Theoretical Approaches to BioInformation Systems

TABIS 2013

Belgrade, Serbia, 17-22 September 2013.

Book of Abstracts

Editor: Branko Dragovich

Institute of Physics - Belgrade, Serbia, 2013

The Conference “Theoretical Approaches to BioInformation Systems - TABIS 2013”

is organized by the Institute of Physics - Belgrade in collaboration with
the Faculty of Mathematics and the Faculty of Biology, University of Belgrade, Serbia

under the auspices of

the Ministry of Education, Science and Technological Development of the Republic of Serbia

International Advisory Committee

Pavle Andjus (Belgrade, Serbia)
Milos Beljanski (Belgrade, Serbia)
Radu Constantinescu (Craiova, Romania)
Janos Kertezs (Budapest, Hugary)
Andrei Khrennikov (Vaxjo, Sweden)
Sergei Kozyrev (Moscow, Russia)
Fionn Murtagh (London, UK)
Zoran Obradovic (Temple, USA)
Sergey Petoukhov (Moscow, Russia)
Natasa Przulj (London, UK)
Zoran Rakic (Belgrade, Serbia)
Miloje Rakocevic (Belgrade, Serbia)
Miljko Sataric (Belgrade, Serbia)
Brunello Tirozzi (Roma, Italy)
Peter Tompa (Brussels, Belgium)
Alessandro Treves (Trieste, Roma)
Edward Trifonov (Haifa, Israel)

Scientific Committee

Marko Djordjevic (Belgrade, Serbia)
Branko Dragovich (Belgrade, Serbia)
Michele Caselle (Torino, Italy)
Radmila Panajotovic (Belgrade, Serbia)
Gordana Pavlovic-Lazetic (Belgrade, Serbia)
Paul Sorba (Annecy, France)
Nenad Svrakic (Belgrade, Serbia)
Bosiljka Tadic (Ljubljana, Slovenija)
Igor Volovich (Moscow, Russia)

Local Organizing Committee

Branko Dragovich (Chairman, Institute of Physics, Belgrade)
Sanja Ćirković (Institute of Physics, Belgrade)
Mihajlo Mudrinić (Vinča Institute of Nuclear Sciences, Belgrade)
Radmila Panajotović (Institute of Physics, Belgrade)
Dragan Savić (Institute of Physics, Belgrade)
Nenad Švrakić (Institute of Physics, Belgrade)
Dejan Timotijević (Institute of Physics, Belgrade)
Radomir Žikić (Institute of Physics, Belgrade)

Foreword

Despite many remarkable successes in biology, living systems are not well understood compared to ordinary physical systems. Living matter is very complex and its study needs a multidisciplinary approach including physics, mathematics, chemistry, computer science, and some other related branches of science, in addition to biology.

The 2nd International Conference "Theoretical Approaches to BioInformation Systems" (TABIS2013) takes place in Belgrade, Serbia, 17 - 22 September 2013. The first TABIS conference was held in Belgrade in 2010, with participation of scientists from different fields of theoretical research – computer science, physics, mathematics, biology and chemistry. The objective of TABIS2013 is to bring together scientists from Europe and beyond and provide a stimulating and pleasant environment for exchange of ideas and discussion of results in the genomics, proteomics, cognitive neuroscience and related networks.

The conference program contains 40, 30 and 20 minutes talks, and poster presentations.

Conference Topics include:

- Structure and function of DNA and RNA
- Structure, function and interaction of proteins
- Gene expression and genetic code
- Life as information processing
- Bioinformatics
- Neurons structure and signal porocessing
- Data mining and machine learning
- Cognitive modeling
- Networks: from biomolecules to global systems
- Related topics in medicine, drug design, psychology, …

Table of Contents

A new paradigm of protein structural organization

Nekrasov A.N.[1], Anashkina A.A.[2], Zinchenko A.A.[1]

[1]*Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia*
[2]*Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia*

anastasya.anashkina@gmail.com

In this research the entropy characteristics of natural polypeptide sequences are studied [1]. Based on these results suggested a new paradigm of the structural organization of proteins. Elementary unit in sequence of protein according to this new paradigm is the group of five neighboring residues. Such a group of five neighboring residues has been called the "information unit". The method of analysis of the information structure (ANIS method) [2] was developed based on this paradigm. ANIS method was enabling to disclose the hierarchical organization of structural information (Fig. 1) contained in amino acid sequence of proteins.

Applications of ANIS method in protein engineering [3, 4], studies of structural organization of enzymes [5, 6] and protein-protein complexes [7] are described.

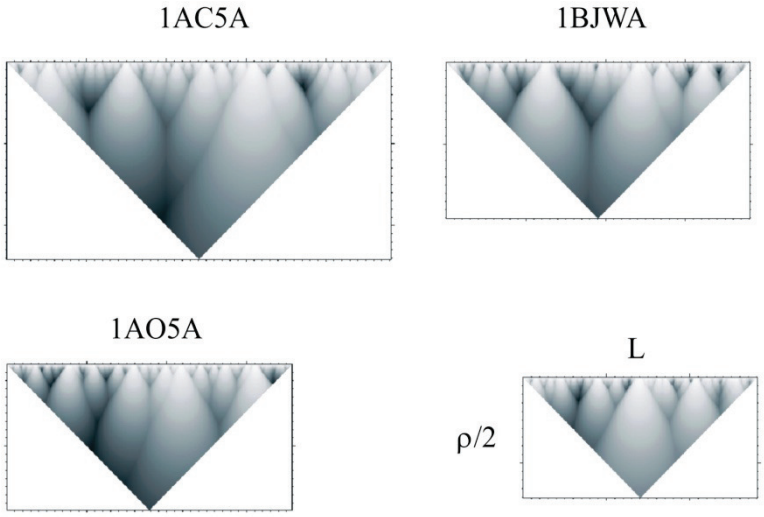


Figure 1 Examples of informational structures of several proteins (1AC5A.PDB, 1BJWA.PDB, 1AO5A.PDB). L is the number of a residue in primary structure of studied proteins. $\rho / 2$ is a value of function half-width, for which a range of interaction of information units has been estimated. Dark areas respect to the regions with high level of correlation of information units.

[1] A.N. Nekrasov, J. Biomol. Struct. Dyn. 20, 87 (2002).
[2] A.N. Nekrasov, J. Biomol. Struct. Dyn. 21, 615 (2004).
[3] A.N. Nekrasov, V.V. Radchenko, T.M. Shuvaeva, V.I. Novoselov, E.E. Fesenko, V.M. Lipkin, J. Biomol. Struct. Dyn. 24, 455 (2007).
[4] A.N. Nekrasov, L.E. Petrovskaya, V.A. Toporova, E.A. Kriukova, A.V. Rodina, E.U. Moskaleva, M.P. Kirpichnikov, Biochemistry (Moscow), 74, 399 (2009).
[5] A.N. Nekrasov, A.A. Zinchenko, J. Biomol. Struct. Dyn. 25, 553 (2008).
[6] T.V. Ostroverkhova, R.V. Chertkova, A.N. Nekrasov, D.A. Dolgikh, M.P. Kirpichnikov, Moscow University Biological Sciences Bulletin, 66, 65 (2011).
[7] A.N. Nekrasov, A.A. Zinchenko, J. Biomol. Struct. Dyn. 28, 85 (2010).

Calcium signalling in health and disease

P.R. Andjus

Faculty of Biology University of Belgrade

pandjus@bio.bg.ac.rs

Calcium can be considered as a physiologically essential metal ion. In the organism as a whole as well as in the brain that is our focus of research numerous intracellular processes are dependent on Ca^{2+} influx from the extracellular space. In the cell Ca^{2+} has an important role as the secondary messenger of various biochemical processes, and in neurons particularly it is the mediator of cellular excitability and synaptic activity. Its low concentration (10^{-7}M) and at the same time its high gradient on the cell membrane is maintained by a complex system of control of the intracellular calcium stores. These Ca^{2+} "depos" can be divided into a) intracellular compartments (sarco-endoplasmic reticulum and mitochondria) and b) cytosolic protein buffers. A particular role in the control of Ca^{2+} homeostasis is played by transporting proteins in the cellular and subcellular membranes. This complex homeostasis can be modeled by different systems of compartmental equations. A Ca^{2+} misbalance i.e. a prolonged rise in its intracellular concentration may cause metabolic changes ultimately leading to cell death. In neurodegenerative phenomena such as the motor neuronal disease – amyotrophic lateral sclerosis (ALS) this process is manifested as excitotoxicity that leads to overactive motor synapses.

Recent years have seen an extensive development of calcium-sensitive probes, transgenic or incubated molecules, that emit fluorescence upon complexing this ion. We will present two examples of applying the probe Fluo 3 or 4 in a confocal microscopy study of the process of pathogenesis in ALS: 1) effect of the egsogenous mutant superoxide dismutase (endogenously present in ALS patients) on the transmembrane flow of Ca^{2+} and its intracellular oscilations; and 2) acute effects of immunoglobulins G (IgGs) from ALS pacients on the Ca^{2+} homeostasis on neural and glial cells isolated and cultured from the brain of the disease model, the mSOD1 G93A rat [1].

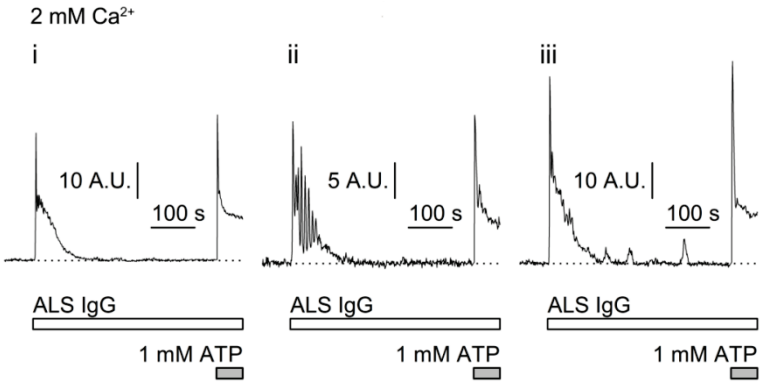


Figure 1 Types of acute effect of ALS IgGs on astrocytes in culture all followed by standard response to ATP (time course indicated by open and closed horizontal bars respectively). i. Single transient elevation. ii. Bursting effect. iii. Superimposed calcium transients. A.U. – arbitrary units of fluorescence signal

Acknowledgment
Presented work is partly supported by MEST RS grant # III41005.

[1] M. Milošević, M. Stenovec, M. Kreft, V. Petrušić, Z. Stević, S. Trkov, P.R. Andjus, R.Zorec, Cell Calcium. 54, 17 (2013).

The crystallographic and quantum-chemical analysis of S···S interactions between cysteine residue

I.S. Antonijević[1], G.V Janjić[2], M. Stojiljković[1] and S.D. Zarić[1,3]

[1] *Department of Chemistry, University of Belgrade, Studentski trg 12-16, Belgrade, Serbia*
[2] *ICTM, University of Belgrade, Njegoševa 12, Belgrade, Serbia*
[3] *Department of Chemistry, Texas A&M University at Qatar, P. O. Box 23874, Doha, Qatar*

szaric@chem.bg.ac.rs

The thiol side chain in cysteine (-CH_2SH) has an important structural role in many proteins and often participates in enzymatic reactions, serving as a nucleophile. However, until now, non-covalent sulfur-sulfur interactions between cysteine residues have not been studied. A suitable method for examining the nature and strength of S···S contacts involves computation of their electronic structures with *ab initio* methods [1].

In this work, the S···S interactions between cysteine residues were studied by analyzing data in crystal structures archived in the Cambridge Structural Database (CSD) and by quantum chemical calculations. In order to find intermolecular S···S interactions between cysteine residues, we performed a CSD search using the model system shown in Figure. We extracted all structures with the distance between two sulfur atoms (d or S_1···S_2 distance) less than 5.0 Å. In this way we found 62 structures, and 92 S···S contacts. Statistical analysis of several geometrical parameters was performed and showed that most of contacts have distance (d) with a maximum of distribution in the region of 3.8 to 4.2 Å. The cysteine residues show preferred values of angle between the mean planes of thiol groups (C-S-H planes) from 0° to 10° (parallel orientation) and orientations with H_1-S_1···S_2-H_2 torsion angle close to 180°.

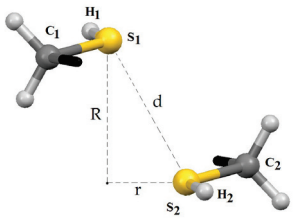


Figure 1 The geometric parameters used for the description of S···S interactions

To examine whether S···S contacts between two -CH_2SH fragments are attractive, and not just the consequence of packing in the crystal structures, we calculated the S···S interaction energies between two methantiole molecules, using the MP2 method and the cc-pVQZ basis set. The calculations were done on four orientations: parallel, planar, normal and for geometry corresponding to the electrostatic model. The results of *ab initio* calculations showed that the interaction are strongest in the case of parallel orientation (-1.67 kcal/mol), while the energies of interactions with normal, planar and orientation corresponding to the electrostatic model are weaker: -0.96, -0.43, -0.14 kcal/mol, respectively. The results of ab initio calculations are in good agreement with results of the CSD search. Namely, the calculated d distances of S···S interactions for all examined model systems are between 3.9 and 4.2 Å.

