: VSDocker 2.0 implements a special algorithm to use computer cluster resources with a maximum efficiency. This algorithm loads all cluster components equally during the whole program operation time and includes several steps: 1) Creation of receptor maps; 2) Unpacking of ligands from multi-ligand containing *.mol2 format into separate files; 3) Converting of *.mol2 to *.pdbqt; 4) Making of separate directories for the work of AutoDock; 5) Docking; 6) Analysis of results. Each of these steps has its own requirements to the system resources. Our algorithm executes steps 2, 3, 4 and 5 in parallel. Steps 1 and 6 are executed before and after main procedures respectively and do not take much time.

Results: VSDocker 2.0 was used for virtual screening of ligands for a human protein Mms2 (PDB ID: 1J7D). As an area of docking we chose Mms2 surface region which is known to interact with a partner protein Ubc13, also present in 1J7D structure. A total of 86775 drug-like ligands from ZINC database was analyzed. We achieved an average throughput of about 420 ligands/CPU/day, which is comparable to the performance of DOVIS in the analogous test. We are further going to verify the obtained results using a surface plasmon resonance method.

Conclusion: VSDocker works both on multiprocessor computing clusters and workstations operated by MPICH2 or MSMPI. It may be very helpful for variety of research tasks based on virtual screening, especially for time consuming applications.

Availability: VSDocker 2.0 is available at http://bio.nnov.ru/projects/vsdocker2.

Acknowledgements: This work was partially supported by RFBR grant № 08-04-01816.

References:

1. G.M.Morris et al. (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility, *J. Comput. Chem.,* **30**: 2785-2791.
2. N.D.Prakhov, A.L.Chernorudskiy, M.R.Gainullin. (2010) VSDocker: a tool for parallel high-throughput virtual screening using AutoDock on Windows-based computer clusters, *Bioinformatics,* (in press).

MECHANISMS OF STRUCTURE FUNCTIONAL RECONSTRUCTION OF SYNAPTIC STRENGTH CHANGING

A.L. Proskura, T.A. Zapara., I.I. Zasypkina., A.S. Ratushnyak*
Design Technological Institute of Digital Techniques SB RAS
e-mail: ratush@kti.nsc.ru
**Corresponding author*

Key words: *LTP, LTD, receptor recycling, dendritic spine, synaptic plasticity*

Motivation and Aim: Lasting increase of synaptic-transmission efficacy and similar plasticity with the opposite effect on the synaptic efficacy induced by correlations in neural activity were discovered in the rat's brain and named long-term potentiation (LTP) and long-term depression (LTD) respectively. Some pre and postsynaptic mechanisms involved in the induction, expression and maintenance of LTP/LTD have already been revealed. New experimental data relating to molecular mechanisms of storing and reproduction of information in a single neuron also require as further experimental investigations so the profound theoretical analysis of available data which itself is one of the important task of bioneuroinformatics. On the base of integration and analysis of various experimental data we are developing the model of subcellular processes, providing the link between structural and functional synaptic plasticity.

Methods and Algorithms: To reconstruct the model of molecular net the GeneNet technology was used. In the context of this investigation we were estimating the temporal dynamics of the glutamate receptor concentrations in dendritic spines.

Results: Using GeneNet technology we reconstructed the graphic model reflecting events relating to synaptic strength changing: the lateral diffusion between synaptic and extrasynaptic compartments, intracellular trafficking and glutamate receptors recycling in dendritic spines. On the base of analysis of experimental data the temporal dynamics of glutamate receptor concentration provoked by LTP or LTD-inducing stimuli has been estimated.

Conclusion: At the present stage a special attention was paid to the short-term processes of changing the efficiency of synaptic strength (up to 30 min). Our graphical model of subcellular processes reflects the basic events of rearrangement of postsynaptic density: input of calcium ion in cell, redistribution of glutamate receptors within the dendritic spines. We examined in detail the processes of lateral diffusion between synaptic and extrasynaptic compartments, exo/endocytosis, intracellular trafficking of endosomes with glutamate receptor by actin-based myosin motors. The regulatory loop provided by BDNF and Arc proteins is of interest. Later on the development of the model at the expense of the processes taking part in the long-term synaptic strength changing is planned.

SYSTEMS BIOLOGY APPROACH TO IDENTIFICATION OF BIOMARKERS FOR METASTATIC PROGRESSION IN CANCER

Andrey Ptitsyn[1*], Douglas Thamm[2]
Colorado State University
[1]Center for Bioinformatics, Department of Microbiology, Immunology and Pathology,
[2] Animal Cancer Center
[*]corresponding author
Email: Andrey.ptitsyn@colostate.edu

Motivation and aim: Metastases are responsible for the majority of cancer fatalities. The molecular mechanisms governing metastasis are poorly understood, hindering early diagnosis and treatment. Previous studies of gene expression patterns in metastasis have concentrated on selection of a small number of "signature" biomarkers.

Methods and Algorithms: We have modified canonic biomarker analysis pipeline to allow application of Systems Biology approach. We apply quantile normalization, permissive J5 metric for inference, Jackknive procedure for computational validation. The entire expression space is partitioned using Natural Taxonomy algorithm from the FOREL family. Clusters are visualized in Spotfire after two-steps dimensionality reduction anchored on cluster centroids. Pathway Analysis is based on cross-validated Indenuity systems IPA and GeneGo Metacore.

Results: We proposed an alternative approach that puts into focus gene interaction networks and molecular pathways rather than separate genes. We have reanalyzed expression data from a large set of primary solid and metastatic tumors originating from different tissues using the latest available tools for normalization, identification of differentially expressed genes and pathway analysis. Our studies indicate that regardless of the tissue of origin, all metastatic tumors share a number of common features related to changes in basic energy metabolism, cell adhesion/cytoskeleton remodeling, antigen presentation and cell cycle regulation. Analysis of multiple independent datasets indicates significantly reduced oxidative phosphorylation in metastases compared to primary solid tumors. The translational implication of our findings is the possibility to reposition a number of recently developed drugs already in clinical trials from controlling rare inoperable tumors to much wider general anti-metastatic application.

Conclusion: Our methods allow identification of robust, although not necessarily highly expressed biomarkers. A systems approach relying on groups of interacting genes rather than single markers is also essential for understanding the cellular processes leading to metastatic progression. We have identified metabolic pathways associated with metastasis that may serve as novel targets for therapeutic intervention.

IDENTIFICATION OF CANDIDATE GENES FOR HYPERTENSIVE PHENOTYPE MANIFESTATION IN ISIAH RATS

T.O. Pylnik*, S.E. Smolenskaya, A.L. Markel, L.N. Ivanova, O.E. Redina
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: smallynx@ngs.ru
**Corresponding author*

Key words: *microarray, QTL analysis, candidate genes for hypertension, ISIAH rats*

Motivation and Aim: The present study was performed to detect the key genes involved in the hypertensive phenotype manifestation in the ISIAH rats with inherited stress-induced arterial hypertension.

Methods: To detect the candidate genes of hypertensive phenotype the microarray analysis (RatRef-12 Expression BeadChips, Illumina, USA) was performed. Data aquisition and analysis was done by BeadStudio software (Illumina, USA) using gene expression module, rank invariant normalization and p value ≤ 0.01. Hypothalamus, pituitary, medulla oblongata, adrenal gland, liver and kidney were taken from hypertensive ISIAH (n=1) and normotensive WAG (n=1) rats to detect the differentially expressed genes. Separated kidney cortex and medulla were tested additionally in ISIAH (n=3) and WAG (n=3) rats. QTL analysis was performed to identify the genetic loci for arterial blood pressure and related traits. The F_2 hybrids (ISIAH x WAG) of 6-month old male rats were used. Total genome was scanned with 151 polymorphic markers. Linkage analysis was done using MAPMAKER/EXP 3.0 and MAPMAKER/QTL 1.1 programs.

Results: Based on microarray analysis the differentially expressed genes were identified. The interstrain comparison revealed 16 genes differentially expressed ($p \leq 0.01$) in all the tissues analyzed. According to their level of expression these genes were divided into two groups: 13 genes were downregulated and 3 were upregulated in ISIAH rats. The combination of the QTL analysis and microarray data showed that the revealed genes are located in the loci associated with the following physiological traits: blood pressure, plasma corticosterone level, adrenal gland weight, concentration of dopamine and norepinephrine in medulla oblongata and norepinephrine in hypothalamus, locomotion and rearing on periphery of the open field arena.

Conclusion: Most of the selected genes differentially expressed in all analyzed tissues of the ISIAH and WAG rats are located in QTL for physiological traits related to the regulation of the blood pressure. These genes may be considered as candidate genes for further analysis.

Acknowledgements: Supported by RFBR N 08-04-00533, 08-04-01048 and partly by a grant from Ministry of Science and Education of the RF (grant 3H-319-09), Presidium Programs of Russian Academy of Sciences: “Molecular and Cellular Biology” (grant 10.12) and “Biodiversity” (grant 23.26).

LOCAL GEOMETRY OF PROTEIN SURFACE IS HIGHLY OPTIMIZED FOR HYDRATION

S.V. Rahmanov*, I.V. Kulakovskiy, L.A. Uroshlev, V.J. Makeev
GosNIIGenetika Research Institute, Moscow, Russia
e-mail: sergeira@inbox.ru
**Corresponding author*

Key words: *protein structure, water solvation, molecular interaction*

Motivation and Aim: It is a common knowledge that water has a very high energy of intermolecular interactions, largely due to a polar nature of its molecules. It is manifested through very high, for its molecular mass, melting and boiling temperatures and enthalpies, heat capacity and surface tension. Water role in folding and organization of biological macromolecules is considered paramount, as many macromolecules form a hydrophobic core of non-polar residues and a polar surface. Up to 90 per cent of the free energy stabilizing the folded native structure in soluble proteins is estimated to be contributed by the hydrophobic interactions [1].

While studying protein hydration at the atomic level, it was found, quite unexpectedly, that it was possible to achieve a successful decoy recognition in most cases of soluble proteins, based on a crudely estimated free energy of a protein structure hydration alone, without any consideration of intramolecular interactions [2]. The present study is motivated by this result, seeking to establish particular aspects of structural organization of proteins underlying this phenomenon. Also it was shown that water in the first hydration layer near macromolecular surface, contrary to previous beliefs, fully retains its structure of the bulk liquid, namely the oxygen atom radial distribution function [2]. These results may indicate that in addition to the common approach where the macromolecular solvation energy is measured by the accessible surface area of polar groups, a local atomic 3D geometry of macromolecule should also be taken into account. In this study we try to answer the question, how much the "surface" of proteins is adapted, in terms of its geometry, for interaction with networks of explicit water molecules.

Methods and Algorithms: We reconstruct atom-level solvation pattern around protein 3D structures using empirical potentials of atomic interaction. The potentials are obtained by statistical analysis of thousands of 3D macromolecular structures determined experimentally by methods such as X-ray crystallography and stored in publically available data bases like PDB. Using empirical potentials allows one to calculate, for any point in the macromolecular context, inputs from all surrounding atoms and groups into estimated local free energy of hydration.

Results: We have found that inputs from nearby protein atoms into local hydration energy are positively correlated for the predicted hydration sites on the protein surfaces.

Shuffling of protein surface atoms, while preserving its shape and the polar\non-polar accessible surface area, leads to a noticeable decrease in the estimated total structure hydration energy.

Conclusion: Spacial geometry of protein surfaces is shown to be highly optimized for hydration.

Availability: http://bioinform.genetika.ru/index.htm, http://line.imb.ac.ru/ion-calculator

Acknowledgements. This research project is supported by a Russian Foundation for Basic Research grant # 10-04-01809-a.

References:

1. Li X, Liang J. (2006) Knowledge-based energy functions for computational studies of proteins. In: *Computational Methods for Protein Structure Prediction and Modeling* Edited by: Xu Y, Xu D, Liang J. (Springer-Verlag).
2. Rakhmanov, S.V. and V.J. Makeev. (2007) Atomic hydration potentials using a Monte Carlo Reference State (MCRS) for protein solvation modeling. *BMC Struct Biol*, 7: p. 19.

THE GENETIC CONTROL OF PHYSIOLOGICAL TRAITS IN HYPERTENSIVE ISIAH RATS

O.E. Redina*, S.E. Smolenskaya, T.O. Pylnik, A.L. Markel
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: oredina@ngs.ru
**Corresponding author*

Key words: *hypertensive ISIAH rats, QTL analysis, arterial blood pressure, body and organ weights, plasma corticosterone, dopamine and norepinephrine concentration in hypothalamus and medulla oblongata*

Motivation and Aim: The ISIAH rat strain was selected for increased response of systolic arterial blood pressure to a mild emotional stress caused by 0.5 h restriction in a cylindrical wire-mesh cage. As a result of the selection, the ISIAH rats acquired the number of characteristic features concerned the hypertensive status. The genetic control of physiological traits related to the hypertensive phenotype in ISIAH rats was studied.

Methods: QTL analysis was performed to identify the genetic loci for arterial blood pressure and related traits: body weight, absolute and relative kidney, heart and adrenal weights, plasma corticosterone, dopamine and norepinephrine concentrations in hypothalamus and medulla oblongata. The F_2 hybrids (ISIAH x WAG) of 6-month old male rats were used to scan genome with 151 polymorphic markers. Linkage analysis was done using MAPMAKER/EXP 3.0 and MAPMAKER/QTL 1.1 programs. The multitrait mapping was performed with QTL Cartographer Version 1.17, JZmapqtl software. To detect the candidate genes in the genetic loci the microarray analysis (Illumina, USA) was used.

Results: Several significant QTL were found: 1) for basal blood pressure with two peaks near markers D1Rat168 (LOD score 3.42) and D1Rat76 (LOD score 3.34) on Chr.1; 2) for stress-induced blood pressure (LOD score 3.08) in the region D1Rat54–D1Rat168 on Chr.1; 3) for body weight on Chr.15 near D15Rat80 marker (LOD score 2.75, effect of ISIAH alleles +31.6 g) and on Chr. X in the region DXRat104-DXMco53 (LOD score 3.39, effect of ISIAH alleles +26.2 g); 4) for the stress-induced plasma corticosterone level (LOD score 3.33) on Chr.9. This QTL explained 15.3% of the phenotypic variability; 5) for the adrenal gland weight on Chr. 6 (LOD score 3.68) which explained 16.8% of the trait phenotypic variability; 6) for dopamine concentration in medulla oblongata on Chr. 8 near marker D8Rat149 (LOD score 6.77). This locus explained 25.4% of the trait variation. The presence of the ISIAH rat alleles in this locus caused the significant decrease in the brain dopamine concentration. Besides, all the traits analyzed were characterized by multiple suggestive QTL. The bivariate analysis helped to define the loci containing genes with the pleiotropic effect on two or more phenotypic traits. The differentially expressed genes located in the QTL were considered as candidate genes and selected for further analysis.

Conclusion: The QTL analysis confirmed the polygenic control of all the traits. Most traits are under the genetic control of multiple loci with small effects. The combination of the QTL and microarray analysis data will help to understand the molecular mechanisms involved in the hypertensive phenotype manifestation.

Acknowledgements: Supported by RFBR N 08-04-00533, 08-04-01048 and partly by a grant from Ministry of Science and Education of the RF (grant 3H-319-09), Presidium Programs of Russian Academy of Sciences: “Molecular and Cellular Biology” (grant 10.12) and “Biodiversity” (grant 23.26).

THE MATHEMATICAL MODELLING OF NITRITE METABOLISM REGULATION IN *ESCHERICHIA COLI* CELL DURING NITRATE RESPIRATION UNDER ANAEROBIC CONDITIONS

Ree N.A[1]*, Likhoshvai V.A.[1, 2], Khlebodarova T.M.[1]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
e-mail: seahourse@mail.ru
** Corresponding author*

Key words: *respiration, mathematical modelling, generalized Hill function*

Motivation and Aim: E.coli possesses different electron acceptors in the electron transport chain under anaerobic conditions, of which the most energetically favorable is nitrate. Nitrate utilization by nitrate reductases results in nitrite formation. As nitrite is extremely toxic for the cell, regulation of genes involved in nitrate respiration chain is highly associated with genes of nitrite metabolism, the main components of which are NrfA and NirB nitrite reductases. The whole pattern of regulatory mechanisms of anaerobic nitrate and nitrite utilization is not yet revealed. This paper is devoted to the modelling of these processes.

Methods and Algorithms: Michaelis–Menten equations and generalized Hill functions were used to describe kinetic processes of nitrate/nitrite metabolism in *E.coli* cell and mechanisms of genetic regulation of nitrate/nitrite reductase gene expressions [1]. Parameters of the model were taken from the literature or evaluated during elementary submodel verification. Calculations were made using our own programs.

Results: The created mathematical model describes genetic regulation of nitrate reductase A, nitrite reductases NrfA and NirB enzyme expressions, exchange of nitrate and nitrite between the cell and environment, and also it describes metabolic reactions of nitrate and nitrite utilization to ammonia. The model adequately fits the data on expression of nitrate/nitrite reductases genes [2, 3] and describes the dynamic of nitrate, nitrite and ammonia concentrations depending on the initial nitrate concentration in environment.

Conclusion: The analysis of dynamical properties of the model with 100-fold higher expression of terminal NrfA nitrite reductase, which uses nitrite as electron acceptor, predicted the level of nitrite in environment in 1.3-3.5 times lower then that in the wild type.

Acknowledgements:

This work was supported by RFBR (08-04-01008-a), the program of the Russian Academy of Sciences Nos. 22.8, 23.29 and 21.26, SBRAS Presidium Nos. 107 and 119 and grant NSh-2447.2008.4.

References:

1. Likhoshvai V., Ratushny A. (2007*) J. Bioinform. Comput. Biol.,* 5(2B):521-531.
2. Wang H., Gunsalus R.P. (2000) *J. Bacteriol.,* 182(20): 5813-5822
3. Wang H., Tseng C.P., Gunsalus R.P. (1999) *J. Bacteriol.* 181(17):5303-5308.

MATHEMATICAL MODELING OF NUCLEOTIDE BIOSYNTHESIS IN *ESCHERICHIA COLI*

M.T. Ree[1]*, T.M. Khlebodarova[1], and V.A. Likhoshvai[1,2]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
E-mail: rim@bionet.nsc.ru
**Corresponding Author*

Key words: *Escherichia coli, mathematical modeling, nucleotide biosynthesis*

Motivation and Aim: The development of the ***in silico cell***, a bioinformatics resource for modeling, simulation and analysis of intracellular processes, is vital for systems biology. In this regard, the modeling of pyrimidine and purine nucleotide metabolism in *E. coli* cells is of not only purely scientific but also practical significance. In addition to the fact that nucleotides are components of DNA and RNA, ATP and GTP are involved in nearly all cell processes as phosphate and energy sources. Formulas on the base of various nucleotides are widely used in medicine and sports. Their industrial production, in particular, by transgenic producers, is still expensive. Mathematical modeling of processes and analysis of model behavior against altered genetic background provide an approach to the solution of pertinent fundamental and applied problems.

Methods and Algorithms: The construction of elementary models invoked the mass-action law, Michaelis–Menten equation, King–Altman method, and generalized Hill functions [1]. Metabolic system models were constructed as systems of ordinary differential equations describing global rates of variations in low-molecular-weight compound concentrations consumed or produced in the system to be modeled. The global rates were calculated from the law of summation of local rates described in the elementary models. Numerical calculations and model analysis were performed with STEP+ software [2].

Results: On the base of earlier results [3, 4], we developed mathematical models describing instantaneous rates of 47 elementary reactions in the nucleotide biosynthesis pathways. These elementary models were assembled into models of metabolic networks of pyrimidine (PyB) and purine biosynthesis (PuB). It was shown that the model operates either in the steady-state mode or in the continuous self-oscillation mode with a short period. Only steady-state modes were detected in the PuB model. The PyB model contains parameters substantially affecting the yields of target products: UTP, CTP, dCTP, and TTP.

Conclusion: Analysis of the operation mode of mathematical models of pyrimidine and purine nucleotide biosynthesis in *E. coli* cells shows that continuous short-period oscillations of substrate/product concentrations can arise in the range of physiological values of model parameters corresponding to the log phase of cell growth in the PyB model but not in PuB. The biological significance of the theoretically predicted qualitative differences in model behavior is discussed.

Acknowledgments: This work was supported by projects of the RAS (Nos. 22.8, 21.26), projects of the Siberian Branch of the RAS (Nos. 107, 119), and grant NSh-2447.2008.4.

References:

1. Likhoshvai V.A., Ratushny A.V. (2007) *J. Bioinform. Comput. Biol.,* **5**:521-531.
2. Fadeev S.I. et. al. (2006) *Proc. of the IC. on BGRS, Novosibirsk,* **2:**118-120.
3. Ratushny A.V., Nedosekina E.A. (2006) *Proc. of the IC on BGRS, Novosibirsk,* **2:**35-39.
4. Nedosekina E.A. (2006) *Proc. of the IC on BGRS, Novosibirsk,* **2:**60-67.

DIRECT MULTIPARTICLE MODELS OF PHOTOSYNTHETIC PROTEINS INTERACTIONS

Galina Riznichenko*, Ilia Kovalenko, Dmitry Ustinin, Andrew Rubin
Biological Faculty of Moscow State University, 119992, Moscow, Russia,
E-mail: *riznich@biophys.msu.ru*; Phone: 495-9390289; Fax: 495-9391116

*Motivation and Aim:*For understanding the regulation mechanisms on the level of whole thylakoid membrane it is useful to integrate kinetic and structural data in the computer model of the complete system. Special interest must be focused on the role of space organization of thylakoid membrane. Taking into account the heterogeneity of thylakoid membrane and the fact that the sizes of mobile carrier's molecules are comparable to the thickness of the membrane and lumen area, it becomes obvious, that chemical kinetics equations are oversimplified.

Methods and Algorithms: In collaboration with the Dept. of Computer Methods in Physics of the Physical faculty of MSU we develop the new approach to the simulation of photosynthetic electron-transport processes. The model presents the 3D-scene, consisting of three compartments: stroma, membrane and lumen. Photosystem I (PS1), Photosystem II (PSII), cytochrome and other complexes are imbedded in membrane according to experimental data. The electron transport inside the complexes is described by differential equations for the state probabilities. Movement of mobile carriers plastoquinone molecules inside the membrane, plastocianin (Pc) molecules in lumen and ferrodoxin (Fd) molecules in stroma is described by the formalism of Brownian dynamics. The interactions between mobile proteins and embedded in the membrane multienzyme complexes includes the electrostatic interactions of the local charges of the molecules with the other molecules and the lipids of photosynthetic membrane.

Results: Direct multiparticle model gives the possibility to follow the travel of individual mobile molecule as well as to present kinetic and statistical characteristics of the system and to visualize the whole scene of the processes observed. The interactions of Pc-cytochrome f, Pc – PSI, PSI – Fd, Fd-FNR in solution and interactions of Pc-Cytf, Pc-PSI in solution and in lumen were simulated.

Conclusion: The comparison of the simulation curves with experimental data for wild type and mutants of these proteins shows that complex geometry of the reaction space and electrostatic forces provide observed efficiency of electron transport for the photosynthetic electron transport steps under consideration.

Supported by RFBR grants NN 08-04-00354, 07-04-00375.

ADAPTIVE TRADE-OFF BETWEEN REPRODUCTION AND SURVIVAL IN MEDITERRANEAN FRUIT FLIES INDUCED BY CHANGING DIETARY CONDITIONS

A.A. Romanyukha[1*], A.S. Karkach[1], A.I. Yashin[2]
[1] Institute of Numerical Mathematics, Russian Academy of Sciences, Gubkina 8, 119333 Moscow, Russia,
[2] Duke University, Center for Demographic Studies, 2117 Campus Drive, Box 90408, Durham, NC 27708-0408, USA
E-mail: eburg@inm.ras.ru

Key words: *adaptation, life history, cost of reproduction, coordinated aging, fitness*

Motivation and Aim: The conception of the cost of reproduction provides an important insight on connection between fertility and life span in living organisms. Despite substantial progress in understanding this connection many important features of fertility-longevity trade-off are masked by confounding factors, and remain poorly understood.

Methods and Algorithms: We performed reanalysis of data from experimental study of fertility and longevity response to different diets in females of Mediterranean fruit fly *C. capitata* [1, 2].

Results: A negative dependence between average fertility and longevity was observed in the long lived part of experimental cohorts as the protein content of the diet changed. In order to explain the observed phenomenon we suggest a mechanistic resource allocation model. The model is further development of the resource allocation model proposed in [3].

Conclusion: The presence of a fertility-longevity trade-off suggests a possibility of existence of some resource used both by reproduction and somatic maintenance in a fly. The trade-off may be a manifestation of metabolic machinery, processes and genetically determined laws of control which define balance between the processes of reproduction and regeneration. We propose and discuss a principle of dynamic resource allocation which explains fertility–longevity data for the long-, intermediate- and short-lived flies. Adaptive allocation of metabolic and other resources allows flies to tailor their life history parameters to the environment. Due to limitations of the physiological adaptation a significant share of the population may be genetically "preadapted" to different environmental conditions thus contributing to population stability and heterogeneity. This may be observed even in relatively homogeneous populations, such as experimental fly cohorts.

Acknowledgements: Work was supported by NIH/NIA Grants R01-AG-028259, R01-AG-027019, P01 AG-08761, Russian Foundation for Basic Research Grant 07-01-00577a, and Max Planck Institute for Demographic Research, Rostock, Germany.

References:

1. J.R.Carey et al. (2002) Food pulses increase longevity and induce cyclical egg production in Mediterranean fruit flies, Functional Ecology 16: 313–325.
2. J.R.Carey et al. (1998) Dual Modes of Aging in Mediterranean Fruit Fly Females, Science, 281: 996–998.
3. A.A.Romanyukha et al. (2004) The impact of diet switching on resource allocation to reproduction and longevity in Mediterranean fruitflies, Proc. R. Soc. Lond. B. 271: 1319–1324.

HOMEOSTASIS MAINTENANCE, TISSUE TURNOVER AND AGING

A.A. Romanyukha
Institute of numerical mathematics RAS, Moscow, Russia
e-mail: eburg@inm.ras.ru

Key words: *homeostasis, adaptation, aging, fitness*

Motivation and Aim: The mechanism underlying the age-related reduction in physical performance is obscure. The maximal aerobic capacity (VO_2max) is the principal determinant of physical performance. The VO_2max value declines at the rate of 5-7% per ten years (1). The assumption that the phenomenon is caused by insufficient regeneration capacity contradicts with the high rate of adaptation to physical exercise and to the increase of VO_2max (2). The aim of the work is to build a conceptual model of age-dependant dynamics of physiological systems and physical performance.

Methods and Algorithms: Model is based on the hypothesis of adaptive nature of homeostasis maintenance and tissue turnover control. It is suggested that the individual's fitness increases as the sum of costs of maintenance and tissue turnover ($\boldsymbol{E_M}$) and the losses on the incomplete regeneration and homeostasis deviations ($\boldsymbol{E_P}$) diminishes. As the common currency for this costs and losses we suggest to use organism's energy spending for corresponding metabolic processes (3).

Results: The type of interdependence of rates of energy spending on maintenance and tissue turnover $\boldsymbol{E_M}$ and of the losses on the incomplete regeneration and homeostasis deviations $\boldsymbol{E_P}$, enables us to conjecture the existence of minima of the total energy consumption $\boldsymbol{E_T}$ = min ($\boldsymbol{E_M} + \boldsymbol{E_P}$). The temporary deviation from the optimal state (for example, if $\boldsymbol{E_{Pt}} > \boldsymbol{E_P}$) may lead to the transition to a new stable state with $\boldsymbol{E_{Tt}} > \boldsymbol{E_T}$. Recovery of the initial level of $\boldsymbol{E_P}$ may lead to some intermediate optimal point $\boldsymbol{E_{PA}}$ ($\boldsymbol{E_{Pt}} > \boldsymbol{E_{PA}} > \boldsymbol{E_P}$).

Conclusion: Thus, the age-related changes in tissues and structural-and-functional characteristics of physiological systems may be considered as a side-effect of adaptation to the alternating environment.

Acknowledgements: Grant RFBR 10-01-00779-a

References:

1. M. L. Pollock et al. (1997) Twenty-year follow-up of aerobic power and body composition of older track athletes**,** *J Appl Physiol,* **82:** 1508 - 1516.
2. G. D'Antona et al. (2003) The effect of ageing and immobilization on structure and function of human skeletal muscle fibres, *J. Physiol. (Lond.).* **552:** 499–511.
3. Romanyukha A.A., Rudnev S.G., Sidorov I.A. (2006) Energy cost of infection burden: An approach to understanding the dynamics of host–pathogen interactions, *J. Theoretical Biology*, **241:** 1-13.

