DYNALIFE

80, 70, 20 Conference
Towards Excellence and Convergence Research in Theoretical Biology

2 – 4 May 2023

Aula Magna Trentin

Ca' Foscari University

Ca' Dolfin, Venice · IT

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the European Union

Università Ca'Foscari Venezia
Department of Environmental Sciences, Informatics and Statistics

unibz
Freie Universität Bozen
Libera Università di Bolzano
Università Liedia de Bulsan

ABOUT DYNALIFE

In the mid-twentieth century two new scientific disciplines emerged forcefully: molecular biology and information-communication theory. At the beginning cross-fertilisation was so deep that the term genetic code was universally accepted for describing the meaning of triplets of mRNA (codons) as amino acids.However, today, such synergy has not take advantage of the vertiginous advances in the two disciplines and presents more challenges than answers. These challenges are not only of great theoretical relevance but also represent unavoidable milestones for next generation biology: from personalized genetic therapy and diagnosis, to artificial life, to the production of biologically active proteins. Moreover, the matter is intimately connected to a paradigm shift needed in theoretical biology, pioneered long time ago in Europe, and that requires combined contributions from disciplines well outside the biological realm. The use of information as a conceptual metaphor needs to be turned into quantitative and predictive models that can be tested empirically and integrated in a unified view. The successful achievement of these tasks requires a wide multidisciplinary approach, and Europe is uniquely placed to construct a world leading network to address such an endeavour. The aim of this Action is to connect involved research groups throughout Europe into a strong network that promotes innovative and high-impact multi and inter-disciplinary research and, at the same time, to develop a strong dissemination activity aimed at breaking the communication barriers between disciplines, at forming young researchers, and at bringing the field closer to a broad general audience.

Sequencing Technologies: Revolutionizing Biology from Theory to Practice with Nanopore Sequencing

Vesselin Baev[1], Mariyana Gozmanova[1], Elena Apostolova[1], Galina Yahubyan[1]

[1]*Faculty of Biology, University of Plovdiv, Tsar Assen 24, Plovdiv, 4000, Bulgaria*

e-mail: baev@uni-plovdiv.bg

Keywords: sequencing, genome, nanopore, ONT, workflow

The field of molecular biology has witnessed a remarkable revolution in recent years, largely due to the incredible advances made in sequencing technologies. Whole genome sequencing has become a key tool in understanding organisms' genetic makeup, behavior, and evolution. Nowadays, researchers are able to sequence and digitalize an organism's entire genome in a matter of hours or days, a feat that would have taken years just a few decades ago.

Sequencing technology advances have not only revolutionized the way we study biology but have also opened new avenues for theoretical biology to establish connections between the principles and mechanisms that govern living systems, starting from the molecular level, and reaching the ecological level. Vast amounts of genomic data have accumulated on organisms, allowing the development of new theoretical models that can explain the complex interactions between genes, cells, and organisms, as well as between organisms and their environment. By sequencing the microbiome, researchers can now understand how microorganisms interact with their hosts and how they can be harnessed to improve human health. This has opened up new avenues for the development of new therapeutic and diagnostic tools.

In this regard, Nanopore (ONT) sequencing technology is a fascinating innovation in the field of molecular biology that allows for rapid and real-time analysis of long-read DNA and RNA. Moreover, Nanopore MiniON devices are highly compact and easy to use, making sequencing technology accessible to every laboratory. Here, we present a workflow (wet lab and dry lab) for bacterial genome sequencing with ONT that goes from the DNA sample to fully digitalized annotated genome within 2 hours. The bioinformatics steps of the pipeline are implemented in the Galaxy framework, and the workflow can be easily shared and deployed among researchers.

Acknowledgements: We thank the COST Action DYNALIFE CA21169. This research was funded by the Bulgarian National Science Fund, grant number KP-06-N-36/3 (КП-06-Н-36/3). The authors are also supported by the project BG-RRP-2.004-0001-C01.

Simultaneous evolution of primitive coding systems

Paweł Błażej[1], Paweł Mackiewicz[1], Dorota Mackiewicz[1]

[1]*Department of Bioinformatics and Genomics, Faculty of Biotechnology, University of Wroclaw, Poland*

e-mail: pawel.blazej@uwr.edu.pl

Keywords: genetic code, amino acid, code evolution

It seems reasonable to assume that the present structure of the standard genetic code (SGC) has emerged from primitive coding systems which encoded a smaller number of amino acids and showed a high probability of translational errors. These systems had evolved concurrently in order to encode new genetic information and improved coding stability. During the evolution, some codes had disappeared at random or due to the inadequate level of coding accuracy. They could also be transformed into more complex coding systems. The SGC seems to be the final stage of this evolutionary process.

In our simulations, we investigated simultaneous evolution of different coding systems. The starting point was a population of genetic codes encoding a small number of different amino acids. They all evolved under simulation constraints such as the rate of mutation, new amino acid introduction and coding system exchange. The evolving codes are selected due to the accuracy of amino acid coding.

The results showed that the coding system which encode 20 amino acid and stop coding signal could emerge from a population of primitive genetic codes encoding only a few amino acids. The codes encoding the 21 labels are the best according to the assumed criteria and dominate in the evolving population, although they are not similar to the SGC in terms of structure.

Quantum noise may limit the mechanosensory sensitivity of cilia in the left--right organizer of the vertebrate bodyplan

Julyan Cartwright

[1]*Instituto Andaluz de Ciencias de la Tierra, CSIC-Universidad de Granada, 18100 Armilla, Granada, Spain*

[2]*Instituto Carlos I de Física Teórica y Computacional, Universidad de Granada, 18071 Granada, Spain*

e-mail: julyan.cartwright@csic.es

Keywords: cilia, left-right organizer, mechanosensing, quantum biology, symmetry breaking

Could nature be harnessing quantum mechanics in cilia to optimize the sensitivity of the mechanism of left--right symmetry breaking during development in vertebrates? I evaluate whether mechanosensing - i.e., the detection of a left-right asymmetric signal through mechanical stimulation of sensory cilia, as opposed to biochemical signalling - might be functioning in the embryonic left--right organizer of the vertebrate bodyplan through quantum mechanics. I conclude that there is a possible role for quantum biology in mechanosensing in cilia. The system may not be limited by classical thermal noise, but instead by quantum noise, with an amplification process providing active cooling.

About the Dissemination and Communication Plan for the COST Action DYNALIFE

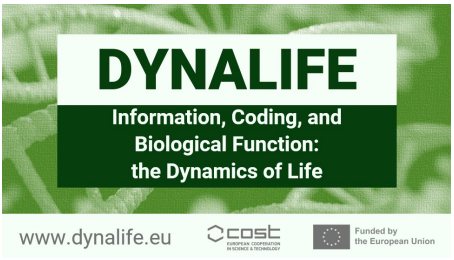
Jitka Čejková[1]

[1]*University of Chemistry and Technology Prague, Czech Republic*

e-mail: dynalife.cost@gmail.com

Keywords: Dynalife; dissemination; communication; network, collaboration.

This abstract outlines the dissemination and communication plan for the COST Action CA21169 INFORMATION, CODING, AND BIOLOGICAL FUNCTION: THE DYNAMICS OF LIFE (DYNALIFE). The plan includes the different tools, channels, and means of communication that will be implemented throughout the Action duration, as well as the target groups of the dissemination strategy. The plan aims to break down barriers between disciplines, stimulate collaboration, communicate important scientific results, and encourage new collaborations. Working Group 4 is dedicated to Dissemination and Communication and coordinates all activities with the Science Communication Coordinator, who works closely with the chair and vice-chair of the Action. The target groups for communication and dissemination include the scientific community, public and private research institutes and universities, the general public, potential industrial partners, hospitals, policy makers, media, and students. To achieve its objectives, the Action will create a panel for dissemination, promote communication and science-art crossover activities, and produce two short videos. It will also establish a forum for communication with potential industrial beneficiaries and stakeholders, organize seminars and conferences, and create digital content. The plan includes milestones and deliverables, including the production of video documentaries and reports on the Action's activities. Dissemination activities will be continuously monitored to ensure relevant information is shared with appropriate audiences by the most effective means.



On a model of energy and information transfer from the “donor” molecular structure to the molecular chain

Chevizovich D.[1], Matic V.[1]

[1]*I Laboratory for Theoretical and Condensed Matter Physics,* ”VINČA” Institute of Nuclear Sciences, National Institute of the Republic of Serbia, University of Belgrade P.O.Box 522, *11001 Belgrade, Serbia*

e-mail: cevizd@vinca.rs

Keywords: energy transfer, information transfer, biomolecular structures, self-trapping

It is thought that molecular chains (such as protein chains with alpha-helical secondary structure, DNA and RNA molecules) can play the role of "bridges" for the highly efficient transfer of vibron excitations or electrons over very long distances (comparable to the length of the molecular chain itself). Due to the interaction with the thermal oscillations of the structure, these excitations can be captured and can form a stable self-trapped (polaron-like) state, which can move through the structure with minimal energy loss. However, the properties of the possibly formed polaron must also be influenced by the presence of the donor molecule.

In the presented work, we discussed the mechanism of excitation transfer from one molecular structure (donor molecule) to the molecular chain. The presence of the donor structure and temperature influence on the energy of self-trapped excitation was considered, in the dependence of the basic energy parameters of the molecular bridge. The obtained results indicate the possibility of the formation of two types of self-trapped states: a quasi-free excitation that can easily move through the molecular bridge and a localized, practically immobile excitation, similar to a non-adiabatic polaron quasiparticle.

Statistical inferring of Rényi transfer entropy significance

Martina Chvosteková

Institute of Measurement Science of Slovak Academy of Sciences, Dúbravská cesta 9, 841 04 Bratislava 4, Slovakia

e-mail: martina.chvostekova@savba.sk

Keywords: causality, conditional mutual information, transfer entropy, Rényi entropy

Transfer entropy is a model-free method detecting the directionality of coupling (i.e., causal connection) from observed data based on information theory. The method involves estimating conditional mutual information. One possible way to write conditional mutual information is in the term of joint Shannon entropies. Transfer entropy is a special case of Rényi transfer entropy of order $\alpha > 0$ for which it holds $\alpha = 1$. Choosing values of α accentuates the causal connection of different probability events. There is only one formula for determining the Rényi transfer entropy based on binning estimation suitable for discrete time series. However, the Rényi transfer entropy estimates from the formula are biased for continuous data. The Rényi transfer entropy for the case of Gaussian variables does not depend on the value of α, i.e., it is equivalent to the transfer entropy. Rényi transfer entropy can be estimated by the k nearest neighbours' technique through the joint Shannon entropies for the system with Gaussian variables, which seems more suitable for continuous data. Surrogate data can be used to test the statistical significance of an estimated Rényi transfer entropy. We select some of the usually used ways to assess transfer entropies and compare them according to the number of correct detections of coupling directionality. We considered two types of synthetic data to properly evaluate the confidence and power of the significance tests of various transfer entropy estimates. We present the results of a simulation study for the case of Gaussian variables.

Acknowledgement: The work was supported by the Slovak Research and Development Agency, project APVV-21-0216, the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences, project VEGA 2/0023/22 and project VEGA 2/0096/21.

Application of molecular-bioinformatic approach in prediction, detection and prevention of epidemics caused by different pathogens: phylogeographic analysis

Valentina Cirkovic[1], Marina Siljic[2], Marko Jankovic[2], Maja Stanojevic[2]

[1] *Group for Medical Entomology, Institute for Medical Research, University of Belgrade, Dr Subotica 4, 11000 Belgrade, Serbia*
[2] *Faculty of Medicine, Institute of Microbiology and Immunology, University of Belgrade, Dr Subotica 1, 11000 Belgrade, Serbia*

e-mail: valentina.cirkovic@imi.bg.ac.rs

Keywords: phylogeography, pathogens, Bayesian analysis

Phylogeographic reconstruction of evolutionary history of different pathogens is aimed to predict the emergence of infectious diseases by detecting the key host species and the geographic areas from which pathogens spread and to analyse the impact of natural reservoirs movement on the spread of viral diseases. Spatiotemporal reconstructions can be assessed by two different phylogeographic approaches; phylogeographic inference in discrete space and phylogeography in continuous space and time. The Bayesian method implemented in the BEAST software package can be used to reconstruct the spatio-temporal evolutionary history. In addition to the appropriate evolutionary model, chosen based on the value of the AIC criteria in the jModelTest program, it is necessary to calibrate other parameters in the BEAST software (molecular clock, demographic model) to obtain the most reliable model for distribution analysis. The aim of this study was to reconstruct the phylogeny of West Nile virus linage 2 (WNV-2) and Usutu virus (USUV) on a spatio-temporal scale in order to estimate the time of origin and patterns of geographical dispersal of the isolates from Serbia.