The study of non-covalent interactions involving sulfur-sulfur contacts may be of great importance for identifying the S···S interactions in biological systems.

[1] R. Silaghi-Dumitrescu, A. Lupan, Cent. Eur. J. Chem. 11, 457 (2013).

On semi-supervised methods to predict patient survival from gene-expression data

S. Bojić

Department for Biostatistics, Medical informatics, and researches in Medicine, Faculty of Medicine, University of Belgrade, Serbia

ceca.bojic@gmail.com

Survival prediction from high-dimensional genomic data is an active field in today's medical research. Most of the proposed prediction methods make use of genomic data alone without considering established clinical covariates that often are available and known to have predictive value. Most of them suffer from the information loss inherent to the defining the outcome as binary ("good" versus "bad").

There is a lack of consensus in the field on which prediction method is the most accurate, as well as which variable-selection method should be used in the dimension reduction step. This ongoing debate is complicated even more with substantial variation regarding the metric used in evaluating the performance of the prediction methods. There is also the limited number of publicly available datasets on which most of the work is done, which evidently biases results.

In the following presentation we'll try to tickle all these issues, as well as to present our own thoughts, problems and results while pursuing the best prediction model.

[1] H.M.Bovelstad, O. Borgan. Assessment of evaluation criteria for survival prediction from genomic data. Biometrical Journal. 53 (2011) 2, 202-216.
[2] E.Bair, R. Tibshirani. Semi-Supervised Methods to Predict Patient Survival from Gene Expression Data. PLoS Biol 2(4):e108. Doi:10.1371/journal.pbio.0020108.
[3] H. Li, Y. Luan. Kernel Cox Regression Models for Linking Gene Expression Profiles to Censored Survival Data. Pacific Symposium on Biocomputing 8:65-76 (2003).

Cognition as a dynamical system: mathematical modeling

E. Cadez[1], M.J Spivey[1] and V.M. Cadez[2]

[1]*Cognitive and Information Sciences, University of California Merced, Merced, CA 95343, USA.*
[2] *Astronomical Observatory, 11160 Belgrade, Serbia.*

vcadez@aob.bg.ac.rs

Dynamical systems are those that change over time and they are widely present in nature. Although these systems appear in a broad variety of fields they may often be described by similar differential equations which introduces elements of interdisciplinarity in models of otherwise entirely unrelated processes. In this contribution, a typical radiative transfer equation for a physical process of radiation propagating through an interactive non-uniform medium is applied in the domain of cognitive science to model a dynamical process of learning. By solving the corresponding evolution equation for various types of free coefficients and initial conditions we obtained models of time evolution of individual memory content subject to given initial state. Parameters that we freely chose at this point are related to description of selected memory decay in time, nonlocal memory effects arising from integral contribution of past events, additional activations in time during the considered process, mutual interactions of events and forming false memories, memory recovery and possible resonances if interactions between individuals are included [1]-[4].

The aim of this work is to provide an insight into possible solutions for the dynamic system of learning process and shed light of significance of freely chosen parameters involved. Such a theoretical understanding can eventually point out the directions of future experimental work in the field of cognition including perception-action, memory, attention, decision making, learning, problem solving, and language [5].

[1] E. Cadez, V. M. Cadez and E. Heit, 43. Ann. Meeting of the Society for Math. Psychology, Portland (2010).
[2] E. Cadez and E. Heit, the 33. Ann. Conf. of the Cognitive Science Society. Boston, MA (2011).
[3] M. A. Riley and J. G. Holden, WIREs Cogn Sci 2012. doi: 10.1002/wcs.1200 (2012).
[4] E. Cadez, E. and E. Heit,. Soc. for Math. Psychology, 46. Ann. Meeting; Potsdam (2013).
[5] E. Cadez, Submitted for publication in Journal of Integrative Neuroscience.

The role and functions of microRNA-mediated circuits in the human regulatory network

M. Caselle

Institute of Physics, University of Turin, Turin, Italy

caselle@to.infn.it

MicroRNAs are endogenous non-coding RNAs which negatively regulate the expression of protein-coding genes in plants and animals. They are known to play an important role in several biological processes and, together with transcription factors, form a complex and highly inter-connected regulatory network. Looking at the structure of this network, it is possible to recognize a few overrepresented motifs which are expected to perform important elementary regulatory functions. Among them, a special role is played by miRNA-mediated feedforward loops in which a master transcription factor regulates a microRNA and, together with it, a set of target genes. We show analytically and through simulations that the incoherent version of this motif can couple the fine-tuning of a target protein level with an efficient noise control, thus conferring precision and stability to the overall gene expression program, especially in the presence of fluctuations in upstream regulators. Among the other results, a nontrivial prediction of our model is that the optimal attenuation of fluctuations coincides with a modest repression of the target expression. This feature is coherent with the expected fine-tuning function and in agreement with experimental observations of the actual effect of a wide class of microRNAs on the protein output of their targets. We also discuss the impact on fine tuning and noise-buffering efficiency of the cross-talk between microRNA targets (the so called "sponge effect") that naturally arises if the microRNA-mediated circuit is not considered as isolated, but embedded in a larger network of regulations. Finally as an example of our results we discuss in detail the miRNA mediated FFLs involving Myc as master Transcription factor.

[1] Osella M, Bosia C, Corá D, Caselle M., The role of incoherent microRNA-mediated feedforward loops in noise buffering, PLoS Comput Biol. Mar;7(3):e1001101 (2011).
[2] El Baroudi M, Corà D, Bosia C, Osella M, Caselle M., A curated database of miRNA mediated feed-forward loops involving MYC as master regulator, PLoS One., Mar 3;6(3):e14742 (2011).
[3] Bosia C, Osella M, El Baroudi M, Corà D, Caselle M, Gene autoregulation via intronic micrornas and its functions, BMC Systems Biology, 6: 131 (2012).

Modeling neural flow through linearization procedures

Radu Constatinescu

University of Craiova, Dept. of Physics, Craiova, Romania

rconsta@yahoo.com

The paper aims to present the structure of some electronic circuits allowing the modeling of the neural flow. The main idea of all the neural models consists in finding electronic circuits which generate signals similar with those registered during the physiologic investigations of the brain. The circuit laws are expressed as nonlinear differential equations that can be solved using various mathematical methods. We shall consider the linearization procedure of the nonlinear Van der Pol oscillators as it was suggested by ChuaÂ's type circuits, ones of the first and largely studied circuits successfully depicting nonlinear dynamics similar to the neural flow. These circuits contain special diodes with negative rezistivity and generate a wide variety of chaotic phenomena. We shall study the generalized form of the equations associated with ChuaÂ's circuits and we shall compare their symmetry properties with the ones of the generalized Lorentz systems. Explicit solutions which fit with the experimental data will be obtained through the similarity reduction technique.

Modeling of bacterial immune systems: processing of CRISPR transcripts

Marko Djordjevic[1], Magdalena Djordjevic[2] and Konstantin Severinov[3,4]

[1]*Faculty of Biology,University of Belgrade, Belgrade, Serbia*
[2]*Institute of Physics, University of Belgrade, Belgrade, Serbia*
[3]*Institutes of Molecular Genetics and Gene Biology, Russian Academy of Sciences, Moscow, Russia*
[4]*Waksman Institute, Rutgers, State University of New Jersey, Piscataway, USA*

magda@ipb.ac.rs

CRISPR/Cas (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated sequences) is a recently discovered prokaryotic defense system against foreign DNA, including plasmids and viruses [1]. CRISPR cassette is transcribed as a continuous transcript (pre-crRNA), which is processed by Cas proteins into small RNA molecules (crRNAs) that are responsible for defense against invading viruses [2]. While CRISPR/Cas is a subject of intensive research, it remains poorly understood how this system is regulated under different conditions.

We concentrate on the mechanism of pre-crRNA processing in *E. coli*, where recent experiments show that overexpression of cas genes generates a large number of crRNAs, from only few pre-crRNAs [3]. Based on these observations, we develop a model of CRISPR processing and show that the system acts as a strong linear amplifier upon cas overexpression [4]. Interestingly, this strong amplification crucially depends on fast non-specific degradation of pre-crRNA by an unidentified nuclease, which suggests that this nuclease is a major control element of CRISPR/Cas response.

We furthermore investigated regulation of CRISPR/Cas system at the level of transcription regulation. We show that overexpression of cas genes above a certain level does not result in a further increase of crRNA, but that this saturation can be relieved if the rate of CRISPR transcription is increased. Moreover, a small increase of CRISPR transcription rate can substantially decrease the extent of cas gene activation necessary to achieve a desired amount of crRNA. We therefore suggest that activation of CRISPR array transcription may be another important control mechanism, as it can either increase the amount of generated crRNAs or substantially decrease the extent of *casE* gene activation necessary to achieve desired crRNA levels.

Inspired by our study of CRISPR/Cas system in *E. coli*, we next investigate design of a synthetic gene circuit that can generate a large amount of useful product from small amounts of potentially toxic substrate [5]. We optimize the circuit, with the goal of generating maximal product amounts, without increase of substrate amounts upon system induction. Our results show that such optimal system can rapidly increase the product amount for up to three orders of magnitude, while keeping the substrate amount at constant level.

Acknowledgment
This work is partially supported by Marie Curie International Reintegration Grants within the 7th European community Framework Programme (PIRG08-GA-2010-276996 and PIRG08-GA-2010-276913) and by the Ministry of Education and Science of the Republic of Serbia under project number ON173052.

[1] Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. Biology Direct 1:7 (2006).
[2] Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. Science 315, 1709 (2007).
[3] Pougach K, Semenova E, Bogdanova E, Datsenko KA, Djordjevic M, Wanner BL, Severinov K. Mol. Microbiol. 77, 1367 (2010).
[4] Djordjevic M, Djrodjevic M, Severinov K. Biology Direct, 7:24 (2012).
[5] Djordjevic M, Djordjevic M, Phys Biol. 9(5):056004 (2012).

Transcription start site identification in bacteria: a modeling approach

M. Djordjevic

Faculty of Biology,University of Belgrade, Belgrade, Serbia

dmarko@bio.bg.ac.rs

Genome-wide predictions of bacterial promoters is a classical bioinformatics problem, where available methods show poor accuracy. Similarly, experimental methods for genome-wide promoter detection (ChIP-chip experiments and its derivatives) display a large number of false positives. To understand the origin of these false positives, we calculated the kinetic parameters of transcription initiation for all segments in E. coli intergenic regions [1]. We found that RNA polymerase (RNAP) DNA-binding domains are designed so as to reduce the number of poised promoters. However, despite this reduction, there is still a significant number of poised promoters in *E. coli* intergenic regions (~30% of the sequences that are strongly bound by RNAP), which in part explains a large number of false positives in the promoter searches. Furthermore, we surprisingly found that sequences of the experimentally confirmed promoters increase the extent of RNAP poising. To explain this finding, we proposed an extension of the mix-and-match model for promoter recognition to kinetic parameters. We are currently developing a novel method for bacterial promoter prediction, which is based on explicit calculation of the kinetic parameters for transcription initiation [1] and on a more accurate description of promoter sequence specificity [2]; preliminary results indicate that this method significantly (for at least 50%) reduces the number of false positives.

Acknowledgment
This work is supported by a Marie Curie International Reintegration Grant within the 7th European community Framework Programme (PIRG08-GA-2010-276996) and by the Ministry of Education and Science of the Republic of Serbia under project number ON173052.

[1] Djordjevic M, Integrative Biology, 5(5):796-806 (2013).
[2] Djordjevic M, Journal of Bacteriology, 193(22): 6305-14 (2011).

Ultrametricity in bioinformation sytems

Branko Dragovich

Institute of Physics, University of Belgrade,Pregrevica 118, 11080 Belgrade,Serbia

dragovich@ipb.ac.rs

We consider p-adic ultrametric structure of the genetic code and its possible extensions to some bioinformation systems. The genetic code is a connection between 64 codons, which are building blocks of the genes, and 20 amino acids, which are building blocks of the proteins. In addition to coding amino acids, a few codons code stop signal, which is at the end of genes and terminates process of protein synthesis. Codons are ordered triples composed of C, A, U (T) and G nucleotides. Each codon presents an information which controls use of one of the 20 standard amino acids or stop signal in synthesis of proteins. We have shown that the space of 64 codons of the vertebrate mitochondria clasterize with respect to 5-adic distance into 16 quadruplets, which each of these quadruplets separates into two doublets under 2-adic distance. Then each of 32 codon doublets is related to one of 20 amino acids or to stop signal. Hence, it follows that p-adic distance describes nearness (similarity) of codons in their bioinformation sense.

Taking into account p-adic distance between constituents of two strings, we introduce modified Hamming distance and consider its possible application in investigation of similarity in bioinformation systems. Some other aspects of ultrametricity and p-adic distance in bioinformation systems will be also discussed.

Acknowledgment
This work is partially supported by the Ministry of Education, Science and Technological Development, Serbia, contracts 173052 and 174012.