INVESTIGATION OF DIRECT INTERACTION OF REPAIR DNA POLYMERASE β AND AUTONOMOUS 3′→5′-EXONUCLEASES TREX1 AND TREX2

N.V. Belyakova, O.K. Legina, N.L. Ronzhina*, I.V. Shevelev, V.M. Krutyakov
Konstantinov Petersburg Nuclear Physics Institute RAS, Gatchina, Leningrad District, Russia
e-mail: ronzhina@list.ru
**Corresponding author*

Key words: *DNA polymerase β, 3′→5′-exonuclease TREX, direct interaction, complex of proteins, DNA synthesis, accuracy*

Motivation and Aim: DNA polymerase β is a key enzyme in base excision repair of DNA, but it is error-prone DNA polymerase and inserts a wrong nucleotide with the probability 10^{-3} [1]. Autonomous (individual) 3′→5′-exonucleases can correct these mistakes and increase the fidelity of DNA synthesis 30-fold [2]. Investigation of direct physical interaction between recombinant proteins DNA polymerase β and 3′→5′-exonucleases TREX1 and TREX2 is important for clarifying of the mechanism of accuracy increasing of DNA synthesis.

Methods and Algorithms: Recombinant proteins hPolβ-his, mTREX1-his and hTREX2-MBP were expressed in cells of *E. coli*: BL21(D3)[pET21b(mTREX1-his)], BL21(D3)[pET21b(hPolβ-his)] and MC1061[pMAL(hTREX2-MBP)]. Fusion proteins were purified on corresponding columns with Ni-agarose or maltose-sepharose. It was not detected contaminant DNA polymerase and nuclease activities in corresponding preparations of TREX1 and TREX2 and hPolβ-his. Gel filtration on the column with Sephacryl S-200 and immunoassays in variants of dots and blotting were used for examine of protein interaction.

Results: The results of gel-filtration show formation of the complex between 3′→5′-exonucleases mTREX1 and hTREX2 and DNA polymerase β: up to 80% of polymerase activity shifts to the high molecular weight protein zone compared to individual hPolβ-his. Analysis of activities of pure enzymes and their complexes detects increasing of DNA polymerase activity 4-fold. Anti-TREX1 or anti-TREX2 antibodies detected correspondingly binding mTREX1-his or hTREX2-MBP with hPolβ-his which was immobilized on membranes. Immunoassays of hPolβ-his with immobilized mTREX1-his and hTREX2-MBP also reveal direct association of these proteins in complex.

Conclusion: The results in vitro demonstrate direct interaction of recombinant proteins of human repair DNA polymerase β with proofreading 3′→5′-exonucleases mTREX1 and hTREX2.

References:

1. L.Servant et al. (2002) A role for DNA polymerase β in mutagenic UV lesion bypass, J. Biol. Chem., **277**: 50046–50053.
2. N.V.Belyakova et al. (2007) Complex of repair DNA polymerase β with autonomous 3′→5′-exonuclease shows increased accuracy of DNA synthesis, Izvestiya Akademii Nauk, Seriya Biologicheskaya, **5**: 517–523.

IMMUNE SYSTEM DEVELOPMENT AND BODY GROWTH: WHAT IS THE RELATIONSHIP?

S.G. Rudnev*, A.A. Romanyukha, A.I. Yashin[1]
Institute of Numerical Mathematics RAS, Moscow, Russia
[1]Center for Demographic Studies, Duke University, USA
e-mail: rudnev@inm.ras.ru
[*]Corresponding author

Key words: *immune defense, energy cost, adaptation, antigen load, basal metabolic rate*

Motivation and Aim: In contrast to other biological tissues and organs, the immune system aging starts early in life and can be irreversible in character. In order to study this phenomenon, the mathematical model of age-related changes in population of peripheral T cells [1] was modified to describe ontogenetic changes of an immune system during all postnatal life [2,3].

Methods and Algorithms: An assumption was used about linear dependence of antigen load from basal metabolic rate, which, in turn, depends on body mass following the allometric 3/4 power scaling law [4,5].

Results: Model-based function of the energy cost of infection burden was constructed and used as a measure of the immune system effectiveness. The model parameters were estimated using differential evolution (DE) algorithm. The dependence of optimal resource allocation from the parameters of antigen load was studied.

Conclusion: A whole body of evidence from animal and human studies supports the hypothesis about the existence of trade-off between immune defense and organism's growth. Our model allows for investigation of possible consequences of such a trade-off and suggests an existence of stabilizing effect of body growth on the immune system dynamics. The results of sensitivity analysis emphasize the importance of the exposure to pathogens at the early period of developing adaptive immunity that may have late consequences for healthy aging.

Acknowledgements: A part of this work was performed at the Max Planck Institute for Demographic Research (Rostock, Germany).

References:

1. A.A.Romanyukha, A.I.Yashin (2003) Age related changes in population of peripheral T cells: towards a model of immunosenescence, *Mech. Aging Dev.*, **124**: 433-443.
2. S.G.Rudnev, A.A.Romanyukha, A.I.Yashin (2006) Modeling of immune life history and body growth: the role of antigen burden. MPIDR Working Paper WP-2006-042, November 2006, Max Planck Institute for Demographic Research, Rostock, Germany. 34p
3. S.G.Rudnev, A.A.Romanyukha, A.I.Yashin (2007) Modelling T cell population development and estimation of resource allocation effectiveness, *Math. Modelling*, **19**: 25-42 (in Russian).
4. M.Kleiber (1932) Body size and metabolism, *Hilgardia*, **6**: 315-353.
5. J.B.West, J.H.Brown (2005) The origin of allometric scaling laws in biology from genomes to ecosystems: towards a quantitative unifying theory of biological structure and organization, *J. Exp. Biol*., **208**: 1575-1592.

MAPPING DNA POLYMORPHISMS ON THE EPIDEMIOLOGY OF HUMAN DISEASES

S.G. Rudnev
Institute of Numerical Mathematics RAS, Moscow, Russia
e-mail: rudnev@inm.ras.ru

Key words: *human diseases, tuberculosis, population genetics, HLA region, relative risk, software, genetic indices of epidemiological risk*

Motivation and Aim: In order to study the role of genetic factors in the observed geographical variations of the incidence of tuberculosis in Russia, the genetic susceptibility index was recently proposed [1]. For this, data on genetic polymorphisms of various ethnic groups, as well as relative risk estimates for TB at the level of low-resolution PCR typing of HLA DRB1 gene in ethnically Russian individuals [2] were utilized. The index correlated well with TB incidence in geographically and genetically distant populations representing Moscow city and Russian ethnic republics [1]. The aim of this study was to extend our approach and to develop general procedure of converting molecular genetics data to genetic indices of epidemiological risk.

Methods and Algorithms: To calculate the values of relative risk at the level of alleles, genotypes, or haplotypes, the IrGene 1.0 software program was used [3].

Results: A scheme relating molecular population genetics data to quantitative predictions of epidemiological risk in terms of disease incidence, morbidity and mortality is suggested.

Conclusion: An approach considered could be used to clarify hereditary component in determining epidemiological parameters of human diseases, as well as to estimate possible consequences of ethnodemographic changes for epidemiology.

Availability: The above software can be assessed at http://immunol.inm.ras.ru/old/IrGene.

Acknowledgements: The author thanks Profs R.P. Selitskaya and M.N. Boldyreva for original data and for valuable discussions.

References:

1. S.G.Rudnev, R.P.Selitskaya, M.N.Boldyreva (2009) On the relative risk concept and TB morbidity in Russia: linking population genetics and epidemiological studies, *Russ. J. Numer. Anal. Math. Modelling,* **24:** 377-384.
2. R.P.Selitskaya et al. (2009) On polymorphism of the HLA-DRB1 locus and susceptibility to tuberculosis, *Immunology*, **30**: 338-341 (in Russian).
3. E.A.Sytin (2009) IrGene 1.0, interface and code for analysis of population genetics data in immunology, *Collected Articles of Young Scientists of CMC Department*. M.V.Lomonosov Moscow State University, **6**: 163-167 (in Russian).

CATARACTOGENESIS IN OXYS RAT: EPITHELIUM DAMAGE, DOWNREGULATION OF CRYSTALLIN EXPRESSION AND ANTIOXIDANT TREATMENT

Rumyantseva Yu.V.
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: sensorik@gmail.com

Key words: *cataract development, expression of α-crystallins, senescent-accelerated OXYS rats*

Motivation and Aim: Cataract is one of the *most common* causes of age-dependent visual impairment and blindness. Its pathogenesis is associated with oxidative stress but remains poorly understood. Cataract declare itself by the coloration of lens, progressive *increase in* oxidation, and insolubilization of lens proteins, up- or down-regulation expression the most part of genes associated with epithelium decline. It is known that antioxidants may delay progression of age-related diseases including cataract. But the correct estimation of these effects *on human* is difficult in connection with individual distinctions of nutrient supply and age-related deficits. It was showed that senescencent- accelerated OXYS rats is suitable models for cataract and can be use for study of pathogenesis and design of new therapeutic targets. The aim of this investigation was comparison age- and disease-related morphological changes of rat`s lens and expression of α-cristallis as key lens's proteins. In addition effects of mitochondria-targeted antioxidant SkQ1 on these parameters were studied.

Methods and Algorithms: Real-time PCR with specific primers to αA- and αB-cristallin gene in lens, parameters of redox-homeostasis and morphological analysis of lens had been performed using 20 days (without clinical signs of cataract), 3 (manifestation the first signs) and 12 months-aged OXYS (pronounced stage of cataract) and even-aged Wistar rats as control. Experimental groups of rats get SkQ1 (250 nmol/kg) orally from 9 to 12 month.

Results: Appearance of the first clinical signs of cataract in OXYS rat occurs under conditions of increased GSH level in the lenses while strong stages of cataract are characterized by dramatic decline of GSH. In the lenses of 20 days-old OXYS rats expression of αA- and αB-crystallin gene were 2-fold higher than in the Wistar rats. But the development of clinical manifestations of early stages of cataract was associated with significant reduction in expression of both this genes in OXYS rat's lenses, which becomes more pronounced during disease progression. Thus, at the age of 3 mo αA- and αB-crystallin expression was 3-fold and 25% lower than in control, respectively, at the age of 12 mo a 21-fold decrease was observed. Changes in gene expression during senile cataract development are associated with damage to lens epithelial cell layer or its enzymatic system. Histological examination revealed in the OXYS rat's lens dystrophic changes associated with cataractogenesis. At the age of 20 days the lenses of OXYS rats distinguished by the presence of cortical and epithelial vacuoles, fiber cells swelling, in the later stages of the disease observed increase in cytoplasmic granularity, and misfolding capsular proteins. The second main of this study was to determine the effects of SkQ1 on the cataract development in OXYS rats. We showed, that SkQ1 supplementation not only prevented cataract development in OXYS rats but decreased pathological changes in lens. Simultaneously, SkQ1 normalized expression of αA-crystallin gene and increased expression of αB-crystallin.

Conclusion: Summarizing all the data obtained in this study as well as previous results, we conclude that cataract development in OXYS rats can occur due to changes in lens epithelium and SkQ1 restores its structural and functional parameters.

Acknowledgements: This work was supported by grants from RFBR 08-04-00722 and by "MSU centre of Mitoengineering," LLC, Moscow State University (MSU).

APPLICATION OF DNA SEQUENCING AND DETECTION OF LEVELS EXPRESSION EPIDERMAL GROWTH FACTOR RECEPTOR (EGFR) IN CANCER PRACTICE

S.M. Sirotkina[1], A.H. Sabirov[2]*
Tyumen State University, Tyumen, Russia[1]
Tyumen Academy of Medicine, Tyumen, Russia[2]
e-mail: sabirov58@mail.ru
[*]Corresponding author

Key words: *epidermal growth factor receptor, DNA sequencing.*

Motivation and Aim: The aim of the research is to study prognostic and diagnostics significance of the detection expression levels of EGFR in serum, as well as the presence of mutations in certain oncogenes of cancer patients.

Methods and Algorithms: Research is based on testing of cancer patients who were having the examination and therapy in Sabirov Center of molecular and genetic diagnostics from 2008 to 2009. The level of loose EGFR forms in serum was tested by means of immunoferment analysis (IFA) in ELISA modification.

Material for molecular genetic studies provided the tumor cells and blood plasma of these patients. The presence of mutations was determined in the oncogenes p53 (exons 5,6,7,8), C-kit, B-raf, APC, K-ras, E-cadherin and p16 by means specially chosen primers. The results verified by direct sequencing of PCR fragments.

Results: The research includes 220 patients at the ages of 25 to 70.

The average value of level of loose EGFR forms in serum of patients with mammary gland cancer (60 p.) was 4,47 fmol/ml. For patients with stomach cancer (75 p.) this level was 4,8 fmol/ml, for lung cancer patients (25 p.) it was 4,77 fmol/ml (an average standard of EGFR is 3,6 fmol/ml).

When molecular genetic studies have observed a significant decrease in the number of mutations in the course of special treatment after 2 courses of chemotherapy (PCT) (7 cases).

After 6 courses of PCT, mutations persisted in 4 patients, i.e. molecular remission was not achieved, indicating a lack of elimination of tumor cells, unreasonable transfer of the patient in 3 clinical group and required PCT need to continue or change the scheme.

Conclusion: 1.There were excess of upper bounds of loose EGFR forms expression level standart (3,6 fmol/ml) (overexpression) in the course of research.

2.So it is reliable that EGFR overexpression is of importance in carcinogenesis. It is a marker for describing tumor behavior and instituting individual treatment for patient with malignant tumors.

3.EGFR overexpression, as well as the presence of mutations in certain oncogenes of cancer patients (p53 (exons 5,6,7,8), C-kit, B-raf, APC, K-ras, E-cadherin and p16) can also be a target for creation of new kinds of anti-tumor therapy pointed at mitosis signal transmission blocking (for targeting therapy).

SYMMETRY IN OLIGONUCLEOTIDE COMPOSITION IN GENOMES AND EXTREMAL PROPERTIES OF DNA SEQUENCES

M.G. Sadovsky
Institute of computational modelling of SB RAS, Krasnoyarsk, Russia
e-mail: msad@icm.krasn.ru

Key words: *(frequency, entropy, information value, sense)*

Studying the statistical and combinatorial properties of DNA still may bring new exciting results. Some of them seem to be quite acquainted from previous data, some are not. Even the results looking rather familiar, may reveal some new, unexpected and very important properties of DNA sequences, thus enhancing system biology development. Any genome exhibits a symmetry in the oligonucleotide composition (frequency dictionary): for word w of the length q with the frequency f_w meets the complementary palindromic word $\overline{w}$ of the very proximal frequency $f_{\overline{w}}$. Here the frequency is calculated over a single strand. Such proximity of the frequencies was known quite a time ago; meanwhile, its meaning is still underestimated. The symmetry follows in the conservation laws execution; two of them are Chargaff's rules. Previously [1], the information value of a word was found to be more sensitive index, in comparison to the frequency. The symmetry mentioned above is also observed for the information value of the words composing the complementary palindrome, while it exhibits more deviations, in comparison to the frequency itself.

To measure the concordance in the complementary palindromes information values, the index

$$\mu = \|\Omega\|^{-1} \sum_{\omega} |p_{\omega} - p_{\overline{\omega}}|$$

has been introduced. Here $\|\Omega\|$ is the cardinality of a dictionary, and p_w, $p_{\overline{w}}$ are the information values of the words composing a complementary palindrome.

To figure out the tentative biological meaning of the symmetry and, more important, the violation of that latter, we studied the discrepancy in μ alongside a sequence. DNA sequence was divided into 1000 equal fragments, and the frequency dictionaries W_q, $3 \le q \le 8$ were developed. Then, the index μ was determined for each fragment. To reveal the biological impact of the variation of the index, the same procedure has been carried out for a surrogate sequence of the same length, and the same mononucleotide composition. The fragments were long enough to neglect a finiteness effects (not less that 10^5 nucleotides).

It was found that real pattern of the index μ variation alongside a genome heavily differs from a similar observed for surrogate sequences. Moreover, while an average (over entire genome) level of the index μ variation for real sequences significantly exceeded that latter observed for surrogate entities, for $3 \le q \le 6$, this level evidently less, for $q=8$.

References:

1. A.N.Gorban et al. (2000 Classification of symbol sequences over their frequency dictionaries: towards the connection between structure and natural taxonomy, *Open Systems & Information Dynamics,* **7(1):** 1-17.

BIOINFORMATICS APPROACH TO UNDERSTAND MECHANISM OF INTERACTION OF VPS34 AND VPS15

S. Sahu[1]*, H. Mohanta[1]
1National Institute of Technology, Durgapur, India,
*e-mail: *sombeet@gmail.com *Corresponding Author*

Introduction: Cell membranes contain the lipid phosphatidylinositol (PtdIns), which can be differentially phosphorylated on the hydroxyl groups present at 3, 4 and 5 position of its inositol ring to give rise to seven other phosphatidylinositol phosphate moieties. Phosphoinositides play pivotal roles in the regulation of a wide variety of cellular processes. PI3Ks have been divided into 3 classes of which class 1 PI3K have been the most studied extensively. Expression patterns and mode of regulation of class 3 PI3Ks are less well understood. Unlike class 1 and claas2, class 3 PI3Ks can phosphorylate only PtdIns. Class 3 subunits have a catalytic unit VPS34 and a regulatory subunit VPS15. Cellular mechanisms of oncogenic cells can be better understodd as class 3 is involved in diverse intracellular trafficking events including autophagy and phagosome formation, internal vescicle formation with multicellular endosomes, retrograde endosome to Golgi transport at nuclear membrane.

Methods and algorithms*:* Bioinformatics approach was used to find the similarities between the genes encoding for these two proteins VPS34 and VPS15 and also between the proteins itself. We seek to evaluate the co-evolution of the two proteins across genomes of organisms. Then we compared the protein sequences of the organisms with protein database to consolidate the domains responsible for the binding of VPS34 and VPS15.

Results*:* Homologues of VPS15 suggest that it contains an N-terminal consensus sequence.VPS15 consists of a predicted protein kinase domain followed by a central region containing multiple heat repeats and a series of C-terminal WD40 domains. It was found that the Viridiplantae, protists,metazoan and bacteria form distinct groups in the phylogenetic trees. Then the domain analysis showed that all of them contain distinct domains of WD40 and Heat repeats. Further analysis involved comparison based on tertiary structure of the proteins as they are the functional one and hence gave interesting similarities. It has been shown that the interacting motifs of class 2 PI3kinases are mostly similar to the class 3 ones.

Inference*:* The WD40 repeats might form a β-propeller like structure, similar to the C-terminus of Gβ subunits. The VPS15 binds to these Gβ subunits. Interestingly VPS15 has not been formally proven to possess protein kinase activity. A comparison of the VPS15 sequence with that of other protein kinases shows significant differences; VPS15 for example lacks the canonical GXGXXG motif involved in ATP binding (the sequence is GSTRFF in VPS15). Significant probability in the comparison based on 3-d structure concluded that the mechanism can be mostly similar to the binding of class 2 regulatory and catalytic subunits.

STRUCTURAL PECULIARITIES OF PLANT PROTEIN PHOSPHATASE INTERACTION WITH OKADAIC ACID

D.A. Samofalova[1,2*], A.Yu. Nyporko[1,2], Ya.B. Blume[1]
[1]*Institute of Food Biotechnology and Genomics, Natl. Academy of Sci. of Ukraine,* Osipovskogo str., 2a, *Kyiv* 04123
[2]*Taras Shevchenko Kiev National University, Kiev, Ukraine*
e-mail: samofalova_dariya@i.ua
[*]*Corresponding author*

Key words: protein phosphatases (PP), okadaic acid (OA), specific interaction

Motivation and Aim. Computational protein-inhibitor analysis as a method of prediction allows to hypothesize about unknown intracellular functions of molecular cascades that can be verified afterwards experimentally. Among these cascades, phosphorylation represents a special interest due to both canonic and uncanonic functions. In spite of understanding of cell role of dephosphorylation the features of spatial structure of plant protein phosphatases (PPs) are very important. Up to now their interaction with specific inhibitors are not studied especially in plant cell cascades. Last investigations revealed that okadaic acid (OA), specific inhibitor for animal PPs can interact efficiently with plant PP1 and PP2A phosphatases, but structural peculiarities of such interaction must be assessed yet.

Methods and Algorithms. The spatial structure of plant (*Arabidopsis thaliana*) PP1 and PP2A protein phosphatases were reconstructed with homology modelling method. Structural features of binding sites for OA in plant and animal PP1 and PP2A phosphatases were compared with spatial fitting method. To verify a stability of obtained complexes they were undergone a molecular dynamics computation for 100 ns interval using GROMACS software.

Results and conclusion. It was shown that OA binding sites of plant and animal PP1 have 80% identity and 92% similar amino acid residues whereas OA binding sites for respective PP2A have 91% identity and 95% similar amino acid residues, respectively. Accordantly to our data, stabilization of molecular oscillations and conformation energy levels of all studied complexes takes a place very rapidly (during first 2 ns of dynamics). Stable levels of conformational energy for OA complexes with plant PP1 and PP2A constist 1498 kJ/mol and 1568 kJ/mol. In the case of OA complexes with animal PP1 and PP2A these levels are equal to 1460 and 1500 kJ/mol, respectively. Both amplitude and average level of OA molecular oscillations were approximately twicely reduced in all studied complexes in comparison with free (water) state. It testifies the stabilization of OA spatial structure in binding state against free state. Conformational energy of OA due to transfer from water to binding sites also decreases for all investigated protein phosphatases (in the range from 240,7 до 684,9 kJ/mol), that is evidence of higher energy preference of the OA binding in all studied sites. Thus, the identity between structural mechanisms of interaction OA with animal and plant PP1 and PP2A protein phosphatases might consider to be proved.

STABLE BORDERS OF GAP GENE EXPRESSION ARE FORMED BY CANALIZATION OF THE BICOID MORPHOGEN VARIABILITY IN THE DROSOPHILA BLASTODERM

Maria Samsonova
St.Petersburg State Polytechnical University, St.Petersburg 195251, Russia
e-mail: samson@spbcas.ru

Key words: *Segmentation, dynamical systems analysis, canalization, Drosophila*

Motivation and Aim. The extensive variation in early gap gene expression in the *Drosophila* blastoderm is reduced over time due to the gap gene cross regulation. This phenomenon is a manifestation of canalization, the ability of an organism to produce a consistent phenotype despite variations in genotype or environment. To explore how canalization arises we applied a theory of dynamical systems.

Methods and algorithms. We undertook the analysis of the gap gene border formation in the model in terms of the phase portrait of the dynamical system, dictated by the configuration of attractors and their basins of attraction in the space of protein concentrations. We also explored the dynamical effects of varying spatial profiles of Bicoid protein concentration on the formation of the expression border of the gap gene *hunchback (hb)*.

Results. The canalization of the gap gene expression can be explained as a result of the action of robust attracting states. The complex patterning of the gap gene system reduces to the three qualitative dynamical mechanisms of (1) movement of attractors, (2) selection of attractors, and (3) selection of states on a one dimensional manifold. The last of the three mechanisms also causes the domain shifts of the gap genes, providing a simple geometric explanation of a transient phenomenon. The effect of the varying spatial profile of Bicoid protein concentration on the formation of the expression border of the gap gene *hb* was tested against a family of Bicoid gradients obtained from individual embryos. We show that for each of the Bicoid profiles the *hb* border formation is associated with intersection of the spatial gradient of the maternal Hunchback protein and a boundary between the attraction basins of two different attractors.. The observed reduction in variability of the *hb* expression border can be explained by specific geometrical properties of the basin boundaries.

Conclusions. Our analysis shows that the complex phenomena of canalization and pattern formation in the *Drosophila* blastoderm can be understood in terms of the qualitative features of the dynamical system. The result confirms the idea that attractors are important for developmental stability and shows a richer variety of dynamical attractors in developmental systems than has been previously recognized.

ISOLATION AND CHARACTERIZATION OF NOVEL CELL SIZE MUTANTS IN *SACCHAROMYCES CEREVISIAE*

A. Manukyan[1, 2], H. Dungrawala[1], L. Abraham[1], K. Nazaryan[2] and BL Schneider[1]*
[1.] *Department of Cell Biology and Biochemistry, Texas Tech University Health Sciences Center, Lubbock, TX, 79430 USA*
[2.] *Institute of Molecular Biology, National Academy of Sciences of Armenia, Yerevan, Armenia*
e-mail brandt.schneider@ttuhsc.edu
** Corresponding author*

Motivation and Aim: Cultures of proliferating cells maintain a constant average cell size. Initial systematic genome-wide genetic screens in yeast have identified several hundred genes intimately involved in cell size homeostasis. To complete these screens, we examined the cell size phenotypes of ~800 new complete ORF deletions and isolated 15 new cell size mutants. Wholesale characterization of new and old cell size mutants has revealed that most of these genes (e.g. *WHI5* or *BCK2*) function by regulating the timing of the expression of the G1-phase cyclins (*CLN*s). Yeast coordinate growth with proliferation by ensuring that cell cycle progression is dependent upon the attainment of a critical cell size. Moreover, progression past START is dependent upon a critical threshold of *CLN* expression. For example, deletion of *CLN* genes dramatically increases cell size whereas *CLN* over-expression makes cells very small. However, the molecular mechanism that sets the *CLN* threshold for START is not well understood. Therefore, we tested the hypothesis that a sub-set of cell size control genes alter cell size by modulating the *CLN* threshold.

Methods and Algorithms: Using quantitative RT-PCR we have assessed how single ORF deletions of a large sub-set of cell size control genes affect *CLN* expression.

Results and Conclusion: Not surprisingly, a number of genes were identified that modulated *CLN* expression levels. Counter intuitively, we find that, in general, mutants that decrease *CLN* expression result in abnormally small cells and vice versa. To investigate the mechanisms responsible for these observations, we used real time microscopic analysis of single cells to determine how altering *CLN* expression levels effects cell size. Results suggest that some of these genes function downstream of *CLN*s and may be involved in setting the Cln threshold for Start.

MOLECULAR COEVOLUTION OF BIRD SCHISTOSOMES AND ITS INTERMEDIATE SNAIL HOSTS

S. Semyenova*, A. Lopatkin, E. Kheidorova[1], V. Vasyliev, M. Voronin[2], S. Be'er[2], A. Shestak, S. Vodyanitskaya[3], N. Yurlova[3], V. Mischenkov[4], L. Akimova[4], O. Zazornova[2], T. Zhukova [4], A. Ryskov, G. Chrisanfova.

*E-mail: seraphimas@mail.ru *Corresponding author*

Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia; [1]Science and Production Association "Scientific and Practical Center of the National Academy of Sciences of Belarus on Biological Resources", Minsk, Republic of Belarus; [2] Center of Parasitology of A. N. Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Moscow, Russia; [3] Institute of Animal Systematic and Ecology, Siberian Division, Russian Academy of Sciences, Novosibirsk, Russia; [4] Belorussian State University, Minsk, Republic of Belarus.

Key words: *trematoda, bird schistosomes, snail hosts, ITS2 rDNA, cox1, coevolution*.

The aim. For the study of host-parasite coevolution of trematodes and intermediate host snails we employed ITS2 rDNA to molecular identification both of cercariae of bird schistosomes and its hosts. Additionally we sequenced *cox1* genes for parasites obtained from different snails and birds.

Material and Methods. Cercariae, eggs and adult worms of bird schistosomes were recovered from snails and waterfowl birds from eight Russian (Moscow and Novosibirsk regions) and three Belorussian (Naroch, Shvakshty, Polonevitchi) water ponds. We sequenced the ITS2 rDNA of 50 naturally infected intermediate snail hosts (*Radix* spp., *Lymneae* spp., *Planorbarius corneus*) and 30 adult worms. Comparative phylogenetic analyses were performed with PAUP and PhyML.