In the present study we analysed NS5 segment sequences of WNV-2 and USUV in two separate datasets, isolated from different geographic areas with the aim to explore spatio-temporal dispersal of these viruses out of Africa. Phylogeographic analysis for both datasets was done in BEAST v1.10.4 software package using relaxed molecular clock and *Gaussian Markov random field* (GMRF) *Skyride* model. The convergence of parameters was assessed through the ESS>200 checked using Tracer v1.6. Phylogenetic tree was visualized using FigTree software v1.4.4.

Bayesian clade credibility (MCC) tree of the WNV-2 NS5 gene showed that majority of Serbian isolates were grouped in a single large clade together with sequences from Greece, thus forming large monophyletic clade. Phylogeographic analysis implied introduction of WNV-2 to Serbia from Hungary and further local spread. Phylogeny of USUV NS5 gene showed that new USUV strain from Serbia belongs to a heterogeneous Central European cluster also consisting of strains from Italy, Hungary, Croatia, the Czech Republic, and Germany. In addition, phylogeographic analysis conformed that USUV entered to Serbia from Italy.

Our results evidence presence and circulation of analyzed pathogens in Serbia at least a decade prior to the first evidenced outbreak of WNV-2 in 2012 and demonstrate the usefulness of bioinformatic methods providing a comprehensive view of the evolutionary history of these pathogens.

New directions for artificial molecular machines and motors

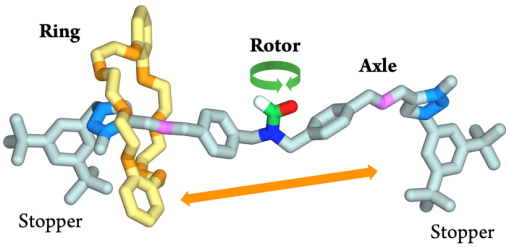
Alberto Credi

Center for Light Activated Nanostructures (CLAN), Dipartimento di Chimica Industriale "Toso Montanari", Alma Mater Studiorum - Università di Bologna, and Consiglio Nazionale delle Ricerche, Bologna, Italy.

e-mail: alberto.credi@unibo.it

The construction of molecular scale devices and machines[1] have formidably stimulated the creativity of chemists in the past three decades.[2,3] The interest on this kind of systems arises from their ability to perform a (useful) function in response to chemical and/or physical signals (e.g., light). Mechanically interlocked molecules exhibit appealing structural and functional properties for the construction of nanoscale devices and machines; molecular shuttles based on rotaxanes constitute common examples.[3]

Here I will describe investigations undertaken in our laboratories, aimed at inducing and controlling nanoscale movements in rotaxanes and related species to perform functions such as transmitting motion between sites[4] (see Figure) and activating mechanically chiral structures for enantioselective guest recognition.[5] From a fundamental viewpoint these systems behave as molecular switches under thermodynamic control. Appropriately designed architectures, however, can exploit an energy harvesting process to operate away from thermodynamic equilibrium.[6] Moreover, by exploiting energy and/or information ratcheting effects, directional and autonomous movement of the molecular components can occur.[1-3] We have combined this strategy with a minimalist chemical design to realize artificial nanoscale pumps powered by light[7] and electricity.[8] Besides their interest for fundamental science, these systems have the potential to bring about radical innovation in catalysis, materials science, energy conversion, robotics and medicine.[9]



Acknowledgements: Support from the European Union (H2020 ERC AdG 'Leaps' 692981, FET-OPEN 'Magnify' 801378, ITN 'ArtMoMa' 860434) and the Ministero dell'Università e Ricerca (Grants 20173L7W8K and R16S9XXKX3) is gratefully acknowledged.

References:

[1] A. Credi, V. Balzani, *Molecular Machines*, 1088 Press, Bologna, 2020. Available for download at 1088press.it
[2] V. Balzani, A. Credi, M. Venturi, *Molecular Devices and Machines – Concepts and Perspectives for the Nano World*, Wiley-VCH, Weinheim, 2008. C. Bruns, J. F. Stoddart, *The Nature of the Mechanical Bond: From Molecules to Machines*, Wiley, Hoboken, 2016.
[3] M. Baroncini *et al.*, *Chem. Rev.* **2020**, *120*, 200.
[4] S. Corra *et al.*, *Chem* **2021**, *7*, 2137.
[5] S. Corra *et al.*, *J. Am. Chem. Soc.* **2019**, *141*, 9129.
[6] F. Nicoli *et al.*, *J. Am. Chem. Soc.* **2022**, *144*, 10180.
[7] G. Ragazzon *et al.*, *Nat. Nanotechnol.* **2015**, *10*, 70. A. Sabatino *et al.*, *Angew. Chem. Int. Ed.* **2019**, *58*, 14341. J. Groppi *et al.*, *Angew. Chem. Int. Ed.* **2020**, *59*, 14825. S. Corra *et al.*, *Chem. Eur. J.* **2021**, *27*, 11076. M. Canton *et al.*, *J. Am. Chem. Soc.* **2021**, *143*, 10890. S. Corra *et al.*, *Nat. Nanotechnol.* **2022**, *17*, 746.
[8] G. Ragazzon *et al.*, *Angew. Chem. Int. Ed.* **2023**, *62*, e202214265.
[9] S. Corra *et al., Adv. Mater.* **2020**, *32*, 1906064.

Epigenetic contribution to the Zea mays L. cold and drought memory

Dudić Dragana[1]

[1]Faculty of Informatics and Computer Science, Union-Nikola Tesla University, Cara Dušana 62-64, Belgrade, Serbia

e-mail: ddudic@unionnikolatesla.edu.rs

Keywords: RNAseq, Zea mays L., DNA methylation, bisulfite sequencing

Chromatin and DNA modifications are epigenetic markers that have influence on gene expression and they are essential for the responses to the environment. Because the germline of plants is formed during the late development, the epigenetic markers are potentially heritable. Zea mays L. is a worldwide important crop with its yield negatively affected by a wide range of abiotic stresses and determination of the phenotypic and epigenetic sources that could lead to the stabilization and incensement of Zea mays L. yield is crucial. Our ongoing study involves publicly available data for 46 inbred lines of Zea mays L. Until now, we conducted leaf whole transcriptome sequencing and complete RNAseq analysis for all 46 inbred lines and identified significant differentially expressed genes (DEGs) and single nucleotide polymorphisms (SNPs) in nuclear, plastid and mitochondrial parts of Zea mays L. genome. We identified 77 DEGs between Lancaster and non-Lancaster groups grown under optimal conditions. Cold and drought tests are conducted on the two subsets of four Lancaster and four non-Lancaster lines. The first subset was formed of cold tolerant and cold sensitive, and six (chloroplast ATP-sulfurylase, photosystem II cytochrome b559 alpha subunit, CIPK serine-threonine protein kinase 15, glutamyl-tRNA reductase, photosystem II reaction center protein I and Calvin cycle CP12 - chloroplastic-like encoding genes) out of seven selected DEGs are marked as significant for cold response in Zea mays L. The second subset was formed from drought tolerant and drought sensitive lines and detection of significant DGEs is an ongoing project and we have several DGEs under the investigation. In order to determine the effect of the epigenetic variation to the cold and drought resistance ability, we plan to extend our study with whole genome bisulfite sequencing of eight Lancaster and eight non-Lancaster to test if the epigenetic variation contributes to the memory of the adaptation of Zea mays L. We will detect differentially methylated regions (DMRs), significant SNPs, and transposable element (TE) insertions and associate DMRs with SNPs and TEs. Also, we plan to repeat the experiment for several generations in order to test the plant adaptation memory.

Joint modelling of multiple omics datasets and outcome variables

Jeanine Houwing-Duistermaat[1,2], Zhujie Gu[3,4], Hae-Won Uh[4], Said el Bouhaddani[4]

[1]Dept. of Mathematics, Radboud University Nijmegen, Netherlands
[2]Dept of Statistics, University of Leeds, UK
[3]Medical Research Council Biostatistics Unit, University of Cambridge, UK
[4]Dept. of Data science & Biostatistics, University Medical Center Utrecht, Netherlands

e-mail: Jeanine.Duistermaat@ru.nl

Keywords: Partial least squares; Generalized linear models; Data integration; Omics

The biological mechanisms underlying human diseases are often complex. Diverse omic datasets represent various aspects of these mechanisms in relation to outcomes. Recent advances in high-throughput technologies have made it affordable to measure these omic levels for many studies. Typically, the relationships between these datasets and an outcome variable are studied one-by-one. Although these studies provide biological insights of diseases on a single omic level, they ignore correlations among the omic levels. Several data integration methods exist; hence one might consider to first integrate the datasets and then in a second stage study the relationship between relevant features and the outcome variable. A drawback of such an approach is that the correlation between the outcome and the omics variables is not considered in the first step and that the uncertainty in the data integration is ignored. We propose a new model for two omic datasets and an outcome variable, where the relationship of the omic datasets with the outcome is modelled via the joint parts of the omic datasets.

Our model builds upon our recently developed PO2PLS model, which considers for each omic dataset a latent variable model. In the PO2PLS model, a subset of the latent variables of the two datasets are linked and represent the joint part. Further data-specific latent variables are included to model the differences between the two datasets. Here, we add a submodel for the outcome by considering the latent variables of the joint part as independent variables in a generalized linear model for the outcome variable. Gaussian distributions are used for the latent variables. The parameters are estimated using an EM algorithm. The asymptotic behaviour is studied. Extensive simulation studies have been performed to study the performance of the model under various settings and to compare the method with existing approaches. Finally, the approach is illustrated by an application in which the joint effect of methylation and glycomics is modelled on Down Syndrome cases and their relatives.

We conclude that while the estimation of the new proposed model is more computational intensive than existing methods, it performs well and can provide more insight into underlying biological mechanisms.

Statistical integration of multi-omics and drug screening data from cell lines

Said el Bouhaddani[1], Hae-Won Uh[1], Jeanine Houwing-Duistermaat[2]

[1]*Dept. Data science & Biostatistics, UMC Utrecht, Heidelberglaan 100, 3584CX, Netherlands*
[2]*Dept. Mathematics, Radboud University Nijmegen, Heyendaalseweg 135, 6525AJ, Netherlands*

e-mail: s.elbouhaddani@umcutrecht.nl

Keywords: POPLS-DA, data integration, FDA drug screen, gene-gene interaction network

The collection of data on multiple molecular levels has become a fundamental aspect of modern biological research. Examples are cell line studies, where multiple omics datasets are measured, as well as interventional data such as high-throughput (drug) screens. A holistic "integrative" approach involving all data can provide new insights in biological systems and interactions.
Our motivational example is a study of multiple system atrophy (MSA) and Parkinson's disease (PD) utilizing cell lines. The aim was to better understand the molecular basis of MSA and PD and find potential disease-modifying drugs. To this end, transcriptomics, proteomics and 1600 FDA-approved drug screening data are measured in human brain cell lines under a disease-inducing and control environment. Challenges include high-dimensionality, non-overlapping datasets, strong correlations between measurements, and heterogeneity across the multi-omics and screening datasets. For an integrative analysis of these data, novel statistical data integration approaches are needed.
We propose a novel statistical data integration workflow to obtain a list of genes and proteins that can best distinguish disease cell lines from controls across all omics datasets and, within this list, identify targets of potential disease-modifying drugs. For the first part of the workflow, we develop Probabilistic OPLS discriminant analysis, POPLS-DA, to model the multi-omics datasets in terms of joint, omics-specific, and residual components. These components consist of weighted linear combinations of genes and proteins. The outcome is included in the model via the joint components to obtain genes and proteins that best distinguish cases and controls across all omics data. All POPLS-DA parameters are simultaneously estimated with maximum likelihood using a memory-efficient EM algorithm. Based on the top genes and proteins, a gene-gene interaction network is built using String-DB. For the second part, we propose a 'direct neighbor' approach to integrate the screening data with the interaction network: we use DrugBank to obtain all interactors of the validated drugs from the drug screening and intersect these with our list of top genes and proteins. An extensive simulation study will be conducted to investigate the performance of POPLS-DA. We apply our approach to transcriptomics, proteomics and screening data measured in the cell lines and construct an integrated interaction network based on all data. This network will highlight a druggable subnetwork that can potentially be used in a novel therapy for MSA and PD.