[1] B. Dragovich and A. Dragovich, "A p-adic model of DNA sequence and genetic code", p-Adic Numbers, Ultrametric Anal. Appl. 1, No. 1, 34--41 (2009); [arXiv:q-bio.GN/0607018v1].
[2] B. Dragovich and A. Dragovich, "p-adic modelling of the genome and the genetic code", Computer Journal 53, No. 4, 432--442. (2010); [arXiv:0707.3043v1 [q-bio.OT]].
[3] B. Dragovich, " p-Adic structure of the genetic code", NeuroQuantology 9, No. 4, 716--727 (2011); [arXiv:1202.2353v1 [q-bio.OT]].

Can we use standard tools to predict functional effects of point variations outside conserved domains? TET2 example

B. Gemovic, V. Perovic, S. Glisic and N. Veljkovic

Centre for Multidisciplinary Research and Engineering, Vinca Institute of Nuclear Sciences, University of Belgrade, Belgrade, Serbia

gemovic@vinca.rs

Half of mutations in the cancer-associated proteins are located outside conserved domains (CDs), making the prediction of their functional effects a significant issue. TET2 is an epigenetic regulator, with two well defined CDs. Mutations in this gene are important molecular markers of myeloid malignancies.

Dataset contained 45 neutral single nucleotide polymorphisms (SNPs) and 121 somatic TET2 mutations. Subset of 42 SNPs and 27 mutations were outside CDs. Functional effects were assessed by: PolyPhen-2, SIFT, MutPred and PhD-SNP. First three tools, with available probability scores, were estimated through area under the ROC curve (AUC). Binary classification of all tools was tested using cross-tabulation and Fisher's Exact test.

Analyses of subset of variations outside CDs showed AUC values of 0.55-0.68. These were 20-31% lower compared with the whole dataset. Accuracies of binary classifications were 0.54, 0.55, 0.62 and 0.61 for PolyPhen-2, SIFT, MutPred and PhD-SNP, respectively. Contrary to the whole dataset, Fisher's test of the subset didn't show significant classification.

This study shows best efficiency of MutPred and PhD-SNP, but they have extremely low rate of true positives (4%). Overall results suggest low efficacy of commonly used tools for predicting functional effects of variations outside CDs, emphasizing the need for new algorithms.

Acknowledgement
This research has been founded by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

[1] I. A. Adzhubei, S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork et al. Nat. Methods. 7, 248 (2010).
[2] E. Capriotti, R. Calabrese, R. Casadio, Bioinformatics. 22, 2729 (2006).
[3] F. Delhommeau, S. Dupont, V. D. Valle, C. James, S. Trannoy, A. Massé et al. N. Engl. J. Med. 360, 2289 (2009).
[4] B. Li, V. G. Krishnan, M. E. Mort, F. Xin, K. K. Kamati, D. N. Cooper et al., Bioinformatics. 25, 2744 (2009).
[5] P. C. Ng, S. Henikoff, Nucleic Acids Res. 31, 3812 (2003).
[6] P. Yue, W.F. Forrest, J.S. Kaminker, S. Lohr, Z. Zhang, G. Cavet, Hum. Mut. 31, 264 (2010).

Bioinformatics analysis of gene-expression strategies of bacterial viruses

Jelena Guzina and Marko Djordjevic

Faculty of Biology,University of Belgrade, Belgrade, Serbia

jeca.282@gmail.com

Bacteriophages are viruses that can specifically destroy a given pathogenic bacteria. Understanding their infection strategies has recently came in focus, in the context of increased resistance of pathogenic bacteria to antibiotics. A most common way of analyzing bacteriophage infection strategies is a labor-intensive combination of biochemical and bioinformatic approaches and macroarray measurements. We here investigate to what extent we can understand gene expression strategies of lytic phages, by directly analyzing their genomes through bioinformatic methods. We address this question on a recently sequenced lytic bacteriophage 7 – 11 that infects bacterium *Salmonella enterica* [1]. This bacteriophage is homologous to another lytic bacteriophage phiEco32, which was experimentally characterized in detail [2]; while we do not use this homology in our analysis, we use it to verify the obtained results.

Our main result is identification of novel promoters for the bacteriophage-encoded sigma factor. Interestingly, standard methods for promoter recognition, which are based on Monte-Carlo procedures, fail to correctly identify the promoters, but a simpler procedure that is based on pairwise alignments of the intergenic regions correctly identifies the desired motifs; we argue that such search strategy is more effective for promoters of bacteriophage-encoded sigma factors that are typically well conserved but appear in low copy numbers. Furthermore, we also identified promoters for bacterial-encoded sigma factor in the bacteriophage genome, by using a recently improved model of specificity of bacterial promoters [3]. Identification of all the promoter elements allows clustering the bacteriophage genes in putative early, middle and late class. All the obtained information is self-consistent, i.e. clustering the genes in the temporal classes is consistent with the gene function and their transcription orientation, while promoter specificity is consistent with that of the homologous phage phiEco32. We therefore find that direct analysis of bacteriophage genome sequences is a plausible first-line approach for understanding bacteriophage transcription strategies.

Acknowledgment
This work is partially supported by Marie Curie International Reintegration Grant within the 7th European community Framework Programme (PIRG08-GA-2010-276996) and by the Ministry of Education and Science of the Republic of Serbia under project number ON173052.

[1] J. Guzina, M. Djordjevic, to be submitted (2013).
[2] O. Pavlova et al., J Mol Biol. 416, 389 (2012).
[3] M. Djordjevic, J Bacteriol. 193, 6305 (2011).

Functional dissection of a disordered plant chaperone ERD14

Zoltan Magos[1], Bianka Szalainé Ágoston[1,2], Anna Felmery[1], Beata Szabo[1], Agnes Tantos[1], Nikoletta Murvai[1], Andras Lang[2], Laszlo Buday[1], Kyou Hoon Han[3,4], Andras Perczel[2], Lajos Kalmar[1], Peter Tompa[1,5]

[1]*Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary*
[2]*Institute of Organic Chemistry, ELTE University, Budapest, Hungary*
[3]*Department of Bioinformatics, University of Science and Technology, Daejeon, Korea*
[4]*Biomedical Translational Research Center, Division of Convergent Biomedical Research, KRIBB, Daejeon, Korea*
[5]*VIB Department of Structural Biology, Vrije Universiteit Brussel, Brussels, Belgium*

lajos.kalmar@ttk.mta.hu

Intrinsically disordered stress proteins of plants and animals, such as A. thaliana ERD14, are potent chaperones in vitro, protecting client proteins against loss of activity and aggregation under denaturing conditions. Extending on prior in vitro observations, our multi-dimensional in-cell NMR of ERD14 in *E. coli* shows that the protein is largely disordered in vivo, whereas it's short conserved regions (Ka, Kb, Kc, ChP and S-segments) which transiently sample helical conformations in vitro, are involved in recognizing partner proteins in the cell. Overexpressed ERD14 raises the viability of cells from 38.9 % to 73.9 % following heat stress (50^{o}C x 15 min) due to generally increasing solubility of the proteome. Deleting Kc, ChP-segments or the hydrophobic region H decreases chaperone activity, whereas scrambling the entire protein fully impairs activity. The intrinsically disordered chaperone is stable in the cell under stress, its protective effect is not saturable, and it does not translocate to the cell membrane, which suggest that its primary effect is a physical protection of proteins localized in the cytoplasm. Our data connect structural disorder and protein function in live cells, leading to a testable molecular hypothesis about the molecular mechanism of this, and possibly other, intrinsically disordered chaperones.

On protein function prediction methods

J.Kovacevic[1] , P. Radivojac[2] and G. Pavlovic-Lazetic[1]

[1]*Faculty of Mathematics, University of Belgrade, Belgrade, Serbia*
[2]*School of Informatics and Computing, Indiana University, Bloomington IN, USA*

jovana@matf.bg.ac.rs

Knowing the function of a protein informs us on its role in the organism. With large number of genomes being sequenced every year, there is a growing number of newly discovered proteins. Fast and accurate information on protein function is especially important in context of human diseases, since many of them can occur due to functional mutation. Current experimental methods for functional annotation of proteins are too slow for such quick income of new data which is why automated prediction of protein function is widely used. The most common form of functional classification is the Gene Ontology (GO), which provides three hierarchical classifications as directed acyclic graphs: molecular function ontology, biological process ontology and cellular component ontology [1]. With this respect, protein function prediction can be assessed as a structural classification problem. Here, we will give an overview on our current research in the field, namely one method for predicting protein function from its amino acid sequence. The method is based on the generalization of support vector machines for complex outputs [2] especially modeled for Gene Ontology hierarchy.

Acknowledgment
The work presented has been supported by the Ministry of Education and Science, Republic of Serbia, Projects No. 174021.

[1] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Nat. Genet. 25(1), 25-9. (2000 May).
[2] I. Tsochantaridis, T. Joachims, T. Hofmann, Y. Altun, JMLR 6, 1453-84 (2005).

Protein subunit association: a social network

Mounia Achoch[1], Giovanni Feverati[2], Laurent Vuillon[3], Kave Salamatian[1] and Claire Lesieur[4]

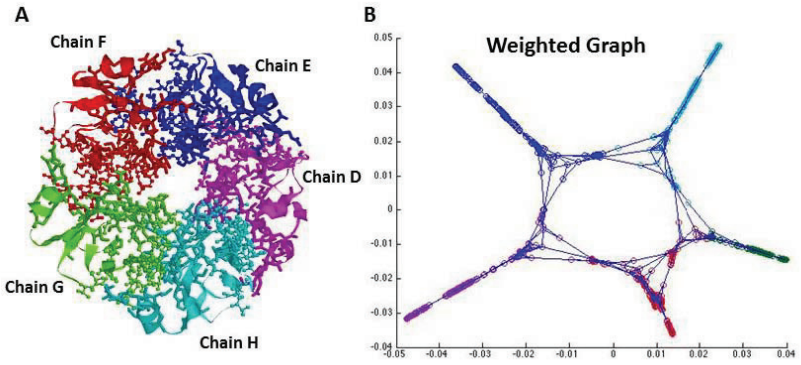
[1]*LISTIC, Université de Savoie, 74941 Annecy le Vieux Cedex, France*
[2]*LAPTH UMR 5108, Université de Savoie, CNRS, BP 110, 74941 Annecy le Vieux Cedex, France;*
[3]*LAMA, Université de Savoie, CNRS UMR 5127, 73376 Le Bourget du Lac, France;*
[4]*Université Joseph Fourier, CNRS FRE 3405, AGIM, Grenoble, France*

claire.lesieur@agim.eu

Most proteins cannot function as single unit but associate subunits via formation of protein interfaces, to be biologically active. Deciphering the amino acids involved in subunit association, the so-called hot spots, and their role in the interface formation is important particularly because subunit association is related to pathologies (e.g. Alzheimer).

We have explored the applicability of the concept of networks to the problem of subunit association (1). To do so a new algorithm built on spectral graph theory, called SpectralPro, has been developed and successfully tested on the cholera toxin B pentamer (Fig. 1A). The use of such methodologies has extended beyond cellular networks (interactomes) over the recent years showing relevancy on protein dynamics and protein functions (e.g. enzyme functional center or binding specificity) (2,3).

SpectralPro describes protein interface as a hot spot interaction network or a graph (Fig. 1B). The nods are the hot spots and the edges are the hot spot interactions. SpectralPro reasonably detects the hot spots compared to other programs. It clusters hot spots according to connectivity, centrality and distance interactions. The biological significance of such network parameters is tested by looking at the effect of the hot spots virtual mutation on the interface stability.



Spectral Pro not only detects hot spots but splits them into categories so hypotheses on their role in the mechanism of interface formation can be put forward.

Figure 1 Cholera toxin B subunit pentamer (CtxB5). A. x-ray structure of CtxB5. Each monomer has a different color. Atoms of the toxin interfaces are in balls and sticks, rest is in ribbons. B. Graph of the toxin interfaces generated by SpectralPro.

Acknowledgment
We thank the federation of research MSIF (Modelization, Simulations, Interaction Fundamental) for supporting our work (http://lapth.in2p3.fr/msif/en/).

[1] Feverati, G., Achoch, M., Zrimi, J., Vuillon, L., and Lesieur, C. (2012) Beta-Strand Interfaces of Non-Dimeric Protein Oligomers Are Characterized by Scattered Charged Residue Patterns. PloS one 7, e32558.
[2] Brinda, K. V., Mitra, N., Surolia, A., and Vishveshwara, S. (2004) Determinants of quaternary association in legume lectins. Protein Sci 13, 1735-1749.
[3] Csermely, P. (2008) Creative elements: network-based predictions of active centres in proteins and cellular and social networks. Trends Biochem Sci 33, 569-576.

Interactions of water with protein aromatic residues – *ab initio* study on model compounds

D.P. Malenov[1], G.V. Janjić[2], D.Ž. Veljković[1], S.D. Zarić[1,3]

[1]*Department of Chemistry, University of Belgrade, Belgrade, Serbia*
[2]*IHTM,University of Belgrade, Belgrade, Serbia*
[3]*Department of Chemistry, Texas A&M University at Qatar, Doha, Qatar*

dusanmalenov@yahoo.com

Water molecules are ubiquitous in protein structures and play important role in protein folding, structure, activity and protein-ligand interactions [1]. Interactions of aromatic residues in proteins with water are of great importance for stabilizing local protein structure [1] and can significantly contribute to hydration energy [2].