Results. Four recently recognized schistosome species (*Bilharziella polonica, Trichobilharzia szidati, T. franki, T. regenti*) and seven snail species (*P. corneus, L. stagnalis, L. turricula, R. auricularia, R. lagotis, R. labiata R. ampla*) were distinguished at studied area. At least three parasite and snail species were recovered from different Moscow ponds - *T. szidati* (in *L. stagnalis*), *T. franki* (in *R. auricularia*) and *T. regenti* (in *R. lagotis*). Only *T. szidati* parasitized on *L. stagnalis* and *L. turricula* were detected in Novosibirsk region. Cercariae of *B. polonica* were found exclusively in *P. corneus* obtained from Belorussian lakes. The eggs of *T. regenti* and adult worms of *B. polonica* were recovered from five species of waterfowl birds from Naroch Lake. The 30 cercarial isolates of the new *Trichobolharzia* species preliminary named as *T. sp.* var. *narochanica* were isolated from the *R. ampla* and *R. labiata* snails from Naroch Lake. The genetic uniqueness of the new species has been support by the variability of mitochondrial *cox1* genes [1].

*Conclusion*s: The studied isolates of bird schistosomes belong to four species of *Trichobilharzia* genera parasitizing multiple snail host species. The intraspecific variability of *cox1* for all five bird schistosome species studied may be useful for "barcoding" of avian schistosomes [1, 2].

Availability. The epidemiological values as well as taxonomic position of studied isolates were discussed.

References:

1. G. G. Chrisanfova et al. (2009) Variability of bird schistosomes (Class Trematoda, Family Schistosomatidae) of Lake Naroch': identification of a new species in the *Trichobilharzia ocellata* group. Doklady Biochemistry and Biophysics 428: 268–272
2. A. A. Lopatkin et al. (2010) *Cox1* polymorphism of cercarial bird schistosome isolates (Trematoda: Schistosomatidae) obtained from Moscow and Moscow region water ponds. Russ J Genetics 7: 34-41.

ALLELE DISTRIBUTION OF C-FMS GENE POLYMORPHISMS DEL425 AND 3'UTR TC/CA IN NATIVE HUMAN POPULATIONS OF WESTERN SIBERIA

N.A. Senkova*, A.G. Romaschenko, L.P. Osipova
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: senk_off@mail.ru
**Corresponding author*

Key words: *gene c-fms polymorphisms del425 and 3'UTR TC/CA, macrophage colony stimulating factor, adaptation, native human populations of western Siberia*

Motivation and Aim: Polymorphism analysis of human *c-fms* gene encoding receptor for macrophage colony stimulating factor is of particular interest because of the high pleiotropic effect and the expression of this gene in many tissues. In some studies, association of *c-fms* gene polymorphisms with bronchopulmonary diseases was shown. In addition, a tendency towards an increase of the del 425 rare allele frequency in direction from south-west to north-east was revealed. Thereby the aim of this study was the investigation of the *c-fms* mutant alleles del425 and 3'UTR TC/CA for genetic diversity estimation in native human populations living in extreme environmental conditions of Northern Siberia, and possible influence of these factors on the adaptation processes.

Methods and Algorithms: Four ethnic samples were studied: Tundra Nenets (N=125), Forest Nenets (N=56), Selkups (N=82) and Komi (N=43) inhabiting the North of Western Siberia (Yamalo-Nenets Autonomous District) and relating to the Ural language family. DNA was extracted by phenol-chloroform method. DNA samples were genetyped by allele-specific PCR.

Results: Allele with deletion (del425) was detected in all ethnic samples with frequency ranging from 28% in Tundra Nenets, 17 % in Selkups, 14% in Komi - to 9 % in Forest Nenets. Interpopulation difference comparative analysis demonstrated the similarity between the representatives of Finno-Ugric (Komi) and Samoyedic (Selkups, Forest Nenets) language groups belonging to the Ural family. The exclusion are Tundra Nenets (relating to Samoyeds too) that reliably differ from the other populations studied ($P<0.05$). As regards the other *c-fms* polymorphism (3'UTR), its rare variant CA was found in all four populations more evenly and with higher frequency varying from 35 % in Komi, 33 % in Forest Nenets and Selkups – to 25 % in Tundra Nenets.

Conclusion: Interpopulation difference comparative analysis of distribution of the *c-fms* gene deletion variant del425 demonstrated a reliable difference between the sample of Tundra Nenets inhabiting higher latitudes of the North and Forest Nenets, Selkups, and Komi samples. Therefore the participation of this variant in adaptation processes is not excluded. The was no reliable difference in *c-fms* 3'UTR polymorphism allele frequencies between the ethnic samples studied.

THE IMPACT OF TWO ABUNDANT TRANSPOSABLE ELEMENT FAMILIES ON GENOME DIFFERENTIATION IN POLYPLOID WHEAT

E.M. Sergeeva*, I.G. Adonina, A.B. Shcherban, E.A. Salina
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: sergeeva@bionet.nsc.ru
**Corresponding author*

Key words: *transposable element, DNA-transposon Caspar, LTR-retrotransposon Fatima, Triticum, Aegilops*

Motivation and Aim: The bread wheat *Triticum aestivum* is allohexaploid (BBAADD, $2n = 6x = 42$) with genome size 16Ч10^9 base pairs. The complete genomic sequence of this species is still unknown. According to different estimations, transposable element (TE) families of different abundance account for 60-70% of the wheat genome. TEs are likely to influence considerably the differentiation of genomes in the allopolyploid nucleus. The analysis of genomic organization and phylogenetic relationships of different TE families will allow us to make the inferences about evolution of genomes in the polyploid wheat and its diploid relatives. We took into analysis two TE families abundant in the wheat genome which belong to the TE classes with differing mechanisms of transposition: DNA-transposon *Caspar* and LTR-retrotransposon *Fatima*.

Methods and Algorithms: We combined two approaches: the phylogenetic analysis of the TE copies available in databases, and fluorescent *in situ* hybridization to the *Triticum* and *Aegilops* metaphase chromosomes to explore the chromosomal distribution of the TEs. The public nucleotide database mining for the TE sequences was conducted using consensus nucleotide sequences coding for transposase (for the DNA-transposon) or polyprotein (the LTR-retrotransposon) with BLASTn algorithm. The multiple nucleotide sequence alignments were performed by the ClustalW program; phylogenetic trees were constructed using the neighbor-joining method by the MEGA4 software package.

Results: We found the DNA-transposon *Caspar* to be localized predominantly in the subtelomeric chromosomal regions of the bread wheat and its diploid relatives. The LTR-retrotransposon *Fatima* is localized mainly on the chromosomes belonging to the B-genome of the polyploid wheats and on those of *Aegilops speltoides* (presumable donor of the genome B). The phylogenetic analysis showed that both the TE families formed distinct genome- and species-specific groups and their proliferation in genomes mainly took place during the divergence of the diploid *Triticum* and *Aegilops* species.

Conclusion: We demonstrated the impact of the DNA-transposon *Caspar* to the formation and differentiation of the subtelomeric chromosomal regions of different genomes in the polyploid wheat; and the impact of the LTR-retrotransposon *Fatima* to the differentiation of the B-genome of the polyploid wheats. Our results point the important role of the TE families in structuring and evolution of grass genomes.

Acknowledgements. We thank the Russian Foundation for Basic Research (08-04-00307) for financial support.

COMPARISON OF METHODS FOR RECONSTRUCTION OF MODELS FOR GENE EXPRESSION REGULATION

A.A. Shadrin[1,*], I.N. Kiselev,[1] F.A. Kolpakov[2,1]
[1]*Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia*
[2] *Institute of Systems Biology, Novosibirsk, Russia*
e-mail: inter.cm@mail.ru
[*]*Corresponding author*

Key words: *gene expression regulation, microarray experiment, experimental data smoothing, parameters optimization.*

Motivation and Aim: DNA microarray technology is widely used nowadays to gain data on gene expression; however, information obtained in this way can not be used directly and needs processing. It would facilitate the simulation of gene interactions and the reconstruction of the process under study, as well as give the opportunity to predict the development of the process at various changes in the conditions of its occurrence. The objectives of this study were: 1) consideration of linear and nonlinear [1] differential models of gene expression regulation, 2) study of various data processing methods on the models behavior, and 3) test and comparison of the considered models on the base of data published in [2].

Methods and Algorithms: for experimental data smoothing we used two different types of smoothing spline, nuclear smoothing with Epanechnikov core and the classic least squares smoothing method. During models parameters optimization we compared three methods: simulated annealing, evolutionary algorithm [3] and advanced method of gradient descent.

Results: two differential models of gene expression regulation were examined in details on the dataset containing yeast cell cycle associated genes expression, measured as amounts of mRNA using microarrays at 18 time points over two cell cycle periods, which was published in [2]. The effect of data smoothing method and parameters optimization algorithm were investigated. The best result for the data smoothing was obtained using a spline with a fixed weighting parameter, in the parameters optimization the best in accuracy and speed proved to be the evolutionary algorithm.

Conclusion: Although the study performed on the example of two models demonstrated that the spline with a fixed weighting parameter and the evolutionary algorithm were the most optimal, it also became apparent, that the choice of methods for data smoothing and parameters optimization could strongly influence the behavior of the model under other conditions, hence, each study requires individual approach to select the most (or more) optimal methods.

References:

1. Tra Thi Vu, Jiri Vohradsky (2007) Nonlinear differential equation model for quantification of transcriptional regulation applied to microarray data of Saccharomyces cerevisiae. *Nucleic Acids Research*, Vol. 35, No. 1.
2. Spellman,P.T., Sherlock,G., Zhang,M., Iyer,V.,Anders,K., Eisen,M., Brown,P., Botstein,D. and Futcher,B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol. *Biol. Cell,* 9, 3273–3297.
3. Thomas P. Runarsson, Xin Yao. Stochastic Ranking for Constrained Evolutionary Optimization. IEEE Transactions on evolutionary computation, vol. 4, No. 3, september 2000.

BIOMOSA: BIOLOGICAL FUNCTIONAL MODULE SEARCH ALGORITHM

P. K. Singh[1], Mohit Jha [1], Madhvi Shakya*[2]
[1]*Department of Bioinformatics MANIT Bhopal, India*
[2]*Department of Mathematics MANIT Bhopal, India*
E-mail: madhvishakya@yahoo.co.in
Corresponding author

Key words: *Functional modules, Interaction network, Subgraphs*

Identification of functional modules in protein interaction networks is the first step in understanding the organization and dynamics of cell functions. Revealing modular structures in biological networks will help us in understanding how cells function (Hartwell et al., 1999 Bork et al., 2004). Interaction networks can be partitioned into functional modules, which accomplish discrete biological functions in isolation from other modules in the networks (Hartwell et al. 1999). Our study includes how different ways of partitioned cellular processes relate to the general definition of modularity in terms of isolated functional entities.

Biological knowledge is represented by networks. These include regulatory and metabolic networks, protein-protein interaction networks, and many others. Exploration of these biological networks may extract meaningful information in term of functional modules from these biological networks. In this work new approach is adopted to integrate protein-protein interaction network datasets of various organisms to predict the functional modules. Here graph topological similarity search algorithms are taken into account to find similar subgraphs. These identified densely connected subgraphs are treated as functional modules, which are rigorously validated.

The complexity of the protein interaction network and regulatory network is very high, and also number of proteins may vary in different organisms. A novel algorithm framework is developed for fast comparison of complex interaction networks for various organisms. The algorithm demonstrates that results are biologically meaningful and relevant functional modules are obtained.

Conclusion: Algorithm demonstrates that method can accurately identify functional modules. Hence it carries the promise to be highly useful in analysis of interaction network in term of functional modules.

References

1. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. (1999) From molecular to modular cell biology, *Nature*, **402**: C47-C52.
2. Bork,P. et al. (2004) Protein interaction networks from yeast to human. *Curr. Opin. Struct. Biol.*, **14**: 292-299.

STRUCTURAL ORGANIZATION OF THE MALARIA MOSQUITO HETEROCHROMATIN: FROM CHROMOSOMAL MORPHOLOGY TO GENOME SEQUENCE

Maria V. Sharakhova[1], Phillip George[1], Irina V. Brusentsova[2], Scotland Leman[3], Christopher D. Smith[4], Igor V. Sharakhov[1*]

[1]Department of Entomology, Virginia Tech, Blacksburg, VA, USA; [2] Department of Molecular and Cellular Biology, Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia; [3]Department of Statistics, Virginia Tech, Blacksburg, VA, USA; [4]Department of Biology, San Francisco State University, San Francisco, CA, USA;
*e-mail:igor@vt.edu; *corresponding author*

Key words: *heterochromatin, malaria mosquito, imunostaning, Bayesian statistics*

Motivations and aim: Heterochromatin plays an important role in chromosome function and gene regulation. The goal of this study was to map and characterize the heterochromatic portion of the malaria mosquito *Anopheles gambiae* genome.

Methods and Algorithms: Imunostaining of *An. gambiae* polytene chromosomes was performed using antibodies against Heterochromatin Protein 1 (HP1) and lamin Dm0 *Drosophila melanogaster.* Physical mapping was done by Cy3, Cy5 labeled PCR amplified genes. The bioinformatics analysis was conducted using C program ATCONTENT, Biomart and SMARTest. We developed novel Bayesian statistical models to analyze genome landscapes in different types of chromatin. A gene ontology (GO) annotation file was generated using Interproscan-assigned GO terms and custom Perl scripts. Segmental duplications were detected using BLAST-based whole-genome assembly comparison limited to putative segmental duplications represented by pairwise alignments with ≥2.5-kb and >90% sequence identity.

Results: Based on chromosomal location, two types of heterochromatin have been identified in *An. gambiae* chromosomes: pericentromeric and intercalary. Intercalary heterochromatin was represented by two morphological forms: compact, block-like type and diffuse, mesh-like type. Pericentromeric heterochromatin in *An. gambiae* has a mesh-like structure. Immunostaining of the *An. gambaie* chromosomes with antibodies against HP1 and lamin Dm_0 identified the major invariable sites of protein localization in regions of diffuse heterochromatin and euchromatic region 9C but not in the compact intercalary heterochromatin. To determine the euchromatin-heterochromatin boundaries in the *An. gambaie* genome, we physically mapped genes to the polytene chromosomes: 16.6 Mb of the mapped portion in the *An. gambiae* genome has been identified as heterochromatin. Unlike *Drosophila*, *Anopheles* possesses three large regions (0.7, 0.8, and 2.9 megabase pairs long) of intercalary heterochromatin. The study determined 230 in the mapped portion of the *An. gambiae* heterochromatin. GO analysis revealed that heterochromatin is enriched by genes with DNA-binding and regulatory activities. Bayesian statistical models showed that heterochromatin and euchromatin differ in gene density and that all chromatin types are distinguishable from each other by the coverage of transposable elements. Conversely, segmental duplication, matrix-associated regions, and simple repeats are distributed regardless of the chromatin type. Finally, we have demonstrated that the 42 Mb assembly of unmapped scaffolds ("unknown" chromosome) have molecular characteristics of heterochromatin.

Conclusion: The study revealed the presence of large blocks of intercalary heterochromatin in *An. gambiae* genome. We conclude that each chromatin type has a unique combination of various classes of genes and repetitive DNA.

GENOME LANDSCAPE AND EVOLUTIONARY PLASTICITY OF CHROMOSOMES IN MALARIA MOSQUITOES

I.V. Sharakhov[1]*, A. Xia[1], M.V. Sharakhova[1], Z. Tu[2], S.C. Leman[3], J.A. Bailey[4], C.D. Smith[5]

[1]*Department of Entomology, Virginia Tech, Blacksburg, VA, USA;*
[2]*Department of Biochemistry, Virginia Tech, Blacksburg, VA, USA;*
[3]*Department of Statistics, Virginia Tech, Blacksburg, VA 24061;*
[4]*Program in Bioinformatics and Integrative Biology and Department of Medicine, University of Massachusetts Medical School, Worcester, MA 01605;*
[5]*Department of Biology, San Francisco State University, San Francisco, CA 94132.*
e-mail: igor@vt.edu
[*]*Corresponding author*

Key words: *genome landscape, chromosome evolution, rearrangements, inversions*

Motivation and Aim: Nonrandom distribution of rearrangements is a common feature of eukaryotic chromosomes that is not well understood in terms of genome organization and evolution. In the major African malaria vector *Anopheles gambiae*, polymorphic inversions are highly nonuniformly distributed among five chromosomal arms and are associated with epidemiologically important adaptations. However, it is not clear whether the genomic content of the chromosomal arms is associated with inversion polymorphism and fixation rates.

Methods and Algorithms: To better understand the evolutionary dynamics of chromosomal inversions, we created a physical map for an Asian malaria mosquito, *Anopheles stephensi*, and compared it with the genome of *An. gambiae* using the Nadeau and Taylor method and the GRIMM algorithm. We also developed and deployed novel Bayesian statistical models to analyze genome landscapes in individual chromosomal arms *An. gambiae.*

Results: Here, we demonstrate that, despite the paucity of inversion polymorphisms on the X chromosome, this chromosome has the fastest rate of inversion fixation and the highest density of transposable elements, simple DNA repeats, and GC content. Moreover, the sex chromosome was enriched in genes encoding for proteins responsible for signal transduction, which suggests their role in mating behavior and speciation. The highly polymorphic and rapidly evolving autosomal 2R arm had overrepresentation of genes involved in cellular response to stress supporting the role of natural selection in maintaining adaptive polymorphic inversions. In addition, the 2R arm had the highest density of regions involved in segmental duplications that clustered in the breakpoint-rich zone of the arm. In contrast, the slower evolving 2L, 3R, and 3L, arms were enriched with matrix-attachment regions that potentially contribute to chromosome stability in the cell nucleus.

Conclusion: These results highlight fundamental differences in evolutionary dynamics of the sex chromosome and autosomes and revealed the strong association between characteristics of the genome landscape and rates of chromosomal evolution. We conclude that a unique combination of various classes of genes and repetitive DNA in each arm, rather than a single type of repetitive element, is likely responsible for arm-specific rates of rearrangements.

RNA CONFORMATION AS A POSSIBLE EVOLUTIONARY CONSTRAIN.

Y.A. Darikova[1], E.A. Nikitina[1,3], D.Yu. Sherbakov[1,2]
[1]*Limnological Institute SB RAS, Irkutsk, Russia*
[2]*Irkutsk State University*
[3] *Irkutsk Regional Center for the Prevention and Fight against AIDS and infectious diseases*
e-mail: sherb@lin.irk.ru
**Corresponding author*

Motivation and aim: Systematic comparative study of non-coding nucleotide sequences at population level or in a set of closely sister species may shed new light on forces influencing evolutionary transitions of genomes other then the fates of polypeptide functions in case of protein-coding sequences. Here we compare patterns of micro-evolution of several introns in a species flock of Baikalian snails belonging to genus *Baikalia* to the patterns of evolution of a protein-coding gene from the same set of species/populations to show that transitions of tertiary structures of RNAs may be become important constrains on possible evolutionary transitions of some loci. Furthermore, we show that at least in one case of mRNA (virus of hepatitis E, HEV) the tertiary structure may play important role in determining it's essential biological properties and thus be of evolutionary importance.

Methods: Molluscan DNA sequences were obtained as described elsewhere (Darikova & Shcherbakov, 2009). Tertiary structure of RNAs were obtained using Vienna RNA package and MFold. Phylogenetic inferences and analyses were performed using package ape in R. HEV full genome sequences were obtained from Genebank.

Results: One of the introns of the phosphofructokinase gene has been sequenced in seven species of endemic Baikalian gastropods of family Baicaliidae. The length of this intron differs even between sister species. Along with relatively few nucleotide substitutions numerous deletions and insertions accumulated. The number of deletions/insertions does not correlate to genetic distance between the species as estimated by comparing sequences of a mitochondrial fragment. Deduced tertiary structure differs markedly even between sister taxa although there is a consensus motif. Long RNA stems appearing and disappearing in course of evolution may contain regulatory RNAs and therefore the evolution of the intron may be non-neutral. There is no significant correlation between mutational distances estimated for *pfk* and from protein coding *Cox*I for the same species. Two more introns were shown to evolve slower and more similar to *Cox*I.

Quantitative comparison of topologies of the consensus between the most probable conformations of paralogous RNAs shows that in case of non-coding sequences there are two distinctive patterns of evolution: conservation of a topology or it's dramatic changes due to relatively few base substitutions. In case of HEV RNA it was shown that the viruses differing by only non-coding substitutions but causing contrasting clinic manifestations differ by their conformations dramatically.

Conclusion: Conformation of RNA of complementary to both coding and non-coding DNA sequences may be independent and important evolutionary constrain and thus may be taken into account in phylogenetic and other inferences.

References:

1. Darikova JA, Shcherbakov D.Y. Mol.Biol. (Moscow) 2009,43(5):838-844.

ALGORITHM OF SEARCH FOR MIRNA IN SEPARATELY TAKEN BACULOVIRUS GENOME REGIONS

T.V. Shirina*, M.T. Bobrovskaja, E.A. Kozlov
Institute of Molecular Biology and Genetics of NAS, Kiev, Ukraine
e-mail: e.a.kozlov@imbg.org.ua
**Corresponding author T.V. Shirina*

Key words: *bioinformatic, baculoviruses, microRNA, algorithm.*

Motivation and Aim: Bombyx mori (bmo) and Autographa californica (aca) nuclear polyhedrosis viruses (NPVs) contain very late genes *ph*, *p10* and complementary to them genes *orf1629*, *p74* respectively. Transcription of *ph* down regulates expression of *orf1629* [1]. We supposed that expression regulation of *orf1629* is realized by microRNAs. The present work was aimed at developing algorithm of search for miRNA in alternative transcripts (alts) synthesized from separately taken baculovirus genome regions. *Methods and Algorithm:* The search of alts into bmo and aca genome regions containing *ph*, *p10*, *orf1629*, *p74,* search stem-loop structures (sls) into alts, RNAfold [2], Microprocessor SVM [3], miPred [4], miRNA SVM [3], mirScan [5] and miRBase [6]. *Results:* With the help of this algorithm 1348 alts were searched to fined unic sls. Of 1348 alts in question 626 belong to bmoNPV and 722 to acaNPV. Of 626 bmoNPV alts 68 contain 73 sls and of 722 acaNPV alts 55 contain 58 sls. Four of 73 bmoNPV sls passed filters of bioinformatic programs mentioned above and are processed into mature miRs (bmoNPV-miR-1ph, - 2ph, - 3p10, - 4p74) while three of them don't passed only filter miRNA SVM. Therefore the sls sequence for these three miR should be annotated as a miRNA candidate (C): bmoNPV- pre-miR-1Cph, - 2C orf1629, -3Cp74. Similarly, were obtained the results for acaNPV sls. That is acaNPV-mir-1ph, - 2ph, -3p10, - 4p74, - 5p74, - 6p74 and acaNPV - pre-miR-1Cph, -2Cph, -3Cph, -4Corf1629, -5Cp74. *Conclusion:* Involvement of miRs in question to reciprocal regulation of genes above-listed is discussed. *Availability:* The algorithm may be valuable to search miRs in genome of different organisms.

References:

1. Ooi B. G. and Miller L. K. (1990) Transcription of the baculovirus polyhedrin gene reduces the levels of an antisense transcript initiated downstream, *J. Virol*. **64**: 3126-3129.
2. Zuker M. (2003) Mfold web server for nucleic acid folding and hybridization prediction, *Nucleic Acid Res*. **31**: 3406-3415.
3. Helvik S. A., Snove O., Saetrom P. (2007) Reliable prediction of Drosha processing sites improves microRNA genes prediction, *Bioinformatics*. **23**: 142-149.
4. Xue C., Li F., He T., Liu G. P., Li Y., Zhang X. (2005) Classification real and pseudo microRNA-precursor using local structure-sequence features and support vector machine, *Bioinformatics*. **6**: 310-316.
5. Lim L. C., Lau N. C., Weinstein E. G., Abdelhakim A., Yekta S., Rhoades M. W., Burge C. B., Bartel D. P. (2003) The microRNAs of Caenorhabditis elegans, *Genes Develop*. **17**: 991 - 1008.
6. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. (2008) miRBase: tools for microRNA genomics, *Nucleic Acids Research*. **36:** D154-D158

IDENTIFICATION OF PROTEIN BINDING MOTIFS IN RNA CONSIDERING BOTH PRIMARY SEQUENCE AND SECONDARY STRUCTURE

D.N. Shtokalo*[1], A.Y. Palyanov[1,2], D. Babiy[1], M. Heydarian[3], G.St. Laurent[4]
[1]Novel Software Systems, Novosibirsk, Russia;
[2]Ershov Institute of Informatics Systems SB RAS, Novosibirsk, Russia;
[3]Johns Hopkins School of Medicine, Baltimore, MD, USA
[4]Brown University, Providence, USA
e-mail: *dmitry@nprog.ru*
*Corresponding author

Key words: *Protein-RNA interaction, primary sequence and secondary structure motif, HuR protein*

Motivation and Aim: Protein-RNA interactions mediate important processes in cells such as posttranscriptional regulation of gene expression (Good, 1995). Identification of protein attracting motifs in RNA lead to both primary sequence and secondary structure features search problem (Lopez de Silanes et al., 2003). Multi alignment of primary sequences together with secondary structures and stochastic context-free grammar modeling are efficient solutions when search motif in a set of dozens of RNA sequences (Havgaard et al. 2005, E.R.Sean et. al. 1994). However increase of sample size lead to hard computational problems. We propose a relatively fast motif identification approach that was tested on a set of HuR-binding RNA transcripts.

Methods and Algorithms: HuR-RNA complexes were immunoprecipitated and 2369 HuR-bound target RNAs as well as 3750 HuR-rejected RNAs were identified on Affymetrix Human Exon 1.0 ST array. HuR-binding RNA set which is AU-rich and HuR-rejecting RNA set which is GC-rich were scanned with RNAFold program to get locally stable secondary structure units. Secondary structure units were clustered according to shape and a cluster prominently abundant in HuR-binding RNA was found. This cluster consists of structures that are two hairpins in 75nt window. A specially designed algorithm SWORD was applied to find statistically overrepresented and underrepresented words in nucleotide sequences of the cluster.

Results: We found a motif: two hairpins in 75nt window with DDH-WW-DHH-WWW-Y-WW-R word present and SVB-BH-VBS-BBS-VV-N-BB word absent in nucleotide content where D=A|U|G, H=A|U|C, W=A|U, Y=U|C, R=A|G, S=G|C, V=A|G|C, B=U|G|C. This motif hits 51% of 3'utrs of HuR-binding transcripts and 5% of HuR-rejecting transcripts. The motif is found in 23% of transcriptome 3'utrs. Motif presence was checked in other sets of HuR-binding RNAs.

Conclusion: An approach of identification of both primary sequence and secondary structure features of RNA targeted by protein is proposed. The method reveals novel HuR-binding motif present in half of HuR targets obtained from ChIP-chip data. Potentially HuR may bind to different motifs (Bolognani et al., 2010) and the search should be continued in the rest of targets.

References:

1. P.J.Good (1995) A conserved family of elav-like genes in vertebrates, *Proc. Natl. Acad. Sci.*, **92**:4557-4561.
2. I.Lopez de Silanes, et al. (2003) Identification of a target RNA motif for RNA-binding protein HuR, *Proc. Natl Acad. Sci.,* **101:**2987-2992.
3. J.H.Havgaard, et al. (2005) The FOLDALIGN web server for pairwise structural RNA alignment and mutual motif search, *Nucleic Acids Research,* **33**:W650-W653.
4. E.R.Sean, R.Durbin (1994), RNA sequence analysis using covariance models, *Nucleic Acids Research*, **22 (11):** 2079-88.
5. F.Bolognani, et al. (2010) Novel recognition motifs and biological functions of the RNA-binding protein HuD revealed by genome-wide identification of its targets, *Nucleic Acids Res.* **38(1)**:117-30

MODIFIED PSWM FOR RNA EDITING SITES SEARCH

D.N. Shtokalo*[1], S.S. Nechkin[1,2], T.Y. Eremina[1], E.S. Cheremushkin[1,2], R. Reenan[3], G.St. Laurent III[3]

[1]*Novel Software Systems, Novosibirsk, Russia*
[2]*Ershov Institute of Informatics Systems SB RAS, Novosibirsk, Russia*
[3]*Brown University, Providence, USA*
e-mail: dmitry@nprog.ru
[*]*Corresponding author*

Key words: *ADAR editing, context score, position weight matrices, PSWM, RNA editing*

Motivation and Aim: In RNA sequence ADAR ferments substitute adenosine (A) with inosine (I) which is read as guanosine (G). ADAR editing phenomenon increases transcriptomic diversity, regulates alternative splicing and RNA stability[1]. The highest ADAR activity was detected in brain and many ADAR editing sites are located within genes related to human brain diseases[2]. There are computational methods for ADAR editing sites search based on knowledge that ADAR acts in dsRNA[1]. Nucleotide context of ADAR editing sites was shown to have a pattern[3]. We study nucleotide context of ADAR editing sites for the purpose of novel sites recognition and reveal patterns in protein-RNA interactions.