Codon usage in rare disease genes shows evolution- and phenotype-driven codon bias fingerprints

Alessandra Ferlini

Unit of Medical Genetics, University of Ferrara, Italy

e-mail: fla@unife.it

Keywords: 3-7 words: codon usage, CUB , rare disease genes, codon optimization

We calculated codon usage (CU) and defined codon usage bias (CUB) in 6 small cohorts of human genes: non-disease-causing genes (N=29, NDC) and disease-causing genes (N=31, DC). All genes are highly expressed in 3 distinct human tissues, kidney, muscle, and skin. We also compared CU across species in 15 mammals. We obtained CUB hierarchical clusters for each gene cohort which showed tissue-specific and disease-specific CUB fingerprints. We showed that DC genes (especially those expressed in muscle) display a low CUB, well recognizable in codon hierarchical clustering. We defined the extremely biased codons as “zero codons” and found that their number is significantly higher in all DC genes, all tissues, and that this trend is conserved across mammals. Based on this calculation in different gene cohorts, we identified 5 codons which are more differentially used across genes and mammals, underlining that some genes have favorite synonymous codons in use. Since of the muscle genes clear clusters, and, among these, dystrophin gene *surprisingly* does not show any “zero codon” we adopted a novel approach to study CUB, we called “mapping-on-codons”. We positioned 2828 dystrophin missense and nonsense pathogenic variations on their respective codon, highlighting that its frequency and occurrence is not dependent on the CU values. We conclude our strategy consents to identify a hierarchical clustering of CU values in a gene cohort-specific fingerprints, with recognizable trend across mammals. In DC muscle genes also a disease-related fingerprint can be observed, allowing discrimination between DC and NDC genes. We propose that using our strategy which studies CU in specific gene cohorts, as rare disease genes, and tissue specific genes, may provide novel information about the CUB role in human and medical genetics, with implications on synonymous variations interpretation and codon optimization algorithms.

Sparse high-dimensional covariance matrix estimation by composite likelihood truncation with applications to large-scale gene association recovery

Alessandro Casa[1], Davide Ferrari[1], Zhendong Huang[2]

[1]*Free University of Bolzano, via Universita' 1, 39100, Bolzano, Italy*
[2]*University of Melbourne, Parkville, Victoria, 3010, Australia*

e-mail: davferrari@unibz.it

Keywords: High-dimensional covariance, composite likelihood estimation, model selection, gene regulatory networks.

A composite likelihood is a combination of low-dimensional likelihood objects useful in applications where the data have complex structure. Although composite likelihood construction is a crucial aspect influencing both computing and statistical properties of the resulting estimator, currently there does not seem to exist a universal rule to combine low-dimensional likelihood objects that is statistically justified and fast in execution. We develop a methodology to estimate high-dimensional ans sparse covariance matrices and combine the most informative low-dimensional likelihoods from a large set of candidates while carrying out parameter estimation. The new procedure minimizes the distance between composite likelihood and full likelihood scores subject to a constraint representing the afforded computing cost. The selected composite likelihood is sparse in the sense that it contains a relatively small number of informative sub-likelihoods while the noisy terms are dropped. The resulting estimator is found to have asymptotic variance close to that of the minimum-variance estimator constructed using all the low-dimensional likelihoods. The procedure is applied to reconstruction of complex gene association networks through the analysis of gene expression data.

Genetic Code Modelling from the Perspective of Quantum Informatics

Elena Fimmel

1 Mannheim University of Applied Sciences, Paul-Wittsack-Str. 10, 68163 Mannheim, Germany

e-mail: e.fimmel@hs-mannheim.de

Keywords: quantum informatics, genetic code

This talk aims to show the possibilities of modelling the information content carried by quantum mechanical DNA-molecules by means of the formalism used in quantum informatics. Such modelling would open new options to reveal nature's information patents and to use them, for instance, in quantum computing and artificial intelligence (A.I.) Moreover, it would give an opportunity to understand the ways of managing information in living organisms. As an empirical base, the open accessible data from GenBank which contains hundreds of millions of long DNA texts collected from thousands of organisms can be used.

The Statistical Mechanics of Self-Assembly

Achille Giacometti[1,2]

[1]*Dipartimento di Scienze Molecolari e Nanosistemi, Universita' Ca' Foscari Venezia*
[2]*European Centre for Living Technology (ECLT)*

e-mail: achille.giacometti@unive.it

Self-Assembly processes are ubiquitous in biological systems, likely as a result of an evolutional pressure that has optimized it resulting in life-as-we-know-it. In this talk I will discuss two paradigmatic examples of such processes. First I will describe an example of entropically driven self-assembly in liquid crystals with a particular emphasis on how the phase chirality is related to molecular chirality. Within these framework, I will show how the experimental liquid-crystal phase behaviour of helical flagella can be rationalized on the basis of these findings. The second example deals with the folding and aggregation of semiflexible polymers in solution and how the resulting phase behaviour depends upon thermodynamic parameters such as temperature, bending rigidity, excluded volume, as well as range of the interactions with the solvent. The importance of these findings for biological processes will be finally emphasized.
In addition to this scientific contribution, I will also briefly introduce the European Centre for Living Technology (ECLT), a network of 18 different international institutions dealing with complex systems, as well as the achievements of the EUTOPIA COST Action which shares with DYNALIFE many common interests.

A role for circular code properties in translation

Simone Giannerini[1], Diego Luis Gonzalez[2,1], Greta Goracci[3], Alberto Danielli[4]

[1]*Department of Statistical Sciences, University of Bologna, Italy*
[2]*CNR-IMM, Institute for Microelectronics and Microsystems, Bologna Unit, Italy*
[3]*Free University of Bolzano, Faculty of Economics and Management, Italia*
[4]*Department of Pharmacy and Biotechnology, University of Bologna, Italy.*

e-mail: simone.giannerini@unibo.it

Keywords: Circular codes; Protein expression; Translation speed; Codon usage; Reading frame; Symmetries

Circular codes represent a form of coding allowing detection/correction of frame-shift errors. Building on recent theoretical advances on circular codes, we provide evidence that protein coding sequences exhibit in-frame circular code marks, that are absent in introns and are intimately linked to the keto-amino transformation of codon bases. These properties strongly correlate with translation speed, codon influence and protein expression levels. Strikingly, circular code marks are absent at the beginning of coding sequences, but stably occur 40 codons after the initiator codon, hinting at the translation elongation process. Finally, we use the lens of circular codes to show that codon influence on translation correlates with the strong-weak dichotomy of the first two bases of the codon. The results provide promising universal tools for sequence indicators and sequence optimization for bioinformatics and biotechnological applications, and can shed light on the molecular mechanisms behind the decoding process.

Oxidoreductases, trace elements and the evolution of biogeochemistry

Donato Giovannelli[1]

[1]*Department of Biology*
University of Naples "Federico II"

e-mail: donato.giovannelli@unina.it

Keywords: Oxidoreductases

Earth's geosphere and biosphere have coevolved over time, influencing each other's stability and keeping our planet habitable for most of its 4.543 billion years of history. Biogeochemical cycles play a key role in controlling this interaction, connecting long-term geological cycles and the much faster evolution of the Earth's outer biologically dominated envelopes. A small set of microbial-encoded proteins containing redox-sensitive transition metals as their core catalytic center carry out the majority of the key biogeochemical reactions. Metals such as Fe, Co, Ni, Zn, Mo, W, V, and Cu are used in these proteins to access diverse redox couples as a function of the changing planetary availability of these elements over time. Despite the importance of this process, the relationship between metal availability and metabolism evolution and diversity has not been investigated in detail. I pose that elucidating the impact of transition metal availability on microbial functional diversity holds the key to understanding the co-evolution of life and our planet, and a key transition in evolution has been coding for the information necessary to control cofactor assembly and redox chemistry.

Why DYNALIFE?

Diego Luis Gonzalez[1,2]

[1]*Institute for Microelectronic and Microsystems - CNR, 40129 Bologna, Italy*
[2]*Department of Statistical Sciences - UNIBO, 40126 Bologna, Italy*

e-mail: gonzalez@bo.imm.cnr.it

Keywords: DYNALIFE, Theoretical Biology, Convergence Research

In this contribution will be given a brief historical introduction about theoretical biology and several previous attempts to inform biology with the hard sciences for developing a quantitative paradigm.

The seminal attempts of Erwin Schrodinger and Conrad Hal Waddington, that are commemorated in 2023 and shaped the title of the Venice Conference, will be recalled together with other milestones in theoretical biology and in molecular biology.

The reasons because past attempts failed together with the scientific novelties that allow to think that this situation can effectively change, will be analysed.

Between these last, particular emphasis will be given to modern developments regarding information theory, modelling of complex systems, and computer science.

Finally, the main challenge of DYNALIFE will be described together with the general strategy envisaged for attaining such goal, that includes aspects related to convergence (research beyond interdisciplinarity), excellence of research products, and formation of young researchers in this new promissory field.

Dichotomic Classes and Entropy Optimization in Coding Sequences

Simone Giannerini[1], Diego L. Gonzalez[1,2], Greta Goracci[3]

[1]Department of Statistical Sciences, University of Bologna, Italy
[2]IMM-CNR, Bologna Unit, Italy
[3]Faculty of Economics and Management, Free University of Bozen-Bolzano, Italy

e-mail: greta.goracci@unibz.it

Keywords: Non-power model of the genetic code; Coding sequences; Dichotomic classes; Entropy measures; Information theory;

In this work we investigate the existence of universal optimizations in coding sequences. Usually, these take the form of correlations between nucleotides that are observed in many organisms and that might be related to error correction and energy optimization. We address the problem in terms of a mathematical model of the genetic code introduced in Gonzalez(2008) and further studied in Gonzalez et al.(2008, 2009, 2016), Giannerini et al.(2012). This new paradigm leads to the definition of dichotomic classes that can be seen as nonlinear functions of the information contained in a dinucleotide. Such classes represent precise biochemical interactions and are used as a binary coding scheme for DNA sequences. We study the entropy structure of dichotomic classes derived from coding sequences under different probabilistic assumptions on the underlying process. The theoretical derivations are tested by using the new probabilistic results of Giannerini and Goracci (2018) on the small sample asymptotics for the entropy measures used. The results indicate that the paradigms of information/communication theory are essential for the understanding of the organization of genetic information.

References:

[1] D. L. Gonzalez. The mathematical structure of the genetic code. In M. Barbieri editor, The Codes of Life: The Rules of Macroevolution, volume 1 of Biosemiotics, chapter 8, pages 111–152. Springer Netherlands, 2008.
[2] D. L. Gonzalez, S. Giannerini, and R. Rosa. Strong short-range correlations and dichotomic codon classes in coding DNA sequences. Physical Review E, 78(5):051918, 2008.
[3] D. L. Gonzalez, S. Giannerini, and R. Rosa. The mathematical structure of the genetic code: a tool for inquiring on the origin of life. Statistica, LXIX(3–4):143–157, 2009.
[4] D. L. Gonzalez, S. Giannerini, and R. Rosa. The non-power model of the genetic code: a paradigm for interpreting genomic information. Phil. Trans. R. Soc. A 2016 374 20150062.
[5] S. Giannerini, D. L. Gonzalez, and R. Rosa. DNA, frame synchronization and dichotomic classes: a quasicrystal framework. Phil. Trans. R. Soc. A, Vol. 370, Number 1969, 2987–3006, 2012.

Computing with a network of interacting chemical oscillators.