Protein hydrophobic pockets can contain limited number of water molecules that interact with aromatic rings. In those pockets, water can form four types of interactions with aromatic rings: OH/π, CH/O, parallel alignment (P) and lone pair/π (LP/π) [3, 4]. Protein Data Bank (PDB) search of interactions between water and aromatic residues showed that phenylalanine residue interacts with two water molecules, tryptophan with 2.5 and tyrosine residue with 3 water molecules on average [5]. Here we present our results on interactions of two water molecules with benzene as model compound.

Ab initio calculations on sixteen water/benzene/water model systems that contain two water/benzene interactions were performed using Møller–Plesset perturbation theory of the second order and cc-pVTZ and cc-pVQZ basis sets. The model systems were investigated by calculating their trimer binding energies and by estimating their synergetic effects [6].

The results show that trimer binding energy is lowest in the system containing two OH/π interactions (-6.04 kcal/mol), which are the strongest among all water/benzene interactions (ΔE = -3.22 kcal/mol). However, synergetic effect is negative in this system, since these two interactions weaken each other by 0.40 kcal/mol. The weakening was also found in the system containing two CH/O interactions, which weaken each other by 0.31 kcal/mol. On the opposite, systems containing one CH/O and one OH/π interaction possess substantial trimer binding energy (-5.17 kcal/mol), but also significant positive synergy, strengthening each other by 0.44 kcal/mol.

Repulsive lone pair/π interactions (ΔE = +0.55 kcal/mol) were shown to become significantly (0.46 kcal/mol) less repulsive when OH/π interaction is also present. However, in model systems with CH/O or another LP/π interaction, LP/π interaction becomes even more repulsive, up to +0.98 kcal/mol. Parallel alignment interactions do not have significant influence energies on other interactions or onto themselves. The obtained synergetic effects are in good agreement with calculated cooperativity energies and they are related to the direction and amount of charge transfer, as shown by Natural Population Analysis [6].

Acknowledgement
This work was supported by the Ministry of Education, Science and Technology of the Republic of Serbia (grant 172065).

[1] J. L. Atwood, F. Hamada, K. D. Robinson, G. W. Orr, R. L. Vincent, Nature 349, 683 (1991).
[2] T. Steiner, A. M. M. Schreurs, J. A. Kanters, J. Kroon, Acta Crystallogr. Sect. D 54, 25 (1998).
[3] G. V. Janjić, D. Ž. Veljković, S. D. Zarić, Cryst. Growth Des. 11, 2680 (2011).
[4] M. Egli, S. Sarkhel, Acc. Chem. Res. 40, 197 (2007).
[5] K. Håkansson, Int. J. Biol. Macromol. 18, 189 (1996).
[6] D. P. Malenov, G. V. Janjić, D. Ž. Veljković, S. D. Zarić, Comp. Theor. Chem. 1018, 59 (2013).

Self-organization as a processing of hidden causal information

M. Milovanović

Mathematical Institute of the Serbian Academy of Sciences and Arts, Belgrade, Serbia

milosm@mi.sanu.ac.rs

The term self-organization refers to processes in which structure or organized behavior emerges without the external influence. A paradigmatic approach to self-organization, presented in [1] and [2], begins with a dynamic random field on the spatio-temporal lattice. Subsequently, one constructs the random field of local causal states which form the Markov spatio-temporal random field and contain minimal information required for optimal prediction of the original field either locally or globally. Eventually, one defines self-organization as increase of local complexity – information contained in local causal state.

We find that this model shares common features with the wavelet-domain hidden Markov model in signal processing [3]. The starting point of this analogy is to consider observations as distorted measurements of another unseen set of state variables that have their own dynamics. Local causal states, conceived as hidden states of the model, however, reveal new type of causality, named *perfective cause*, with which local causality implies both prediction and retrodiction. This property of the model is called *temporal irrelevance*. This means that separation onto the future and the past becomes irrelevant for interpretation of causality, but the arrow of time is determined by increase of local complexity measuring system's acausality i.e. entropy of causal state variable.

In that way, information of the signal decomposes in two terms – causal structural information and added noise information [4]. The first one, which is stored in hidden causal structure, is conceived as the measure of system's self-organization given that it is associated with increase of local complexity in temporal domain. This term, the so-called global complexity, characterizes complex systems as those overcoming causality i.e. in which acausal properties come to the fore over time. Applicability of the method to the broad class of signals, including non-linear dynamic systems, plasma turbulence data etc., discloses life as a general physical phenomenon related to processing of causal information.

Acknowledgment
The author acknowledges support by the Serbian Ministry of Education and Science through the projects OI 174014 and III 44006.

[1] Shalizi C. R., Shalizi K. L. and Haslinger R., Phys.Rev. Lett., 93 (2004) 118701.
[2] Shalizi C. R., Discrete Math. Theor. Comput. Sci. Proc.,AB (2003) 11.
[3] Crouse M., Nowak R. and Baraniuk R., IEEE Trans. Signal Process., 46 (1998) 886.
[4] Milovanović M., Rajković M., Europhysics Lett., 102 (2013) 40004.

From the genetic code to fractional systems/control

N.Ž. Mišić

Lola Institute, Kneza Višeslava 70a, Belgrade, Serbia

nmisic@afrodita.rcub.bg.ac.rs

A link between *information* and *spacetime topology*, also considered as a revival of Aristotle's dynamic dualism of *potential* and *act*, is a very actual and prospective problem both in physics and biology. On a more phenomenological level, this link in biology is the focus of *biosemiotics*, where the biological coding/computing is considered as the basic mechanism of macroevolution, most efficiently based on a *nested principle* [1], emphasizing the importance of the first biocode/biocomputation – *genetic code*/*translation* [2], not only as the origin of life, but also as the link between the physical and biological world.

Shcherbak's revealing arithmetic inside the universal genetic code based on *decimal number* 037 as a packing nucleon quantum of genetic code constituents [3] has opened particularly interesting mathematical approach. We showed [4] that the basic property of 037 is a *cyclic equivariability* (both for the *multipliers* and *digits equidistance*), and that its generalized numbers, *Self-similar or Shcherbak Numbers* ($\mathcal{S}$), have a simple form $\mathcal{S} = \mathcal{S}_n(q) = \frac{R_n(q)}{n}$ for $n|q-1$, where $R_n(q) = \sum_{0\le i<n} q^i$ are the *generalized Niven repunits* of length n in the number system of radix q ($n, q \in \mathbb{N}_{>1}$). Since also $R_n(q) = \frac{q^n-1}{q-1} = \prod_{d|n,d>1} \Phi_d(q)$ where $\Phi_d(q)$ is dth cyclotomic polynomial, follows a relation of $\mathcal{S}$ to the nth roots of unity and thus to the cyclotomic lattices, indicating a uniform space filling and the *Bravais lattices*, especially for a square lattice ($n = 4$) and an equilateral triangular (hexagonal) lattice ($n = 3{,}6$) [4,6]. In the context of $\mathcal{S}$, the minimal number systems for their realization are $q = 5$ and $q = 4{,}7$, respectively. But if we observe the decimal system as the *biquinary* system ($q = 2 \cdot 5$) then $q = 10$ is a minimal system for realization of both symmetries (037 is the *closed-pack circular cluster* for both square and hexagonal lattice [5]). Consequently, it is sensible to analyze the symmetries of decimal system not only for the *triplets* ($n = 3m$), but also for the *doublets* ($n = 2m$) (the genetic code is arranged as a *nested code* – a *"coarse code" on the level of doublets* and a *"fine code" on the level of triplets* [4]). Then a maximal length of doublets and triplets for $n < q = 10$ is respectively 2^3 and 3^2, which gives the interrelations for the pairs of cyclic numbers, the pair $(73{,}137)$ as the factors of $10^{2^3} - 1$ and the pair $(27{,}37)$ as the factors of $10^{3^2} - 1$, based on the polynomial $x^2 \mp x - 10 = 0$ and its solutions $\psi = \pm 3{,}7015\ldots$ and $\psi' = \mp 2{,}7015\ldots$ The importance of this *ψ-polynomial is in its perception as the simplest modular extension of the Golden Mean or ϕ-polynomial* $x^2 \mp 1x - 1 = 0$, so that the *constant ψ is an analogue of selfsimilarity constant ϕ for the measurement base extension from number* 1 *to number* $10 \equiv 1 (\mathrm{mod}\ 9)$ [6].

Among other correspondences between ϕ, ψ and $\mathcal{S}$, as well as relations of $\mathcal{S}$ to the numerical representation of space (for instance, $\mathcal{S}_9(10) = 012345679 = 037 \cdot 000333667$ is the period of number $R_\infty^2(10)$ while also $8\mathcal{S}_9(10) = 098765432$ [4,6]), particular importance potentially belongs to the fact that for ψ both $\psi \approx \mathcal{S}_3(10)/10 = 037/10$ and $\psi^2 = \psi + 10 \approx \alpha/10$ are valid, where α is the *fine structure constant*, which generally indicates that all these constants with their numerical/arithmetical/geometrical properties *harmonize scaling and shifting trough selfsimilarity*, and as such they could be a part of gauge system (field) which results in the fractional systems/control.

Acknowledgment
This research has been partially funded by the Ministry of Science and Technological Development of the Republic of Serbia, through Projects TR-32040 and TR-35023.

[1] B. Barbieri, Ed., Introduction to Biosemiotics, Springer (2007); The Codes of Life, Springer (2008).
[2] C.R. Woese, Proc. Natl. Acad. Sci. USA 54, 1546 (1965); Proc. Natl. Acad. Sci. USA 99, 8742 (2002).
[3] V.I. Shcherbak, J. Theor. Biol. 166, 475 (1994); BioSystems 70, 187 (2003).
[4] N.Ž. Mišić, Proceedings of NEUREL, 97 (2010); Proceedings of DOGS, 132 (2010).
[5] N.J.A. Sloane, B.K. Teo, J Chem Phys 83, 6520 (1984).
[6] N.Ž. Mišić, Neuroquantology 9, 702, (2011); Ph.D. dissertation, University of Belgrade (2006).

Determining correlation of T-cell epitope location and order/disorder protein structure

N.S. Mitić[1], M.D. Pavlović[2], D.R. Jandrlić[3], S. Malkov[1]

[1]*University of Belgrade, Faculty of Mathematics, Belgrade, Serbia*
[2]*University of Belgrade, Institute of General and Physical Chemistry, Belgrade, Serbia*
[3]*University of Belgrade, Faculty of Mechanical Engineering, Belgrade, Serbia*

nenad@matf.bg.ac.rs

Highly disordered protein regions are prevalently hydrophilic, extremely sensitive to proteolysis in vitro, and are expected to be under-represented as T-cell epitopes [1]. According to the NetMHCpan and NetMHCIIpan T-cell epitope predictors [2] and several publicly available disorder prediction algorithms [3], frequency of epitopes presented by human leukocyte antigens (HLA) class-I and -II was found to be more than 2.5 times higher in ordered than in disordered protein regions (dependending on the disorder predictor). HLA class-II-binding epitopes were more frequent in both ordered and disordered regions, compared to the HLA class-I epitopes. Both HLA class-I and HLA class-II binding epitopes are prevalently hydrophilic in disordered and prevalently hydrophobic in ordered protein regions and epitopes recognized by HLA-II alleles are more hydrophobic than those recognized by HLA-I ones. Epitopes, predicted to bind with high affinity to the both classes of HLA molecules, display more hydrophobicity than low affinity-binding epitopes. Promiscuous epitopes were also predicted to be more hydrophobic than unique epitopes. These data suggest that reverse vaccinology is oriented towards prevalently hydrophobic, ordered regions. Discovered relations were also valid if proteins were grouped according to the main taxonomic categories, or if alleles were grouped according to the HLA class-I and HLA class-II supertypes (except for the class-I supertype A3, in which the main part of recognized epitopes was prevalently hydrophilic). Distribution of epitopes in ordered and disordered protein regions revealed that the curves of order-epitope distribution were convex-like while the curves of disorder-epitope distribution were concave-like and that the trend of distribution of HLA-I- and HLA-II- binding epitopes was opposite, for all analyzed disorder predictors. The comparison between predicted and experimentally evaluated epitopes of a tumor-associated antigen of MAGE-A34 [4] extensively studied for immunotherapy, confirmed that majority of epitopes presented by HLA class-I and HLA-II molecules was located in ordered and disorder/order boundary protein regions, which may have implications on epitope mapping.

Acknowledgement
The work presented has been financially supported by the Ministry of Education, Science and Technological Development, Republic of Serbia, Projects No. 174021,174002 and TR 31055.

[1] P. L.Carl, B.R. S.Temple, P. L .Cohen, Arthritis Res.Ther. 7, R1360 (2005).
[2] http://www.cbs.dtu.dk/services/
[3] Z. Dosztanyi, J. Chen, A.K. Dunker, I. Simon, P. Tompa, Brief. Bioinform. 11, 225 (2009).
[4] Cancer immunity database [http://www.cancerimmunity.org/].