Methods and Algorithms: The context of ADAR editing sites doesn't have enough positions with prominent nucleotide frequencies therefore more sensitive than Position-Specific Weight Matrices (PSWMs) model should be used to recognize site in a given sequence. The modification of PSWM model that consider correlations between nucleotides in different positions, called "correlation matrix" (CM), was developed. CM consist of positive (-log(p-value)) and negative (log(p-value)) scores calculated for statistically overrepresented and underrepresented occurrences of pairs of nucleotides in every two positions in context sequences. CM was constructed using known human ADAR editing sites [2] taking +/-20 nt context. Sum of FP and FN error rates for recognized sites was minimized to FP=0.195 and FN=0.259. To make sure CM model is not overfitted we check it loose recognition power if substitute learning sample of real sites with non-sites.

Results: The constructed CM was applied to filter possible ADAR editing sites in brain deep sequencing data obtained with Helicos sequencer. 950 ADAR editing sites were found that have A-to-G evidence in both deep-seq and EST data, and have good Context Score.

Conclusion: The fast computational method for ADAR editing sites search is created and applied to discovery of potential sites in deep sequencing data. The proposed method modifies classical PSWM model and reveals nucleotide pattern in protein-RNA interaction.

Availability: Correlation matrix http://nprog.iis.nsk.su/supplementary1.doc, novel ADAR editing sites http://nprog.iis.nsk.su/supplementary2.xls

References:

1. E.Y. Levanon, et al. (2004) Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat Biotechnol*, **22**: 1001-1005.
2. J.B. Li, et al. (2009) Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science*, **324**: 1210-1213.
3. T.R. Dawson, et al. (2004) Structure and sequence determinants required for the RNA editing of ADAR2 substrates. *J Biol Chem*. **279**: 4941-4951.

GENE NETWORKS MODELLING. LIMITING TRANSITIONS IN PROCESSES OF SYNTHESIS.

D.N. Shtokalo*[1,3], V.A. Likhoshvai[1,2], S.I. Fadeev[1,3]
[1]*Novosibirsk State University, Novosibirsk, Russia;*
[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia;*
[3] *Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia;*
e-mail: dmitry@nprog.ru
[*]*Corresponding author*

Motivation and Aim: Live of a cell stand for permanent synthesis of proteins, RNAs, DNAs and interactions of internal cell products and external compounds. Genes being DNA regions encode internal cell products such as RNAs and proteins thus underlie main cellular processes: development, amplification, specialization, response on external signals. To understand cellular processes, networks of genes and their products interactions have to be studied. Mathematical modeling of gene networks lead to hard resolve problems. Modeling of several hierarchical levels at once multiply dimension and complexity of system of equations by many times. The question we ask in this work, proceeding with [1], is how crucial to consider elementary reactions taking place in subsystems for adequate modeling of supersystem. In other words, can we ignore elementary biochemical reactions on matrix stages of elongation during polymer molecule synthesis when model networks of complete molecules?

Methods and Algorithms: The simplest gene network that consist of one self-regulating element $y = y(n,t)$ was modeled with a system of ordinary differential equations of dimension $n+1$:

$$\begin{cases} \frac{dx_1}{dt} = f(t,y) - f^+_{n,1}(x_1) + f^-_{n,2}(x_2) - s_{n,1}(x_1) \\ \frac{dx_{i+1}}{dt} = f^+_{n,i}(x_i) - f^+_{n,i+1}(x_{i+1}) - f^-_{n,i+1}(x_{i+1}) + f^-_{n,i+2}(x_{i+2}) - s_{n,i+1}(x_{i+1}) \ , \ i = \overline{1,n-2} \\ \frac{dx_n}{dt} = f^+_{n,n-1}(x_{i-1}) - f^+_{n,n}(x_n) - f^-_{n,n}(x_i) - s_{n,n}(x_n) \\ \frac{dy}{dt} = f^+_{n,n}(x_n) - g(y). \end{cases},$$

where x_i - intermediate stages of synthesis of product y, $f(t,y)$ - speed of synthesis initiation, $f^+_{n,i}(x_i)$ – speed of direct process that translate substance from stage i to $i+1$, $f^-_{n,i}(x_i)$ – speed of reverse from stage i to $i-1$, $s_{n,i}(x_i)$ - sinks, $g(y)$ – product utilization. We put $f(t,y) = \bar{f}(y) = \alpha/(1+\beta y^\gamma)$, $f^{+/-}_{n,i}(x_i) = n/\tau_i^{+/-} \cdot x_i/(1+\rho x_i^\xi)$, $s_{n,i}(x_i) = \omega_i x_i$, where $\alpha \ge 0, \beta > 0, \gamma > 0$, $\tau_i^+ > 0, \tau_i^- > 0, \rho \ge 0, \xi > 0, \omega_i \ge 0$ – parameters.

Results: It was discovered numerically that under particular conditions solution y of the system is approached by the solution z of the delayed differential equation $dz(t)/dt = e^{-d} f(t-\tau, z(t-\tau)) - g(z(t))$, where $\tau > 0$ – delay, $\tau > 0$ – sink defect. In some cases limiting theorems has been proofed i.e. if $\tau_i^+ = \tau^+, \tau_i^- = \tau^-, \rho = 0, \omega_i = \omega$ and $\tau^- > \tau^+$ then $\max_{t<T} |z(t) - y(n,t)| \to 0$, $T > 0$, where $\tau = \frac{\tau^-\tau^+}{\tau^- - \tau^+}$, $d = \tau\omega$.

Conclusion: If number of elementary reactions is big enough and direct process is faster than reverse process then model may ignore details and time of every single reaction and consider only integral time and integral sink that appear as delay and defect respectively in corresponding model with delayed equation. Limiting transition from the system to the delay equation bears slight parameters variations and variation of initial data. Theorems put a number on connection of micro and macro levels in live systems.

References:

1. V.A.Likhoshvai, et al. (2004) Modeling by a Delay Equation of Multi-Stage Synthesis of a Sample Without Bifurcation., *Sib. J. of Industrial mathematics,* **7(1):** 73-94.

REGULATORY EFFECTS OF FIBROBLASTIC AND VASOENDOTHELIAL GROWTH FACTORS IN EXPERIMENTAL MYOCARDIAL INFARCTION

M.G. Shurygin*, I.A. Shurygina, N.N. Dremina, O.V. Kanya
Scientific Center of Reconstructive and Restorative Surgery SB RAMS, Bortsov Revolutsii 1, 664003, Irkutsk, Russia
e-mail: shurygin@rambler.ru
**Corresponding author*

Key words: *VEGF, FGF, myocardial infarction*

Motivation and Aim. The aim is to investigate regulatory effects of vasoendothelial (VEGF) and fibroblastic growth factors (FGF) in a reparation process by an example of experimental myocardial infarction.

Methods and Algorithms. Female Wistar rats (N=126, weight 220-250 g) at the age of 9 month were selected in accordance with the "Guide for the Care and Use of Laboratory Animals" (NIH publication 85-23, revised 1985). The experiment was carried out in accordance with the principles of humane treatment of animals set out in the directives of the European Community (86/609/EES) and the Helsinki Declaration. We used myocardial infarction model with diathermocoagulation of coronary artery. The study comprised three groups of animals: "FGF" (N=41), "VEGF" (N=40), "Control" (N=45). Animals from "FGF" group had 100 ng FGF2 (Sigma F0291 Lot 124K0797), animals from "VEGF" group had 100 ng VEGF164 (Sigma V3638-10UG Lot 064K1239) injection into left ventricle at a 1.5 h period after operation. We used methods of light microscopy, immunohistochemistry, morphometry.
Results. We found out that the high FGF2 or VEGF level extended infiltrative phase of inflammation. FGF2 has cytoprotective effect on endotheliocytes. We have determined that the increasing of FGF2 level takes effect on the early collagen synthesis in infarction and board regions (from 3 days in group with FGF2 injection vs 7 days in control group).

VEGF also have increased the synthesis of collagen. The collagen volume on 30 day in "VEGF" group was higher than in "Control" group (58.56 vs 40.51%, $p<0.05$). At the same time, excessively high concentrations of VEGF led to the development of focal cardiosclerosis in the intact myocardium, which is a by - effect of this growth factor. VEGF provided a significant effect on angiogenesis. We were the first who found that the injection of intracardiac VEGF experimental animals had increased survivability of cardiomyocytes in the infarction zone.

Conclusion. Thus, VEGF and FGF are important regulators of the reparation process in a myocardial infarction.

Acknowledgements. This study was supported by a grant number P1250 from the Russian Federal Purposive Program "Scientific and scientific-pedagogical cadres innovation Russia" for 2009-2013.

INGIBITOR OF p38 MAPK AND INFLAMMATION AFTER SURGICAL WOUND

I.A. Shurygina*, M.G. Shurygin, N.V. Zelenin, G.B. Granina, S.A. Lepekhova
Scientific Center of Reconstructive and Restorative Surgery SB RAMS, Bortsov Revolutsii 1, 664003, Irkutsk, Russia
e-mail: irinashurygina@gmail.com
**Corresponding author*

Key words: *mitogen activated protein kinase, p38, inflammation*

Motivation and Aim. To study the role of p38 MAP(mitogen activated protein)-kinase mechanisms in the regulation of inflammatory response and on this basis to develop ways to control inflammation.

Methods and Algorithms. Male Wistar rats (N=60, weight 220-250 g) at the age of 9 month were selected in accordance with the "Guide for the Care and Use of Laboratory Animals" (NIH publication 85-23, revised 1985). The experiment was carried out in accordance with the principles of humane treatment of animals set out in the directives of the European Community (86/609/EES) and the Helsinki Declaration. The study was approved by the Local Ethics Committee. We studied the effect of an inhibitor of p38 MAP-kinase SB203580 in the development of inflammatory response in models of skin-muscle wounds. We used methods of light microscopy, immunohistochemistry, tensiometry.

Results. While reviewing the literature [1, 2, 3], we hypothesized that the development of inflammatory response involved mainly JNK (c-Jun N-terminal kinase) and p38 MAP-kinase cascade. Our study confirmed this hypothesis. We revealed that the inhibition of p38 MAP-kinase cascade significantly affects the duration and severity of the phases of inflammation, modifies the formation of connective tissue scar in the wound process.

Conclusion. The results show the importance of studying regulators of cell growth and differentiation as potential drugs significantly affecting the outcome of the pathological processes.

Acknowledgements. This study was supported by a grant number P803 from the Russian Federal Purposive Program "Scientific and scientific-pedagogical cadres innovation Russia" for 2009-2013.

References:

1. U. Oltmanns, R. Issa, M. Sukkar et al. (2003) Role of c-jun N-terminal kinase in the induced release of GM-CSF, RANTES and IL-8 from human airway smooth muscle cells, Br. J. Pharmacol., 6: 1228-1234.
2. M. Li, D. Georgakopoulos, G. Lu et al. (2005) p38 MAP kinase mediates inflammatory cytokine induction in cardiomyocytes and extracellular matrix remodeling in heart, Circulation, 111: 2494-2502.
3. W.I. Boer (2005) Perspectives for cytokine antagonist therapy in COPD, DDT, 2: 93-106.

PREDICTION OF VIRAL MICRORNAS USING CLASSIFICATION BASED MACHINE LEARNING APPROACH

Vaibhav T.[1], AlexGopanenko K.[2], SivaPrasad C.V.S.[1,*]
1-Indian Institute of Information Technology, Deoghat, Jhalwa, Allahabad,India and
2-Dr.Sci. Principle investigator, Koltzov institute of Developmental Science, Russian Academy of Science, Moscow, Russia
e-mail: shiva@iiita.ac.in
** Corresponding author.*

Motivation and Aim: MicroRNAs are known as strong candidates for gene expression regulation. In case of viral genomes, identification of miRNAs by using experimental methods is quite tedious and expensive. Hence in silico prediction of viral precursor miRNAs (pre-miRNAs) has became as a major inexpensive choice, but Very little work has been carried out in the areas of prediction of mature miRNAs [1]. The aim of our study is to develop a classification based machine learning approach (SVM) for prediction of viral mature miRNAs.

Methods and Algorithms: This work has been developed on machine learning based model, and it can recognize mature miRNAs in the given pre-miRNA stem loop secondary structure. The pre-miRNA stem loop structures from the training dataset were collected up from the MIRBASE database and we stored the experimentally validated precursor microRNAs along with the mature miRNAs for a list of viruses available in the database [2]. miRNAs from the stem loop structures of pre-miRNAs were extracted keeping the secondary structures (bulges and loops) intact. A set of features from these datasets were extracted in the numerical form and stored as a training dataset. These numerical values with class labels (+1 for positive set and -1 for negative set) are used as inputs for a Support Vector Machine model (SVM) [3]. The prediction efficiency was determined by a 5-fold cross-validation or more commonly the leave-one-out cross-validation (LOOCV) procedure. Genome wide search was carried out to predict the miRNAs from three viruses [4].

Results: A preliminary test, keeping most of the negative set data unused, to determine the efficacy of the study showed 90% accuracy. Later a complete exhaustive set of negative data was applied to the SVM model (in order to keep all the experimental data available to us), and the accuracy was found to be 85.3%. We also carried out the predictions in three viral genomes and showed a list of mature miRNAs. Our method showed a better accuracy than other existing mature miRNA prediction methods. It is the first program developed to predict the mature micro-RNAs by using viral genomes.

Conclusion: Proposed model helps to predict a handful of miRNAs from the viral genomes, which may help to find out various gene interactions in pathogenicity.

Reference:

1. Ying Sheng et al. Mammalian MicroRNA Prediction through a Support Vector Machine Model of Sequence and Structure, September 7, 2007, PlosOne.
2. http://www.mirbase.org/ hosted in the Faculty of Life Sciences, University of Manchester.
3. A guide for beginners: C.-W. Hsu, C.-C. Chang, C.-J. Lin. A practical guide to support vector classification.
4. Shiva Kumar et al. Prediction of viral microRNA precursors based on human microRNA precursor sequence and structural features. August 20, 2009, Virology Journal.

CORRELATIONS BETWEEN IS6110 RFLP GENOTYPES AND DRUG RESISTANCE OF MYCOBACTERIUM TUBERCULOSIS STRAINS

A.Yu. Sivkov*, S.I. Tatkov, V.A. Mordvinov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
sivkov@bionet.nsc.ru
**Corresponding author*

Key words: *(Mycobacterium tuberculosis, MDR, XDR genotyping, IS6110 RFLP.)*

Motivation and Aim: Multidrug-resistant tuberculosis (MDR-TB) and Extensively Drug-Resistant (XDR-TB) are a life-threatening form of TB affecting half a million people every year. M/XDR TB is more difficult and costly to treat than drug-susceptible TB (1). Number of MDR-TB patients in Novosibirsk region was constantly increased last years. Analysis of correlations between genetic and drug resistant characteristics of M. *tuberculosis* strains could shed a light on the reasons of MDR-TB strain dissemination. Aim: Study of Novosibirsk region TB strain samples collection using IS6110 RFLP genotyping and analysis of correlations between genetic and drug resistant characteristics of strains under investigation.

Methods and Algorithms: Strains of *M. tuberculosis* were isolated from primary TB patients, residences of Novosibirsk region, who were registered as TB patients since 1 January 2003 to 31 December 2003. Drug resistances were determined by routine microbial methods and independently by homemade biochips. For IS6110 RFLP genotyping (2) 230 strains were selected randomly from 607 isolates of *M. tuberculosis* in our collection. Cluster and statistical analysis of IS6110 RFLP data were performed by the BioNumerics software program (Applied Maths) and Statcalc module of Epiinfo software.

Results: 173 strains of *M. tuberculosis* were genotyped with IS6110 RFLP. A phylogenic tree was created with the help of UPGMA procedure. Tree consists of 157 genotypes; 24% of all strains were distributed into 16 clusters, containing 2-7 isolates. In the phylogenic tree the branch of Beijing genotype of M. *tuberculosis* was detected by IS6110 RFLP and spoligotyping methods. In our collection MDR-strains were detected in 21,4%. In the Beijing branch one cluster (28 isolates) included 60,7% MDR isolates significantly more than in others ($p<0,00001$, chi-square criteria). For this genotype MDR-TB risk factor is in 9,96 times than for any others RFLP genotypes.

Conclusion: It is well known that insertion elements IS6110 can change the local genome organization and its functioning. Strains, belonging to Beijing MDR branch, have from 10 to 25 IS6110 elements and unique patterns of their distribution in genome. We propose that specific location of IS6110 elements in MDR strains of M. *tuberculosis* contributes to preferential accumulation of mutations leading to resistance.

References:

1. World Health Organization (2009) A ministerial meeting of high M/XDR burden countries, 1-3 april 2009, Beijing, China, , Page 2
2. Cohn DL, O'Brien (1998) The use of restriction fragment length polymorphism (RFLP) analysis for epidemiological studies of tuberculosis in developing countries, Int J Tuberc Lung Dis. Jan;2(1):16-26.

BUILDING OF CELL-ORIENTED MODELS OF REGULATION OF STRUCTURE OF 2-DIMENSIONAL GROWTH CELL TISSUE

Pavel Smal*, Sergey Nicolaev
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: pasha.smal@gmail.com
**Corresponding author*

Key words: *shoot apical meristem, mathematical model, .DL-system formalism*

Motivation and Aim: Studying of mechanisms of regulation cellular structure of tissue is one of important task in theory of morphogenesis. And mathematical modeling is important instrument solving this problem.

However, the little number of experimental data make difficulties for modeling. Other difficulty is various character of processes in tissue. We consider that chemistry-kinetic processes and mechanic deformation are continuous, and cell division is discretely process. This supposition bring us to making discrete-continuous model and versatility system for modeling.

Methods and Algorithms: In our model we consider that cell is a polygon with internal edges. External edges correspond cell's walls, and internal edges offer cytoskeleton. In given singly connected area we built Voronoi diagram and suppose that Voronoi polygons correspond cells. To compute mechanical deformation with finite elements method we consider received cell ensemble as flat articulate truss composed of elastic shank.

In consideration chemistry-kinetic processes we consider that diffusion inside cell much quickly then diffusion between cells. And mathematical model this process presented by pseudoviscosity method for system of differential equation solvable by semi-implicit method.

For description cell division we use L-system. It's known that L-system is a good formalism for description line or tree structure. We offers method of description 2-dimensional cell tissue in terms of dL-system. Alphabet consist of two type of modules *Cell* and *Edge*. String of these modules describes geometric structure of tissue. And rewriting rules have enough natural form.

Results: Consideration chemistry-kinetic model by the example of model of shoot apical meristem of *Arabidopsis* reveal problems in choosing model parameters. But solution identification problem allows us to find parameters corresponding biologically distribution of morphogens.

Availability: We expect that built system for modeling allow to verify different complex models and facilitate process of modeling. And building different models could help us to fully appreciate processes of morphogenesis.

References:

1. S.V. Nikolaev, A.V. Penenko, V.V. Lavreha, E.D. Mjolsness, N.A. Kolchanov. (2007) A Model Study of the Role of Proteins CLV1, CLV2, CLV3, and WUS in Regulation of the Structure of the Shoot Apical Meristem, *Russian Journal of Developmental Biology*, **38:** 383–388.
2. P. Prusinkiewicz, A. Lindenmayer (1990) The algorithmic beauty of plants. *N.Y.:Springer-Verlag*

RED AND FAR-RED LIGHT SIGNAL TRANSDUCTION IN EARLY PLANT DEVELOPMENT

O.G. Smirnova
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: planta@bionet.nsc.ru

Key words: *gene networks, photoreception, signal transduction, phytochrome*

Motivation and Aim: Light is one of the most important regulators of plant development. Phytochromes A-E (Phys) are photoreceptors responsible for red/far-red light perception. Phys measure red to far-red light ratio, allowing the plant to assess the quantity of active light for photosynthesis. Phys control plant development by regulating expression patterns of approximately 10% of the plant transcriptome. Elucidation of the molecular mechanisms underlying Phy-controlled signal transduction has therefore attracted considerable attention.

Methods and Algorithms: The GeneNet Input interface specially designed for this database allows an automatic direct translation of the input annotated information into the GeneNet database format. Formalized data on the structure–function organization of the gene network is displayed as a diagram [1].

Results: Phy-controlled signaling is a complex cellular process. It starts with the light-induced conformational change of Phys and includes regulated partitioning and degradation of signaling components and of the photoreceptors themselves. Phosphorylated phyA accumulates in the nucleus and preferentially associates with the COP1/SPA1 complex. In contrast, underphosphorylated PhyA predominantly associates with the PhyA-signaling intermediates FHY3 and FHY1, which protect PhyA from being recognized by the COP1/SPA complex [2]. The gene network includes functioning of a COP1/SPA1 repression protein complex in darkness, light activation and nuclear partition of Phys, deactivation of the repression complex in light, and formation of a response in the form of inhibition of hypocotyl growth, development of cotyledons and chloroplasts, and anthocyanin accumulation with participation of LAF1, HFR1, HY5 and PIFs transcription factors.

Conclusion: The gene network reconstruction and predictive analysis provide insight into the molecular-genetic aspects of the plant development under red/far-red light conditions. It was shown that along with positive regulation, negative regulation plays an important role in photomorphogenesis.

Availability: The diagram of the gene network regulating of plant photomorphogenesis is available through the GeneNet viewer at http://wwwmgs.bionet.nsc.ru/systems/mgl/genenet

Acknowledgements: This work was supported by RF Ministry of Science and Education (grant 2.1.1/6382), SB RAS Integration Project (no. 28) and RAS Program “Biodiversity” (no. 23.22).

References:

1. E.A Ananko et al. (2005) GeneNet in 2005. *Nucleic Acids Res.,* **33**: 425-427.
2. Y.Saijo. (2008) Arabidopsis COP1/SPA1 complex and FHY1/FHY3 associate with distinct phosphorylated forms of phytochrome A in balancing light signaling. *Mol. Cell*, **31**: 607-613.

MODELING OF THE INDIVIDUAL STRUCTURAL TEMPLATE OF PROTEIN ON DETERMINING IT NUCLEOTIDE SEQUENCES

V.V. Sokolik
SI « Institute of neurology, psychiatry and narcology of the AMSU», Kharkov, Ukraine
e-mail: sokolik67@rambler.ru

Key words: *a configuration of peptide bond, genetic code of spatial structure of protein, personal structural templates of major actin and α-actin 1*

Motivation and Aim: The novel way of modeling of spatial structure of protein on determining it nucleotide sequences is elaborated. On analogies with rotamers on C—C σ-bound a concept configurations of peptide bound (CPB) is entered as basic element three-dimensional structure of protein, encoded in a genome along with amino acid.

Methods and Algorithms: On ringlateral models of polyalanine it is shown, those three discrete variant configurations of peptide bound (R, 0 and L) determine three different second structures of polypeptide chain: the right spiral, the β-strand and the left spiral. On the basis of the empirical table of a composite genetic [1] code the author of the given research had been made the **table of a genetic code of three-dimensional structure of protein**. The important advantage of a method is that it is possible to construct a structural pattern individually for any unknown protein only "having read through" determining it nucleotide sequence in according with the table of a genetic code of spatial structure of protein, and also to decode presence and position of fragments of secondary structure of it.

Results: In this worker were modeled a structural templates of two homologues major actin and α-actin 1. Its visualized is possible in the graphic editor Ggenedit.exe.

Conclusion: Thus, two homologous proteins (the major actin and α-actin 1) the identical size (375 a.a.) with identity degree amino acid sequences of 91 % are characterized only 35 % of similarity of sequences of a CPB. It means that such homologues possess various three-dimensional structures and about any general template for its modeling of speech cannot go.

References:

1. Kushelev A.U., Polischuk S.E., Nedel'ko E.V. (2002) Construction of scale model of structure of protein. Issues of the day of modern science, 2: 236—240.

FUNCTIONAL APPROACH FOR MODELING OF CELL VOLUME REGULATION IN HYPOTONIC MEDIUM (RVD).

E.I. Solenov, A.P. Ershov, D.A. Medvedev, D.I. Karpov, A.V. Ilyaskin, G.S. Baturina
Institute of Cytology and Genetics SB RAS, Lavrentyev Institute of Hydrodynamics SB RAS, Novosibirsk State University, Novosibirsk, Russia
e-mail: eugsol@bionet.nsc.ru

*Motivation and Aim***.** Cells contact with hypotonic fluid which osmolality varies significantly and the cells continually expose to osmotic stress. To avoid damage and perform their functions cells require an effective volume regulation mechanism. Despite its importance, very little is known about cell-volume regulation.

The purpose of this study was mathematical modeling of the volume-regulatory response in the principal cells of rat kidney collecting duct (CD) epithelia during acute swelling in hypotonic medium.

Methods. The model is based on experimental data obtained in experiments with principal cells of micro dissected CD fragments from rat kidney. Hypotonic shock was created by PBS diluted with 50% of water. Changers in cell volume were measured with calcein quenching method. Digital integration of the equations was done by Runge Kutta method of 4^{th} order.

Results. The model created as a system of differential equations based on the equations of Goldman, Hodgkin, Katz. System includes equations for calculation of ion currents across cell membrane, including currents of sodium and potassium through Na,K-ATPase, and control function defines the values of permeability cell membrane for water and ions during the process of RVD. It is shown that the model could be adjusted to a good agreement with experimental results.

*Conclusion***.** Supposition is made that parameters of control function reflect the time course of regulatory processes of RVD in real cell.

ANALYSIS OF PROTEIN LENGTH ISOFORMS CO-OCCURRENCE IN PROKARYOTES FOR SINGLE-DOMAIN PROTEINS

P.A. Starostina*[1], D.A. Afonnikov[1,2], K.V. Gunbin[1]
[1]- *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]- *Novosibirsk State University, Novosibirsk, Russia*
e-mail: spolina@bionet.nsc.ru
[*]*Corresponding author*

Key words: *gene duplication, protein co-occurrence, isoform length, bacterial genomes*

Motivation and Aim: Gene duplication is one of the basic steps of changing of protein function and evolution of genetic networks [1, 2]. Some of the proteins after duplication can acquire short peptides at their termini that could be important for new function. Example is two proteins, YfiA and HPF, from *Escherichia coli* that mediate transition of the cell from growth to stationary phase [1]. They have conserved RNA binding domain with 40% sequence similarity and differ in length by 17 amino acids. Despite their similarity they have opposite functions in 70S ribosomal particle formation by the mechanism of concurrent binding [3]. These proteins tend to co-occur in bacterial genomes. It is believed that such systems generated by gene duplication can serve as fundamental blocks of biochemical adaptation mechanisms [4].

Methods and Algorithms: We implement the approach for finding such type of genetic systems based on homologous proteins length isoforms co-occurrence in bacterial genomes. The method combine protein length isoform identification in the sets of homologous sequences from bacterial genomes detected by BLAST homology search with NCBI CDD and Pfam domain identification.

Results: We analyzed set of 706 prokaryotic genomes. For the set of single-domain proteins represented in E.coli we identified homologs with distinctive length isoforms (corresponding to high peaks in length distribution histogram). For about of 600 of them we identified pattern of isoform co-occurrence in genomes that satisfy the 0.01 threshold according to the Fisher exact test. The COG class enrichment analysis based on permutation test showed that these co-occurred isoforms overrepresented in “Amino acid transport and metabolism”, “Transcription” and “Defense mechanisms” COG classes.

Conclusion: Thus, results obtained can be useful for further analysis of bacterial systems ecology and evolution.

Availability: Available upon request.

Acknowledgements: This work was supported by RFBR grant No. 09-04-01641-a, Biosphere Origin and Evolution program.

References:

1. J. Zhang (2003). Evolution by gene duplication: an update, TRENDS in Ecology and Evolution Vol.18 No.6 June 2003: 292-298.
2. M. M. Babu et al. (2004). Structure and evolution of transcriptional regulatory networks, Current Opinion in Structural Biology 2004, **14**: 283–291.
3. M. Ueta et al. (2008). Role of HPF (Hibernation Promoting Factor) in Translational Activity in Escherichia coli, *J Biochem*, **143**: 425-433.
4. W. Ma et al. (2009) Defining Network Topologies that Can Achieve Biochemical Adaptation, Cell. 138: 760-773.