Jerzy Gorecki

Institute of Physical Chemistry
Polish Academy of Sciences
Kasprzaka 44/52, 01-224 Warsaw, Poland

e-mail: jgorecki@ichf.edu.pl

Keywords: chemical computing, networks, oscillations

We have observed unprecedented success of semiconductor computing technology in recent years, and the Moore law reflects it. On the other hand, artificial chemical computing based on chemical reactions is still at a very early stage of development. Obviously, animals and humans can process a considerable amount of information and solve complex problems at low energy costs with their nervous systems and brains. In my opinion, the bottom-up design strategy from logic gates to complex information processing devices is one of the factors that slow down the progress of chemical computing technology because it does not allow for the optimum use of the computing potential of chemical media.
In my presentation, I follow the idea of computation with reaction networks proposed 30 years ago by Ross and collaborators. Contrary to the earlier work, I focus attention on reaction nodes demonstrating highly nonlinear behavior such as chemical oscillation [1]. I present the idea of a computing network, define how the input information can be introduced, and discuss how the output information can be extracted. For a specific problem, the parameters of a computing network can be optimized using a top-down design. A few examples of networks solving academic and real-life problems will be shown [2].

References:
[1] Jerzy Gorecki, Information Processing Using Networks of Chemical Oscillators, Entropy 2022, 24(8), 1054; https://doi.org/10.3390/e24081054
[2] Gizynski K.; Gorecki J., Cancer classification with a network of chemical oscillators, Physical Chemistry Chemical Physics, 2017, 19, 28808–28819.
[3] Bose A.; Dittrich P.; Gorecki J., The Concilium of Information Processing Networks of Chemical Oscillators for Determining Drug Response in Patients With Multiple Myeloma. Front. Chem., 2022, 10; doi: 10.3389/fchem.2022.901918

Analysis of cyclic properties of networks with Ihara's zeta function: Signaling networks in biological processes

Clara Grácio

University of Évora, Largo dos Colegiais, Nº 2, 7004-516 Évora, Portugal

e-mail: mgracio@uevora.pt

Keywords: Ihara function; cyclic structure, biological networks.

Network representations are good tools for characterizing interaction patterns between the constituents of large complex biological systems. They translate and simplify complexity of such systems to the topological properties of their associated graphs. The structure of existing cycles in complex networks strongly affects the behavior of processes supported by these networks. In this work, we examine the cyclic properties of networks and, in particular the use of Ihara's zeta function for counting cycles in networks.

On application of computational methods in network-based biological research

Milana Grbić[1], Dragan Matić[1],

[1]*Faculty of Natural Sciences and Mathematics, University of Banja Luka, Mladena Stojanovića 2, Bosnia and Herzegovina*

e-mail: milana.grbic@pmf.unibl.org

Keywords: PPI networks, protein complexes, variable neighbourhood search, communitydetection, graph similarity

In one of his lectures, Paul Nurse stated "*If we can understand cells, then we are very close to understand life.*" Proteins are responsible for many process in the cell. Proteins which mutually interact at the same time and at the same place make specific groups – *protein complexes*. In computational studies, proteins and interaction between proteins are represented as Protein – Protein Interaction (PPI) networks. In a PPI network, nodes are proteins and the edge (link) between two nodes exists if the corresponding proteins interact. In some PPI networks, the existence of PPIs is based on curated biological data, while the others combine information retrieved from different sources, including the findings obtained by various computational methods. In that sense, development and application of computational algorithms, which can confirm the existing or predict new PPIs is of a crucial importance for better understanding processes in the cell.

We analyse the relationship between PPI networks and protein complexes. In literature, there is an assumption that protein complexes usually form dense connected subnetworks in a PPI network. On the other hand, some recent researches show that this is not always the case. One of the problems considered in this research is adding the minimal number of PPIs in a PPI network in order to connect considered protein complexes. Another important question is identification of protein complexes in PPI networks. This problem may be connected to the well-known problem of community detection. We analyse capabilities of various community detection algorithms to identify protein complexes in PPI networks. In the process of protein complex identification, it is important to determine which proteins make a ‘’core” of a protein complex. In order to answer on this question, local properties, like node centrality and similarity of subnetworks induced by core proteins, may be of use.

Parameterization and validation of an accurate force-field for molecular dynamics simulations of biomolecular systems

Sonja Grubišić

University of Belgrade, Institute of Chemistry, Technology, and Metallurgy, National Institute of Republic of Serbia, Njegoševa 12, Belgrade, 11000 Serbia

e-mail: sonja.grubisic@ihtm.bg.ac.rs

Keywords: Amber Force field, Amino acids, α-aminoisobutyric acid (Aib), Molecular Dynamics, TOAC amino acid

Proper simulation of the structure and dynamics of large molecular systems rests on accurate and reliable force fields (FF) which can reproduce their behaviour both in the gas phase and in condensed phases at a reasonable computational cost. α,α-Dialkylated amino acid residues have gained significant importance as an effective way for introducing backbone conformation constraints in synthetic peptides. As it is well known, the behaviour of such rigid amino acids differs substantially from that of standard mono-substituted α-amino acids, and correspondingly their description at the atomistic level is not entirely obvious. This talk is concerned with the results of the development and validation of force field (FF) parameters for accurate description of aminoxyl radicals (nitroxide probes) incorporated in peptides/proteins. Recently, force field for the study of large nitroxides in condensed phases has been developed [1]. In addition, the popular biomolecular AMBER force field (FF) has been extended with new parameters for the simulations of peptides containing α,α dialkylated residues with cyclic and linear side chains [2-5]. Together with the recent set of nitroxide parameters this extension allows treating the TOAC residue (TOAC, 2,2,6,6-tetramethylpiperidine-1-oxyl-4-amino-4-carboxylic acid) widely used as a spin label in protein studies. The residue is stable under ordinary conditions, has a well localized unpaired electron, and can provide access to useful conformational information through well-established magneto-structural relationships. By comparison between simulated and available experimental data (EPR), we have demonstrated the reliability and accuracy of the force field for predicting stable secondary structures.

A computer analysis of the wobble-effect and its impact on genetic code variations optimized for the robustness against point mutations

Markus Gumbel[1] (joined work with Elena Fimmel[1], Martin Starman[1] and Lutz Strüngmann[1])

[1]*Center for Algorithmic and Mathematical Methods in Medicine, Biology, and Biotechnology, Mannheim University of Applied Sciences, 68163 Mannheim, Germany*

e-mail: m.gumbel@hs-mannheim.de

Keywords: genetic code; point mutations; wobble effect; evolutionary algorithm

The assignments of codons to amino acids within the standard genetic code might facilitate to reduce the problems caused by point mutations. The code's robustness to point mutations can be described using the so-called conductance measure [1] - a weighted graph-based method. We analyze the impact of the wobble effect on genetic code tables and search for the optimal robustness using an evolutionary algorithm [2, 3] that optimizes the weights of the conductance graph. We show that the robustness is least affected by mutations in the third position, as in the wobble effect. The results clearly show that point mutations in the first and, more importantly, in the second base of a codon have a very large impact on the robustness of the genetic code. These results are placed in the context of single nucleotide variants (SNV) in coding sequences. This talk will discuss what structure of a genetic code evolves from random code tables when robustness is maximized. The results suggest that the evolving code tables are very similar to the standard genetic code and that robustness to point mutations appears to be an important factor in the evolution of the standard genetic code.

References:

[1] Błażej, P., Kowalski, D.R., Mackiewicz, D., Wnetrzak, M., Aloqalaa, D.A., Mackiewicz, P., 2018. The structure of the genetic code as an optimal graph clustering problem (preprint). bioRxiv. https://doi.org/10.1101/332478

[2] Fimmel, E., Gumbel, M., Starman, M., Strüngmann, L., 2021. Robustness against point mutations of genetic code extensions under consideration of wobble-like effects. Biosystems 208, 104485. https://doi.org/10.1016/j.biosystems.2021.104485

[3] E. Fimmel, M. Gumbel, M. Starman, L. Strüngmann: Computational Analysis of Genetic Code Variations Optimized for the Robustness against Point Mutations with Wobble-Like Effects. Life (2021), 11, 1338. https://doi.org/10.3390/life11121338

Protocells and Information

Silvia Holler

[1]*CIBIO, UniTN, Trento, Italy*

e-mail: silvia.holler@unitn.it

Keywords: protocells, droplets, DNA, information

Information is hidden in all the entities that surround us in everyday life. It can be found in cells and in their genetic code, but also in varying properties and/or behaviours of other types of systems. An exploratory study is performed where information are linked with protocells systems. Protocells are specifically coupled with DNA and organized depending on its quantity and base pairing. Population aggregation allows the collection of chemical polymers. Competing base pairing also guarantees the population disaggregation and the release of compounds hidden in the aggregates. Unconventional type of information will also be presented. They will be directly related to the variation of physio chemical properties of protocells. Many of the emergent phenomena of life may have arisen within classes of physio-chemical systems that can be composed of a diverse range of material going from: vesicles, oil droplets, cellular automata and reaction-diffusion systems. Given the stark differences in composition between these systems, it is clear that life like behaviors are not unique to human life but occur in many scenarios: information can be found and stored in all of them.

A marriage between Biology and the Environmental Sciences

Herbert E. Huppert Sc. D. FRS

Cambridge University, United Kingdom

e-mail: heh1@cam.ac.uk

I have worked extensively on applying various aspects of fluid mechanics to understand the world in which we live. Thus I have written papers on atmospheric motions and been involved in CCS (Carbon Capture and Storage) and was the chair of the special European Committee reporting to the European Commissioners on CCS. I have also written numerous papers on oceanic motions and at the moment am particularly interested in the importance of ocean waves over coral layers, being responsible for biological life within corals. I have worked extensively on geologically oriented problems, including the flow of magma through the Earth's crust, its explosive eruption into the atmosphere, endangering birds life of birds as well as influencing human activity. Understanding these processes has suggested to me a way of defending against tsunamis, which in the past have caused enormous loss of life and money. I would describe this work briefly and discuss the biological implications.

Statistical challenges and solutions in human microbiome data analysis

Eliana Ibrahimi[1], Alise Ponsero[2]

[1] *University of Tirana, Bulv Zogu I 25/1, Tirana, Albania*
[2] *University of Helsinki, Haartmaninkatu 3, Helsinki, Finland*

e-mail: eliana.ibrahimi@fshn.edu.al

Keywords: Microbiome data, statistical methods, data transformation, sparsity.

The human microbiome is the complex community of commensal, pathogenic, and symbiotic microorganisms that literally share our body space. The human microbiome has emerged in the past few decades as central to human health and disease. Studies have shown that the human microbiome is involved in various physiological processes, including digestion, metabolism, immune, and neurological functions. Critically, the human microbiome is dynamic, highly individual-specific, and influenced by both host genetics and environmental exposures. Therefore understanding this complex interaction between the human microbiome and human health has only been made possible by the advent of next-generation sequencing technology and the development of dedicated statistical methods. Typical human microbiome research goals are focused on investigating the mechanisms by which the microbiome influences human health, the impact of new therapies for various diseases and conditions, and understanding the impact of external factors, such as diet, age, and antibiotic use, on the human microbiome. Microbiome datasets are typically highly dimensional, sparse, overdispersed, zero-inflated and compositional. Additionally, technical and/or resource limitations can often lead to small sampling sizes. These unique features and the complex nature of this data make the statistical analysis of microbiome datasets a challenge. Therefore, the development and validation of data transformation techniques and statistical approaches needed to analyze such complex data are under research and still being established. This presentation discusses the current hopes and challenges in analyzing microbiome datasets and presents the statistical methods currently applied and their limitations in answering microbiome research questions.