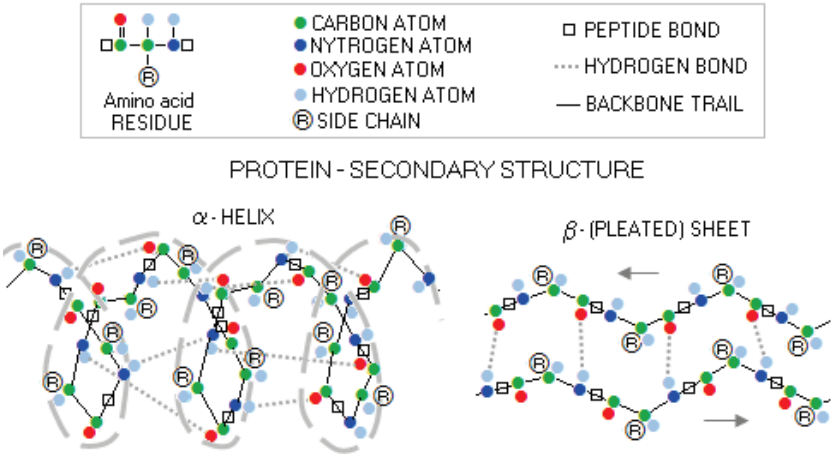
Local protein structure prediction by Bayesian probabilistic approach principle

Mihajlo Mudrinić

Vinca Institute of Nuclear Sciences, University of Belgrade, Belgrade, Serbia

m.mudrinic@vinca.rs

The task of understanding and predicting how to translate the information coded in the amino acid sequence of proteins into knowledge of how such protein would fold, is one of the most important problems in biochemistry. Consequently, we want to understand how the primary structure (the sequence of residues) gives rise to tertiary structure (the folded state). The work is focused on intermediate structure between the two, the secondary structure (patterns or motifs like helices). First step known as "data fusion" in methodology for visualization will be presented. Data fusion entails integration of paired analysis (based on the relation protein block-amino acid propensity) into knowledge based on estimation theory including statistical pattern recognition and multivariate analysis.



Acknowledgment
The research has been supported by the Ministry of Education, Science, and Technological Development of Serbia through the project Physics and Chemistry with Ion Beams, Project No. III 45006.

[1] Udo von Toussaint, Rev. Mod. Phys. 83, 943 (2011).

Machine learning methods for optimization of sepsis therapy and drug repositioning

Zoran Obradovic

Data Analytics and Biomedical Informatics Center Temple University, Philadelphia, USA

zoran.obradovic@temple.edu

The first part of this lecture will describe a recent result [1] of our ongoing DARPA DLT project aimed to reduce mortality rate in septic shock, which is the leading cause of death in intensive care units. This is a challenging objective due to the fast progression and complex multi-stage nature of acute inflammation. Our approach, shown to rescue more patients than standard therapy, is based on early diagnosis and optimized blood purification that relies on analysis of temporal dependencies in high dimensional multi-source data. The second part of the lecture will discuss our integrative computational framework to predict novel drug indications for both approved drugs and clinical molecules by integrating chemical, biological and phenotypic data sources [2]. Individual similarity measures for each of these data sources are defined and a weighted k-nearest neighbor algorithm is utilized to transfer similarities of nearest neighbors to prediction scores for a given compound. A large margin method was used to combine individual metrics into a global metric. A large-scale study was conducted to repurpose 1,007 drugs against 719 diseases. Experimental results showed that the proposed algorithm outperformed previously developed computational drug repositioning approaches.

Moreover, the new algorithm also ranked drug information sources based on their contributions to the prediction, thus paving the way for prioritizing multiple data sources and building more reliable drug repositioning models. In particular, integrating knowledge from all three sources was beneficial for drug indication prediction, and the weights suggest which data sources and similarity comparisons need improvement.

Acknowledgment

The first part of this work is joint research with Dr. Mohamed Ghalwash and Dr. Vladan Radosavljevic who are postdoctoral associates at my laboratory while the second part is joint work with Dr. Ping Zhang from IBM T.J. Watson Labs and Dr. Pankaj Agarwal from GlaxoSmithKline R&D. Funding for this research is provided in part by DARPA-N66001-11-1-4183 and by GlaxoSmithKline LLC.

[1] Ghalwash, M., Radosavljevic, V., and Obradovic, Z., "Early Diagnosis and Its Benefits in Sepsis Blood Purification Treatment," Proc. 2013 IEEE International Conference on Healthcare Informatics, International Workshop on Data Mining for Healthcare, Philadelphia, PA, USA, Sept. 2013.

[2] Zhang, P., Agarwal, P. and Obradovic, Z. "Computational Drug Repositioning by Ranking and Integrating Multiple Data Sources," Proc. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Prague, Czech Republic, Part III, LNAI 8190, 579-594, 2013, Springer-Verlag, Heidelberg, Sept. 2013.

Radiation effects of slow electrons on biomolecules - where the experiment and theory meet

Radmila Panajotović

Institute of Physics,University of Belgrade, Belgrade, Serbia

radmila@ipb.ac.rs

Low-energy electrons are proved to be the most abundant secondary species created in the irradiation of the living tissue by high-energy ionizing radiation (X- and g - rays, ions, etc).Their impact can result in formation of dangerous molecular fragments from nucleic bases [1] and breaking the strands of the DNA [2]. They also proved to be an efficient tool in production of bio-chips on the SAM structure [3]. Investigation, both experimental and theoretical, of the vibrational and electronic excitation of nucleic bases and nucleosides by low-energy electron impact leading to dissociation through formation of temporary negative ions aims at understanding the correlation of nucleic basesugar moiety conformational coupling and its consequences on the bond cleavage in DNA. For example, the conformation of the 2'-deoxyribose moiety with respect to the base is expected to influence which species are formed upon exposure of nucleosides to ionizing radiation.

Furthermore, owing to the vast number of metabolic processes that can lead to potentially damaging radiation effects, the cell membrane with its highly ordered structure and complex functionality relying on the activity of various proteins, presents another important target for cancer research (for example, DPPC is a major component of the highly radiation sensitive lung tissue,) and a perfect "live model" of a biosensor. Radiation damage of molecules that compose the cell membrane, such as phospholipids, proteins and polysaccharides, can be crucial for the operation of cell's ion transport and signaling. It is a challenging task for both experiment [4] and theory [5], mostly due to the difficulty in recreating a reproducible, stable and "true-to-life" sample for various experimental techniques (KPFM, STM, FT-IR, XPS, NEXAFS, HRLEES, etc.) to be applied and, in the case of theoretical modeling, properly including the forces that govern the interaction between the individual components in a bio-molecular complex while keeping the computation time relatively short.

In this talk it will be discussed where the limits of both experiment and theory of radiation damage in biomolecules lie, as well as where the results of both complement each other.

[1] Radmila Panajotovic and Leon Sanche, "From DNA to nucleic bases - the effects of low-energy electron impact", Journal of Physics: Conference Series, 88 (2007) 012074.
[2] Panajotovic, M Michaud and L Sanche, "Cross sections for low-energy electron scattering from adenine in the condensed phase", Phys. Chem. Chem. Phys. 9 (2007) 138.
[3] R Panajotovic, F Martin, P Cloutier, D Hunting, and L Sanche, "Effective Cross Sections for Single Strand Break Production in Plasmid DNA by to 4.7 eV electrons", Radiation Research, 165 (2006) 452-459.
[4] Panajotovic, R, Schnietz M, Turchanin A, Mason N and Gölzhäuser A, "XPS Study on Effects of Electron-Beam Irradiation of Thin Condensed DPPC Films", Book of Abstracts, ECAMP X, 4.-10. June 2010, Salamanca, Spain.
[5] S. L. Dudarev, "Density Functional Theory Models for Radiation Damage", Annual Review of Materials Research, 43 (2013) 35-61.

Mining associations for organism characteristics in prokaryotes – an integrative approach

G. Pavlovic-Lazetic[1], V. Pajic[2], N. Mitic[1], J. Kovacevic[1] and M. Beljanski[3]

[1]*Faculty of Mathematics,University of Belgrade, Belgrade, Serbia*
[2]*Faculty of Agriculture,University of Belgrade, Belgrade, Serbia*
[3]*Institute of General and Physical Chemistry, Serbia*

gordana@matf.bg.ac.rs

Correlation between specific organism characteristics (such as genome size, genome GC content, optimal growth temperature, habitat, oxygen requirements) has been the subject of many studies with different outcomes regarding different taxonomic categories [1]-[3]. We reconsider such correlations for superkingdoms Archaea and Bacteria organisms and extend the study in a number of ways. We use a larger dataset of prokaryotes by integrating several existing databases with data obtained by literature mining [4]. We apply algorithms for association rule mining in order to identify the most confident associations between specific modalities of the characteristics considered. The finite state method for literature mining as well as the results of both literature and association rules mining [5],[6] are presented and commented on.

Acknowledgment
The work presented has been supported by the Ministry of Education and Science, Republic of Serbia, Projects No. 174021.

[1] PV Burra, L. Kalmar, P. Tompa, PLoS ONE 5(8): e12069. oi:10.1371/journal.pone.0012069 (2010) .
[2] NJ MacDonald, RB Beiko, Bioinformatics, 26, 15 (2010).
[3] M. Tamur, P. D'haeseleer, Bioinformatics, 24, 13 (2008) .
[4] V. Pajic, GM Pavlovic-Lažetic, MV Beljanski, BW Brandt, MB Pajic, Int. J. Data Mining and Bioinformatics, 7, 2 (2013).
[5] V. Pajic, G. Pavlovic Lažetic, M. Pajic, Lecture Notes in Computer Science, 6807, ISSN 0302-9743, Springer Berlin Heidelberg (2011) .
[6] GM Pavlovic-Lažetic, NS Mitic, JJ Kovacevic, Z.Obradovic, SN Malkov, MV Beljanski, BMC Bioinformatics 12:66 (2011) http://www.biomedcentral.com/1471-2105/12/66.

Matrix genetics: algebra of projection operators, cyclic groups and inherited ensembles of biological cycles

S. Petoukhov

Mechanical Engineering Research Institute, Russian Academy of Sciences, Moscow, Russia

spetoukhov@gmail.com

This lecture is devoted to applications of projection operators to study molecular-genetic systems and some inherited physiological phenomena. The author analyzes matrix representations of ensembles of molecular-genetic elements in a form of a Kronecker family of matrices $[C\ U;\ A\ G]^{(n)}$ where C means cytosine, U – uracil, A – adenine, G – guanine, (n) – Kronecker power. This analysis develops significantly a matrix approach for studying the genetic code, which was begun in a classical work [1]. The (8*8)-matrix $[C\ U;\ A\ G]^{(3)}$ contains all 64 triplets in a strong order. The nature has divided the set of 64 triplets into two equal subsets: 32 triplets have "strong roots" and 32 triplets have "weak roots" (these names are taken from [1]). A disposition of triplets with strong roots and weak roots in the matrix $[C\ U;\ A\ G]^{(3)}$ gives such symmetrical mosaic that a mosaic of each column of this matrix has meander-like character and coincides with one of Rademacher functions. By replacing each of triplets with a strong (weak) root with number "+1" ("-1"), we receive a numeric "Rademacher representation" R of the symbolic matrix $[C\ U;\ A\ G]^{(3)}$. Another phenomenological fact (a unique status of uracil U) leads to a "Hadamard representation" H of the matrix $[C\ U;\ A\ G]^{(3)}$.

These two (8*8)-matrices R and H are main mathematical objects of the lecture. Each of them is a sum of such 8 sparse matrices that each sparse matrix is an oblique projection operator (it satisfies the criterion $P^2 = P$). Combinations of these "genetic" (8*8)-projectors in a form of sums of two or more projectors lead to sets of cyclic groups and other interesting mathematical objects, which are used by the author to model inherited ensembles of biological phenomena [2-3]. The lecture specially describes modeling ensembles of cyclic processes because a living body is a huge chorus of inherited cyclic processes. The cyclic groups, which arise on the basis of exponentiation of certain sums of the genetic projectors, have a useful property of a selective control (or coding) of cyclic changes of vectors inside appropriate sub-spaces of an 8-dimensional vector space. The author shows an algorithm to extend the Rademacher and Hadamard (8*8)-matrices R and H into $(2^n * 2^n)$-matrices with similar properties to simulate big ensembles of interrelated cyclic processes. This topic relates with fundamental problems of a biological watch, bio-rythmology and chrono-medicine.

The lecture also demonstrates connections of different sums of the genetic (8*8)-projectors with complex numbers and split-complex numbers, Hamilton quaternions and split-quaternion by J.Cockle. The author uses exponentiation of different sums of these genetic projectors to simulate some other inherited biological phenomena: color perception, properties of which correspond to Newton's color circle; ensembles of phyllotaxis patterns, etc. He puts forward a "projection conception", which analyses living bodies as colonies of projection operators and constructions on a basis of direct sums of vector sub-spaces. Projectors are used widely in mathematics, physics (including quantum mechanics), chemistry, informatics, logics, etc. It seems that the nature likes projectors. For example, electromagnetic vectors are represented as sums of their projections in a form of electric and magnetic vectors. Our results give evidences that the notion "projection operators" can be one of useful notions and instruments in the field of mathematical and theoretical biology. In the author's laboratory the described results are used for developing algebraic biology and biotechnical applications. We believe that living matter is an algebraic essence in its informational fundamentals.