COMPUTER ANALYSIS OF STRESS RESPONSE NETWORK *E.COLI*

I.L. Stepanenko[1,2]*, I.I. Titov [1,2]
[1]*Institute of Cytology & Genetics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
e-mail: stepan@bionet.nsc.ru;
** Corresponding author*

Key words: *stress response, gene network, topology, cluster, motif*

Motivation and Aim: Advances in high-throughput technologies and computational genomics allow reconstruction of transcriptional regulatory networks *in silico*. Studies of local network topology and dynamics can be used to investigate as well to predict the activity of biological processes and the system's responses to environmental factors. Methods for rational decomposition the network into relatively independent functional modules are essential for better understanding of wide-genome network organization. In the present study we reconstruct the stress response *E. coli* networks and analyze their statistical and dynamical properties.

Methods and Algorithms: The GeneNet technology was used for description the regulation of osmotic and oxidative stress, starvation, DNA damage in *E.coli*. GeneSensor database accumulates information on the sensitivities of the *E.coli* genes to stress stimuli. The database GenSensor contains the data about the structure of bacterial promoters that are activated and expressed in response to external stimuli, the transcription factors; and conditions that provoke the maximal response to the given type of stimulus. The database is available at http://srs6.bionet.nsc.ru/srs6/. The GeneNet scheme Stress response (E.coli) is available at http://wwwmgs.bionet.nsc.ru/mgs/gnw/genenet/. The computational analysis of statistical and dynamical properties was performed by the program package GeNet [1].

Results: We found that stress response gene network is organized in a modular, hierarchical manner. Module 1 is composed of genes related to respiratory functions (low level O_2.). Module 2 is controlled by oxidative stress factor SoxS. Module 3 includes genes regulated by the sigma factor S. RpoS is the master regulator in a complex regulatory network that governs the expression of many stationary phase-induced and osmotically regulated genes in *E.coli*. Module 4 is controlled by OxyR. It induces transcription of a set of antioxidant genes, including *katG*, *ahpCF*, *dps, grxA*, and *oxyS*. Module 5 is composed of the genes induced by DNA damage. Module 6 is composed of factors regulating rpoS, gene controlling RpoS transcription factor.

Interestingly, of three methods of network clustering only one resembles the network natural decomposition where each subnet corresponds to a different type of response. The network motifs found (of size from 3 to 9 nodes) form the modules of cassette signal propagation; bigger motifs typically include smaller ones. Dynamical analysis allowed us to enumerate the stationary states and to find the key network regulators.

Conclusion: This work allowed the identification of six interconnected stress response regulatory network modules. In most case, each of these modules includes genes physiologically related functions involved in stress response *E.coli.*

Acknowledgements

The work was supported by RFFI 08-04-01008-a and SD RAS Multidisciplinary Projects #113 and 119.

References:

1. Titov et al. (2010) GeNet: a program package for computational analysis of gene network statistics and dynamics (this issue).

CONTROL OF *ESCHERICHIA COLI DPS* GENE EXPRESSION IN RESPONSE TO THE TOXIC ACTION OF CADMIUM

T.Y. Stepanova[1*], V.A. Likhoshvai[1,2], I.V. Babkin[1], and T.M. Khlebodarova[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
e-mail: stepanova.tayu@gmail.com
**Corresponding author*

Key words: *Escherichia coli, dps*, *gfp, mathematical modeling, transcription regulation*

Motivation and Aim: The *Escherichia coli dps* gene encodes a DNA-binding protein, which protects the cell from oxidative and some other stresses. Earlier experiments with a genosensor based on the promoter of the *dps* gene and the reporter Gfp protein showed that the promoter was upregulated by Cd^{2+} ions and that the time course of *E.coli*/pDps-gfp genosensor operation depended nonlinearly on Cd^{2+} concentration [1]. The mechanism by which the genosensor is regulated is unknown.

Methods: The models were constructed on the base of generalized Hill functions [2]. Mutant variants of the *dps* promoter were synthesized by PCR with oligonucleotides bearing nucleotide substitutions at RpoE and RpoD polymerase binding sites. Cell fluorescence was measured as in [3].

Results: Mathematical models were constructed for Cd^{2+}-dependent pathways regulating the operation of the *E.coli*/pDps-gfp genosensor and the nonspecific change in the overall metabolic activity in cells exposed to cadmium. The calculated genosensor activity curves were in good agreement with previous experimental results [1]. Two hypotheses are put forward as to the pathway regulating *E.coli*/pDps-gfp activity. In one of them the genosensor operation is controlled by interaction between RpoD and stress polymerase RpoE. The other hypothesis suggests that the nonlinear mode of genosensor expression is determined by association of two or more transcription factors (TFs). The presence of corresponding binding sites is in agreement with both hypotheses. To choose between them, genosensor variants with mutations in the –35 and –10 regions of binding sites for RpoE and RpoD were constructed, and their expression with the presence of Cd^{2+} ions was studied.

Conclusion: In general, the results indicate that the complex response of *E.coli*/pDps-gfp to Cd^{2+} is mediated by RpoD polymerase alone. It may be related to activation of two TFs, Rob and OxyR. Rob is responsible for heavy metal tolerance in the cell, and OxyR is an oxidative stress sensor. Its induction may be related to the ability of Cd^{2+} to produce reactive oxygen species.

Acknowledgments: This work was supported by the Russian Foundation for Basic Research, project 08-04-01008a; projects of the Russian Academy of Sciences Nos. 22.8 and 23.29; projects of the Siberian Branch of the RAS Nos. 107 and 119; and grant NSh-2447.2008.4.

References:

1. Likhoshvai V.A. et al. (2009) *Vestnik VOGIS*, **13**:731-740.
2. Likhoshvai V.A., Ratushny A.V. (2007) *J. Bioinform. Comput. Biol.,* **5**:521-531.
3. Khlebodarova, T. M. et al. (2007) *J. Bioinform. Comput. Biol*., 5:507-520.

USING LOGISTIC REGRESSION AND MULTIFACTORIAL DIMENSIONALITY REDUCTION SOFTWARE FOR DETECTING GENETIC PREDISPOSITION IN CHILDREN TO ESSENTIAL ARTERIAL HYPERTENSION: RESULT OF 12 SNP's ANALYSES

D.A. Stroy*, E.A. Nastenko, V.L. Gurianova, V.E. Dosenko, A.A. Moybenko
Bogomoletz Institute of Physiology, Kiev, Ukraine
e-mail: addictive.signals@gmail.com

Motivation and aim. Augmented possibilities for single nucleotide polymorphism (SNP) detection in people with various polygenic diseases and implementation of the international program "Human Variom" leads to necessity for development of new approaches and finding optimal algorithms of statistical methods which make possible to analyze accumulated huge blocks of data.

In the stream of this tendency, the limelight of cardiovascular diseases investigation is the study of its genetic mechanisms, in particular, essential arterial hypertension (EAH), the polygenic etiology of which is without doubt.

The aim of our study was to create the combined approach for application of parametric and non-parametric statistical methods for detecting genetic predisposition to EAH based on results of 12 SNP's analysis in groups of children with EAH comparing to practically healthy donors.

Methods and algorithms. Single nucleotide polymorphism was determined using polymerase chain reaction with restriction fragments length analyses or polymerase chain reaction in real time using Custom Taq-Man SNP genotyping assays in 154 children suffering from EAH and 140 practically healthy donors. Statistical analyses was conducted using logistic regression in statistical package SPSS ver. 17.0 and multifactorial dimensionality reduction with the help of MDR software ver. 2.0.

Results. MDR analyses indicated that the best factor of prediction among all 12 SNP's was the polymorphism of eNOS gene (Glu298/Asp). Moreover, genotypes algorithmization showed that heterozygote in the model created by MDR plays the role of defender and has a protective effect and decreases the risk of EAH onset while the frequent and rare variants do not have such effect. Logistic regression let us to estimate the significance of such a hypothesis and to evaluate odds ration with the 95% confidence intervals ($P = 0.004$, odds ratio 5.122 (95% CI 1.67 – 15.75). The best 4 component model included 2 SNP's of NOS3 gene (T786/C and Glu289/Asp), AGT and ATR1 genes.

All of the first three SNP's showed significance ($P < 0.05$) and AGT SNP maximizes the model's predicting potential in a way of creating synergetic and correlation relationship with the other model's components.

Conclusion. In our work the necessity of combined application of parametric and non-parametric statistical methods was shown and one- and 4 components models were created. Correct analyses and interpretation of this data makes possible to evaluate genetic risk factors forming predisposition to EAH and estimate the risk of development of this pathology.

IN SILICO PROGNOSIS OF THE TATA BOX POLYMORPHISMS PHENOTYPIC EFFECTS.

V.V. Suslov*, P.M. Ponomarenko[1], M.P. Ponomarenko[2], L.K. Savinkova[2], N.A. Kolchanov[2]
[1]*Novosibirsk State University, Novosibirsk, Russia*
[2]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
e-mail: valya@bionet.nsc.ru
[*]*Corresponding author*

Key words: *polymorphism, TATA-box, TBP/TATA binding, in silico test system*

Motivation and Aim: The association of polymorphisms in regulatory DNA regions with mutant phenotypes is poorly studied. One of the challenging problem of post-genomic bioinformatics is development of the test systems that would enable *in silico* prognosis of the polymorphism effect on the genes expression and phenotype based on a small number of experiments. With this aim we *in silico* analyzed the effects of the 38 TATA box polymorphisms in 27 genes of commercial and laboratory animals and plants on the TATA box/TATA-binding protein (TBP) affinity and compared polymorphism's impacts on the value of affinity with known phenotipical effects of polymorphisms.

Methods: Using standard Entrez and Scirus key searches, polymorphisms were found. The data on gene expression and mutant phenotypes were taken from original articles or GenBank. Value of the TBP/TATA affinity for normal TATA box (S^0) and mutants $S^{\#}$ was *in silico* estimated using the original equilibrium equation for the four subsequent steps of TBP/TATA box binding [1]. The change of this affinity was evaluated using formula,

$\delta(S^0 \rightarrow S^{\#}) = -\ln[K_{D,\,TATA}](S^{\#}) - \{-\ln[K_{D,TATA}](S^0)\}$.

Statistical significance of the value was tested by Student's *t* test at $\alpha < 0.05$, as defined [1].

Results: It was hypothesized that biochemical manifestation of the mutation $S^0 \rightarrow S^{\#}$, in case that $\delta(S^0 \rightarrow S^{\#}) > 0$, would be the excess of the protein, relative to the normal state, $\delta = 0$, and in case that $\delta(S^0 \rightarrow S^{\#}) < 0$, it would be the protein deficiency. The prognosis were found to be statistically significant for 13 out of 38 polymorphisms upon the expected frequency of $p_0 = 0.05$, established by Student's t test. This excess over the expected level was statistically significant ($\alpha < 10^{-7}$, binomial law) and pointed to sufficient sensitivity of equilibrium equation of TBP/TATA box binding [1] for analysis of the TATA box mutations. In the literature data, we found the confirmation for either the excess ($\delta > 0$) or deficiency ($\delta < 0$) of the products of corresponding genes for 10 out of 13 (76%) statistically significant TATA box mutation prognoses (upon the frequency of $p_0 = 50\%$ of the expectation of random concordance of binary prognoses).

Conclusion: Significant congruence of *in silico* prognosis with mutant phenotypes validates the developing of the test system for prediction of the effects of TATA box mutations on phenotype.

Acknowledgements: The work is supported: RAS programs 21, 22, 23, "Biosphere origin and geo-biological system evolution"; Integration project 119, State Contract № P721, RFBR 09-04-01641-a

References:

1. P.M Ponomarenko et al. (2009) Prognosis of affinity change of the TATA-binding protein to TATA-boxes upon polymorphisms of the human gene promoter TATA boxes, Mol Biol (Mosk), 43(3), 512-520 (see in this issue V.V. Suslov et al. SNPs in the HIV-1 TATA Box and the AIDS Pandemic).

PHYLOGENETIC DECOMPOSITION OF THE GENE NETWORKS REGULATING THE LEVEL OF THYROID HORMONES.

V.V. Suslov*, K.V. Gunbin, V.S. Timonov
[2]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
e-mail: valya@bionet.nsc.ru
[*]*Corresponding author*

Key words: *databases, ES-TRRD, thyroid system, gene networks, GeneNet, phylogenetic decomposition*

Motivation and Aim: Being a derivative of the endostyle, thyroid system is one of the earliest chordate-specific endocrine systems. In the graph of gene networks regulating the level of thyroid hormones we can identify several relatively autonomous subunits (TTF-1-dependent regulation, TSH-1-dependent regulation, para- and autocrine regulation, subunit of the H_2O_2 - level control, subunit of the thyroid follicle growth, etc.) [1]. Thyroid system regulates ontogenesis and basic metabolic level in Chordata in close cooperation with other endocrine systems (insulin regulatory system, somatotropic regulatory system, etc.). With increasing distance from the root of the Chordata phylogenetic tree, the extent of this cooperation is growing. In connection with this "absorbing" evolution trend it is of interest to test this gene network by phylogenetic decomposition, thereby pointing to gene duplications which led to complication of the graph.

Methods and Algorithms: Phylogenetic decomposition is the sequential joining of gene-network node pairs (paralogous genes/proteins). Joining order is determined in accordance with the duplication number from the top of the phylogenetic tree of certain protein superfamily to its root [2].

Results: The mammalian gene network of thyroid hormones contains 809 regulatory cycles, whereas the ancestral Chordata gene network of thyroid hormones, which reconstructed using greedy phylogenetic decomposition, contains only some dozens of regulatory cycles. The evolution of gene network of thyroid hormones can be described by two distinct trends. The first trend is the complication due to gene duplications. The second trend is the increasing complexity of regulatory circuits by the integrating of different subunits through the formation of complex nested circuits. The molecular basis of this evolution is the duplication of the transcription factor binding sites in regulatory regions of genes.

Availability: available upon request.

Acknowledgements: The work is supported: RAS programs 21, 22, 23, "Biosphere origin and evolution"; Integration project 119, State Contract № P721, RFBR 09-04-01641-a

References:

1. V.V. Suslov, E.V. Ignat'eva (2002) Molecular genetic mechanisms regulating the Thyroid System: description in the TRRD and GENENET databases. In: Proceedings of BGRS'2002, vol. 2, N.A. Kolchanov et al. (Eds.), 83-86 (Novosibirsk, ICG RAS).
2. V.S. Timonov et al. (2010) Regulatory Circuits and Phylogenetic Decomposition in Gene Networks Evolution Research. In this issue.

SNPS IN THE HIV-1 TATA BOX AND THE AIDS PANDEMIC

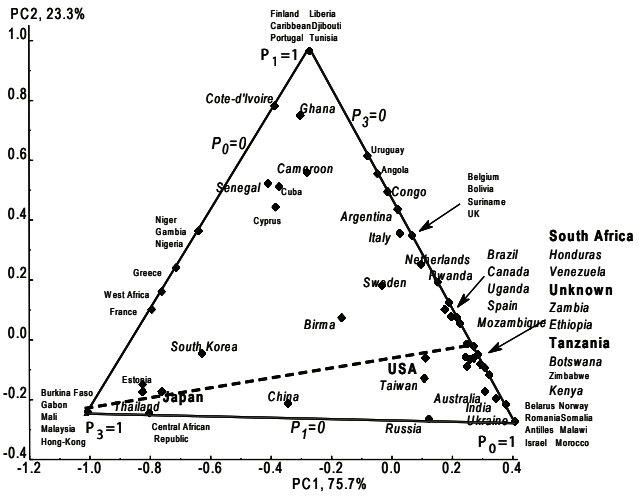
V.V. Suslov*, P.M. Ponomarenko[1], V.M. Efimov[2], M.P. Ponomarenko[2],
L.K. Savinkova[2], N.A. Kolchanov[2]
[1]*Novosibirsk State University, Novosibirsk, Russia*
[2]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
e-mail: valya@bionet.nsc.ru
[*]*Corresponding author*

Key words: *HIV-1, evolution, TATA-box, SNP, TBP/TATA binding*

Motivation and Aim: We analyzed the HIV-1's TATA box functional evolution.

Methods: Among 2662 TATA boxes (2311 HIV-1 isolates) within GenBank 146 alleles of 26 bp in length, S, were found. We denoted *agatgctgCATATAAgcagctgcttt* as norm (59%), S^0. ,The Second in frequency of occurrence (11%), *agatgctgCATAAAAgcagccgcttt*, appeared to be of the core *CATAAAA* associated with HIV-1 subtype E prevailing in South-East Asia. The affinity of the TATA-binding protein (TBP) to TATA box was estimated:

$$-ln[K_{D,TATA}(S)]=10.9-0.23ln[K_{D,TBP/dsDNA}(S)]+0.15PWM_{TATA,Bucher}(S)-0.2ln[K_{D,TBP/ssDNA}(S)],$$



where 10.9 is the nonspecific TBP/DNA affinity; $-\ln[K_{D,TBP/dsDNA}]$ is the contribution of TBP sliding along DNA; $PWM_{TATA,Bucher}$ is the one of TBP/TATA recognition; $-\ln[K_{D,TBP/ssDNA}]$ is the one of the stabilization of the TBP/TATA complex; 0.23, 0.15, 0.2 are stoichiometric coefficients. The significance of deviations was assessed by the Student *t* test. In each of 70 countries we found the frequency of occurrence weighted by the number, *n*, of the associated S: P_0 is the S^0; P_1 is the non-S^0 variants with S^0 affinity; P_2 is the high-affinity S; P_3, is the low-affinity S including core *CATAAAA*; P_4 is the other S.

Results: Analysis of prevalence of the 146 variants revealed principal components PC1 (75.7% variance) and PC2 (23.3% variance). The countries with n<10 are typed in plain; those with 10≤n<99, in italics; those with n≥100, in bold. All the countries fall within a triangle with the vertices $\mathbf{P_0=1}$, $\mathbf{P_1=1}$, $\mathbf{P_3=1}$ and the sides $P_0=0$, $P_1=0$, $P_3=0$. The countries lie close to the P_3 was interpreted as in neutral drift around the S^0, the ones lie close to the P_0 and P_1 was interpreted as under selection towards low-expressing forms or neutral drift around the core *CATAAAA,* the ones lie close vertex $\mathbf{P_3}=1$ to the side P_3 (under broken line) was interpreted as under selection against S^0-like and high-affinity forms. So, each of these countries is specifically associated with one of three trends in HIV-1 evolution: neutral drift around the TATA box norm; neutral drift around the slowly replicating TATA box core *CATAAAA* (phylogenetic inertia); an adaptive increase in the frequency of *CATAAAA*.

Acknowledgements: The work is supported: RAS programs 21, 22, 23, "Biosphere origin and evolution"; Integration project 119, State Contract № P721, RFBR 09-04-01641-a

MOLECULAR GENETICS OF PIG AGGRESSION

E. Terenina[1]*, D. Bazovkina[2], S. Rousseau[1], F. Salin[3], S. Monllor[3], R. D'Eath[4], S. Turner[4], A. Kulikov[2], P. Mormede[1]

[1]*INRA PsyNuGen, 33076 Bordeaux Cedex, France*

[2]*Institute of Cytology and Genetics RAS, Laboratory of Behavioral Neurogenomics, 630090 Novosibirsk, Russia*

[3]*INRA UMR Biodiversitŭ Gunes et Ecosystumes, 33612 Cestas Cedex, France*

[4]*Sustainable Livestock Systems, SAC, Edinburgh, UK*

e-mail: elena.terenina@bordeaux.inra.fr

**Corresponding author*

Key words: *aggression, genes, serotoninergic system, SNP, molecular genetics, pig*

Motivation and Aim: Aggression between pigs occurs when unfamiliar pigs are mixed (on farms; in trucks carrying animals to the slaughterhouse, and during lairage). Aggression affects pig welfare (social stress, biting leading to skin lesions), production efficiency, and product quality (carcass yield because of skin lesions, meat quality).

A genetic component to individual aggressiveness has been described in pigs, mice and other species (e.g. Turner et al., 2006[1]; Maxson and Canastar, 2007[2]). It is therefore possible to consider genetic selection against excessive aggressive tendencies but direct phenotypic selection is difficult in the context of commercial facilities. An alternative strategy is to search for molecular genetic polymorphisms associated with aggressive tendencies and use these for marker-assisted selection. We sought polymorphisms in genes known from the literature to be associated with aggressive behavior. Our primary interest is the components of the brain serotonergic system that has been identified as a primary neurochemical system involved in the regulation of aggressive behavior in pigs.

Methods and Algorithms: Individual aggressive tendencies (aggressiveness) were measured after weaning at five weeks of age following a standardized mix as described and validated by Turner et al. (2006)[1]. Phenotypic information, DNA and pedigree are available for 500 animals. Candidate genes, selected from the literature, are involved in the regulation of the serotonergic system (HTR1A, HTR1B, HTR2A, SLC6A4, COMT, MAOA, MAOB), the dopaminergic system (DRD1) and vasopressin (AVP, AVPR1A).

The sequences of candidate genes were obtained from the most recent databases (http://www.ensembl.org/Sus_scrofa/Info/Index) and primers for sequencing were designed with adapted software (Primer Express and Primer 3). The softwares CodonCode Aligner and BioEdit were used for sequence assembly and polymorphism detection. All animals were genotyped for these SNPs by the best available technique, such as sequencing or high-resolution melting (HRM). Association study was performed using the R statistical software system

Results and Conclusion: Numerous new polymorphisms have been detected in candidate genes, and associated with aggressive tendencies. These studies open the way towards efficient marker-assisted selection in farm animal species.

References:

1. S.P.Turner, I.M.S.White, S.Brotherstone, M.J.Farnworth, P.W.Knap, P.Penny, M.Mendl, A.B.Lawrence (2006). Heritability of post-mixing aggressiveness in grower-stage pigs and its relationship with production traitsAnimal Science, **82:** 615–620.
2. S.C.Maxson, A.Canastar (2007). Aggression/ Concepts and Mathods Relevant to Genetic Analyses in Mice and Humans. In: Neurobehavioral Genetics. Methods and Applications, B.C.Jones, P.Mormede (Eds.), 281–289.

SEARCH FOR GENETIC LOCI ASSOCIATED WITH THE MANIFESTATION OF PHYSIOLOGICAL CHARACTERISTICS OF RATS BASED ON INCOMPLETE EXPERIMENTAL DATA WITH THE USAGE OF ARTIFICIAL NEURAL NETWORKS

A.A. Tikhonov*, O.E. Redina, V.M. Efimov, A.L. Markel
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: hallow@ngs.ru
**Corresponding author*

Key words: *arterial blood pressure, genetic loci, data gaps, neural networks.*

Motivation and Aim: Increasing both complexity of experiments and the number of measured parameters leads to more or less large amount of failed measurements in the experimental data. There are two possible ways of treating this issue. In the first case, the damaged data may be not used by researchers at all. In the second case, the data may be reconstructed to be taken into account. The former approach doesn't require any comments while the latter may be realized at least in 3 ways: missed values may be substituted with the mean of possible values, or with the random numbers, otherwise, they may be found "postscriptum" as the values that minimize the output error of a constructed goal function. The first two ways do not consider experimental data as the set of parameters describing one certain phenomenon. The third way supposes that the damaged data are not used in the constructing of the goal function [1]. The aim of this work is to demonstrate a two-step research process when first of all the damaged data is reconstructed before general investigation with the usage of all available information about the experimental object, and then, the repaired information is undergone further analysis. To demonstrate these we have taken as an example the experimental animals' genome and the animals' physiological characteristics [2].

Methods and Algorithms: In our study we used an ensemble of the one-layered back-propagation networks with different parameters. The bootstrap multiplicative technique was applied to the base dataset. The resulting bootstrap copies were used as alternative training datasets to provide reliable control on the adequacy for each network in the ensemble.

Results: A program was written and the series of computational experiments were carried out to obtain the values of missed data and to make sure that the suggested technique of reconstructing damaged data gives adequate results. The simulated missed values were used for detecting the genetic loci of interest.

Conclusion: The proposed neural network algorithm is promising to search for genetic loci associated with the physiological characteristics.

References:

1. Zs.J.Viharos et al. (2002) Training and application of artificial neural networks with incomplete data, In: *Developments in Applied Artificial Intelligence,* 649-659 Springer Berlin / Heidelberg.
2. O.E.Redina et al. (2006) Rats with inherited stress-induced arterial hypertension (ISIAH strain) display specific quantitative trait loci for blood pressure and for body and kidney weight on chromosome 1, *Clinical and Experimental Pharmacology and Physiology,* **33(5-6):** 456 – 464.

REGULATORY CIRCUITS AND PHYLOGENETIC DECOMPOSITION IN GENE NETWORKS EVOLUTION RESEARCH

V.S. Timonov [1,2,3,*], K.V. Gunbin [1], I.I. Turnaev [1], M.A. Genaev[1] and D.S. Miginsky [1,2,3]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *GeN Ltd., Novosibirsk, Russia*
e-mail: vtimonov@bionet.nsc.ru
[*] *Corresponding author*

Key words: *gene networks analysis, phylogenetic decomposition, regulatory circuits, computer system*

Motivation and Aim: As well known, the E2F, pRB, cyclin, CDK and CDI paralogous proteins comprise the regulatory core of cell cycle gene network. With search for regulatory circuits and phylogenetic decomposition (PD) of gene network we tried to understand why the complexity of the metazoan cell cycle gene network increased in evolution.

Methods and Algorithms: GeneNetStudio tool developed on ARANEUS toolkit [1] has been used for support of this research. It is new client tool which used for visual reconstruction and analysis of gene networks of GeneNet system [2]. Phylogenetic decomposition is the sequential joining of gene-network node pairs (paralogous genes/proteins). Joining order is determined in accordance with the duplication number from the top of the phylogenetic tree of certain protein superfamily [3] to its root. It was developed the problem-oriented software module for processing of the necessary text information.

Results: The mammalian cell cycle gene network contains more than 95000 regulatory cycles, whereas the ancestral eukaryotic cell cycle gene network, which reconstructed using greedy PD, contains only 17 regulatory cycles. The stepwise decomposition resulted that over 80% of evolutionary event in this gene network appears to be a kind of gene duplications followed by origin of concurrent inhibitor or activator of their ancestors.

Conclusion: The specialized algorithm and program module for GeneNetStudio has been developed. They used for research in areas of regulatory circuits and phylogenetic decomposition. It carried on perfection of problem-oriented software module for results visualization on gene networks. Improvement of a technique of information processing is spent.

Availability: Available upon request.

Acknowledgements: Work was supported by RFBR (No. 09-04-01641-a) “Computer research of molecular genes evolution and molecular-genetic systems of multicellular animals”.

References:

1. Patent No. 2009611761 RUS. (2009) Toolkit for the development of network editors (ARANEUS) / ICG SB RAS.
2. E.A. Ananko et al. (2005) GeneNet in 2005, *Nucleic Acids Res*, **33**:425–427.
3. D. Wilson et al. (2009) SUPERFAMILY - Comparative Genomics, Datamining and Sophisticated Visualization, *Nucleic Acids Res*, **37**:380-386.

SCIENCE STRUCTURAL INFERENCE

I. Titov*[1,2], K. Rudnichenko[3], F. Fursov[2], P. Krutov[2], A. Kulikov[2,3]
1-Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
2-Novosibirsk State University, Novosibirsk, Russia
3-Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia
e-mail: titov@bionet.nsc.ru
**Corresponding author*

Key words: *text mining, stochastic networks, bibliometrics*

Motivation and Aim: Recent fast growth of biomedical databases motivated development of numerous approaches to automatic extraction of information [1]. More generally, inferring science structure and dynamics from mining of scientific publications could make it possible to reveal hot topics, to find promising co-authorship opportunities and to reduce the activity overlap. Meanwhile in the last decade it became clear that many complex dynamical systems (food and ecological webs, word, metabolic, social and gene networks) share the common structural and dynamical properties.

Methods and Algorithms: In this paper we have analyzed evolution of PubMed co-authorship network and have found the network architecture to be time-invariant. Then, to explore the landscape of biomedical science we have constructed biomedical thesaurus using the PubMed abstracts and created term network based on simultaneous term occurrence. Further we have compared abstract classifications (reflecting differentiation of the research themes) which were based either on the term network or the word network built from MeshWord usage.

The statistical characteristics of these two networks were compared with ones of the word network of literature texts. Finally, to investigate the evolution of biomedical research we have employed hidden Markov model to search for the time peaks of term usage and finding out the hot topics of biomedical science.

Acknowledgements

The work was supported by SD RAS Multidisciplinary Project #111.