Charge and biological function: a peptide story

Nevena Ilieva[1], Peicho Petkov[2], Elena Lilkova[1], Leandar Litov[2]

[1]*Institute of Information and Communication Technologies*
at the Bulgarian Academy of Sciences, 2, Acad. G. Bonchev Str., Sofia, Bulgaria
[2]*Physics Faculty at the Sofia University "St. Kl. Ohridski", 5, J. Bourchier Blvd., Sofia, Bulgaria*

e-mail: nevena.ilieva@iict.bas.bg

Keywords: antimicrobial peptides, peptide charge, molecular dynamics, metadynamics

Antimicrobial peptides (AMPs) have attracted significant scientific interest over the past few decades due to their role as the host's primary defense against microbial invasion. Investigations on AMPs are a growing area of research that has the potential to provide new insights into the mechanisms by which the respective species protect themselves from disease and infection, but also to inspire new treatments for human health conditions amid gradually increasing bacterial resistance to traditional antimicrobials. It is generally accepted that the mechanism of action of AMPs is based on their cationic and amphiphilic nature, which allows them to interact with negatively charged bacterial surfaces and membranes by causing membrane rupture or interfering with metabolic processes. Although most of AMPs are cationic, non-cationic AMPs exist. Moreover, as they are secreted as part of complex multicomponent substances with antimicrobial activity, the presence there of anionic, cationic, and also neutral components has to be investigated in connection with the biological activity of the substance in question. In addition, it is an open question when AMPs assume their biologically active form and how they approach the target membrane. We present *in silico* perspective on the possible modes of action and biological role of non-cationic peptides on the example of newly isolated peptides from the lightest fraction of the mucus of the garden snail *Helix aspersa* with only known amino acid sequence. Based on large-scale molecular dynamics and metadynamics simulations, we argue that linear AMPs self-associate in the body fluids in nanoscale aggregates that represent the perfect transport system. Furthermore, the peptide folding promoted by the amphiphilic structure in the aggregate allows a sufficiently high local concentration of AMPs to be delivered to the target membrane in a functionally active conformation. Our investigations also show that the effect of the AMP charge on the energetics of the peptide-membrane interaction is not as straightforward as assumed *a priori*. While the positive charge is required to electrostatically conduct the AMP to the target membrane, the presence of neutral hydrophobic domains in the AMP molecule is necessary to drive membrane embedding via the hydrophobic effect, and a negatively charged domain can be beneficial to enhance membrane penetration and pore formation.

Acknowledgements: This reserach was supported in part by the Bulgarian Science Fund (Grant KP-06-OPR 03-10/2018).

Spatial host genome rearrangement in influenza infection

Ince Ikbal Agah*[1], Shokriyan Masuma[1], Delmas Bernard[2], Beaujean Nathalie[3,]

[1]Acıbadem University, Istanbul, Turkey; [2] INRA, UR892, France; [3]Inserm, U1208, INRA USC1361, France

e-mail: İkbal.agah.ince@gmail.com

Keywords: Epigenetics; Influenza virus; Histone; Acetylation; Methylation; DNA methylation,

Background: Influenza is a global problem in humans and livestock, and the virus is continuously evaluated. The current mitigation strategies against influenza are not yet successful in eliminating the virus. Therefore, there are high economic losses every year due to the inefficiency of available strategies. Epigenetic mechanisms may significantly influence the effectiveness of anti-viral therapies. Understanding epigenetic mechanisms in host-pathogen systems becomes crucial before designing disease prediction and prevention strategies. This study is designed to characterise the ujmodulation and the possible implication of the epigenetic status of influenza infection. The knowledge may contribute to mapping the critical molecules for defining candidates for biomarker and drug development.

Material/methods: In vitro infection models were tested to determine whether the epigenetic status of infected cells is modulated during influenza infection. The human A549 alveolar epithelial cell line was infected with H1N1 influenza subtype. During the course of influenza infections, various histone modifications (H3K9 acetylation and methylation) were analysed by high-resolution microscopy using fluorescent immunodetection of epigenetic marks. We also determined the global content of DNA methylation using an enzymatic approach called the Luminometric Methylation Assay (LUMA). We have extended our analysis to map the infected cell transcriptome to have an overview of the up/down regulated gene profiles involving epigenetic mechanisms. The transcriptomic analysis provided more evidence of the genes involved in epigenetic regulation.

Results: The LUMA analysis showed clear DNA hypomethylation of the genome of H1N1 infected A549 cells (non-infected; 55.6%, H1N1 infected; 52%). Transcriptomic studies evidenced a modulation of the expression of genes associated with DNA methylation. In particular, A549 cells, the host group, showed that expression of MDB2 is exacerbated during infection. MDB2 binds to and directs the Nurd complex to methylated DNA, while DNMT1 copies existing methylation patterns following DNA replication. DNMT3A and DNMT3B establish new DNA methylation patterns.

Conclusions: We successfully revealed that the influenza infection is accompanied by specific transient histone modifications involving chromosomal rearrangement in flu infection models by immunofluorescent microscopy, biochemical detection of methylation status and transcriptomic analyses of the infected cell model.

Geometric bounds on information flows in evolutionary dynamical systems

Vladimir Jaćimović

Faculty of Natural Sciences and Mathematics,

University of Montenegro, Cetinjski put, 2, 81000 Podgorica, Montenegro

e-mail: vladimirj@ucg.ac.me

Keywords: Fisher information; The fundamental theorem of natural selection; Evolutionary games; Uncertainty relation; Information geometry

The Fisher's Fundamental Theorem of Natural Selection (FFTNS) is a matter of longstanding debate in the community of mathematical biologists. The original Fisher's formulation, published in 1930. in his work „The Genetical Theory of Natural Selection“ was stated as follows:

The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time.

However, the exposition preceeding this statement lacks the mathematical rigor. In such an ambiguous context, it is difficult to say if the Fisher's statement is correct or false, meaningful or trivial. Subsequently, many researchers proposed completions and clarifications of the original Fisher's statement. Recently, John Carlos Baez provided a rigorous formulation of FFTNS for evolutionary games with a finite strategy set. Baez' result is based on the information-geometric approach to evolutionary games. We pursue this idea further and study FFTNS in the context of those mathematical theories in which evolutionary dynamics play an important role: game theory and optimization. We report rigorous formulations of FFTNS for evolutionary games with a continuous trait space and evolutionary optimization. Our results demonstrate that, in general, the words „is equal“ in the original Fisher's statement should be replaced by „does not exceed“, thus turning equality into an inequality. This inequality is saturated only in some (pretty exceptional) cases. We further discuss the information-theoretic meaning of FFTNS, based on the relation between the Kullback-Leibler divergence and the Fisher information. In the new light, FFTNS appears as a universal principle, analogous to the uncertainty relation for energy and time in physics. This unveils a novel aspect of FFTNS that we find the most insightful: it is the statement about an intrinsic time scale of evolutionary dynamics. It imposes the bound on the amount of information that a given evolutionary system is able to gain during an infinitesimally small time step. In other words, it tells us how fast a certain biological system is able to learn.

The dynamics of flowering under heat stress: interaction between genes, miRNAs, circRNAs and lncRNAs

Musa Kavas[1], Bayram Ali Yerlikaya[1], Seher Yerlikaya[1], Nisa Nur Yılmaz[1], Kubilay Yıldırım[2]

[1]Faculty of Agriculture, Department of Agricultural Biotechnology, Ondokuz Mayıs University, 55270, Samsun, Turkey
[2]Faculty of Science, Department of Molecular Biology and Genetics, Ondokuz Mayıs University, 55270, Samsun, Turkey

e-mail: musa.kavas@omu.edu.tr

Keywords: Common bean, heat stress, flowering, lncRNA, circRNA, miRNA

Legumes, such as beans, have an important role in sustainable agriculture due to their ability to fix nitrogen with rhizobium bacteria. Beans are a significant source of protein, micronutrients, and calories, meeting a large part of daily human dietary needs. However, plants are exposed to various environmental stresses, including abiotic stresses that limit their growth, development, and productivity. Abiotic stresses include high temperatures, drought, salinity, heavy metals, and high light intensity. Plants have developed various mechanisms to cope with these stresses, including regulating gene expression. This study focused on determining the transcriptome, miRNA, circRNA, and lncRNA profiles of two varieties of plants, Perola (able to flower and be fertilized under high temperatures) and Karacaşehir-90 (unable to flower under high temperatures), and identifying the genes targeted by miRNAs, circRNAs, and lncRNAs under high-temperature conditions using next-generation sequencing systems and bioinformatics tools. The activity of DEGs, miRNAs and their targets, lncRNAs, and interacted circRNAs were also confirmed with qRTPCR. In conclusion, transcriptome analysis can provide insights into important transcription factors and molecular mechanisms involved in the biotic and abiotic stress response of plants and can be an economical and effective tool for research.

What is life?: Open quantum systems approach

Andrei Khrennikov

Linnaeus University, International Center for Mathematical Modeling in Physics and Cognitive Sciences, Växjö, SE-351 95, Sweden

e-mail: Andrei.Khrennikov@lnu.se

Keywords: Open quantum systems, biosystems, order stability, entropy dynamics, quantum master equation, adaptation to environment, camel-like shape of entropy

Recently the quantum formalism and methodology started to be applied to modeling of information processing in biosystems, mainly to the process of decision making and psychological behavior (but some applications to microbiology and genetics are considered as well). Since a living system is fundamentally open (an isolated biosystem is dead), the theory of open quantum systems is the most powerful tool for life-modeling. In this presentation, we turn to the famous Schrödinger book ``*What is life?*'' and reformulate his speculations in terms of this theory. Schrödinger pointed out to order preservation as one of the main distinguishing features of biosystems. Entropy has the tendency to increase (*Second Law of Thermodynamics* for isolated classical systems and dissipation in open classical and quantum systems). Schrödinger emphasized the ability of biosystems to beat this tendency. We demonstrate that systems processing information in the quantum-like way can preserve the order-structure expressed by the quantum (von Neumann or linear) entropy. We emphasize the role of the special class of quantum dynamics and initial states generating *the camel-like graphs for entropy-evolution* in the process of interaction with a new environment E:

1) entropy (disorder) increasing in the process of adaptation to the specific features of E};

2) entropy decreasing (order increasing) resulting from adaptation;

3) the restoration of order or even its increase for limiting steady state. In the latter case the steady state entropy can be even lower than the entropy of the initial state.

Such quantum entropy dynamics is illustrated by graphs obtained via numerical simulation for quantum master equation. For simplicity of modelling we consider only quantum Markov dynamics. But the real dynamics of biosystems' states is non-Markovean.

Distinguishing Biological versus Abiotic Mineral Structures

Pamela Knoll[1,2], Julyan H. E. Cartwright[2]

[1]*University of Edinburgh, Edinburgh, UK*
[2]*IACT-CSIC, Granada, Spain*

e-mail: pknoll@ed.ac.uk

Keywords: biomineralization, biomimetic, minerals

Living organisms use abundant and nearby minerals to create structures through biomineralization. This can range from highly controlled processes where the product is regulated down to the molecular level to the modification of chemical environments to promote precipitation in the vicinity of the organism. The final products are smoothly-curved shapes that can be simple in form such as the hollow filaments of iron-oxidizing bacteria to intricate, complex structures marvelled for their design such as the silica cage of diatoms. These curvilinear shapes were previously perceived to only be achievable by life and therefore used as a biosignature for evaluating fossils in ancient rocks. However, there have been long known examples of purely abiotic experiments precipitating similar shapes and in recent years many more new systems coming to light. One prominent laboratory experiment forming hollow tubes are chemical gardens. They were first documented in the 17th century and linked to biology due to their plant-like growth. Since then, their physicochemical dynamics have been the main focus of study and have become understood and even applied in technology. What remains elusive is a more comprehensive understanding of the complex dynamics, even in these simple filamentous microbial structures, biomineralization might play in promoting survival of modern living systems and, perhaps, the role of mineralogical spaces in the origin of life.

Network Intrusion Detection by hashed pattern matching using DNA bases encoding

Thaer Hani, Muhammad Kazim, Stefan Kuhn[1,2], Mustafa Kaiiali

[1]*De Montfort University;The Gateway;Leicester;LE1 9BH;United Kingdom*
[2]*Tartu University, Tartu, Estonia*

e-mail: stefan.kuhn@ut.ee

Keywords: Network Intrusion Detection Systems (NIDS); DNA encoding; Protein signatures

Network Intrusion Detection Systems (NIDS) are essential for detecting attacks against computer networks. They can be broadly classified into two types: Misuse (signature-based) NIDS and Anomaly NIDS. Misuse NIDS uses pattern matching methods against a pre-defined database of attack DNA signatures, while Anomaly NIDS detects attacks by observing abnormal user activities or patterns. In this research, we propose a hybrid NIDS that utilizes both types.