[1] B.Konopelchenko, Y.Rumer, Doklady Akademii Nauk SSSR, 223(2), 471 (1975).
[2] S.Petoukhov, http://arxiv.org/abs/1307.7882
[3] S.Petoukhov, http://symmetry.hu/scs_online/SCS_23_3-4.pdf

Mining real-world networks: from biology to economics

Natasa Przulj

Department of Computing, Imperial College London, SW72AZ, UK

natasha@imperial.ac.uk

Comparing networks is important for understanding the organizational principles and tracking the dynamics in a broad spectrum of areas that involve many interacting objects. Examples include relationships between people, the global climate system, the world's economic system, bindings between bio-molecules in a cell, and much else. We introduce new network analysis, comparison and alignment methods, based on counts of local subnetworks, which are easy to compute and produce meaningful and robust results along a wide array of networks. We validate them on simple models, for which, unlike for real world, we know the answers and hence can check the validity of the methods. Also, we apply them to real-world networks from several domains and uncover new domain-specific knowledge.

Acknowledgment
This work was supported by the European Research Council (ERC) Starting Independent Researcher Grant 278212, the National Science Foundation (NSF) Cyber-Enabled Discovery and Innovation (CDI) OIA-1028394, GlaxoSmithKline (GSK) Research & Development Ltd, and the Serbian Ministry of Education and Science Project III44006.

[1] A. Sarajlic, V. Janjic, Neda Stojkovic, Djordje Radak, and N. Przulj, Network Topology Reveals Key Cardiovascular Disease Genes, PLoS ONE, 8(8): e71537, 2013.
[2] W. Hayes, K. Sun, and N. Przulj, On the suitability of graphlet-based measures for biological network comparison, Bioinformatics, 29(4): 483-491, 2013.
[3] V. Janjic, and N. Przulj, Biological function through network topology: a survey of the human diseasome, Briefings in Functional Genomics, 11(6):522-532, 2012.
[4] V. Janjic, and N. Przulj, The Core Diseasome, a special issue on Emerging Investigators, Molecular BioSystems, 8:2614-2625, 2012.
[5] V. Memisevic, and N. Przulj, C-GRAAL: Common-neighbors-based Global Graph Alignment of Biological Networks, Integrative Biology, 4:734-743, 2012.
[6] T. Milenkovic, V. Memisevic and N. Przulj, Dominating Biological Networks, PLoS ONE, 6(8):e23016, 2011.
[7] Arabidopsis Interactome Mapping Consortium, Evidence for Network Evolution in an Arabidopsis Interactome Map, Science, 333:601-607, July 29, 2011.
[8] O. Kuchaiev and N. Przulj, Integrative Network Alignment Reveals Large Regions of Global Network Similarity in Yeast and Human, Bioinformatics, 27(10): 1390-1396, 2011.
[9] O. Kuchaiev, A. Stefanovic, W. Hayes, and N. Przulj, GraphCrunch 2: Software tool for network modeling, alignment and clustering, BMC Bioinformatics, 12(24):1-13, 2011.
[10] N. Przulj, Protein-protein interactions: making sense of networks via graph-theoretic modeling, Bioessays, 33(2), February, 2011.
[11] T. Milenkovic, W.L. Wong, W. Hayes, and N. Przulj, Optimal network alignment with graphlet degree vectors, Cancer Informatics, 9:121-37, June 30, 2010.
[12] H. Ho, T. Milenkovic, V. Memisevic, J. Aruri, N. Przulj, and A. K. Ganesan, Protein Interaction Network Uncovers Melanogenesis Regulatory Network Components within Functional Genomics Datasets, BMC Systems Biology, 4:84, June 15, 2010.

Golden and harmonic mean in the genetic code

Miloje M. Rakočević

Department of Chemistry, Faculty of Science, University of Niš, Serbia

milemirkov@open.telekom.rs

In a previous work we have shown that golden mean is a characteristic determinant of the genetic code (GC), regarding on the codons binary tree, 0 – 63 (Biosystems 46, 283-291, 1998). In a second one we showed a splitting of Genetic Code Table (GCT) into three equal and significant parts, using the harmonic mean [$H(a, b) = 2ab/(a + b)$; $a = 63$, $b = 31.5$] (arXiv:1305.5103v4 [q-bio.OT]). In this communication, however, we will show that a specific *unity* of golden mean and harmonic mean appears to be the determinant of Rumer's Table of 16 nucleotide doublets (Tables 1 – 2).

01. G	GG (6)	**02. F**	UU (4)	**03. L**
04. P	CC (6)	**05. N**	AA (4)	**06. K**
07. R	CG (6)	**08. I**	AU (4)	**09. M**
10. A	GC (6)	**11. Y**	UA (4)	**12. St.**
13. T	**AC** (5)	**14. H**	**CA** (5)	**15. Q**
16. V	GU (5)	**17. C**	UG (5)	**18. W**
19. S	UC (5)	**20. D**	GA (5)	**21. E**
22. L	CU (5)	**23. S**	AG (5)	**24. R**

Table 1. Rumer's Table of nucleotide doublets

01. G	GG (6)	**02. F**	UU (4)	**03. L**
04. P	CC (6)	**05. N**	AA (4)	**06. K**
07. A	GC (6)	**08. Y**	UA (4)	**09. St.**
10. R	CG (6)	**11. I**	AU (4)	**12. M**
13. V	**GU** (5)	**14. C**	**UG** (5)	**15. W**
16. T	AC (5)	**17. H**	CA (5)	**18. Q**
19. L	CU (5)	**20. S**	AG (5)	**21. R**
22. S	UC (5)	**23. D**	GA (5)	**24. E**

Table 2. The modified Rumer's Table

As we have shown, golden mean "falls" between the 38th and 39th codon (38. CAA, 39. CAG), which code for glutamine (Q), a more complex of only two amide amino acids (AAs); two codons, adjacent to the codons (40.UGU, 41.UGC), which code for one of the only two sulfur AAs, cysteine (C). This "harmonization" of diversity is increased by the harmonic mean, in position 42 on the sequence 0-63. The harmonization extends further to "stop" codon (42.UGA) and to codon (43.UGG) that codes for most complex AA, tryptophan (W). [The "42" as ending position on the "Golden route" (with Fibonacci numbers) on the Farey tree, corresponding with six-bit GC binary tree.]

On the other side, the splitting of GCT into three parts through harmonic mean makes that AAs are distinguished on the basis of the validity of the evident regularities of key parameters, such as polarity, hydrophobicity and enzyme-mediated amino acid classification.

And now, the last determination. With a minimal modification of Rumer's nucleotide doublets Table[1] follows the next result. With the four "first" doublets we have four outer "small squares", i.e. codon families (n_1 = {GG,UU,GU,UG}), which code for nonpolar AAs; the four "second" doublets give four inner codon families (n_2 = {CC,AA,AC,CA}), which code for polar AAs. With the four "third" (n_3 = {GC,UA,CU,AG}) and four "fourth" (n_4 = {CG,AU,UC,GA}) doublets are chosen eight intermediate codon families, the first three code for nonpolar AAs [(CU,AU,G~~C~~)→ (L,I,M,A)], and the second five code for polar AAs [(UC,UA,CG,AG,GA)→ (S,Y,R,D,E)]. [2]

[1] If at the beginning of first sub-system, with 6/4 hydrogen bonds, are GG/UU doublets, chemical reasons require GU/UG doublets at the beginning of the second sub-system, with 5/5 hydrogen bonds, instead of AC/CA. From the same reasons, we have the changes: CG/GC & UC/CU on the left and AU/UA & GA/AG on the right.

[2] The polarity/non-polarity after *Cloister energy,* as in Swanson's article (Bull. Math. Biol. 46, 187-207, 1984).

This result leads us to the following conclusion. Within the set of all *n*-gons, where *n* is even number, the case $n = 4$ is only and one case where harmonic mean of "golden whole" ($n^2 - n$) and its half [$(n^2–n)/2$] equals $2n$, and $n^2–n = 3n$. [This ratio 2:3 shows harmonic mean within harmonic mean and the sequence "$n_1 – (n_3$ or $n_4) – n_2$" corresponds with the Cantorian triadic set.] Moreover, such a harmonic mean appears to be corresponding with the number of "small squares" within intermediate space in form of only one "ring" as it follows: [(2 + 2) + (4 x $\underline{0}$) = 4]; [**(4 + 4) + (8 x $\underline{1}$) = 16**]; [(6 + 6) + (12 x $\underline{2}$) = 36]; [(8 + 8) + (16 x $\underline{3}$) = 64] etc. As it is self-evident, the symmetrical "out – middle – in" arrangement is not possible for $n \neq 4$, neither for *n*-gons nor for *n*-letter alphabets.

Codon-anticodon interaction and genetic code evolution

A.Sciarrino[1] and P.Sorba[2]

[1] *Dipartimento dio Fisica, Università di Napoli "Federico 11"Complesso Universitario di Monte S.Angelo Via Cinthia, I-80126 Napoli, Italy*
[2] *LAPTH, Laboratoire d'Annecy-le-Vieux de Physique Théorique, CNRS, Université de Savoie, B.P.110, 71941Annecy-le-Vieux, France*

paul.sorba@lapth.cnrs.fr

Imposing a minimum principle for the interaction in the framework of the Crystal Basis Model of the Genetic Code [1, 2] – which will be defined and commented – we determine the structure of the anticodon in the Ancient, Archetypal and early Genetic codes, that are all reconciled in a unique frame. Most of our results agree with the generally accepted scheme [3, 4].

[1] L.Frappat, A.Sciarrino and P.Sorba, Phys.Lett. A 250, 214-221 (1998).
[2] L.Frappat, A.Sciarrino and P.Sorba, J.Biol.Phys.27, 1-34 (2001).
[3] A.Sciarrino and P.Sorba, BioSystems, 107,113-119 (2011).
[4] A.Sciarrino and P.Sorba, BioSystems, 111,175-180 (2013).

Bifurcation analysis of HPA axis dynamic states under cholesterol regulation

A. Stanojević[1], V.M. Marković[1], S. Maćešić[1], V. Vukojević[2], Ž. Ćupić[3] and Lj. Kolar-Anić[1,3]

[1]*Faculty of Physical Chemistry, University of Belgrade, Belgrade, Serbia*
[2]*Karolinska Institutet, Department of Clinical Neuroscience, Stockholm, Sweden*
[3]*Institute of Chemistry, Technology and Metallurgy, University of Belgrade, Department of Catalysis and Chemical Engineering, Belgrade, Serbia*

anastann@yahoo.com

Hypothalamic-pituitary-adrenal (HPA) axis constitutes a neuroendocrine system of vital importance for maintaining the body's homeostasis in both basal physiology and under various stress conditions. Proper circadian and ultradian dynamics of concentrations of its hormones has proven to be essential for its normal functioning and preserving health [1,2]. Cholesterol, the precursor of HPA axis' chief acting hormone cortisol, has a substantial impact upon HPA axis oscillatory dynamics, altering levels or oscillation amplitudes of cortisol blood plasma concentration [3,4]. However, precise underlying mechanisms and physiological effects of cholesterol's influence on HPA axis are not satisfactorily understood. Investigation of intricate and often counterintuitive interactions between cholesterol and HPA axis dynamic regulation can be substantially aided and predicted by employing modeling and nonlinear systems theory methodologies. To this end, we use the recently proposed five dimensional stoichiometric model of HPA axis activity [5], which beside four key HPA axis hormones (CRH, ACTH, cortisol and aldosterone) incorporates cholesterol as an additional dynamic variable. By altering the model parameters regulating cholesterol's kinetics (rate constants of cholesterol inflow/outflow in/from the system), we systematically investigate the transitions of cortisol dynamic states. Bifurcation analysis reveals that changes in cholesterol levels, realized due to alterations in cholesterol inflow or outflow, critically modify dynamics of HPA axis, shifting it from oscillatory states to stable steady states. These transitions to stable steady states occur through significant transformations in levels and amplitudes of circadian and ultradian cortisol oscillations, leading to exceedingly robust oscillatory dynamic states, and thus diminished capacity of the organism to adequately respond to and subsequently recover from various stress stimuli. The obtained results indicate to a delicate dynamic regulation of the HPA axis by cholesterol, underlining the importance of upholding cholesterol levels within its referenced values for human health. Additionally, the bifurcation analysis of cortisol's concentration dynamic states under cholesterol regulation might help in elucidating alterations in HPA axis dynamics observed in subjects on diet with high fat and cholesterol content, as well in some psychiatric and metabolic conditions, such as eating disorders (anorexia nervosa, night eating syndrome), metabolic syndrome and major depression.

Acknowledgment
The support of Ministry of Education, Science and Technological Development of the Republic of Serbia, grants No. 172015 and 45001, Karolinska Institute Research Funds, the Swedish Research Council and the Knut and Alice Wallenberg Foundation is gratefully acknowledged.