References:

1. Mack R, Hehenberger M. (2002) Text-based knowledge discovery: search and mining of life-sciences documents. *Drug Discov. Today.* 2002 Jun 1;**7**(11 Suppl):S89-98.

MIRNA ANALYSIS WITHIN THE WEB-SERVER GARNA

I. Titov*[1,2] , S. Vorozheikin[1]
[1]*Novosibirsk State University, Novosibirsk, Russia*
[2]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
e-mail: titov@bionet.nsc.ru
[*]*Corresponding author*

Key words: *miRNA, miRNA gene, RNA secondary structure, hidden Markov model*

Motivation and Aim: miRNAs are small non-coding RNAs that control gene expression by negatively regulating translation through binding to mRNA transcripts. Computational prediction of pre-miRNAs, miRNA and their sites is a constantly growing topic of bioinformatics. Nam et al. [1] have developed a pioneering algorithm for both human pre-miRNA and miRNA prediction. Here we present the update of web-server GArna for RNA structure calculation by the programs for miRNA analysis.

Availability: The programs described can be found on the GArna site http://wwwmgs.bionet.nsc.ru/mgs/programs/garna/

Methods and Algorithms: The HMM algorithms of Nam et al. [1] were modified to increase the accuracy of human miRNA and pre-miRNA prediction. The genetic algorithm http://wwwmgs.bionet.nsc.ru/mgs/programs/garna/ or the rnafold program of Vienna package http://rna.tbi.univie.ac.at/ were implemented for RNA secondary structure calculation. For miRNA binding efficiency calculation a new algorithm was developed based on HMM.

Results and Conclusion: Two simplest methods are presented for miRNA site prediction: search by 5' context similarity and calculation of binding affinity. The latter algorithm is generalized by using HMM and taking into account the miRNA overlapping.

We modified the algorithms of Nam et al. [1] based on new HMM model and found the accuracy improvement either for pre-miRNA (Table.1) and *ab initio* miRNA (about 2 nt more accurate miRNA boundary) prediction.

Table 1. Comparison of the efficiency of human miRNA precursor predictions by HMM algorithms.

	Sensitivity	Specificity
ProMiR [1]	0.73	0.96
This work (MIRbase release 12.0)	0.86	0.96
This work (MIRbase release 2.2)	0.93	0.96

References:

1. J.-W. Nam, K.R. Shin, J.J. Han, Y.T. Lee, V.N. Kim and B.T.Zhang. (2005) Human miRNA prediction through probabilistic co-learning of sequence and structure. *Nucleic Acids Research*, **33**(11):3570-3581.

3D-MODELLING OF RNA STRUCTURE

K. Barsov[1], A. Kulikov[1,2], I. Titov*[1,3]
[1]*Novosibirsk State University, Novosibirsk, Russia*
[2]*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*
[3]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
e-mail: titov@bionet.nsc.ru
[*]*Corresponding author*

Key words: *RNA secondary structure, unfolding, design, elastic model, genetic algorithm*

Motivation and Aim: While RNA secondary structure becomes increasingly better predicted, less computational approaches have being developed for 3D RNA structure calculation.

Methods and Algorithms: For 3D RNA structure simulation we used the finite-element elastic model. The tertiary structure energy was calculated as elastic energy expanded by Darbu vector perturbation

$$E = \int D(\vec{w} - \vec{w}_0)(\vec{w} - \vec{w}_0)$$

The energy minimization was performed by the genetic algorithm. The elastic parameters were fitted by the optimal tRNA and X-ray structure comparison.

Results and Conclusion: First we present the results of 3D-structure design of artificial RNAs. Next the computational experiments were performed to simulate the artificial square-lattice extensions and tRNA tertiary structure unfolding.

Acknowledgements

The work was supported by SD RAS Multidisciplinary Project #109.

STRUCTURE OF COLLAGEN FIBRILS DERMIS AND ALGORITHM OF CALCULATION OF ITS STRUCTURAL CHARACTERISTICS

A.O. Titov, M.O. Titov, I.I. Titova, O.P. Titov
GOU VPO East Siberian State Technological University, Ulan-Ude, Russia.
E-mail: fibrilla45@mail.ru

Key words: *collagen molecule, fibril, fiber, internal element, molecular, fibrillar, fiber space.*

Motivation and Aim: Determination of the internal structure of capillary porous materials has presents considerable difficulty. Typically, the basis of the porous structure of the material is laid of the distribution of pore volume on the radius without regard this space with the elements of the solid structure of the material. We have developed a model of collagen fibrils, which allows to calculate the internal structure of the dermis of the elements of its structure: the molecular porosity of fibrillar porosity, fiber porosity. Moreover, calculations can be performed both on virtual objects, and for natural objects tentatively identified two indicators - the moisture content of the sample completely flooded, and the shrinkage of the sample during welding. [1] *Methods and Algorithms*; We developed a model of collagen fibrils, [2] allows to calculate the internal structure of the dermis of the elements of its structure: the molecular porosity of fibrillar porosity, fiber porosity. Moreover, calculations can be performed both on virtual objects, and for natural objects tentatively identified two indicators - the moisture content of the sample completely flooded, and the shrinkage of the sample during welding.

Algorithm of the calculation is based on determining the amount of absorbed moisture (water density adopted in 1000 kg/m3) to change the volume of the molecule (molecular moisture) with a water content of oven-dry condition to a fully flooded (changing diameter of the molecule). By changing the internal dimensions of the spaces in the fibrils with a water content (the intervals between the ends of the molecules, and the volume of the internal element) (fibrillar moisture). On the difference between the total moisture content and the sum of molecular and fibrillar moisture determined mezhfibrillyarnaya or fiber moisture. Shrinkage of the sample of collagen in the dermis of welding gives the amount of space between the ends of the molecules in the fibrils.

Results: In determining on natural samples were obtained the following results

The porosity (volume of the spaces between the molecules), %	26.0
The porosity (volume of space in the fibrils), %	28.4
The porosity (volume of space in the fibers), %	45.6
The surface area of all the molecules in the 0,1 kg, m^2	710,1
The surface area of all fibers in 0,1 kg, m^2	401.8

Conclusion: The results show that the distribution of spaces on the elements of the structure and surface area of structural elements correspond to given in the literature [1]

References:

1. A.N. Mikhailov, Collagen of the skin and the basis of its processing. Moscow, Light Industry 1971. with. 528, ill.
2. A.O. Titov, O.P. Titov, On the structure of collagen fibrils. Leather and fur in the XXI century. The technology, quality, environment, education: Proceedings of the conference. - Ulan-Ude: Izd ESSTU, 2006. - 352.

A COMPUTER ANALYSIS OF FUNCTIONAL RELATIONSHIPS BETWEEN GENES ASSOCIATED WITH MULTIFACTORIAL DISEASES: PRE-ECLAMPSIA AS AN EXAMPLE

E. Tiys[1]*, P. Demenkov[1], E. Vashukova[2,3], A. Glotov[2,3], V. Ivanisenko[1]
[1]*The Institute of Cytology and Genetics, Novosibirsk, Russia*
[2]*D.O. Ott Research Institute of Obstetrics and Gynecology, Russia*
[3]*St. Petersburg State University, St. Petersburg, Russia*
e-mail:tiys@bionet.nsc.ru
[*]*Corresponding author*

Key words: *association networks, diseasome, genotyping, pre-eclampsia, personalized medicine.*

Motivation and Aim: Search for new genetic markers associated with a risk of multifactorial diseases gained new impact through comprehensive genetic analysis of disorders with similar clinical manifestations. The information on this topic still is scattered through numerous reports. Being represented as a single network, such information could be of great value for personalized medicine as it assist in reduction the cost of biochemical and genetic tests. The first task of our work is focused on automated collection all information on genetic nature of the disease into a single system. The second one concerns estimation the contribution of thrombophilia, diabetes and other disorders in development of the pre-eclampsia syndrome.

Methods and Algorithms: We used the associative knowledge networks, reconstructed by the text-mining technology (AND) [1] applied to PubMed abstracts and data retrieved from the databases. A fragment of a full-length associative network, the diseasome [2] was taken and the pre-eclampsia syndrome as the central position was chosen. Afterwards, the preeclampsisome (the part of the diseasome that surrounds the pre-eclampsia syndrome) was reconstructed. Three types of genetic associations were distinguished. Type I, genes whose expression varied in accompanying pathologies. Type II, genetic polymorphisms associated with pathological conditions. Type III, genes of intimate associations between diseases. After automated reconstruction, expert filtration and sorting according to types was performed.

Analysis of the reconstructed network is based on the identification of functionally related genetic groups determining similar etiology. For each gene linking pathologies the vector of topological features was constructed. The coordinates of the vector reflect the number of neighbors, the presence of associations between the gene under consideration and the disease included in the preeclampsisome. These vectors were used for clustering of the genes.

Results: We reconstructed preeclampsisomes of three above types and estimated contributions of accompanying diseases to the preeclampsia syndrome progression. We found functionally interrelated genetic groups and groups of diseases of similar origin

Conclusion: The diseasome reconstruction approach makes possible prediction of potential disease markers by similarities in topological features of vertices. Moreover, these similarities can be useful in personalized medicine in selection of methods of diagnostics appropriate for a patient with manifest phenotypic characters. Analysis of the preeclampsisome-preeclampsia relationships allows to discriminate the accompanying pathologies with outmost contribution to the etiology of pre-eclampsia and suggest precise processes that predispose to it. Suggested approach can be used in relevant studies of other multifactorial disorders too.

*Acknowledgments:*State contract FASI №02.514.11.4123, Interdisciplinary integration project SB RAS №111, №94, Program RAS 22. Molecular and cellular biology.

Availability: By authors request.

References:

1. Demenkov PS et al. Associative network discovery (AND) -computer system to automatic reconstruction of the associative knowledge networks about molecular-genetic interactions (2008) Computational technology. **13, N 2.:** 15-19.
2. Goh KI et al. (2007) The human disease network, *Proc Natl Acad Sci U S A*, vol. 104, no. 21: 8685–8690.
3. V.S.Baranov. The Genetic Pass as a background for of individual and predictive medicine: Pub. N-L, 2009. 528 P. [in Russian]

REGULATORY EFFECTS OF GENES FOR BEHAVIOR ON THE EXPRESSION OF THE S^K COAT COLOR GENE IN AMERICAN MINK (*MUSTELA VISON* SCHREBER, 1777)

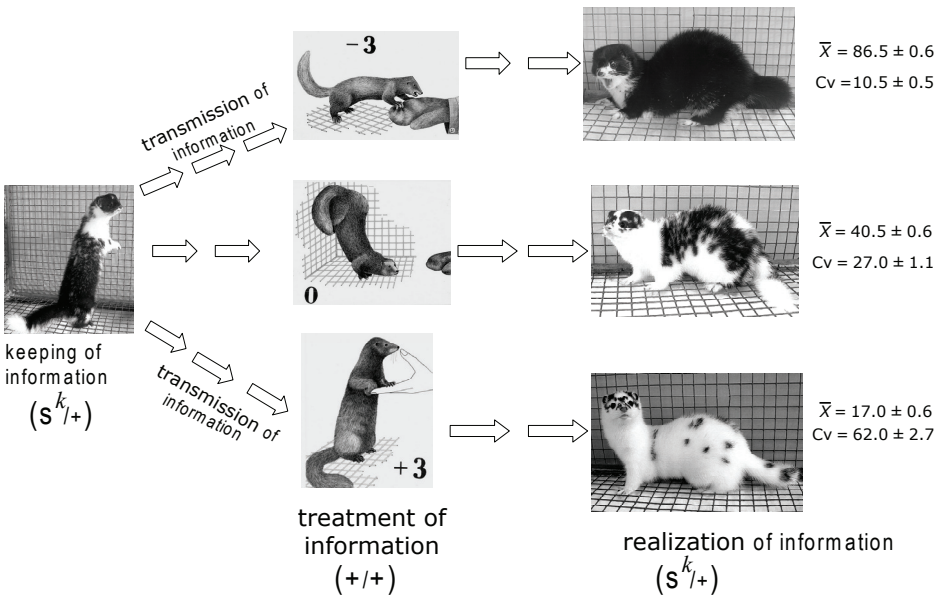
O.V. Trapezov*, L.I. Trapezova
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: *trapezov@bionet.nsc.ru*
*Corresponding author

Key words: *genes for behavior, regulatory effects, coat color gene, American mink, Mustela vison*

Motivation and Aim: Hair coat pigmentation is under the control of dozens of gene. With this in mind, melanogenesis would serve as a good model for studying the mechanisms of gene expression.

Methods and Algorithms: Cross-breeding experiments were performed using male minks with the $S^k/+$ phenotype and *standard* (+/+) female minks of tame (+3 scores), fright (0 scores), and aggressive (–3 scores) lines bred and maintained at the Experimental farm of this Institute.

Results:



Conclusion: The expression and specificity of the coat color phenotypes of mink carrying the $S^k/+$ mutation are under the regulatory effects of genes controlling both aggressive and tame behavior. The inheritance is that the same gene may convey information of different kind, depending on its environment, the same gene may convey one or another completely different meaning.

Availability: available as a commercial package.

References:

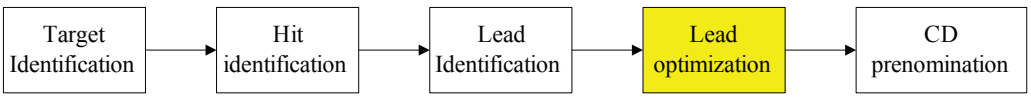
1. D.K.Belyaev, L.N.Trut. (1982) Accelerating evolution, *Science in the USSR*, **5**: 24-64.
2. A.A.Lyapunov. (1980) On the relation of materia, energy and information, In: *Problems of theoretical and applied cybernetics*, 320-323 (Nauka).
3. R.A.Sturm. (2006) A golden age of human pigmentation genetics, *Trends genet*. **20**: 464-468.

NEW APPROACHES TO COMBINATORIAL LIBRARIES GENERATION

S.V. Trepalin*, A.V. Yarkov
Institute of Physiologically Active Compounds RAS, Chernogolovka, Moscow Region, Russia
e-mail: trep@chemical-block.com

Key words: *Drug; Combinatorial chemistry; Structure generation; Molecular design;*

Motivation and Aim: The modern drug discovery process has a number of stages:



Lead optimization includes experimental or *in silico* study of many compounds with structures, similar to lead found. Traditional source of similar compounds is a combinatorial libraries generation. The usual way of such generation is the definition of substituent list for a scaffold and enumeration of all possible combination. It is impossible to generate new cycles as to change scaffold structure using traditional approach also.

Methods and Algorithms: To avoid above limitations two approaches are used. First is the definition of multiply connections points at substituent. The second approach is the creation of transformation rules-mapping of initial and final fragments. First approach leads to new cycles generation. A scaffold and list of substituent are input parameters, as for traditional combinatorial libraries generation. The second approach is used for transformation of one combinatorial library into another.

Results: Significant chemical space increase results in implementation of these approaches. In turn, one may select more diverse dataset for *in silico* study or experimentsl screening.
Conclusion and Availability:. CheD program [1] is used for algorithms realization. As a result, SD files with new combinatorial libraries, are generated. Algorithm [2] is used to make high-quality images of polycyclic chemical structures.

References:

1. S.V.Trepalin, A.V.Yarkov (2001) CheD-Chemical Database Compilation Tool, Internet Server and Client for SQL servers, J.Chem.Inf.Comput.Sci., 41: 100-107
2. S.V.Trepalin, A.V.Yarkov, I.V.Pletnev, A.A.Gakh (2006) A Java Chemical Structure Editor Supporting the modular Chemical Descriptor Language, Molecules, 11: 129-141

OPTIMIZATION OF CRYOBANK CONFIGURATION: RUSSIAN EXPERIENCE

I.S. Trukshin*, R.S. Moskalyuk
Cryogentech Co. Ltd, Saint.-Petersburg, Russia
e-mail: info@cryogentech.ru
**Corresponding author*

Key words: *methods of cryopreservation, cryogenic storage, equipment, history of cryobanking*

Motivation and Aim: The aim of this retrospective study was to follow the history of developing of cryogenic medical technique and equipment and the history of cryobanks in Russian Federation

Methods and Algorithms: Analyzing the experience of Cryogentech Co Ltd in creating cryobanks for such Institutes as Sklifosovsky Medical Institute in Moscow, Scientific Institute of Transfuziology in Saint-Petersburg, Blood Center in Moscow as well as our planning some future big cryobanks in the cities of Russia and some small cryobanks fo the IVF needs in Moscow, Spb, Kazan city etc. Analyzing the experience of Cryogentech Co Ltd with the use of the cryogenic equipment of such well known companies as Taylor-Wharton, MVE, Planer, Cryo Diffusion, AirLiquid etc, comparison of small and large cryobanks and attempt to optimize the functioning of cryobank units.

Results: There are 3 types of the standard cryobanks in Russia (big, medium and small). Well organized "big" cryobank should consist of 4-7 fully equipped robotic cryostorages with supplying vessels, that all connected through vacuum-jacket cryogenic piping to the tank outside the building, programmable freezer with big cryogenic chamber, computer with preinstalled special programs for the whole unit control and monitoring. In some most updated cryobanks we install special equipment for archivation data and traffic control on the basis of Assure24Sevem system developed by Planer Company. Also automatic supplying system should be installed of the solenoid valves, control panels etc. Small cryobanks may consist of 5-10 small cryostorages without automatics and control systems, few dewar vessels filled with Liquid Nitrogen and programmable freezer with small or medium cryochamber. The amount of the stored material may vary from few thousand to few hundred thousands of the samples. It depends on the quantity of cryostorages installed. For opening the cryobank the customer should prepare rooms according to the special requirements – if it is a big cryobank, the ventilation system should be organized, the floors and the walls should be prepared, the heating system should be separated from the equipment. All documents for the equipment, schemes, requirements are certificated in an official way. It is necessary to have certificates for all the main parts (basic equipment) of the particular cryo-unit, that will be useful in the process of its functioning.

Conclusion: Nowadays there is a great demand for cryopreservation units in Russian Federation, e.g. among leading medical centers. Russian cryobanks develop rapidly previous years because of the importance of cryoarchiving the samples and keeping them for the future grows-up from day to day. Russian cryopreservation units (cryobanks) still differ from the European ones by their configuration, but the amount of stored material approaches to the European analogues more and more. Cryogentech Co. Ltd tries to perfect the configuration of cryobanks built up by our company and to approach Russian cryobanks to the International standards.

Acknowledgements: Thanks to Cryo Diffusion for providing necessary materials about European cryobanks

THE RELATIONSHIP BETWEEN EVOLUTIONARY CHANGES IN CYCLINS AND INCREASING THE COMPLEXITY OF EUKARYOTES

I.I. Turnaev[1,*], K.V. Gunbin[1], V.V. Suslov[1], D.A. Afonnikov[1,2]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
e-mail: turn@bionet.nsc.ru
**Corresponding author*

Key words: *cyclins*, *organism complexity, physical-chemical properties*

Motivation and Aim: Eukaryotic evolution resulted in the increasing of organism complexity (OC). The molecular mechanisms of this process are not clear yet. We investigated the relationship between evolutionary changes in physical-chemical properties (PCPs) of a number of cyclin family proteins and OC of eukaryotes.

Methods and Algorithms: Assessment of change of PCPs of the cyclins was based on the data of 531 amino acids characteristics from the AAIndex database. We develop novel computer system (SAMEM), which implements the permutational test for comparison of proteins molecular evolution model with real data. Real data on changes in PCPs of the cyclins estimated using reconstructed ancestral sequences of proteins by ANCESCON (WAG amino acid substitution matrix), FASTML server (LG matrix), AAML from PAML 4.4. package (specific substitutional matrix for cyclins). We used the data published in [1] to estimate OC. The estimation of the relationships between OC and PCPs performed using Spearman and Goodman-Kruskal Gamma nonparametric statistics.

Results: The results demonstrated a correlation between evolutionary changes in some PCPs of cyclins and increasing the OC. For all reconstruction methods it was found that OC and PCPs significantly correlated for 24.29 % AAIndex characteristics for cyclin B in fungi, 28.25 % characteristics for cyclin D and 27.12 % characteristics for cyclin E in animals. The relationship between OC and PCPs for cyclin A and B in animals is weaker (8.29 % and 0 % of characteristics).

Conclusion: The results suggested that the evolutionary changes of cyclin B in fungi and E, D in animals could be involved in molecular mechanisms of the OC increasing during evolution.

Availability: Available upon request.

Acknowledgements: This work was supported by RFBR grant No. 09-04-01641-a; program of RAS 23, biological diversity, project 29; grant НШ-2447.2008.4; contract RAS №10104-37/П-15/110-327/150409-001.

References:

1. C.Vogel, C.Chothia. (2006) Protein family expansions and biological complexity, *PLoS Comput Biol.*, **2**: e48
2. S.A. Glantz. (1997) Analysis of relationships, In: Primer of Biostatistics, S.A. Glantz et al. (Eds.4) (New York).

GPGPU-COMPUTING FOR PREDICTION OF SMALL LIGAND BINDING SITES IN PROTEINS

L.A. Uroshlev[1], I.V. Kulakovskiy[2], S.V. Rakhmanov[2], V.M. Makeev[2]
[1]*Moscow State Forest University, Mytischi, Russia*
[2]*Research Institute for Genetics and Selection of Industrial Microorganisms, Moscow, Russia*
e-mail: leoniduroshlev@gmail.com

Key words: *empirical potentials, ion binding, GPGPU-computing*

Motivation and Aim: At this time there are over 60 thousand known protein 3D structures and this number is ever increasing. One can use this data to obtain empirical potentials for further prediction of small ligand binding sites in a protein structure of interest [1]. This approach is computationally expensive which often limits its practical usage to small molecular structures.

Methods and Algorithms: Here we present an effective strategy to use GPGPU for prediction of small ligand binding sites in proteins. We took the method from [2], which was successfully applied to predict water- and calcium ions in a number of different protein structures. We developed a parallel version of the algorithm and implemented it using CUDA [3] programming model.

For each protein structure CUDA architecture allows to divide it into hundreds of independent blocks for parallel processing. Taking a list of protein structures containing a ligand of interest we compute all contants between the ligand and all other atoms of the structure and store this data in the simple relational database. Then it can be used to evaluate observed and expected distance frequencies and therefore to produce an empirical potentials for the selected ligand.

Taking an independent protein structure the obtained empirical potentials can be futher used to predict possible ligand binding sites. Here we use CUDA to calculate estimates of local ion binding probability for either randomly placed points (Monte-Carlo simulation mode) or fixed-step grid in a parallel mode.

Results: We present a software PIONCA (Protein-ION CAlculator) which uses effective GPGPU implementation of the empirical potential-based method for small ligand binding prediction. We tested it using data for different metal ions and water.

Conclusion: We present an effective computational tool which shows significant performance gain over initial implementation [1].

Availability: The software is available online at the following URL http://line.imb.ac.ru/ion-calculator.

References:

1. S. Rahmanov, V. Makeev (2007) Atomic hydratation potentials using a Monte Carlo Reference State (MCRS) for protein salvation modeling, *BMC Structure Biology,* **7:19**.
2. S. Rahmanov, I. Kulakovskiy, L. Uroshlev, V. Makeev. (2009) Empirical potential for ion binding in proteins, *Journal of Bioinformatics and Computation Biology*. In press
3. NVIDIA CUDA™ Programming Guide SDK, http://developer.download.nvidia.com/compute/cuda/2_2/toolkit/docs/NVIDIA_CUDA_Programming_Guide_2.2.pdf

WEB-BASED GENOME BROWSER USING AJAX AND CANVAS TECHNOLOGIES

T.F. Valeev[1,2*], N. Tolstykh[1], F.A. Kolpakov[1,3]
[1]*Institute of System Biology, Novosibirsk, Russia*
[2]*Institute of Informatics Systems, SB RAS, Novosibirsk, Russia*
[3]*Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia*
e-mail: lan@iis.nsk.su
[*]*Corresponding author*

Key words: *genome browser, web, BioUML*

Motivation and Aim: Currently many genome browsers are available either as part of standalone applications or as web-applications. Web-based genome browsers such as UCSC genome browser [1] have distinct advantage of being available everywhere without the necessity to install or configure additional software aside from web-browser. Unfortunately due to limited capabilities of web technologies most of genome browsers have lack of interactivity, and images are loaded from server side increasing network traffic and response time.

Methods and Algorithms: New standard of HTML language, HTML5, provides a powerful technology called 'canvas' [2] for rendering graphics on client side using JavaScript. This provides new level of interactivity and fast genome browsing over the web. Source data provided by server is transferred via AJAX technology in JSON format which can be easily converted into JavaScript objects. Despite HTML5 is still not officially released (as for March 2010), the 'canvas' technology supported by all major web-browsers except Microsoft Internet Explorer, where it can be emulated using third-party JavaScript libraries. There are already some canvas-based genome browser implementations like AnnoJ [3], but they don't support Internet Explorer, as well as custom tracks and DAS [4] protocol.

Results: We implemented the version of canvas-based genome-browser as proof of concept, and this technology showed to be appropriate for the task. User can browse tracks from Ensembl, use dragging, semantic zoom and get sites information. Obtaining tracks from DAS servers is also supported.

Availability: Canvas-based genome browser is bundled with BioUML web edition [5] and available on {{http://server.biouml.org/bioumlweb/}}. Using anonymous access you can navigate to databases/Ensembl/Sequences and open any chromosome there. DAS tracks example can be found in databases/UCSC.hg19.

Acknowledgements. This work was supported by European Committee grant №037590 "Net2Drug", № 202272 "LipidomicNet" and interdisciplinary project №17 of SB RAS.

References:

1. W. J. Kent et al. (2002) The human genome browser at UCSC, Genome Res., 2002 Jun;12(6): 996-1006.
2. I. Hickson, D. Hyatt (Eds) (2010) HTML5 standard working draft, chapter 4.8.10 {{http://www.w3.org/TR/2010/WD-html5-20100304/the-canvas-element.html}}
3. {{http://www.annoj.org/}}
4. R. D. Dowell et al. (2001). The distributed annotation system, BMC Bioinformatics 2: 7.
5. {{http://www.biouml.org/}}

WIDESPREAD EPIGENOMIC AND TRANSCRIPTIONAL CHANGES IN HUMAN ENDOTHELIAL CELLS REPROGRAMMED INTO PLURIPOTENT STATE

E.M. Vassina[1*], M.A. Lagarkova[1], E.A. Glazov[2], A.M. Mazur[3], E.B. Prokhorchouk[3], S.L. Kiselev[1]

1. Vavilov Institute of General Genetics RAS, Moscow, Russia

2. Diamantina Institute for Cancer; Immunology and Metabolic Medicine; The University of Queensland; QLD Australia

3. Centre "Bioengineering," Russian Academy of Sciences, Moscow, Russia

e-mail: vassinak@vigg.ru

**Corresponding author*

Key words: *DNA methylation and gene expression analysis, reprogramming, induced pluripotent stem cells, embryonic stem cells.*

Motivation and Aim: Induced pluripotent cells (iPS cells) are derived by epigenetic reprogramming, however their genome wide DNA methylation /gene expression patterns, and importantly their underlying association remain largely undefined.

Methods and Algorithms: To address this issue in our study we used two recently generated iPS cell lines derived from human umbical vein endothelial cells (HUVEC); their parental and one non parental HUVEC cell lines; other somatic primary cell lines and three human embryonic cell lines as their pluripotent counterparts. We conducted the genome-wide promoter methylation (Illumina Human Methylation27 BeadChip) and gene expression (Illumina HumanRef-8 Expression BeadChip) microarray analyses on these eight lines. Hierarchical clustering, K-means clustering and gene ontology analyses of differentially expressed genes and differentially methylated promoter CpGs were performed. Most of data analyses and visualization were produced using GenePattern software package available in open source. Merging of methylation and gene expression data were performed using Illumina GenomStudio Software Methylation module.

Results: During reprogramming of endo-iPS cells gene promoter elements and expression of lineage specific endothelial genes became "silenced". Clusters of transcriptionaly active genes were enriched with pluripotency-relatad genes in a fashion similar to ESC lines. Minor differences between iPS and ESCs cells were also observed and mostly consisted of genes which demethylation was not fully completed. Overall genome scale CpG methylation of promoter regions in the pluripotent cells was higher than in somatic. Genome wide gene expression profiling of pluripotent and somatic lines and degree of its association with promoter CpG methylation levels is in progress and will be discussed further.