To achieve efficient pattern matching results, we use DNA encoding, which involves transforming plain text into DNA sequences or Amino Acids. This approach is inspired by the process of encoding genetic information in living cells. We leverage bio-inspired operations and sequence alignment algorithms to search suspicious network transactions for attack signatures.
Our idea is to imitate the viral infection detection to humans in nature. For this, we build a database of protein signatures for known attacks using essential amino acids labels and a Vigesimal numbering system. We then build a neural network model which we train on our protein database to learn common viral proteins patterns by utilizing protein secondary structure similarities and functional domain areas to group similar attacks. Finally, we end up with trained model to identify suspicious Network transactions in real time with minimal matching time and acceptable false positive ratios.

Our initial experiments show high accuracy compared to previous literature work.

Information theoretic view of the genetic code

Ádám Kun[1], Ádám Radványi[2]

[1]*Department of Plant Systematics, Ecology and Theoretical Biology, Institute of Biology, Eötvös University, Pázmány Péter sétány 1/C, Budapest, H-1117, Hungary*
[2]*Centre for Data Science and Digital Development, Moholy-Nagy University of Art and Design, Zugligeti út 9–25, H-1121 Budapest, Hungary*

e-mail: kunadam@elte.hu

Keywords: origin of the genetic code, mutation, distortion, fitness effect, fitness landscape

The copying of genetic information and its translation to peptide sequences can be seen as a communication system and studied with information theory. Protein sequences can be regarded as the messages decoded at translation, while the encoded information (signal) is the nucleic acid sequence. Here, the channel and its fidelity are defined by replication and translation, where "noise" ends up as mutations, mainly due to genomic mutations and mistranslation. Noise produces distortion in the message, but distortion is more than just mutation. Mostly we see mutations as changes at certain places in the genome, but how much it distorts the message depends also on the structure of genetic code, the distribution of codon usage and the effect of the change on protein function. In this talk, I would like to briefly present what we know about mutations, mutation rates and fitness effects of those changes and how it relates to the research on the origin of the genetic code.

Linking biophysical features of radiation with large-scale gene expression changes across normal and tumor tissues

Stella Logotheti[1], Zacharenia Nikitaki[1], Vassiliki Zanni[1], Ioanna Tremi[1], Maria Souli[1], Athanasia Pavlopoulou[2] and Alexandros G. Georgakilas[1]*

[1]*DNA Damage Laboratory, Physics Department, School of Applied Mathematical and Physical Sciences, National Technical University of Athens (NTUA), Zografou, 15780 Athens, Greece*
[2]*Izmir Biomedicine and Genome Center (IBG), Balcova, Izmir 35340, Turkey*

e-mail: alexg@mail.ntua.gr

Keywords: radiation response; transcriptomics; radiobiology database; high LET; low LET; DNA damage response and repair

Ionizing radiation (IR) is a genuine genotoxic agent and a major modality in cancer treatment. IR disrupts DNA sequences and exerts mutagenic and/or cytotoxic properties that not only alter critical cellular functions but also impact tissues proximal and distal to the irradiated site. Unveiling the molecular events governing the diverse effects of IR at the cellular and organismal levels is relevant for both, radiotherapy and radiation protection. Herein, we address changes in the expression of mammalian genes induced after exposure of a wide range of tissues to various radiation types with distinct biophysical characteristics. We constructed a publicly available database, termed RadBioBase (http://radbiodb.physics.ntua.gr/), which includes comprehensive transcriptomes of mammalian cells across healthy and diseased tissues that respond to a range of radiation types and doses. Using integrative bioinformatics, functional enrichment analysis, and machine learning techniques, we unveiled the characteristic biological pathways related to specific radiation types and their association with various diseases. This approach allowed us to decipher the effects of high-versus-low linear energy transfer (LET) radiation on cell transcriptomes and to identify distinct gene signatures for different types and doses of radiation. We further show that different radiation types generate specific lesions that can in turn activate diverse components of the DNA damage response and repair machinery. The intricate interplay between the biophysical features of radiation and the DNA lesions triggers cell-intrinsic and cell-extrinsic pathways which are associated with inflammatory and immunomodulatory responses and are distinct for each radiation type. The detailed characterization of these molecular events at the organismal level is essential in order to improve the efficacy of radiotherapy and reduce systemic radiotoxicity.

Biologically informed deep learning in precision medicine

Andigoni Malousi[1,2]

[1]*Lab of Biological Chemistry, Medical School, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece*
[2]*Genomics and Epigenomics Translational Research Group, Center for Interdisciplinary Research and Innovation, GR-57001 Thessaloniki, Greece.*

e-mail: andigoni@auth.gr

Keywords: deep learning, biologically informed learning, bioinformatics, precision medicine

Biologically informed deep learning (BIDL) is an emerging analytical approach in human -omics studies that uses prior knowledge of biological systems to guide the development of predictive models for precision medicine. The goal is to improve the accuracy and interpretability of these models, and ultimately to personalize diagnosis, therapeutic decisions, and disease prevention.
While contemporary machine learning evolves to minimize the need to know the data, by definition, BIDL relies on the theoretical knowledge to encode relevant feature space, to define architectural components and to interpret outputs. BIDL is often confused with explainable deep learning. Whereas both aim to deliver biological interpretability of predictive models, they apply different approaches to achieve this goal. Typically, BIDL leverage biological systems into models to make them more biologically reasonable and easier to understand, while explainable deep learning helps in the post-hoc understanding the models' inner workings and decision-making. So far in genomics, BIDL has been applied to infer gene activity in single cells, drug response, patient stratifications, survival prediction etc. Among different architectures, graph convolutional neural networks are suitable for biologically informed learning as they operate on graph-structured data, enabling the formulation of a complex feature space of phenotype-specific molecular pathways and protein-protein interactions. Graph-based neural networks are often combined with layer-wise relevance propagation and attention mechanisms to highlight features that contribute most to a prediction. In gene-centric analyses single point variants or indels, transcript profiles, DNA methylation sites etc. are mapped to genes in a way that spatiotemporal interactions can be inferred in graph-based networks. In pathway-centric or protein-centric studies the interacting entities include cellular processes linked with gene sets and post-translational modifications. Despite the progress, there is still limited deployment of BIDL in precision medicine. The main issues are, on one hand, the inherent heterogeneity, sparsity and bias of the -omics data, and on the other the knowledge gap between theoretical biology and deep learning. The convergence of the technological advancements in deep learning with the domain knowledge in phenotype-specific studies through biologically informed models is challenging, yet worth investing for a growing spectrum of bioinformatics applications in precision medicine.

Hydrodynamic coupling in the absence of inertia: the first intercellular communications

Iván Marqués Campillo

Departament de Física, Universitat de les Illes Balears, E-07071 Palma de Mallorca, Spain

e-mail: Iwmarquescampillo197@gmail.com

Keywords: dynamics, overdamped, oscillators, viscous, fronts, self-sustained, tonotopy

Inspired by the problem of how the earliest (proto)cells could have communicated along filaments, we investigate the dynamics of chains of overdamped (inertialess) mechanical self-oscillators mutually coupled to their neighbors by the viscous forces induced by the surrounding fluid. Modeling this arrangement leads to an unusual form of extended dynamical system that has not been thoroughly studied until now. In the absence of inertia, the dynamics effectively transforms a locally defined coupling into a global coupling described by an integral spatial operator with an exponentially weakening kernel. This mutation has a counterintuitive effect on the propagation of perturbations through the medium, as well as on the synchronization behavior of the oscillators in comparison with the more familiar cases described by reaction–diffusion systems.

From the earliest fossils, stromatolites, which are calcareous mounds built up by biofilms of populations of cells (cyanobacteria) we know that populations of cells started to cluster on the origin of life. Such biofilms might seem to be 2D populations, but these bacterial mats are felt-like accumulations of filaments: one-dimensional chains of cells. Thus, it is of interest to understand 1D configurations of coupled motile cells. Even before the evolution of motile organelles, cilia and flagela, cells can have used movement of the cell membrane itself in order to affect their external environment. Once one has cellular movement plus a viscous fluid medium, the movement of one cell will affect its neighbours via hydrodynamic coupling. This, then, inspired our study.

Our contribution to the conference will be of a physical-mathematical nature. Firstly, we will show that in the simplest setup of a chain of linear oscillators, a non-intuitive behaviour appears for the decay of local perturbations, compared to diffusion. Secondly, we will study the propagation of fronts in nonlinear systems that may exist in two different mechanical equilibria in contiguous regions. We will address how mechanical information propagates along the chain and we will determine that the front velocity undergoes a transition from an attracting static solution to a propagating front. Lastly, we will explore the effect of self-sustained mechanical oscillators, i.e., a chain of coupled Van Der Pol oscillators, which can exhibit tonotopy (spatial arrangement of frequencies).

AI methods for biological networks and sequences

Dragan Matić[1], Milana Grbić[1]

[1]*Faculty of Natural Sciences and Mathematics, University of Banja Luka, Mladena Stojanovića 2, Bosnia and Herzegovina*

e-mail: dragan.matic@pmf.unibl.org

Keywords: biological networks, biological sequences, network partitioning, longest common subsequence, combinatorial optimization, metaheuristic algorithms

Nowadays, computational methods are widely used for extracting new information from different biological structures. Many complex problems from biological domain can be modelled by mathematical structures and, therefore, can be solved by using mathematical methods and artificial intelligence algorithms. In this research, we focus on two important biological structures: biological networks and biological sequences in order to analyse some of their properties. Biological networks can be presented in the form of graphs, where nodes are biological entities, and edges (links) between two nodes represent the relationship between corresponding entities. Biological sequences, like DNA or protein sequences, can be represented as strings of characters. Problems defined on these data, like network partitioning and finding common subsequence with different constrains, usually, fall into the class of NP hard problems, which are time consuming and, therefore, very hard to solve. Here we present several such problems and methods for their solving.

Biological networks are usually very large scaled structures containing thousands of nodes and links. In such big structures, it is very hard to find patterns or some other important properties. The common way for their analysing is to partition them into smaller components which still hold important properties. The problem of network partitioning can be solved by different computational methods, including the methods from the area of combinatorial optimization.

Among many important problems related to biological sequences, we focus on measuring sequences similarity by finding the longest common subsequence with some additional biological motivated constrains. This kind of problems can also be solved by metaheuristic algorithms and we present some of them.

Biological information and mathematical structures

Nataša Ž. Mišić, Sascha Cvetkovic

R&D Institute Lola, Kneza Višeslava 70a, Belgrade, Serbia

e-mail: nmisic@rcub.bg.ac.rs

Keywords: genetic code, origin of life, mathematical eigenforms, self-reference, *p*-adic modelling

Biological information acts through coherent multidimensional communication systems based on the coupling of biophysical and biochemical processes. It behaves as an effective non-locality and an integrating background that connects micro and macro spatiotemporal features. Insights from sequence homology indicate that biological information content ("meaning") can only be recognized through direct evolutionary analysis, and not from a single sequence. This is one of the prominent examples that surpasses the Shannon information concept [1] and points towards a general question of the inherent factors and decision mechanisms that, together with environmental factors, favorize certain information regarding the origin and evolution of life [2].
Successful mathematical modeling of both biological and physical phenomena has raised speculative ideas about the mathematical nature of reality, particularly in relation to metrics and measures. In such an approach, the mathematical eigenforms that encompass all of their structure-preserving transformations are important formalisms for concepts of self-reference and reflexivity [3]. These concepts are close to autopoietic (self-producing) systems that are applied to the description of living organisms [4].
In our presentation, we discuss examples of mathematical structures mainly based on mathematical eigenforms that faithfully describe the essential properties of the genetic code [5,6] as the first biological code that represents the informational basis of the origin of life.