[1] S. L. Lightman, B. L. Conway-Campbell, Nat. Rev. Neurosci. 11, 710 (2010).
[2] V. M. Marković, Ž. Ĉupić, V. Vukojević, Lj. Kolar-Anić, Endocr. J. 58, 889 (2011).
[3] H. E. Auvinen, J. A.Romijn, N. R. Biermasz, H. Pijl, L. M. Havekes, J. W. Smit, P.C. Rensen, A. M. Pereira, J. Endocrinol. 214, 191 (2012).
[4] A. R. Tyrka, O. C. Walters, L. H. Price, G. M. Anderson, L. L. Carpenter, Horm. Metab. Res. 44, 543 (2012).
[5] V. M. Marković, S. Maćešić, S. Damjanović, Lj. Kolar-Anić, Physical Chemistry 2012 (S. Anić and Ž. Ĉupić, eds.), Society of Physical Chemists of Serbia, Belgrade, p. 264 (2012).

From genotype to phenotype: *in silico* modelling of serotonergic system

S. Svikovic[1], M. Brkusanin[1], J. Karanovic[1], M. Pantovic[2], S. Djurica[1], G. Brajuskovic[1], S. Romac[1], M. Ivkovic[2] and D. Savic-Pavicevic[1]

[1]*Centre for Human Molecular Genetics, Faculty of Biology, University of Belgrade, Belgrade, Serbia*
[2]*Clinic for Psychiatry, Clinical Centre Serbia, Belgrade, Serbia*

svikovic@gmail.com

Serotonin is a neurotransmitter involved in the regulation of variety of behaviours, including food intake, aggression and mood. Dysregulation of serotonergic transmission in nervous system had been proposed in several neuropsychiatric disorders. A multitude of functional polymorphisms in the genes involved in serotonergic transmission have been associated with these disorders. Due to the underlying genetic and environmental complexity, probable mechanisms of these disorders could be proposed using system approach describing kinetics of serotonergic transmission [1].

We developed computational model of serotonergic synapse using CellDesigner v4.2 and Systems Biology Workbench (SBW) software package using data obtained from BRENDA and BioGPS databases. In parallel, we have genotyped a total of 9 polymorphisms in 6 genes involved in serotonergic transmission (tryptophan hydroxylase 2–*TPH2*, monoamine oxidase A–*MAOA*, serotonin transporter–*SERT*, serotonin receptor 1A–*5HTR1A*, serotonin receptor 2C–*5HTR2C* and brain-derived neurotrophic factor–*BDNF*) on 59 patients with major depression (MDD) and 81 control subjects. MDD patients were diagnosed and classified as melancholic and atypical in accordance with DSM-IV criteria. Additionally, data on patients response to selective serotonin reuptake inhibitors (SSRIs) were available. After the computational model of serotonergic synapse was optimized, we inputted the genotypes of patients and controls to ascertain whether there are significant differences in kinetic parameters and stability of serotonergic transmission in MDD patients compared to controls, between different types of MDD, as well as with response to SSRI.

The results we obtained following model optimization show the significance of negative feedback loop formed by presynaptic serotonin autoreceptor 1A (5HTR1A) in stabilizing the serotonin efflux into the synaptic cleft. Secondly, we have shown the importance of synaptic plasticity of serotonergic synapse promoted by postsynaptic *BDNF*, which increases the sensitivity of postsynaptic receptors. The decrease of binding affinity of serotonin to postsynaptic receptors dramatically decreases the robustness of the system, making it more unstable in increased neuronal firing rate.

We performed stability analysis utilizing Nyquist [2]. We have shown significant differences in measured kinetic parameters of serotonergic transmission during 10 second excitation period between MDD and control subjects: serotonin release into synaptic cleft ($p<0.001$) and downregulation of monoamine oxidase A activity ($p<0.05$). We have shown that MDD patients with melancholic type differ significantly from those with atypical type in serotonin synthesis kinetics, packaging into synaptic vesicles and serotonin degradation ($p<0.001$). In addition, we have shown significant differences in response to mood regulators in parameters associated with serotonin synthesis, serotonin degradation and serotonin autoreceptor trafficking ($p<0.05$).

Our results support the results we obtained from the association studies, emphasizing the role of redundant epistatic interactions in canalizing phenotypes with similar kinetics of serotonergic transmission.

Acknowledgement
This work has been supported by the research project OI173016 sponsored by the Ministry of Education, Science and Technology of the Republic of Serbia.

[1] J. Best, H.F. Nijhout, M. Reed. Theoretical Biology & Medical Modelling 7:34, 1 (2010).
[2] L. Stark, P.M. Sherman. J. Neurophys. 20:17 (1957).

Genomic analysis of complex phenotypes

Nenad Svrakic

Institute of Physics,University of Belgrade, Belgrade, Serbia

nenads@ipb.ac.rs

In this talk we systematize existing data and review new findings on the cause of schizophrenia and other complex disorders and outline an improved mixed model of their risk.

Recent findings

Multiple and variable genetic and environmental factors interact to influence the risk of complex disorders. Both rare variants with large effect and common variants with small effect contribute to genetic risk of such phenotypes, with no indication for differential impact on their clinical features. Accumulating evidence supports a genetic architecture of neurodevelopmental disorders with multiple scenarios, including additive polygenic, heterogeneity, and mixed polygenic-heterogeneity. The epigenetic mechanisms that mediate gene–environment (GxE) interactions provide a framework to incorporate environmental factors into models of schizophrenia risk. Environmental pathogens with small effect on risk have robust effects in the context of family history of schizophrenia and other disorders. Hence, genetic risk for mental disorders and other complex phenotypes may be expressed in part as sensitivity to environmental factors.

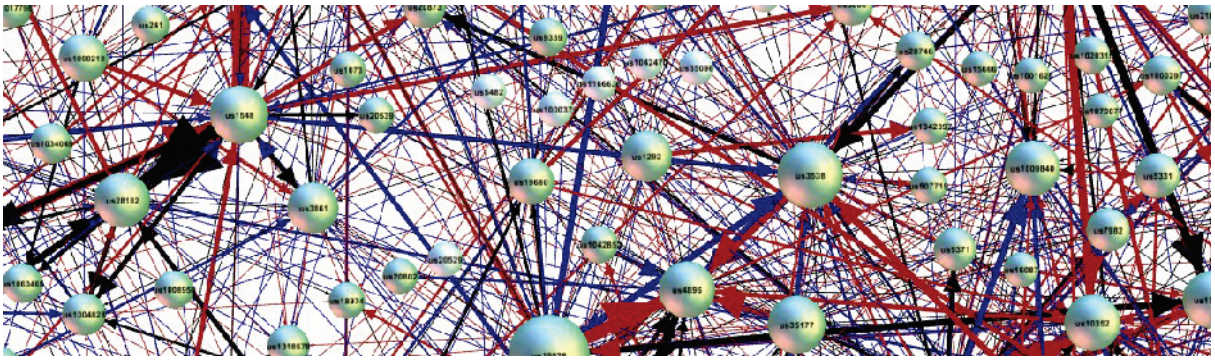
Online social networks: the structure of emotional dialogs

B. Tadic[1], V. Gligorijevic[1] and M. Suvakov[2]

[1]*Department of Theoretical Physics, Jozef Stefan Institute, Ljubljana, Slovenia*
[2]*Institute of Physics, University of Belgrade, Belgrade, Serbia*

bosiljka.tadic@ijs.si

In human behavior emotions play an important role from the level of brain activity of each individual to large-scale social dynamics. Recent developments in the quantitative study of emotion based on Russell model in psychology [1] and machine-learning methods of text interpretation, emotion components can be retrieved from text messages in online communications among users [2,3]. Quantitative science of human dynamics on the Web in the framework of statistical physics of complex systems, supplied with the machine-learning methods of text analysis, strives to reveal the role of emotions in collective behaviors of users [2,4]. Applying these approaches for the empirical data analysis and agent-based modeling [5], we have studied several online communication systems (survey is available online [4]); our results indicate that, by various mechanisms in the dynamics, the emotional dialogs can lead to bursting events involving many users, and to the emergence of specific sorts of social networks. In this lecture, we will present two archetype social structures which are based on emotional communications, specifically the dialogs in MySpace social network [6] and in online chats on Ubuntu channel [7,8]. We will give a detailed analysis of their topology, focusing on the role of emotion components (valence and arousal) in building particular substructures, as well as the network resilience. By fractal analysis of the time series of emotional messages, in underlying self-organized dynamics, further distinctions between these emergent social structures will be pointed out.



Acknowledgment
Supported by the program P1-0044 by the Research agency of the Republic of Slovenia, FP7 project Cyberemotions, and the Ministry of Education, Science and Technology of the Republic of Serbia.

[1] J.A. Russel, A circumplex model of affect, J. Pers. Soc. Psychology 39:1161-1178 (1980).
[2] CYBEREMOTIONS FP7 project: http://www.cyberemotions.eu/
[3] R.A. Calvo, S. D'Mello, Affect detection, and interdisciolinary review of models, methods, and their applications, IEEE Trns. Affect. Comp. 1:18-37 (2010).
[4] http://www-f1.ijs.si/~tadic/projects/cybere_.html
[5] B. Tadic, M. Suvakov, Can human-like Bots affect collective mood: Agent-based simulations of online chats, arxiv:1305.2741 (2013).
[6] M. Suvakov, M. Mitrovic, V. Gligorijevic, B. Tadic, How the online social networks are used: Dialogs-based structure of MySpace, J. Roy. Soc. Interface 10:20120819 (2012).
[7] V. Gligorijevic, M. Skowron, B. Tadic, Structure and stability of online chat networks built on emotion-carrying links, Physica A 392:538-543 (2013).
[8] V. Gligorijevic, M. Suvakov, B. Tadic, Building online social networks with users, agents and Bots, in Complex Networks and their Applications, H. Cherifi, Ed., Cambridge Scholars Pub. (2013).

Associative memory of the hippocampus: storage and retrieval of ultrametric patterns

D. Bianchi, M. Piersanti and B. Tirozzi

Department of Physics, University of Rome, "La Sapienza", Rome, Italy

b.tirozzi@libero.it

Associative memory has been analyzed since the celebrated Hopfield model [1] in many papers based on theoretical arguments. The configurations of the system are sequences of bits (0,1) and some configurations are stored including the in the synaptic interaction via the Hebb rule. The retrieval is obtained giving some configuration not very different from the stored patterns as an initial condition and letting the system evolve according to the threshold dynamic. If the evolving configuration gets blocked in some of the stored patterns then this pattern is retrieved. These ideas have been extrapolated from the system of biological neurons and put in a theoretical well defined model, which in some cases has been solved analytically. In this work we try to apply these ideas to a particular system of real neurons, the CA1 pyramidal neurons of the hippocampus, a region of the brain where the memory is formed. The input are the impulses coming from the Enthorinal cortex and the CA3 system. The patterns are presented to the network of CA1 neurons as a set of impulses corresponding to the neurons active in the pattern. The storage and retrieval processes are different from the Hopfield model, but we also organized the patterns in human and animals in a hierarchical structure with the ancestors and descendants. We study the retrieval and storage of ultrametric patterns with different degrees of ultrametricity We analyze also the capacity of the network in healthy conditions of the neurons comparing with situation of unpaired neurons as it happens in the case Alzheimer. We also checked that the presence of CREB improves the memory functions in the case of unpaired neurons while there is no big difference in the control conditions.

[1] J.J Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proc. National Ac. of Sciences. 79,8, 2554-2558, (1982).

Self-organizing non-Euclidean representations in the brain

F. Troiani[1], E. Urdapilleta[1], F. Stella[1] and A. Treves[2]

[1]*Cognitive Neuroscience, SISSA, Trieste, Italy*
[2]*Science Office, Embassy of Italy, Tel Aviv, Israel*

ale@sissa.it

So-called *grid cells* have been discovered in the rodent brain, in 2005 [1], that express with their activity a remarkably regular metric tessellation of a flat surface - a triangular grid. Their spiking activity is concentrated when the animal is at the nodes of an imaginary grid, different from neuron to neuron. Later found also in crawling bats [2], it is not clear yet whether in flying bats grid units will provide a regular 3D tiling - such as a face-cubic-centered or hexagonally-compact-packing arrangement. Most computational models of this phenomenon are based on various forms of wiring instructions [3], whereas we have been analyzing a model [4] that demonstrates how grid units can self-organize, spontaneously, during animal development. Our model predicts that if rats are raised in a cage with non-flat geometry, such as a ball [5] or a hyperbolic surface of negative curvature, grid units will express a non-Euclidean regular tiling, with firing rate peaks that have e.g. 5 or 7 nearest neighbors instead of the Euclidean 6 so far observed with flat cages. Experiments are underway to test this prediction.