Conclusion: DNA methylation and gene expression profiling on a genome-wide level offers new criteria for establishment of iPS cell lines. Accumulation of these data also provides a valuable recourse for computational biology as basis for comparative epigenome/transcriptome analysis.

COGNITIVE RESEARCH AT THE INTERSECTION OF INFO, BIO AND NANO

B.M. Velichkovsky[1,3]*, J.R. Helmert[1], S. Pannasch[2] and S.A. Shevchik[3]
[1]Institute of Psychology, Faculty of Mathematics and Natural Sciences, Dresden University of Technology, Germany; [2]Brain Research Unit, Low Temperature Lab, Helsinki University of Technology, Finland; [3]Institute of Cognitive Studies, Kurchatov Research Center, Moscow, Russia
[*]Corresponding author, e-mail: velich@psychologie.tu-dresden.de

Key words: *Brain Imaging, Cognition, Evolution, Eyetracking, Perception, Communication*

Motivation and Aim: In the last two decades, research on human cognition immensely benefited from application of methods and models of natural sciences. Within the NBIC-framework [1], one could expect a strong feedback from cognitive studies to the fields of computer sciences, molecular biology and nanotechnology. To find these synergies we investigated brain mechanisms of active vision and their implications for the development of innovative human-computer interfaces.

Methods and Algorithms: Fast video-based eyetracking is the method of choice in analyses of visual cognition. It is important from the point of view of ecological validity and practical applications. However, it is of significance for neurobiological research as well. As a matter of fact, eye movements are a common output of a number of philogenetically evolved brain systems. In the talk, we demonstrate how eyetracking analysis disentangles these influences on task solution completing data obtained with neuroimaging methods. Four groups of tasks are considered, with corresponding paradigms of eyetracking and neuroimaging. The first paradigm consists in the analysis of distractor influences on the duration of fixations in free exploration of complex images. The second investigates the role of attention in the perception of static and dynamic visual scenes. The third dissociates the subjective focus of visual work and the physical location of fixations. Finally, the fourth paradigm is related to an analysis of the role of social gaze in processes of communication.

Results and Conclusion: Up to five different brain systems are at play in regulating parameters of eye movements in these experimental situations, from low-level midbrain structures to those of medial prefrontal cortex. Therefore, by simply analysing human eye movements one can conclude on the work of these brain systems and on the task at hand. We analyse implications of these findings for understanding of biological mechanisms of cognition and designing the next generation of cognitive interface technologies such as adaptive assistance systems as well as attentive and multimodal interfaces (see [2, 3]).

Acknowledgements: The work has been supported by grants from ERC (#232946), EU Commission (PERCEPT), BMBF (ZOOM) and RFBR (ofi-m 09-06-12003).

References:

1. W.S.Bainbridge, M.C.Roco (2006). *Managing nano-bio-info-cogno innovations: Converging technologies for improving human performance*. Dordrecht: Springer.
2. B.M.Velichkovsky et al. (2010 in press). Measurement-related issues in investigation of active vision, In *Measurement with person,* B.Berglund et al. (Eds.), NY: Taylor & Francis.
3. U.Vogel et al. (2009). Bi-directional OLED microdisplay for interactive see-through HMDs, *Journal of the Society for Information Display,* **17(3)**: 175-184.

ANALYSIS OF THE DEGENERATE MOTIFS IN PROMOTERS OF miRNA GENES EXPRESSED IN DIFFERENT TISSUES OF PRIMATES

O.V. Vishnevsky[*1,2], K.V. Gunbin[1], A.V. Bocharnikov[2], V.V. Suslov[1], E.V. Berezikov[1,3]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
[3]*Hubrecht Institute, RNAAS, Utrecht, Netherlands*
e-mail: oleg@bionet.nsc.ru
[*]*Corresponding author*

Key words: *miRNA genes, promoters, oligonucleotide motifs*

Motivation and Aim: Numerous miRNAs play an important role in translation regulation, modulating embryo development, stem cells proliferation and tissue differentiation. Aberrant miRNA expression has been associated with deceases like cancer, microcephaly and schizophrenia. But it is too little known about regulation of miRNA expression.

The aim of our research is to reveal and analyze the regulatory signals in promoter regions of miRNA genes expressed in different tissues of primates.

Methods and Algorithms: The sets of promoters of miRNA genes, expressed in brain, excretory system and intestine from human, macaque, pan and pongo in the [-1000; +1] region relative to the transcription start site are obtained from DB FANTOM 4. The ARGO system [1] is used for detection of degenerate region-specific oligonucleotide motifs.

Results: A few hundreds of high-significant degenerate motifs are obtained and classified using TRRD and TRANSFAC DB. Some of them are categorized as potential transcription factor binding sites (TFBS), another motifs could be related to some still unknown species-specific TFBF or to some structural features of promoters, like short polyA-polyT runs are known to induce DNA curvature or to be "easily melting" sites.

We estimate the oligonucleotide similarity (H_{Oli}) of human promoters and promoters of evolutionary close and far species using the approach we offered [1]. An average similarity (H_{Align}) for promoters of the same species is estimated using pairwise alignment. It was shown that the H_{Oli} level in brain is very high for related to human species (pan (0.98) and pongo (0.95)), some lower for evolutionary distant species (macaque (0.86)), and very low for random sequences (0.2). Comparison of H_{Align} with H_{Oli} demonstrates high conservatism of brain-specific regulatory signals in promoters of primates (0.92), against a background of lowered average homology of its promoters (0.86). The lowest H_{Oli} level (average value is 0.82) between human and all others obtained for promoters of genes expressed in excretory system could demonstrate the significant differences in sweat secretion and corresponding changes in K^+/Na^+ metabolism regulation.

Conclusion: We revealed potential regulatory signals and transcription factor binding sites in promoters of miRNA genes expressed in different tissues of primates, which can be the targets for the following experimental analysis.

References:

1. O.V.Vishnevsky, N.A.Kolchanov (2005) ARGO: a web system for the detection of degenerate motifs and large-scale recognition of eukaryotic promoters, *Nucleic Acids Research*, **33**, 417-422.

ANALYSIS OF THE DEGENERATE MOTIFS IN REGIONS OF FOXA-BINDING SITES

O.V. Vishnevsky*[1,2], T.I. Merkulova[1], A.G. Ananko[2], M.M. Lavrentiev[2]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
e-mail: oleg@bionet.nsc.ru
**Corresponding author*

Key words: *FoxA2-binding sites, ChIP-Seq, oligonucleotide motifs*

Motivation and Aim: Foxhead box A is the family of transcription factors, whose expression is critical to the development and homeostasis of liver. Information about regions of FoxA binding sites (BS) localization, obtained in massively parallel *in vivo* and *in vitro* experiments requires analysis using the high-performance computer systems.

Methods and Algorithms: The set (1) of non-overlapping [-100;100] regions relative to the experimentally proved FoxA binding sites is obtained from TRRD. Another set (2) is formed of the sequences, derived in ChIP-Seq experiment [1], in region [-100;100] relative to positions of maximum peak height. The next advance of ARGO system [2] is used for detection of degenerate region-specific oligonucleotide motifs. This computer package is based on the FPGA and GPU technologies to process the massive data. We developed the ARGO_CEL [3] system to find potential composite regulatory elements (CE).

Results: All high-significant degenerate motifs obtained in every set are classified using TRRD and TRANSFAC DB. About a third of them are categorized as potential BS of FoxA, others were related to BS of another transcription factors. The significant CEs, consisting of FoxA BS- like motifs and of the motifs related to BS of other transcription factors are revealed in the set (1). Different subtypes of FoxA BS- like motifs were located in sequences, containing experimentally proved multiple FoxA BSs. The most of FoxA BS- like motifs revealed in the set (2) demonstrate significant correlation with the peak height (max. 14%). The analysis of (2) doesn't reveal any significant CEs, and moreover the CEs revealed in the set (1) of FoxA BSs don't demonstrate significance in the set (2). It is shown, that only the FoxA BS- like motifs revealed in the set (1) are over-represented in the set (2), but others are not significantly presented in the set (2).

Conclusion: We suppose that the sequences of BSs experimentally proved *in vitro* and obtained from functionally related regulatory regions, could not correspond exactly to all kinds of BSs revealed in massively parallel sequencing *in vivo* experiments.

Acknowledgements This work was supported by the SB RAS Integration Project No. 119

References:

1. E.D.Wederell, 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.
2. O.V.Vishnevsky, N.A.Kolchanov (2005) ARGO: a web system for the detection of degenerate motifs and large-scale recognition of eukaryotic promoters, *Nucleic Acids Research*, **33**, 417-422.
3. D.Y.Oshepkov, et al. (2008) Regulatory DNA Sequences: Computer-Assisted Research, In: *Computational system biology*, N.A.Kolchanov et al., 38-126, Publishing house of SB RAS, Novosibirsk.

ANALYSIS OF THE DEGENERATE MOTIFS IN PROMOTERS OF AUXIN RESPONSIVE GENES

O.V. Vishnevsky*[1,2], V.V. Mironova[1], A.G. Ananko[2], M.M. Lavrentiev[2]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
e-mail: oleg@bionet.nsc.ru
**Corresponding author*

Motivation and Aim: The plant hormone auxin plays a key role in plant development. In promoters of early auxin responsive genes the Auxin Response Elements (AuxREs) were found that are specifically bound by transcription factors of ARF (Auxin Response Factor) family. Experiments display that AuxRE often operates as a part of composite element (CE) [1]. We developed the ARGO_CEL [2] program for recognition of potential CEs in regulatory regions of auxin responsive genes.

Methods and Algorithms: Two sets of sequences were created using published data: (1) [-100;100] regions relative to the experimentally proved AuxREs; (2) [-2000;1] promoter regions of auxin responsive genes [3]. The ARGO system [4] based on the Field-Programmable Gate Array and Graphic Processor Unit technologies is used for detection of degenerate region-specific oligonucleotide motifs.

Results: About dozen of high-significant degenerate motifs are obtained per every set. The most significant motif **TGTCNC**, found in all sites, corresponds to the well known TGTCTC-like AuxRE sequence. A few other motifs, overlapping with it describe alternative variants of the flanking regions. Other significant motifs, revealed in set (1) don't relate to TGTCTC-like context, and are located in regions surrounding **TGTCNC.** Some of them co-present with alternative motifs of **TGTCNC** flanks, significantly anti-correlate with other alternative motifs of **TGTCNC** flanks forming potential CEs (like AuxRE/MYBcore). The most significant motifs revealed in the set (2) locate in core promoters and correspond to transcription factor binding sites, like TATA-box *etc*. Groups of co-presenting motifs form potential CEs. We demonstrate that the CEs found in the set (1) are significantly over-represented in core promoters of early auxin response genes.

Conclusion: We revealed potential CEs covering experimentally proved AuxREs, and in promoters of early auxin regulated genes. These CEs consist of ARFs binding site and of motifs, corresponding to binding sites of another transcription factors. The data can be used for the experimental analysis of ARF targets and for recognition of auxin regulated genes.

Acknowledgements: SB RAS Integration Project 119 and RFBR (№ 08-04-01214-a).

References:

1. Shin R. et al. (2007) The Arabidopsis Transcription Factor MYB77 Modulates Auxin Signal Transduction, *The Plant Cell,* 19: 2440–2453.
2. D.Y.Oshepkov, et al. (2008) Regulatory DNA Sequences: Computer-Assisted Research, In: Computational system biology, N.A.Kolchanov et al., 38-126, Publishing house of SB RAS, Novosibirsk.
3. Mironova et al. Proximal promoters are enriched with auxin responsive elements in early auxin induced genes. In this issue.
4. O.V.Vishnevsky, N.A.Kolchanov (2005) ARGO: a web system for the detection of degenerate motifs and large-scale recognition of eukaryotic promoters, Nucleic Acids Research, 33, 417-422.

COLLAGEN-RELATED PATTERNS IN GENOMES: RECOGNITION AND ANALYSIS

A.V. Vlasova*, P.K. Vlasov, F.A. Kondrashov, N.G. Esipova, V.G. Tumanyan
Engelhardt Institute of Molecular Biology RAS, Moscow, Russia
e-mail: *vlasova.av@gmail.com*
*Corresponding author

Key words: *collagen, pattern recognition*

Motivation and Aim: Many genomes contain genes with collagen-like (CLS) patterns not only collagen genes associated with various functions. Genes of such a kind were found in taxons − in eukaryotes, prokaryotes and even in viruses. It is some evidences about existence of collagen-containing genes in plant genomes. A specific accurate method for recognition of CLS segments in genome sequences would give us new candidates to genes with collagen pattern. The standard universal gene search programs have some restrictions, which complicate adequate annotation of collagen-containing genes. As a result collagen sequence can be discarded as a kind of repeat.

Methods and Algorithms: There is inherent nucleotide periodicity in collagen-containing genes corresponding to the Gly-X-Y pattern in the amino acid sequence. Our program scans input sequence to mark all CLS on it; some CLS are discarded in view of amino-acid restrictions in annotated collagens. Then, we calculate frequency distribution along the sequence in accord to obtained marking. Additionally, this simple procedure for fast and accurate CLS-annotation permits finding collagen insertions into different genes.

Results: During scanning of human genome we found all known genes with collagen pattern. Above them many isolated CLS are out of coding regions. CLS in the classic collagen genes are characterizing by especially high density. Thus, we demonstrate that it is possible to use simple criteria to identify collagen genes. Interestingly, scan of the *A. thaliana* genome revealed many CLS and some CLS may be treated as exons of the collagen-containing proteins.

Conclusion: Our finding on human genome scan suggests that criteria used make possible the accurate annotation of the collagen genes, as well as possible prediction of new collagen-containing protein genes. In this respect, CLS groups obtained in Arabidopsis genome looks like candidates to be exons of the collagen-containing protein genes, and it is an aim of experimentalists to confirm our suggestions.

Availability: We catalogue all CLS found in human genome in the database available on the web (http://bsb.imb.ac.ru/cgi-bin/clp.cgi).

Acknowledgements: This work was supported by grant from the Russian Foundation for Basic Research (No 08-04-01770a) and grant on Fundamental Sciences – Medicine of Presidium RAS.

References:

1. R. Hieta, J. Myllyharju. (2002) Cloning and characterization of a low molecular weight prolyl 4-hydroxylase from Arabidopsis thaliana. *J. Biol. Chem*. **277(26):**23965-71.
2. M. Rasmussen et al. (2003) Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *J. Biol. Chem.* **278(34):**32313-6.
3. Y. Xu et al. (2002) Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. *J. Biol. Chem.* **277(30):**27312-8.

CELLULAR CASPASES: NEW TARGETS FOR THE ACTION OF PHARMACOLOGICAL AGENTS

T.O. Volkova, O.V. Kurmyshkina*, N.N. Nemova
Petrozavodsk Stat University, Petrozavodsk, Russia
e-mail: VolkovaTO@yandex.ru
**Corresponding author*

Key words: *caspases, apoptosis, differentiation, myeloid cancer cell lines.*

Motivation and Aim: The search for new intracellular targets for the action of pharmacological agents is one of the central problems of modern biology and medicine. Enzymes playing a key role in signal transduction during the induction of proliferation, differentiation and apoptosis are considered the most promising among target molecules. Cellular caspases are one of such regulatory enzymes.

The present study aimed to investigate alteration of caspases 3, 6 and 9 gene expression as well as change in enzyme activity in erythromyeloleukosis cells K562 under the treatment with inducers of erythroid (cytosar, adriamycin) and myelocytic (dexamethasone, phorbol 12-myristate 13-acetate – PMA) differentiation. The extent of apoptosis induction in cell cultures was evaluated in parallel.

Methods and Algorithms: Differentiation of cells was determined using indirect immunofluorescence assay with FITC-labeled monoclonal antibodies (CD11c, CD14, GpA). Gene expression was assessed by real-time PCR, with SYBR Green as fluorophore. Amplification was performed on iQ5 using kits combined with reverse transcription (iScript One-Step RT-PCR Kit). Caspase activity was measured according to the manufacturer's protocol (BioRad, USA) with AFC-labeled substrates. Caspase substrates were as follows: 3 – DEVD (Asp-Glu-Val-Asp), 6 – VEID (Val-Glu-Ile-Asp), 9 – LEHD (Leu-Glu-His-Asp). DNA fragmentation (apoptosis) was estimated by electrophoresis in 2% agarose gel.

Results: Treatment of K562 cells with 2 mM cytosar or 1 μM adriamycin for 2 days resulted in enhanced caspase 3 gene expression and protein activity. The number of GpA-positive cells in cultures increased by 3.2–3.5 times. Apoptosis was only detected by the forth day against preliminary elevation of caspase 9 activity. Dexamethasone (5 μM) and PMA (100 nM) raised the number of CD11c-, CD14-positive cells by the forth day of treatment. An increase of caspase 3 and 6 gene expression and enzyme activity was observed in cultures with PMA, apoptosis was not detected. In cells incubated with dexamethasone caspases 3, 6, 9 were induced as well as pronounced DNA fragmentation. When a combination of PMA + cytosar (or adriamycin) was added to cell cultures simultaneously a redistribution of caspase activities was observed as compared with PMA-treated cells: caspase 6 activity decreased, caspases 3 and 9 activity dramatically increased, apoptosis was induced.

Conclusion: 1. Different caspase cascades are activated at the induction of distinct cell differentiation pathways, while apoptosis may not be observed.

2. Treatment of cells by combination of inducers leads to the redistribution of caspase functional activity. This fact should be taken into account when developing new-generation drugs specifically inhibiting caspases.

This work was supported by a President Program grant for the Leading Scientific Schools of Russian Federation №306.2008.4.

SELF-ASSOCIATION OF TRANSMEMBRANE HELICES: STRUCTURAL INSIGHT FROM COMPUTER SIMULATIONS.

P.E. Volynsky* A.A. Polyansky, D.E. Nolde, R.G. Efremov
Institute of Bioorganic Chemistry RAS, Moscow, Russia
e-mail: volynski@yandex.ru
**Corresponding author*

Key words: *oligomerization of membrane proteins, molecular modeling, free-energy calculations*

Motivation and Aim: Association of membrane proteins plays a very important role in cell life. Often, protein-protein interactions have a regulatory function. Thus, understanding on the atomic level the driving forces of membrane protein association is crucial for deciphering of their mechanisms of action. One of the most common elements of membrane proteins fold is an alpha-helix intersecting the membrane (transmembrane helix, TMH). So, it is rational to start the solution of MP's association from the simplest case – namely from dimerization of isolated TMH. In spite of its simplicity, such a system includes all important characteristics encountered in more complex MPs. Moreover, interaction of individual TMH has a serious biological impact – for instance, receptor tyrosine kinases are activated in answer to dimerization, which in turn depends on association of their single TMH. Therefore, delineation of molecular mechanisms of such processes will facilitate prediction of membrane protein structure and make possible design of membrane proteins with predefined structure/activity.

Methods and Algorithms: In this work TMH dimerization is studied via computer simulations. To predict the spatial structure of TMH dimers several approaches are used: (1) ab initio prediction via fitting of spatial and polar properties of monomers' surfaces, (2) Monte-Carlo conformational search in implicit membrane and (3) molecular dynamics in full-atom bilayers. Often to collect the reliable structure combinations of aforementioned methods are needed. The assessment of quality of the predicted structural models performed through the calculations of free energy of dimerization using umbrella sampling via restrained molecular dynamics in water-cyclohexane slab.

Results: Application of the aforementioned approaches to study of biologically relevant dimers (TMH of proapoptosis protein Bnip3, receptor tyrosine kinases EphA1, ErbB) resulted in prediction of structural models close to native structure. Moreover, it also provides important information about molecular determinants of dimerization: dynamic behavior of different groups, protein-lipid interactions and so on. Calculations of the dimerization free energy permitted qualitative assessment of importance of different groups in stability of the dimers.

Conclusion: Application of computer simulation techniques to study of dimerization of TMH provides an important information about their structure and energetic. Results, obtained for the dimers with experimentally known structure demonstrate that such analysis is an effective tool for structure prediction and for design of new TMH with predefined dimerization ability.

SPATIAL RESOLVED PROTEOMICS IN CANCER RESEARCH

Ferdinand von Eggeling*, Giinther Ernst, Christian Melle
Core Unit Chip Application (CUCA), Institute of Human Genetics, Jena University Hospital, Jena, Germany
e-mail: fegg@mti.uni-jena.de

Key words: *Proteomics, SELDI, MALDI imaging, microdissection, cancer, biomarker*

Biomarkers, especially diagnostic useful ones, are needed for different diseases. This is in particular true for tumor marker. In nearly all neoplasias the early detection is of high importance for adequate therapies and therefore for the survival of the patients. Until now most found protein markers are lacking from sensitivity and specificity and/or that they are not able to detect tumours in early stages. Therefore all efforts have to be directed to find highly specific early biomarkers, which have also to be inexpensive that they can be applied in a screening or routine diagnostic.

We believe that these low abundant markers can only be identified directly in tissue with a precise information about their localisation. Therefore, in this talk we want to illustrate our way from SELDI analysis of microdissected tissue to techniques that allow spatial resolved proteomics in cancer research.

References:

1. von Eggeling F, Melle C, Ernst G (2007) Microdissecting the proteome. Proteomics 7:2729-2737

NEW METHODS AND PROGRAM TOOLS FOR IN SILICO DOCKING AND VALIDATION OF LIGAND BINDING WITH PROTEIN TARGETS

Y.N. Vorobjev*, A.V. Popov
Institute of chemical biology and fundamental medicine of Siberian Branch of RAS, Novosibirsk, Russia
**Corresponding author ynvorob@niboch.nsc.ru*

Motivation and Aim. A rapidly growing number of experimental and theoretically predicted protein structures require blind docking methods for characterization of binding sites of proteins. *Methods and Algorithms* Friendly user interface with a set of program tools for preparation of data for protein target and ligand molecules and in silico blind docking tools are developed: the protein PDB file validator, the ligand topology file builder and the blind hierarchical docking method bhDock. The report describes its implementation and test results of accuracy assessment. The bhDock method utilizes two step algorithm, (1) analysis of entire protein surface, calculation of a comprehensive set of low-resolution binding sites, and their ranking by a simple score function, which is equal to a number of low-resolution contacts between low-resolution ligand model, resented as a sphere of radius 3.0-3.5 E and atoms of a protein, (2) refinement of ligand binding pose for a small set of high ranked low-resolution binding sites via a molecular dynamics based method of global optimization. The global optimization method utilizes (i) a set of initial ligand orientations uniformly distributed in the orientation phase space, (ii) simulated annealing coupled with guided force field deformation of protein-ligand interactions.

Results. Assessment of the bhDock method on the set of 37-protein-ligand complexes has shown the success rate of predictions of 78%, which is better than the rate reported for the most cited docking methods: AutoDock, DOCK, GOLD and FlexX on the same set of complexes . The method of estimation of an absolute free energy of binding is based on the two state approach, i.e. initial state with unbound protein and ligand in solvent and the final protein-ligand bound state. The free energy is calculated for a representative set of microstates of molecular dynamic trajectory for a free protein/ligand and for bound protein-ligand complex. A solvation energy are calculated by a state of art Fambe-pH[1,2] method taking into account pH dependence of solvation energy in bound and unbound state of the ligand. It is shown a good correlation between calculated and experimental data for a set of azole ligands of cytochrome p450 CYP51.

Conclusion The developed tools open the door to unified and accurate approach toward in silico drug discovery.

Acknowledgements: This research was supported by RFBR 09-04-00136, SB RAS integration projects (№16, №119, 2009-2011) and Russian Ministry of Education and Science (02.740.11.0079/NS-652.2008.4).

References:

1. Vorobjev Y.N., Vila J.A., Scheraga H.A. J Phys Chem B 2008: 112, 11122-11136.
2. Yelena A. Arnautova, Yury N. Vorobjev, Jorge A. Vila and Harold H. Scheraga. Identifying native-like protein structures with scoring functions based on all-atom ECEPP force fields, implicit solvent models and structure relaxation. PROTEINS: Structure , Function, and Bioinformatics **2009**, 77: 38-51

EVIDENCE OF NEGATIVE SELECTION AGAINST MIRNA EXPANSION OVER HUMAN GENOME

S. Vorozheikin*[1], A. Kulikov[1,2], I. Titov[1,3]
1-Novosibirsk State University, Novosibirsk, Russia
2-Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia
3-Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: Pavel.Vorozheykin@gmail.com
**Corresponding author*

Key words: *miRNA, miRNA gene, RNA secondary structure, single substitution, evolution*

Motivation and Aim: While miRNA functioning becomes increasingly better investigated, little is known about the origin and evolution of miRNA.

Methods and Algorithms: In the study we used the sequences of known human mature miRNAs, pre-miRNAs (http://www.mirbase.org/, release 12), and human genome (http://www.ncbi.nlm.nih.gov/, release 37.1).

Results and Conclusion: For the 692 known human miRNAs we have found a surprisingly huge number of 2122 homologs which may be considered as "young" or "candidate" miRNAs. 785 of them retain complete sequence similarity to the original miRNA genes, in accordance with the inverted duplication hypothesis [1]. Of the rest 1278 young miRNA genes we have chosen those which had accumulated from 1 to 8 single substitutions and have compared their secondary structure with the structure of origins. The number of substitutions which preserve original structure was found to be significantly less than expected by Monte Carlo test taking into account the rates of single substitutions in human genome [2], independently of number of substitutions. This suggests the negative selection against fast miRNA expansion over human genome.

Finally we have found that either miRNA or young miRNA nearest-neibor-distance distributions obey the similar Weibull laws. We have performed the simulations of miRNA inverted duplications and have fitted the Weibull parameters.

Acknowledgements

The work was supported by RAS program "Biosphere origin and evolution" (GC №10104-37/П-15/110-327/150409-001) and SD RAS Multidisciplinary Project #119.

References:

1. Allen, E., Xie, Z., Gustafson, A.M., Sung, G.H., Spatafora, J.W., and Carrington, J.C. (2004) Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. *Nat. Genet*. **36**: 1282–1290.
2. M. Krawczak, E.V. Balla and D.N. Coopera (1998) Neighboring-Nucleotide Effects on the Rates of Germ-Line Single-Base-Pair Substitution in Human Genes. *Am. J. Hum. Genet.* V. **63**, pp.474–488.

THE DATABASE ON TRANSCRIPTION FACTOR BINDING SITES DERIVED FROM CHIP-SEQ EXPEIMENTS

I.S. Yevhsin[1,2,*], Y.V. Kondrakhin[1,3], R.N. Sharipov[1,4], F.A. Kolpakov[1,3]
1Institute of Systems Biology, Novosibirsk, Russia; 2Novosibirsk State University, Novosibirsk, Russia; 3Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia; 4Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia;
e-mail: ivan@systemsbiology.ru
**Corresponding author*

Key words: *transcription factor binding site, ChIP-Seq, database*

Motivation and Aim: Precise identification of transcription factor binding sites (TFBS) is an important challenge of bioinformatics. Recently developed ChIP-seq method combines chromatin immunoprecipitation (ChIP) with high-throughput massively parallel sequencing and allows identification of *in vivo* TFBS in the genome scale. However ChIP-seq data requires further analysis for accurate localization of TFBS.

Methods and Algorithms: We built the comprehensive collection of human TFBS identified using published data of ChIP-seq experiments. We used bowtie software [1] for alignment of short reads to the human genome. Since different algorithms for identification of genome regions enriched by reads can generate quite distinguished sets of fragments we applied two most popular methods - SISSRs [2] and MACS [3] to each dataset. Most of the published ChIP-seq data were processed in the same fashion and resulting genome regions were stored in MySQL database along with experiment and computational analysis descriptions. For precise localization of binding sites in identified genome regions we built sequence motif models using extended position weight matrix (PWM) method [4] and predicted exact locations of binding sites.

Results and Conclusion: Information from 44 published ChIP-seq experiments on 28 distinct human transcription factors was gathered in specialized database. The database supports queries by transcription factors, gene names and identifiers, genome coordinates, cell types and conditions. Query result can be displayed in tabular form as well as using BioUML genome browser (http://server.biouml.org/bioumlweb). Further collected data will be combined with microarray data for reconstruction gene networks and simulation gene expression.

Acknowledgements: This work was supported by EU grants FP6 №037590 “Net2Drug” and FP7 №202272 “LipidomicNet”.

References:

1. B.Langmead et al. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, *Genome Biology,* **10(3):** 212.
2. R.Jothi et al. (2008) Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data, *Nucleic Acids Research,* **36(16):** 5221-5231.
3. Y.Zhang et al. (2008) Model-based analysis of ChIP-Seq (MACS), *Genome Biology*, **9(9):** R137.
4. E.Ananko et al. (2007) Recognition of interferon-inducible sites, promoters, and enhancers, *BMC Bioinformatics,* **8(56)**.