References:

[1] Koonin, E.V. **2016.** The meaning of biological information. *Philos. Trans. R. Soc. A* 374(2063), 20150065.
[2] Wills, P.R. **2016.** The generation of meaningful information in molecular systems. *Philos. Trans. R. Soc. A* 374(2063), 20150066.
[3] Kauffman L.H. **2017.** Eigenform and reflexivity. *Constr. Found.* 12(3), 246–252.
[4] Varela, F., Maturana, H.R., Uribe, R. **1974.** Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems* 5, 187–196;
[5] Mišić, N.Ž. **2016.** Standard genetic code: *p*-Adic modelling, nucleon balances and selfsimilarity. *FU Phys. Chem. Tech.* 14, 275–298.
[6] Dragovich, B., Mišić, N.Ž. **2019.** *p*-Adic hierarchical properties of the genetic code. *Biosystems* 185, 104017.

Non-globular proteins in the era of Machine Learning (ML4NGP) – COST Action CA21160

Alexander Miguel Monzon[1]

[1]*Department of Information Engineering, University of Padova*
Via Giovanni Gradenigo 6/B - 35131 - Padova, Italy

e-mail: alexander.monzon@unipd.it

Key words: non-globular proteins, bioinformatics, machine learning

Protein structure prediction has long been considered the “Holy Grail” of structural biology. The recent success of AlphaFold has ushered in a new era of highly accurate structure prediction, bringing to light the secrets hidden in the three-dimensional structures of globular proteins, increasing our understanding about their structural features and molecular function. However, a large proportion of the proteomes from all domains of life are rich in sequences that do not fold into regular structures, commonly known as non-globular proteins (NGPs). NGPs comprise intrinsically disordered regions, repeats, low-complexity sequences, aggregation-prone and phase-separating sequences, and are implicated in a range of age-related diseases. Their heterogeneous structural states and low sequence complexity challenge current experimental structure determination techniques and machine learning (ML) methods for structure prediction, making the molecular understanding of their sequence-structure-dynamics-function relationship difficult. The recent improvements of ML approaches and advances in determining NGP structural ensembles call for a timely re-assessment of the interplay between experiments and computation. The ML4NGP Action aims to establish an interdisciplinary pan-European network to favour this interplay, fostering experimental frameworks designed to provide information to computational methods, and novel computational methods developed, trained and benchmarked with experimental data. The main aim and objective of the Action is to describe the function and properties of NGPs by combining experimental data and novel machine learning ML approaches.

Towards neural modulation using battery- and electronics free implantable microdevices.

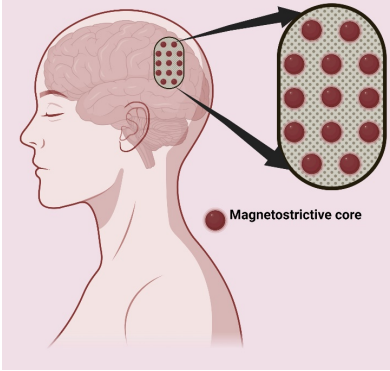
Ram Prasadh Narayanan[1], Ali Khaleghi[1,2], Ilangko Balasingham[1,2]

[1]*Institute of Electronic Systems, Norwegian University of Science and Technology, Norway*
[2]*Intervention Center, Oslo University Hospital, Norway*

e-mail: ram.p.narayanan@ntnu.no

Keywords: Electrical biostimulation, Magnetoelectric coreshells, wireless implants

Neurogenesis and neuroplasticity are essential for brain health, allowing us to form memories, learn new skills, and recover from injury or disease. To have two-way communication with stimulation possiblity within the neural system, we are developing Brain-Machine-Interface and connectivity, as part of the onging Horizon Europe Pathfinder Open proeject, B-CRATOS, as well as wireless control of microdevices. Recently, we have conceptualized the use of implantable hybrid composites made from magneto-electric (ME) microstructures for remote and localized electrical stimulation of neurons. These composites are batteryless and electronics-free, utilizing the non-linear material property of MEs to stimulate neural activity. This involves an array of multi-functional ME elements integrated into a flexible patch for collective sensing of low-frequency electric fields generated by local neural activity. This patch can wirelessly communicate with a readout system using backscatter-based communication, where MEs function as micro antennas. By analyzing the spatiotemporal collective signal from individual MEs and by using deep learning (DL) models, we can build a tomogram map of neural signaling pathways. This research could have potential applications in treating paralysis, Parkinson's, and palliative care.



References

[1] Singer, Amanda, et al. "Magnetoelectric materials for miniature, wireless neural stimulation at therapeutic frequencies."*Neuron*107.4 (2020): 631-643.

[2] Jang, Jinhyeong, and Chan Beum Park. "Magnetoelectric dissociation of Alzheimer's β-amyloid aggregates."*Science Advances* 8.19 (2022)

[3] Pardo, Marta, and Sakhrat Khizroev. "Where do we stand now regarding treatment of psychiatric and neurodegenerative disorders? Considerations in using magnetoelectric nanoparticles as an innovative approach."*Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology* 14.3 (2022).

From Feynman's ratchet to timecrystalline molecular motors

Antti J. Niemi

Nordita, Stockholm University and Uppsala University, Stockholm, Sweden

e-mail: Antti.Niemi@su.se

Keywords: Molecular motors, shape-space geometry, timecrystalline dynamics

Cats have an instinctive ability to use the connection governing parallel transport in the space of shapes to land safely on their feet. Here we argue that the concept of connection that is extensively used in general relativity and other parts of theoretical physics, also explains the impressive performance of molecular motors by enabling molecules to evade conclusions of Feynman's ratchet- and-pawl analysis. We first demonstrate, using simple molecular models, how directed rotational motion can emerge from shape changes even without angular momentum. We then computation- ally design knotted polyalanine molecules and show how their shape space holonomy connection organizes individual atom thermal vibrations into collective rotational motion, independently of angular momentum. Our simulations show that rotational motion arises effortlessly even in am- bient water, making the molecule an effective theory time crystal. Our findings have potential for practical molecular motor design and engineering and can be verified through high-precision nuclear magnetic resonance measurements.

.

The synergy between deep learning and partial differential equations; Example: 3D-MRI neuroanatomical segmentation based on PDE-Graph algorithm

Andrej Novak[1,2], Domjan Barić[1], Lovre Kardum[1]

[1]*Department of Physics, Faculty of Science, University of Zagreb, Bijenicka c. 32, Zagreb, Croatia*
[2]*Dubrava University Hospital, Avenija Gojka Šuška 6, Zagreb, Croatia*

e-mail: andrej.novak@phy.hr

Keywords: brain segmentation, deep neural networks, partial differential equations

Abstract text. The process of digital image segmentation divides an image into regions that are homogeneous with respect to some characteristic, such as pixel intensity or texture. This is an essential process in neuroimaging, particularly in the quantification of total brain volume (TBV) and the detection of structural changes over time. With the development of computer science and neuroimaging methods, deep learning has become a popular approach for medical image segmentation, with U-net being one of the most popular models. However, deep learning methods lack interpretability and require large amounts of annotated data, which can be scarce and expensive. To address these issues, the study proposes a PDE-graph based algorithm that can match the quality of image segmentation results obtained by deep learning methods. Mechanistic models, such as PDE-graph based models, encapsulate dynamic phenomena and are scalable, interpretable, and allow for the establishment of basic process dynamics and initial model identification. The manual labelling of data is expensive, requires expert annotators, and is inconsistent across annotators. In this contribution, we consider how PDE-graph models can help in overcoming this limitation and achieve accuracy comparable to that of manual labelling.

Decoding semiotic minimal genome, a non-genocentric approach

Carolina Gómez-Márquez[1], J. Alejandro Morales[1], Teresa Romero-Gutiérrez[1,2], Omar Paredes[1,3] and Ernesto Borrayo[1]

[1] *Translational Bioengineering Department, Exact Sciences and Engineering University Center, Universidad de Guadalajara, Guadalajara 44430, México*
[2] *Departamento de Ciencias de la Salud - Enfermedad como Proceso Individual, Tlajomulco University Center, Universidad de Guadalajara, Guadalajara, México*
[3] *Mechatronics Department, School of Engineering and Sciences, ITESM Tecnologico de Monterrey, Monterrey 64849, México*

e-mail: omar.paredes@academicos.udg.mx

Keywords: Minimum Genome; Non-genocentric; Biological information flow

The search for the minimum information required for an organism to sustain a cellular-system-network has rendered the identification of a fixed number of known genes, as well as genes which their function remains to be identified. The approaches used in such search generally focus their analysis on coding genomic regions --based on the genome to proteic-product perspective-- which leave other fundamental processes aside, mainly those that include higher-level information management. To cope with this limitation, a non-genocentric approach based on genomic sequence analysis by language processing tools, along with gene ontology, may prove effective in the identification of those fundamental genomic elements for life autonomy, as it will provide an integrative analysis of the information value of all genomic elements, regardless of their coding status.

Multimedia presentations on the human genome.

Implementation and assessment of a teaching program for the introduction to genome science using a poster and animations.

Christoforos Pavlakis

Technical University of Crete, Greece

e-mail: chrispavlakis8@hotmail.com

Keywords: Human genome, genome science, multimedia presentation, poster, illustration, animation, teaching program, high school.

Genome science, including topics such as gene recombination, cloning, genetic tests, and gene therapy, is now an established part of our daily lives; thus we need to learn genome science to better equip ourselves for the present day. Learning from topics directly related to the human has been suggested to be more effective than learning from Mendel's peas not only because many students do not understand that plants are organisms, but also because human biology contains important social and health issues. Therefore, we have developed a teaching program for the introduction to genome science, whose subjects are focused on the human genome. This program comprises mixed multimedia presentations: a large poster with illustrations and text on the human genome (a human genome map for every home), and animations on the basics of genome science. We implemented and assessed this program at four high schools. Our results indicate that students felt that they learned about the human genome from the program and some increases in students' understanding were observed with longer exposure to the mixed multimedia presentations.

The *ceteris paribus* dilemma: medicinal chemistry struggling with complex systems

Stefano Piotto[1]

[1]*Department of Pharmacy, University of Salerno, via Giovanni Paolo II, 132 – 84084 Fisciano (SA), Italy*

e-mail: piotto@unisa.it

Keywords: Drug behaviour; Complex systems

The term "*ceteris paribus*" means "all else being equal," highlighting the assumption that drugs act in isolation and without external influences. However, drugs interact with complex biological systems in reality, making it challenging to predict their efficacy and side effects. Biological systems have multiple levels of organization, from molecules and cells to tissues and organs; each level is interconnected and influences the others. Therefore, when a drug is introduced into the body, it can interact with multiple targets and pathways, leading to a cascade of physiological responses that can be difficult to predict.

The complex and interconnected nature of biological systems means that drugs cannot be viewed in isolation, and their behaviour must be studied in the context of the whole organism.

This presentation explores the limitations of traditional drug development strategies and the emerging approaches to address the *ceteris paribus* dilemma. These include systems biology, network pharmacology, and artificial intelligence to model and predict drug behaviour in complex systems. The presentation also highlights the need for interdisciplinary collaboration among scientists to address this challenge and improve drug development outcomes.

Chaotic neural spiking as a candidate for coded inteneural communication

Oreste Piro[1,2]

[1]IDepartment of Physics, University of Balearic Islands, Carretera Valldemossa km 7,5 Palma de Mallorca, Spain
[2]Department of Ecology and Marine Resources, Mediterranean Institute for Advanced Studies, IMEDEA (CSIC–UIB), Esporles, Spain

e-mail: oreste.piro@uib.es

Keywords: chaotic firings, interneural communication, intemittence

The van der Pol–FitzHugh–Nagumo neuron model with inertia was shown to exhibit a chaotic mixed-mode dynamics composed of large-amplitude spikes separated by an irregular number of small-amplitude chaotic oscillations. In contrast to the standard 2D van der Pol–FitzHugh Nagumo model driven by noise, the interspik eintervals distribution displays a complex arrangement of sharp peaks related to the unstable periodic orbits of the chaotic attractor. For many ranges of parameters controlling the excitability of the system, we observe that chaotic mixed-mode states consist of lapses of nearly regular spiking interleaved by others of highly irregular one. We explore here the emergence of these structures and show their correspondence to the intermittent transitions to chaos. In fact, the average residence times in the nearly-periodic firing state, obey the same scaling law – as a function of the control parameter – than the one at the onset of type I intermittency for dynamical systems in the vicinity of a saddle node bifurcation. We hypothesize that this scenario is also present in a variety of slow-fast neuron models characterized by the coexistence of a two-dimensional fast manifold and a one-dimensional slow one. We also show experimental evidence that the behaviour described above is present in a class of neurons and we finally speculate on the possible functional role that these complex temporal patterns of the neuron firing might have on the codification of intenerual communication through shared nerve bundles.