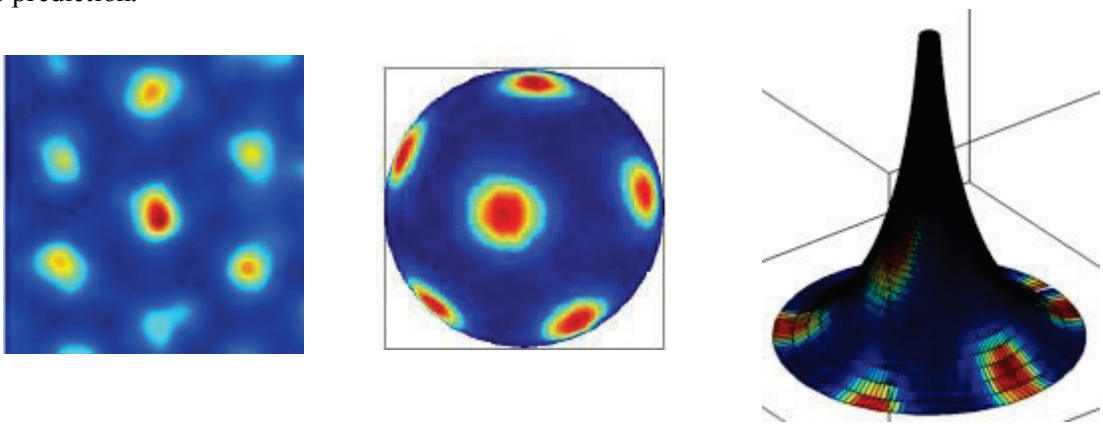


Figure 1 Grids may reflect environmental geometry
Examples of a real grid recorded in a rat in a flat box (left), of one simulated on a ball, with coordination number 5 (center, [5]), and of one simulated on a pseudosphere, with coordination number 7 (right).

Acknowledgment
Partial support from the SPACEBRAIN and then from the GRIDMAP EU projects. Extensive discussions with members of both collaborations, and the hospitality of the Ulanovsky lab at the Weizmann Institute of Science, are gratefully acknowledged.

[1] T. Hafting et al. Microstructure of a spatial map in the entorhinal cortex. Nature 436 (7052): 801-806 (2005). doi:10.1038/nature 03721.
[2] M. Yartsev et al., Grid cells without theta oscillations in the enthorinal cortex of bats. Nature 479 (7371): 103-107 (2011).
[3] E. Zilli. Models of grid cell spatial firing published 2005-2011. Frontiers Neural Circuits 6: art. 16 (2012). doi:10.3389/fncir.2012.00016.
[4] E. Kropff, A.Treves. The emergence of grid cells: intelligent design or just adaptation? Hippocampus 18:1256-1269 (2008).
[5] F. Stella, B. Si, E. Kropff, A. Treves. Grid cells on the ball. Journal of Statistical Mechanics, P03013 (2013), http://dx.doiorg/10.1088/1742-5468/2013/03/P03013.

Simple physics and bioinformatics of nucleosome positioning

Edward N. Trifonov

Institute of Evolution, University of Haifa, Israel

trifonov@research.haifa.ac.il

The problem of sequence-specific nucleosome positioning exists since the work of Trifonov and Sussman, 1980,when first evidence has been obtained that chromatin DNA has a hidden 10-11 base periodicity, apparently associated with DNA bending in the nucleosome. The observation was made by positional autocorrelation analysis – technique equivalent to the one normally used for extraction of weak signalsin time series. Since chromatin community has no background in signal processing,nor in physics, the field of nucleosome positioning suffered three decades of mistrust, confusion and misconceptions. One especially damaging wrong idea was mirror symmetry of the nucleosome positioning sequence pattern (Satchwell et al., 1986), mistaken for complementary dyad symmetry. It is still frequently referred to by massively confused chromatin community, although the physically correct picture has been described many times since 1980, as well as the sequence patterns consistent with the physics of DNA deformation. Recent discovery of strong nucleosome DNA sequences (Salih and Trifonov, 2013), with clearly visible rather than hidden sequence periodicity, puts an end to the misunderstanding. Deformation of DNA in the nucleosome is largely guided by the RR/YY dinucleotide stacks. The purine residues are harder to unstack, therefore they are placed towards DNA-histone interface, while pyrimidine-pyrimidine stacks are easier to deform and because of this are oriented outwards where deformation is the largest. This makes the alternation R_5Y_5 an ideal sequence for DNA bending in the nucleosome. In addition, the YR dinucleotides are known to be preferentially located in minor grooves of DNA facing the histone octamer. All previous techniques used for the extraction of the nucleosome DNA periodical pattern also ended with the dominating $(RRRRRYYYYY)_n$ sequence, since 1983, with 10 or 11 bases within parentheses, to satisfy average 10.4 base periodicity. The whole length nucleosome DNA positioning pattern oscillates with the period, also known since 1979 but accepted by the chromatin community only very recently. The whole length pattern allows accurate mapping of the nucleosomes along the sequences with single-base resolution, so that orientation and accessibility of sequence elements of interest on the surface of the nucleosome are also specified.

Genome and language – two scripts of heredity

Edward N. Trifonov

Institute of Evolution, University of Haifa, Israel

trifonov@research.haifa.ac.il

One striking manifestation of life is invention of linear script, twice in evolution, first as genomic sequences, and then – as language writings. Both possess their specific alphabets (4-letter and 20-letter for genetic texts, and 18 to 59 letters for languages). Both, of course, are converted to 0`s and 1`s in internal computer encoding of the texts, and in this form are indistinguishable. Question to theory of informatics: is the simple linear script the best possible way to store and transmit the information, so that in the hardships of evolution of species and H. sapiens specifically, it had to take over both in genomes and in languages? Evolution of both started, likely, from simple repetitions (TGTGTG…, GCCGCCGCC…, ma-ma-ma, da-da-da,…) (Frenkel and Trifonov, 2012; Trifonov and Zemkova, in preparation). Later the repetitions continued to enter the texts, but simultaneously they accumulated useful mutational changes, turning in more complex strings, already not recognizable as repeats. There are plenty of characteristics shared by the two texts of different(?) nature, such as frequency vocabularies, rules of "pronounceability" (alternations of polar and non-polar residues in proteins, and vowels/consonants in speech), contrast words, Shannon N-gram extension patterns, variation of linguistic complexity and more. There are, of course, some remarkable differences as well. One of them is overlapping character of the multiple codes in genomes, while such overlapping in languages is only rare, exotic phenomenon (like acrostics). Another important difference is strict rule of almost ideal identity in replication, while the identity in writings is, essentially, kept only for consecutive editions of canonical and authored texts. Diversity in human writings is, virtually, unlimited, while only certain degree of diversity is allowed in evolution of organisms. Questions to social sciences: May one consider language and the whole corpus of writings not just a cultural heredity but also as subject of biological evolution? Is there pressure of natural selection acting on ideas and canonical texts? Will the canonical texts (basic human knowledge, religious writings, theories) eventually crystallize in frozen invariants of all cultures? The tempting, perhaps, non-orthodox thought is: both types of texts are products and subjects of evolution, inseparable manifestations of life, both belonging to domain of biology.

Different statistical methods for calculations of amino acid propensities toward certain secondary protein structure types

D.Ž. Veljković[1], H. Grubor[1], S.N. Malkov[2] and S.D. Zarić[1, 3]

[1]*Department of Chemistry, University of Belgrade, Studentski trg 16, P.O.Box 158, 11000 Belgarde,Serbia*
[2]*Department of Mathematics, University of Belgrade, Studentski trg 16, 11000 Belgrade, Serbia*
[3]*Department of Chemistry, Texas A & M University at Qatar, P.O. Box 23874, Doha, Qatar*

vdusan@chem.bg.ac.rs

Understanding relation of primary and secondary structure of proteins is essential for predicting protein structure. The conformational preferences of amino acids are very important for understanding conformational interactions in proteins. Moreover, when used as propensities they can be helpful in predicting secondary and tertiary structures of proteins. Although a lot is known about secondary and tertiary protein structure and protein folding, there are still some open questions, like molecular mechanism of protein self-assembly which is not understood completely. Propensity for secondary structures represents an intrinsic property of amino acid, and it is now known that different amino acids have distinct propensities for the adoption of helical, strand, and random coil conformations. Several statistical studies were performed in order to calculate amino acid propensities [1,2,3].

In our previous work we carried out study of propensities using statistical method [4,5]. The calculated correlations of amino acids with secondary structure types from a large data set from Protein Data Bank enabled us to determine more clearly an amino acid`s tendency to participate in a particular secondary structure type. Based on the study clear preferences of amino acids towards certain secondary structures classify amino acids into four groups: α-helix preferrers, strand preferrers, turn and bend preferrers, and His and Cys (the latter two amino acids show no clear preference for any secondary structure). Amino acids in the same group have similar structural characteristics at their Cβ and Cγ atoms that predicts their preference for a particular secondary structure.

In this work we compare other statistical methods for amino acid propensities with the statistical method which we used in our previous work. Comparison of other statistical methods for calculation of amino acid propensities with statistical method which we used previously was made on the basis of correlation coefficients ($\rho(s,p)$). The results show that although methods are similar, there are some significant differences, resulting in a more explicit connection of our classification to amino acid chemical structure. Also, application of our statistical approach allows for stricter conclusions, without misjudgment on the amino acid's preferences.

Acknowledgment
This work was supported by Ministry of Education and Science of Republic of Serbia, grant No. 172065.

[1] P. Y. Chou, G. D. Fasman, Biochemistry 13 (2), 211 (1974).
[2] M. Levitt, Biochemistry 17, 4277 (1978).
[3] V. I. Lim, J. Mol. Biol. 873, 88 (1974).
[4] S. N. Malkov, M. V. Živković, M. V. Beljanski, M. B. Hall, S. D. Zarić, J. Mol. Model. 14, 769 (2008).
[5] S. N. Malkov, M. V. Živković, M. V. Beljanski, S. Đ. Stojanović, S. D. Zarić, The Protein Journal, 28, 74-86 (2009).

Kink solitons and breathers in microtubules

S. Kuzmanović[1] and S. Zdravković[2]

[1]*University of Priština, Kosovska Mitrovica, Serbia*
[2]*Vinča Institute, University of Belgrade, Belgrade, Serbia*

szdjidji@vinca.rs

Microtubules (MTs) are a major part of cytoskeleton and represent a network for motor proteins. They are holly cylinders formed by 13 long structures called protofilaments (PFs). Elementary units of PFs are long electric dipoles called dimers. Known models of MTs assume only one degree of freedom per dimer. According to the chosen coordinate the models can be called as longitudinal [1,2] and radial [3]. We start from the MT Hamiltonians which yield partial differential equations. Basically, two analytical approaches for solving these nonlinear equations are known. These are continuum [1-3] and semi-discrete approximations [4-6]. According to the first one kink and antikink solitons move along PFs. An example, coming from the longitudinal model [1], is shown in Fig. 1. The semi-discrete approximation brings about localized modulated waves usually called breathers. An example, describing the radial model [6], is shown in Fig. 2. In both cases ψ represent rescaled coordinates. The parameter σ is proportional to internal electric field strength and n denotes a position of the dimer in PF for an arbitrary time. The coordinate $\xi = \kappa x - \omega t$ is introduced to switch from the partial to the ordinary differential equation.

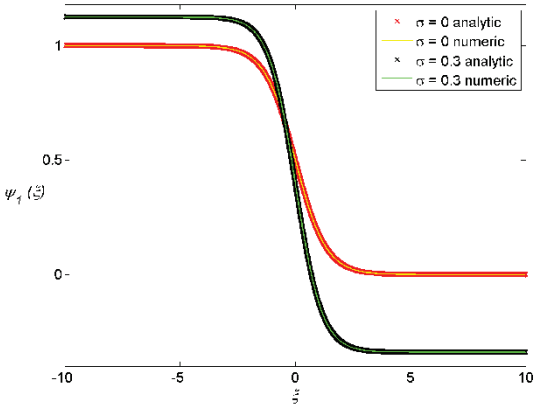


Fig. 1. The function $\psi_1(\xi)$

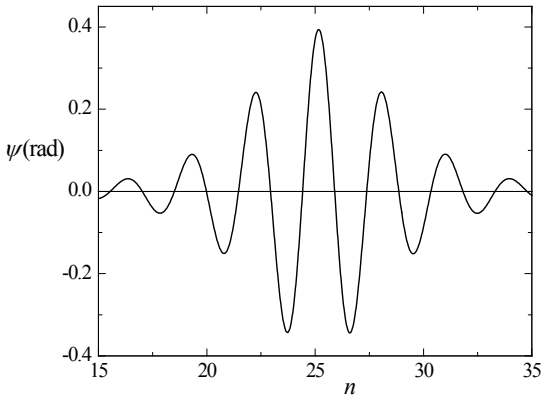


Fig. 2. The function $\psi(n)$ for $t = 3$ns

Acknowledgment
This research has been sponsored by the Ministry of Education, Science and Technology of the Republic of Serbia, grant no. III45010.

[1] S. Zdravković, L. Kavitha, M.V. Satarić, S. Zeković, J. Petrović, Chaos, Solitons Fract. 45, 1378 (2012).
[2] S. Zdravković, M.V. Satarić, S. Zeković, Europhys. Lett. 102, 38002 (2013).
[3] S. Zdravković, M.V. Satarić, A. Maluckov, A. Balaž, Submitted to Eur. Phys. J. E.
[4] M. Remoissenet, Phys. Rev. B 33, 2386 (1986).
[5] T. Kawahara, J. Phys. Soc. Japan 35, 1537 (1973).
[6] S. Zdravković, A.N. Bugay, Localized modulated waves in microtubules. Work in preparation.

Author Index