MATHEMATICAL MODELING OF ANTIOXIDANT SYSTEM IN RATS WITH ASCITIC ZAJDEL HEPATOMA

Yu. Shatalin[1], I. Yevshin[2,3,*], T. Sukhomlin[1], G. Ermakov[1], R. Sharipov[3,4], G. Dymshits[4,2], A. Naumov[1], M. Potselueva[1]

[1]Institute of Theoretical and Experimental Biophysics RAS, Pushchino, Russia; [2]Novosibirsk State University, Novosibirsk, Russia; [3]Institute of Systems Biology, Ltd, Novosibirsk, Russia; [4]Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

e-mail: ivan@systemsbiology.ru

**Corresponding author*

Key words: *Zajdel hepatoma, ascites, oxidative stress, antioxidant system*

Motivation and Aim: It is known that tumor development leads to activation of innate immunity and oxidative stress which affect many physiological functions. In this study we measured the dynamics of antioxidant components in blood plasma and ascitic fluid in the process of tumor development. As the model of tumor development we used ascitic Zajdel hepatoma transplanted into peritoneal cavity of Wistar rats.

Methods and Algorithms: Recently elaborated biochemical, spectral and immune assays methods were applied to follow concentrations of individual antioxidants, reactive oxygen species, metabolites and regulatory protein in dynamics. BioUML workbench (http://www.biouml.org) was used for the formal description and simulation.

Results and Conclusion: We observed a reverse dynamics of several parameters (reactive oxygen species generation capacity of phagocytes; transferrin, ceruloplasmin, uric acid, alpha-tocopherol concentrations) related to oxidative stress in blood and ascitic fluid [1]. We revealed experimentally that oxidative stress induced enhancement of antioxidant defence (e.g., ceruloplasmin increase) in blood, but not in ascites. We hypothesised that such dynamics can be at least partially explained by the fluid redistribution and passive transport between blood plasma and ascites. To check this hypothesis we developed the mathematical model of passive transport between blood plasma and ascites. It takes into account the transport of fluids, proteins (e.g., transferrin, ceruloplasmin) and low molecular weight antioxidants (uric acid and alpha-tocopherol). All data collected for modeling including semantic diagrams and mathematical models were deposited in the BMOND database (http://bmond.biouml.org, DGR0355_1-DGR0355_9). The values of model parameters were either obtained from literature or calculated from experimental data. The time courses of ceruloplasmin, transferrin and uric acid in plasma and in ascites were simulated. The trends of simulated time courses represent qualitatively the experimental dynamics. Further model development is needed to describe the dependence of the plasma-ascites transfer on dynamics of inflammatory stimuli, oxidative stress and antioxidant capacity of the physiological fluids studied.

Availability: The model and related data are available at the BMOND database (http://bmond.biouml.org).

Ackowledgements: This work was supported by the grant №2.1.1.6663 of the program "Development of scientific potential of higher school (2009-2010 years)".

References:

1. *M.M.Potselueva et al*. (1999) Generation of reactive forms of oxygen by polymorphonuclear leukocytes during hepatoma growth in the peritoneal cavity of animals, *Tsitologiia,* **41:** 162-166.

A NEW MODE OF TFR2-DEPENDENT IRON DEPOSITION IN CELLS UNDER OXIDATIVE STRESS CONDITIONS

I. Yevshin[1,2,*], R. Sharipov[2,3], Yu. Shatalin[4], G. Ermakov[4], G. Dymshits[3,1], M. Potselueva[4], A. Naumov[4], T. Sukhomlin[4]

[1]Novosibirsk State University, Novosibirsk, Russia; [2]Institute of Systems Biology, Ltd, Novosibirsk, Russia; [3]Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia; [4]Institute of Theoretical and Experimental Biophysics RAS, Pushchino, Russia;
e-mail: ivan@systemsbiology.ru
**Corresponding author*

Key words: *iron metabolism, transferrin receptors, oxidative stress, mathematical modeling*

Motivation and Aim: Iron (Fe) ions are necessary for cell proliferation, metabolism and many other cellular functions. Cells uptake Fe in a complex with transferrin (Tf) via two receptors TfR1 and TfR2. TfR1 mediates a well-known way whereas the role of TfR2 is poorly known. TfR2 demonstrates some exclusive abilities: stabilization by holotransferrin molecules and biphasic Tf uptake. In oxidative stress and inflammation the increase of Fe level (both free and bound by Tf) is often observed, while the level of free Tf is decreased. We assumed that the reason of this phenomenon is the elevated Tf uptake by cells specialized for Fe deposition and/or transport mediated by specific properties of TfR2. To check this hypothesis we build mathematical model of Fe and Tf uptake by TfR2(+)-cells and investigated its dynamical properties.

Methods and Algorithms: For formal description and simulation of the model of transferrin uptake by a cell we used BioUML workbench (http://www.biouml.org). Fe homeostasis system was divided into several interacting modules: (1) Fe uptake regulation (including TfR1- and TfR2-induced endocytosis of different forms of Tf, vesicular transport, endosomal sorting and Fe transport from endosomes to cytoplasm; (2) ferroportin-dependent Fe export from a cell; (3) Fe deposition in a complex with ferritin; (4) regulation of Fe regulatory proteins (IRP); (5) posttranscriptional regulation by IRP system; (6) regulation of TfR2 expression; (7) regulation of ferritin expression. For description of molecular interactions we used chemical kinetics equations, for description of gene expression the Hill equation was applied. Some model parameters were extracted from literature other was fitted on the base of obtained by authors experimental data. Corresponding system of differential equations was solved using BioUML workbench.

Results and Conclusion: The results of modeling demonstrated that Tf concentration may decrease due to biphasic uptake of Tf by TfR2(+)-cells – a new mode of iron uptake by this type of cells. Increased saturation of Tf with Fe as a consequence of high level of Fe and low level of free Tf may lead to forcing of TfR1-dependent transport of Fe as a result of higher TfR1 to Tf:2Fe affinity (in comparison with free Tf and Tf:1Fe) and activation of TfR2-dependent uptake of Tf:2Fe due to stabilization of TfR2 receptor.

Availability: The model and related data are available at the BMOND database (http://bmond.biouml.org).

Acknowledgements: This work was supported by the grant №2.1.1.6663 of the program "Development of scientific potential of higher school (2009-2010 years)".

BIOINFORMATIC ANALYSIS OF RARE COMBINATIONS OF SINGLE NUCLEOTIDE POLYMORPHISMS IN THE HUMAN GENOME

N.S. Yudin*, E.V. Ignatieva, and V.M. Efimov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: yudin@bionet.nsc.ru
**Corresponding author*

Key words: *single nucleotide polymorphism, gene, human, HapMap project, inter-SNP distance*

Motivation and Aim. Recently, we have found two single-nucleotide polymorphisms (SNPs) separated by one nucleotide in an intron of the RYR1 gene in the Kemerovo pig breed [1]. This combination of SNPs is likely to have arisen in the animal population as a result of two independent mutations, and therefore it should be a very rare event. The aim of this study is to assess the frequency of similar combinations of SNPs (two SNPs separated by one nucleotide in an intron of a gene, detected experimentally at different frequencies among representatives of the same population) in the human genome, as in this mammalian genome SNPs have been studied best.

Methods and Algorithms. The data extracted from the HapMap data set were screened for pairs of intronic SNPs separated by one nucleotide. The analysis was performed on SNPs mapped on the shortest human chromosome 22. The data were obtained using the UCSC table browser. The inter-SNP distances between neighboring SNPs were calculated in Microsoft Excel. The selection of combinations of SNPs, which, according to the HapMap data, had different frequencies in the same population, was based on queries to dbSNP.

Results. Combinations of two intronic SNPs separated by one nucleotide that varied in frequency within the same HapMap population (US residents with Northern and Western European ancestry, Yoruba, Chinese or Japanese) were found only 15 times (0.05%) of the 58,128 SNPs genotyped in 578 genes on human chromosome 22. Three of these combinations were found in the same gene (*LARGE*), and two combinations were found in each of the *EMID1* and *ATXN10* genes. The results of combinatorial analysis showed that the null hypothesis for the observed distribution of 15 pairwise combinations of SNPs in 11 genes among 578 genes on chromosome 22 was rejected at $P < 10^{-5}$. It can be speculated that the particular pattern of genetic variation in the *LARGE*, *ATXN10* and *EMID1* genes is a result of the tendency of the DNA regions within these genes to mutate more frequently. A number of single nucleotide mutations causing congenital muscular dystrophy have been reported in the human *LARGE* gene.

Conclusions. Combinations of two SNPs separated by one nucleotide in intron of a gene, detected experimentally with different frequencies among representatives of the same population, are very rare in the human genome. Single-nucleotide polymorphisms that meet this criterion constitute 0.05% of all SNPs localized on human chromosome 22. A non-random concentration of these unique combinations in certain genes is observed.

References

1. N.S. Yudin et al. (2010) A rare combination of two new single nucleotide polymorphisms in intron 16 of the RYR1 gene in the Kemerovo pig breed. VOGiS Herald. In press.

DIAGNOSTICS OF CANCER DISEASES ON GENE EXPRESSION

N.G. Zagoruiko*, I.B. Borisova, V.V. Dyubanov, O.A. Kutnenko
Institute of Mathematics SB RAS, Novosibirsk, Russia
e-mail: zag@math.nsc.ru
**Corresponding author*

Key words: *gene expression, pattern recognition, feature selection, leukemia*

Motivation and Aim: It is required to choose a subset of genes, on which expression it is possible to distinguish two kinds of leukemia. Complexity of task is in a fact that initial number of attributes (genes) on 2-3 orders of magnitude larger than number of objects (patients). The purpose is in developing such methods of informative attributes selection and constructions of effective decision rules which provide high reliability of disease diagnostic.

Methods and Algorithms: The directed search of attributes is done by iterative algorithm FRiS-GRAD, as criterion for attributes selection the compactness of the training dataset is used, decision rules are constructed by the algorithm FRiS-Stolp. All these algorithms are based on use of new measure of similarity between objects *F* in the form of function of rival similarity (FRiS-function): if the distance from recognized object *z* to the standard of pattern *A* is equal r_1, to the standard of pattern *B* is equal r_2, similarity *z* with pattern *A* in competition with pattern *B* equals to $F(z,A|B) = (r_2-r_1) / (r_2+r_1)$ [1]. Compactness of pattern is determined as average value of rival similarity between all pairs of objects [2].

Results: The specified algorithms were used for some medical tasks solving: diagnostics of prostate cancer on spectra of fibers (162 objects, 17135 attributes), of diabetes of II type on genes expression (43 objects, 5127 attributes) and diagnostics of two types of a leukemia (38 objects, 7143 attributes). It is possible to find some subsets in each task consisting of 3-10 attributes on which 97-100 % accuracy of the diagnosis is achieved. For example, in task of two types of leukemia (AML and ALL) recognition 27 informative subsets on 4-6 genes were selected from 7143 genes, on which expression control objects are recognized without errors It takes 30 seconds of work of the Pentium computer. Best of earlier published results for this task [3] is one subsystem from 128 genes which recognizes control dataset with reliability of 97 %.

Conclusion: The algorithms of recognition based on FRiS-function are developed which allows solving effectively tasks with different proportion of objects and attributes. While analyzing information received from biochips, these algorithms using allows to select low volume subsets of genes, which is essential for different classes of diseases diagnose.

Availability: The specified algorithms are presented on page www.math.nsc.ru/~www.zag Concerning use of programs it is necessary to address to corresponding author.__

Acknowledgements; The work was supported by RFBR, grant 08-01-00040.

References:

1. 1. N.G. Zagoruiko, I.A. Borisova, V.V. Dyubanov, and O.A. Kutnenko. (2008). Methods of Recognition Based on the Function of Rival Similarity//*Pattern Recognition and Pattern Analysis*, Vol. 18. No.1, pp.1-6.
2. 2. Zagoruiko N.G. Measure of Similarity and Compactness in Competitive Space // Advance in Intelligent Data Analysis. Springer-Verlag: Berlin, Heidelberg. 2009. pp. 369-380.
3. 3. Isabelle Guyon, Jason Weston, Stephen Barnhill, Vladimir Vapnik. (2002). Gene Selection for Cancer Classification using Support Vector Machines // *Machine Learning*, 46 (1-3): pp. 389-422.

SUBSTRATE SPECIFICITY OF DNA REPAIR ENZYMES AND ITS INFLUENCE ON MUTATION SPECTRA

D.O. Zharkov
SB RAS Institute of Chemical Biology and Fundamental medicine, Novosibirsk, Russia
dzharkov@niboch.nsc.ru

Key words: *DNA damage, DNA glycosylases, substrate specificity*

Motivation and Aim: DNA and the cellular pool of deoxyribonucleotide triphosphates are exposed to many damaging factors, including UV and ionizing radiation, reactive oxygen species and other reactive metabolites, environmental chemicals, etc. Damaged bases that appear in DNA are the source of mutations if left unrepaired. Normally, damaged bases are excised by dedicated enzymes, DNA glycosylases, in the process called base excision repair. The damaged dNMP is then replaced by canonical dNMP in the process directed by the undamaged base opposite the site of the lesion. Therefore, if DNA glycosylases remove the damaged base from a lesion-containing mismatch, a mutation will occur.

Methods: We have investigated the propensity of several DNA glycosylases (Fpg and Nei from *E. coli*, OGG1 and NEIL1 from mammals) to excise damaged bases from lesion-containing mismatches using the methods of steady-state and pre-steady-state enzyme kinetics.

Results: All DNA glycosylases were to some extent proficient in the removal of their preferred substrate lesions from mismatches. Both Fpg and OGG1 normally excise 8-oxoguanine (oxoG) from pairs with C and discriminate against pre-mutagenic excision from oxoG:A pairs. However, in terms of the specificity constant, this discrimination was only ~20-fold for Fpg and ~5–10-fold for OGG1. A naturally occurring D322N OGG1 variant was 2.3-fold more specific than the wild-type enzyme. The specificity of OGG1 was improved by its interactions with AP endonuclease, the enzyme that acts in base excision repair after DNA glycosylases. Nei and NEIL1 revealed little opposite-base discrimination when processing their natural substrates, oxidized pyrimidines. However, both NEIL1 and OGG1 were able to excise a damaged purine, 8-oxoadenine, exclusively from pre-mutagenic pairs with C. Yet DNA polymerases were unable to form this pair, suggesting that cells are protected from potentially dangerous glycosylase reaction at the level of formation of the lesion-containing mismatch. Investigation of DNA glycosylases by stopped-flow kinetics with fluorescent detection revealed that there are several conformational transitions during the catalytic cycle, some of which provide the basis for discrimination between normal and mismatched substrates.

Conclusion: Although DNA glycosylases are indispensable for removal of damaged bases, their substrate specificity is often insufficient to prevent pre-mutagenic base excision from damage-containing mismatches. Prevention of mutagenesis arising from such mismatches likely requires additional contribution by DNA polymerases and other enzymes participating in base excision repair.

Acknowledgements: This research was supported by the RAS Presidium (Molecular and Cellular Biology Program, 22.14), RFBR (08-04-00596), National Cancer Institute (CA017395-33), and Russian Ministry of Education and Science (02.740.11.0079/NSh-652.2008.4).

NOVEL SEQUENCING METHODS IN EPIGENETICS: DATA GENERATION AND BIOINFORMATICS

S.V. Zhenilo
ZAO "Genoanalitika",Moscow, Russia
e-mail: *szhenilo@gmail.com*

Key words: *next-generation sequencing, ChIP-seq, epigenetics, methyl-DNA-binding proteins*

Motivation and Aim: Owing to rapid technological developments in next-generation sequencing, the available genomic assays have been increased. Chromatin immunoprecipitation followed by genome-wide sequencing is now the main tool for assaying protein-DNA binding in vivo. DNA methylation is one of the epigenetic mechanisms that regulate gene expression. My aim was to analyze the real binding sites of methyl-DNA binding protein Kaiso, which uses a three-zinc-finger motif to bind methylated CGCGs (1). The zinc fingers of Kaiso have a dual specificity in vitro, as they can bind either DNA sequences containing methylated CGCG or the consensus Kaiso binding site (KBS), TCCTGCNA(2). Using ChIP-seq I decided to resolve the question about dual specificity of Kaiso.

Methods and Algorithms: ChIP-seq using different cell lines (mouse fibroblasts wild type and Kaiso-knockout, Kaiso-resin); bioinformatics analyses: base calling, mapping to the genome, algorithms for peak finding, algorithms for analyses of peak distribution in genome, algorithms for consensus site findings.

Results: Using ChIP-seq technology it was shown that Kaiso interacts with some genomic regions that are mainly located at gene promoters and genic regions in CpG islands. Also Kaiso binds throughout human and mouse genome as other methyl-DNA binding protein MeCP2. The analyses of sequencing data have shown that KBS site is not overrepresented in obtained peaks. On the other hand all peaks contained CG dinucleotides and CGCG site is overrepresented in comparison with MBD2 transcription factor.

Conclusion: ChIP-seq technology allows us to resolve the main question about dual specificity of Kaiso. We can conclude that Kaiso first of all methyl-DNA binding protein and binds mainly promoter and genic regions.

Availability: Software used in this work is open source.

References:

1. Prokhortchouk A. et al.(2001) The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev.* Jul 1;15(13):1613-8.
2. Daniel JM. et al. (2002) The p120(ctn)-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides. Nucleic Acids Res. Jul 1;30(13):2911-9.

DYNAMICAL MODELING OF MICRORNA MECHANISMS

A. Zinovyev *[1,2,3,6], N. Morozova[4], N. Nonne[4] , E. Barillot[1,2,3], A. Harel-Bellan[4] and A.N. Gorban[5,6]

[1]*Institut Curie, Paris, France*
[2]*INSERM, U900, Paris, F-75248 France*
[3]*Mines ParisTech, Fontainebleau, F-77300 France*
[4]*CNRS FRE 2944, Villejuif, France*
[5]*University of Leicester, Center for Mathematical Modeling, Leicester, UK*
[6]*Institute of Computational Modeling SB RAS, Krasnoyarsk, Russia*
e-mail: andrei.zinovyev@curie.fr
[*]*Corresponding author*

Key words: *microRNA, translation, mathematical models, limiting steps, dominant systems*

Motivation and Aim: Protein translation is a multistep process which can be modeled as a cascade of biochemical reactions (initiation, ribosome assembly, elongation, etc.), the rate of which can be regulated by small non-coding microRNAs through multiple mechanisms. It remains unclear what mechanisms of microRNA action are the most dominant: moreover, many experimental reports deliver controversal messages on what is the concrete mechanism actually observed in the experiment. Nissan and Parker [1] have recently demonstrated that it might be impossible to distinguish alternative biological hypotheses using the steady state data on the rate of protein synthesis.

Methods and Algorithms: For rigorous analysis of the dynamical models of protein translation, we used asymptotic approach which allowed us to derive semi-analytical solutions of kinetic equations [2].

Results: In contrary to the study by Nissan and Parker, we show that dynamical data allow discriminating some of the mechanisms of microRNA action. We formulate a hypothesis that the effect of microRNA action is measurable and observable only if it affects the dominant system (generalization of the limiting step notion for complex networks) of the protein translation machinery.

Conclusion: Our analysis of the transient protein translation dynamics shows that it gives enough information to verify or reject a hypothesis about a particular molecular mechanism of microRNA action on protein translation. For multiscale systems only that action of microRNA is distinguishable which affects the parameters of dominant system (critical parameters). Asymptotic approach to kinetic models allows to put in order diverse experimental observations in complex situations when many alternative hypotheses co-exist.

Acknowledgements: We thank Vitaly Volpert and Laurence Calzone for inspiring and useful discussions.

References:

1. T. Nissan, R. Parker (2008) Computational analysis of miRNA-mediated repression of translation: Implications for models of translation initiation inhibition. *RNA*, **14**:1480-1491.
2. A. Zinovyev et al. (2010) Dynamical modeling of microRNA action on the protein translation process, *BMC Syst Biol,* **24:** 4(1):13.

ON MODEL PHENOMENOLOGY OF VASCULAR CAMBIUM ACTIVITY

U.S. Zubairova*, S.V. Nikolaev
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: ulyanochka@bionet.nsc.ru
**Corresponding author*

Key words: *vascular cambium, parameterized L-system, mitotic index*

Motivation and Aim: The vascular cambium is a type of meristem – tissue consisting of embryonic cells from which the cells of vascular tissue originate. Study of the vascular cambium activity is a difficult experimental task. In this work we are interested in modeling dynamics of radial distribution of mitotic index in cambium – a measure of cell growth activity. We propose a one-dimensional model of cellular structure of cambium built in the formalism of parameterized L-systems. Cambium cells along the radial axis of the tree trunk form a one-dimensional cell array of the model. Cells can grow and divide, creating a stream from some primary cell inside the tree trunk. Rate of cell growth depends on position of the cell in the cambium, and may depend on external and internal factors.

Methods and Algorithms: For simulation of cell division we used the formalism of parameterized L-systems [2]. Alphabet of the L-system consists of three letters: *Sn* – cell in interphase, *Sm* – cells that are undergoing mitosis (nuclear division), *Sd* – differentiated non dividing cells. Cells of the *Sn* and *Sm* types belong to cambial zone, cells of the *Sd* type leaved this zone. Parameter vector contains phase of the cell cycle, sequence number in the array, time when the cell appeared and moved into a state of mitosis and others. Rules of the L-system reflect the cell cycle progression.

Results: The model is developed as a *Mathematica* program which can be used for computer experimentations to study dynamics of radial distribution of mitotic index in dependence of duration of cell cycle phases and its initial distribution. The last is the axiom of the L-system, and it can be generated in different ways, including a one based on "natural way": string length of 15 cells is generated by L-system rules from a single cell.

Conclusion: The duration of the cell cycle can be estimated from measurements of mitotic index, defined as ratio of the number of cells in cambium zone are in mitosis to the total number of cells in the cambium zone. The computer experiments were performed with the next protocol: for a hundred trees, we simulated dynamics of the cambium cells growth and divisions in course of "a season of vegetation" and the mitotic index was calculated. A set of parameters exist, which supply a periodicity in dynamics of the mitotic index distribution.

Acknowledgements: This work was supported by the RFBR (08-04-01214-a), SB RAS Integration project 119, RAS program №22 (project 8).

References:

1. E.A. Vaganov, U.K. Dzhanseitov. (2000) A phenomenological approach to the analysis of kinetic characteristics of cambium zone in the trunks of coniferous trees, *The successes of modern biology,* **120 #2:** 190-200.
2. P. Prusinkiewicz, A. Lindenmayer (1990) The algorithmic beauty of plants. *N.Y.:Springer-Verlag*.

Author index

C

D

E

F

G

H

I

J

K

L

M

N

O

P

R

S

T

U

V

W

X

Y

Z

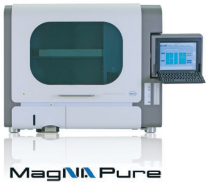
Roche Applied Science

Комплексное решение для клеточных и генетических лабораторий

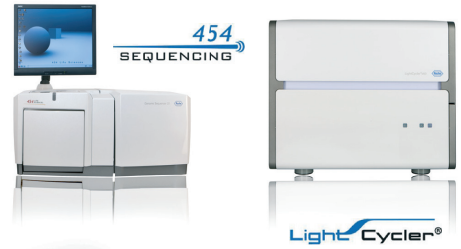
ООО «Рош Диагностика Рус» представляет в России мирового лидера в области молекулярной диагностики, швейцарскую компанию F. Hoffmann-La Roche, в том числе ее подразделение Roche Applied Science. Мы предлагаем диагностическим и научно-исследовательским лабораториям инновационные высокотехнологичные системы для клеточного анализа, экстракции нуклеиновых кислот, ПЦР в реальном времени и геномного секвенирования нового поколения.

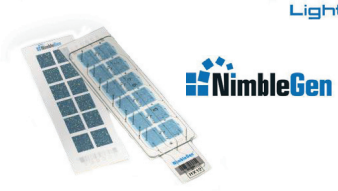
Клеточный анализ

Экстракция нуклеиновых кислот



Генетический анализ





Системы Innovatis и xCELLigence для клеточного анализа

- Подсчет и анализ жизнеспособности клеток
- Анализ морфологии и степени агрегированности клеток
- Анализ пролиферации клеток
- Анализ цитотоксических эффектов в реальном времени
- Анализ миграции / инвазии клеток и т.д.

Системы MagNA Pure для автоматического выделения нуклеиновых кислот из:

- Сыворотки, плазмы и цельной крови
- Тканей, биопсийного материала и гистологических срезов
- Соскобов, БАЖ, мочи, кала и др.

ПЦР-анализаторы LightCycler с детекцией в режиме реального времени

- Количественный анализ нуклеиновых кислот
- Анализ экспрессии генов
- Генотипирование, обнаружение точечных мутаций и метилирования (анализ кривых плавления, в том числе HRM)

Геномные секвенаторы нового поколения GS FLX и GS Junior

- Полное De Novo секвенирование и ресеквенирование геномов
- Таргетное ресеквенирование областей, ассоциированных с заболеваниями
- Секвенирование ампликонов (поиск редких соматических мутаций, гаплотипирование, HLA-типирование, резистентность ВИЧ) и т.д.

ДНК-чипы NimbleGen

- Сравнительная геномика
- Исследование метилирования ДНК
- Анализ экспрессии генов
- Исследование хроматиновой иммунопреципитации и т.д.

ООО "Рош Диагностика Рус"
Россия, 107031 Москва, Трубная площадь, 2
Тел./факс: +7 (495) 229 69 99 / 229 79 99
www.roche.ru
www.roche-applied.science.com

О компании HP

Компания HP является мировым лидером в сфере информационных технологий. Портфель продуктов и решений HP включает в себя принтеры, персональные компьютеры, программное обеспечение, ИТ-решения и сервисы, которые помогают как конечным пользователям, так и компаниям сделать работу с информационными технологиями максимально простой и удобной. С дополнительной информацией об HP (NYSE, Nasdaq: HPQ) можно ознакомиться на сайте www.hp.ru.

Genoanalytica Ltd. specializes in microarray analysis of DNA and RNA samples, as well as genome, transcriptome and ChIP sequencing.

Our staff consists of highly professional specialists with various academic degrees. We had performed hundreds of analyses for our customers' scientific and commercial needs.

Our services include:

MICROARRAYS

- SNP detection and statistical analysis (whole genome or targeted one);
- CpG islands methylation analysis and correlation of it with gene activity;
- Whole-genome expression analysis

We are using high-quality microarrays from the industry leader – Illumina. You can find the list of available array types on their website: www.illumina.com.

SEQUENCING

- Resequencing, for genomes with known reference sequences;
- *De novo* sequencing, e.g. acquiring genome contigs for a species with unknown reference sequences;
- Identification of histone modifications via ChIP-seq experiments.

BIOINFORMATICS

- Mapping of the sequence reads to the reference genome to determine their genomic coordinates and to locate simple polymorphisms in sample;
- Discovering of intraspecific genetic polymorphisms;
- *De novo* assembly of new genomes and transcriptomes;
- ChIP-seq analysis;
- Bisulphite sequencing analysis.

CONSULTING

We will help you to efficiently utilize your resources and advise you on various technological platforms and methods to choose. We can optimize your expenses in genomic studies and select the most advantageous options in your case.

For additional information please visit our website: www.genoanalytica.ru. Also you are welcome to call us: +7 (495) 930 8353, +7 (495) 764 6591; or write a letter to info@genoanalytica.ru.

We would be delighted to help with your experiments.

Научное издание

Труды седьмой международной конференции «Биоинформатика регуляции и структуры генома\ Системная биология»

на английском языке

Proceedings of the Seventh International Conference on Bioinformatics of Genome Regulation and Structure\ Systems Biology

Abstracts have been printed without editing
as received from the authors

Подготовлено к печати
в редакционно-издательском отделе
Института цитологии и генетики СО РАН
630090, Новосибирск, пр. акад. М.А. Лаврентьева, 10

Дизайн и компьютерная верстка: А.В. Харкевич

Подписано к печати 2. 06. 2010 г.
Формат бумаги 70 108 1/16. Печ. л. 29,2. Уч.-изд. л. 32,5
Тираж 370. Заказ 201

Отпечатано в типографии Издательства СО РАН
630090, Новосибирск, Морской пр., 2