Algebraic geometry of disease: the microRNA world.

Michel Planat

CNRS, Institut FEMTO-ST, Université de Franche-Comté, Besançon, France

e-mail: michel.planat@femto-st.fr

Keywords: group representations, algebraic surfaces, disease, microRNA, epigenetics

In addition to the inheritable genetic code that maps DNA nucleotides (nt) to amino acids (aa), there exists a mainly non-heritable 'epigenetic code' that controls gene expression, i.e., the activation or silencing of genes. Understanding the abnormal functioning of this code is crucial, as it is correlated with most diseases. Central to the regulation of gene expression is the RNA-Induced Silencing Complex (RISC), which utilizes a short single strand of non-coding RNA (such as a microRNA, or miR) as a template for the transcript of a complementary messenger RNA (mRNA). Upon binding, an Argonaute protein within RISC activates and cleaves the mRNA.

A miR is a ~22 nt stem-loop structure containing a ~7 nt seed sequence, which is sufficient for recognizing the target genes. In the epigenetic code, the mapping from miRs to genes is non-injective and miR-dependent, similar to the mapping from codons to aa in the genetic code. Our previous work demonstrated that representation theory (the characters of a relevant finite group) offers a model of the genetic code (10.3390/sym12121993). For miRs, we use the representation theory of the infinite group G, generated by the seed (10.3390/sym15030770). The SL(2,C) character variety of G, specifically a Groebner basis B of G, comprises a set of polynomials whose singular structure is crucial for predicting and potentially silencing a disease.

In summary, we found that a disease induced by miR deregulation may occur if the subgroup structure of G is not close to the corresponding free group Fr on r generators (with r+1 being the number of distinct nt in the seed sequence) or if the basis B contains surfaces with isolated singularities (of the A-D-E type). In my presentation, I will focus on oncomirs and miRs related to diseases of the nervous system.

As an illustration, dysregulation of miR-146a-5p (also an oncomir) induces misfolded proteins (prions) and leads to neuropathogenesis. The seed is GAGAAC, and the subgroup structure of the generated group is that of the free group F2. A relevant surface contained in B is f(x,y,z)=-z^3+xy+2yz+2x+4z, which features a triple of isolated singular points of type A2.

Optimal Homotopy Asymptotic Method: Application to dynamical systems

Nicolina Pop[1], Remus Daniel Ene[2]

[1] *Department of Physical Foundations of Engineering, Politehnica University of Timisoara, 2 Vasile Parvan Blvd, 300223 Timisoara, Romania*
[2] *Department of Mathematics, Politehnica University of Timisoara, 2 Victoria Square, 300006 Timisoara, Romania*

e-mail: nicolina.pop@upt.ro

Keywords: Optimal homotopy asymptotic method; boundary layer flow; symmetries; Hamilton–Poisson realization

The synchronization or optimization of nonlinear system performance with applications in medicine or electrical engineering are based on the study of dynamical systems.
Based on the mathematical model development in [`], the Optimal Homotopy Asymptotic Method (OHAM) technique [2] is used to obtain effective and accurate dual analytic approximate solutions taking into account of the thermal effects and to investigate the chemically reactive solute transfer problem in a viscous fluid over an exponentially stretching sheet [3].
The heat transfer problem is analytically explored by using the modified OHAM). By similarity, the motion equations are reduced to a set of nonlinear ordinary differential equations. Based on the numerical results, there are dual analytic approximate solutions within mass and heat transfer problem.
The aim of this work is to investigate the effective and accurate dual analytic approximate solutions taking into account of the thermal effects. The influence of the physical parameters (the Prandtl number and the temperature distribution parameter) over the temperature profile is analytically explored for both solutions: the first approximate solution and the corresponding dual solution [4].
The advantage of this procedure consists in independence of small or large parameters, and provides us with a simple way to optimally control the convergence of the approximate solutions.
Obtained results are in a good agreement with the numerical results and show that our procedure is effective, accurate and easy to use in applications.

References:

[1] Bararnia H., Gorji M., Domairry G. , *Acta Appl Math.* 2009;106:125–133
[2] Ene R.D., Marinca V. *Appl Math Comput*. 2015;269:389–401.
[3] Ene R.D., Pop N. *Wave Random Complex*. 2021; 1–23; DOI: 10.1080/17455030.2021.1971328
[4] Ene R.D., Pop N. *Mathematics,*2023,DOI: 10.3390/math11051124.

Combinatorial co-factor, energy transduction, and origin of information

Steen Rasmussen[1 2 3], Kristoffer Thomsen[1], Marco Tuccio[4]

[1]University of Southern Denmark, Odense, Denmark; *[2]Santa Fe Institute, NM USA; [3]Centre for Living Technology, Venice, Italy; [4]University of Turin, Turin, Italy.*

e-mail: steen@sdu.dk

Keywords: functional information, selection, co-factor, energy transduction, proto-metabolism

We propose that selection between combinatorial co-factors, which modulate an energy transduction process that turns resources into building blocks, can cause the origin of biological information in a simple protocellular system. It is experimentally demonstrated that when a combinatorial co-factor (including 8-oxo-guanine) and the energy transducer $(Ru^{2+})bpy_3$ anchored to a fatty acid vesicle surface can transform resources (picolinieum ester and protected DNA oligomers) into building blocks (decanoic acid and functional DNA oligomers). [1 2 3] This proto-metabolism enables the vesicle container to grow and divide[4] as well as oligomers to ligate into a full DNA strand. In simulation we demonstrate that anchored co-factor replication is possible based on lesion induced DNA amplification (LIDA) without the use of enzymes. [5 6 7 8] Further, we demonstrate in simulation that 8-oxo-guanine integrated within a DNA duplex can still act as an electron donor for the $(Ru^{2+})bpy_3$ energy transducer due to internal DNA charge (hole) transfer properties[7] which are sequence dependent. Our simulations also indicate that the (2D) surface anchoring of the involved molecular complexes tends to speed up the reaction rates compared to reactions in bulk (3D) although crowding factors also impact the reaction rates.[8]
Based on our simulations we can select suitable co-factor DNA strands for optimal metabolic rates as charge transfer is sequence dependent. Our simulations can also estimate optimal sequences dependent replication rates. However, different sequence motifs respectively enhance charge transfer and replication, so it is non-trivial to select optimal co-factors with good, combined charge transfer and replication properties.[7] Thus, the co-factor sequence/composition can be interpreted as primitive biological (functional) information when selection from a combinatorial set of co-factors is possible.

References:

[1] DeClue et al., *JACS* (2009);
[2] Maurer et al., *Chem. Phys. Chem.* (2011);
[3] Cape et al., *Biocon. Chem.* (2012);
[4] Albertsen et al., *Chem. Comm*. (2014);
[5] Bournebush et al., *Art. Life Proc.* (2020);
[6] Engelhardt et al. *Art. Life. Proc.* (2020);
[7] Thomsen, *U. of South. Den., MS Thesis* (2022);
[8] Tuccio, *U. of Turin, MS Thesis*, (2023).

In-silico Model of Extracellular Vesicle-mediated Intercellular Communication

Hamid Khoshfekr Rudsari[1], Mladen Veletic[1,2], Martin Damrath[2], Mohammad Zoofaghari[1], Ilangko Balasingham[1,2]

[1]*Oslo University Hospital, Sognsvannsveien 20, 0372 Oslo, Norway*
[2] *Norwegian University of Science and Technology (NTNU), N-7491 Trondheim, Norway*

e-mail:h.k.rudsari@studmed.uio.no

Keywords: Extracellular vesicle, biogenesis, uptake, theoretical biology, information theory.

Extracellular vesicles (EVs) are small membrane-bound vesicles that play crucial roles in intercellular communication in both normal and disease states. Despite extensive *in-vitro* and *in-vivo* studies, there is still a need for computational and theoretical biology methods of *in-silico* models to gain a deeper understanding of their functions and roles in intercellular communication. In this abstract, we propose to evaluate three key aspects of EV-mediated intercellular communication from an information and communication theoretics perspective: transmitter, channel, and receiver. The transmitter in an EV-mediated intercellular communication involves the biogenesis of EVs, which vary depending on their size and origin. Novel *in-silico* models are required to study the EV release from cells. We utilize intracellular mechanisms such as intracellular calcium signaling to develop *in-silico* models for the release of microvesicles, a type of EVs, from cardiac cells. The transport of EVs after release is also critical to our analytical models, as current *in-vivo* biodistribution models often fail to accurately capture EV transport in non-invasive methods. We model the channel of the communication link for the EV transport by considering their main routes of movement, such as circulation system and extracellular space in tissues. We model the channel for the EV transport in the extracellular space that happens in the extracellular matrix of a tissue by using Navier-Stokes equations of advection-diffusion. The channel modeling is accompanied by communication theories such as finding the capacity of the communication link. EVs have the special ability to target specific cells while they can be engineered for targeted drug delivery applications. After the EV transport, they reach their target cells, the receiver side, and are taken up or internalized via various mechanisms such as endocytosis and fusion. We model the receiver by second-order chemical reactions and demodulation schemes from information theory. In this way, we can find the amount of information transmitted from the transmitter cells, carried out by EVs, and decoded by the recipient cells. Our results present the secretion rate of EVs from cells as well as the EV biodistribution patterns in the extracellular matrix. Also, our results indicate the EV uptake dynamics at the recipient cells for various uptake mechanisms. Developing such *in-silico* models will aid in predicting their potential for therapeutic and diagnostic applications and will be useful for various scientific fields such as biology, biomedical engineering, nanobiotechnology, pharmacology, and medicine.

How fougerite converted pH/redox disequilibria to autotrophic metabolism

Michael Russell

Dipartimento di Chimica, Università degli Studi di Torino, 10125 Turin, Italy

Michaeljrussell80@gmail.com

Keywords: alkaline vent, denitrifying methanotrophic acetogenesis, double layer oxyhydroxides

Abstract. Enzymes are not catalysts – or not just catalysts. They are dynamic disequilibria converters/bio-nanoengines, arranged to work in series *and* parallel as metabolons, now produced and ordered by the genetic code. The enzymes involve binding pockets and binding-change and escapement mechanisms controlled by electrostatic/conformational gates. The submarine alkaline vent theory (Russell, 2023 *Front. Microbiol.* 14, 1083) presumes that the disequilibria driving the autotrophic emergence of life were broadly similar to those driving chemosynthetic life today. It follows that mineral nanocrytals must have been available at the vent to engineer such mechanisms involving: i) the hydrogenation of CO_2 to formate, ii) the oxidation of methane with nitrogen oxides to a methyl group, the concomitant reduction of those oxides to ammonium, iii) the synthesis of the carboxylic and keto acids and their amination to glycine and pyruvate, iv) their condensation to peptides in an introduction to a peptide/amyloid world that v) could capture the inorganic elements that bring organic chemistry to life and vi) act as the first cell membranes/walls and channels condensing ortho to pyro-phosphate. These six points adumbrate a missing Chapter 1 to the life sciences. They are informed by astronomical, geological, microbiological and nano-technological observations underpinned by a stochastic thermodynamics. The 2D mineral mackinawite (FeS), situated in an inorganic membrane, enables the hydrogenation of the CO_2 in the ocean to formate involving the natural proton motive force focused across the inorganic membrane precipitated at the alkaline vent. Thence the double layer oxyhydroxide fougerite $[Fe^{II}_4Fe^{III}_2(OH)_{12}CO_3.3H_2O]$ – the mineral equivalent of green rust – provides the majority of the precipitate membranes (as natural chemical gardens) and acts as a protometabolon involving protean enzymatic behaviours. It is unique owing to its soft properties, its extraordinary electro-mechanochemical capabilities, its bilaterally active pH-, redox-sensitive and flexible cationic ordered layers in which the degrees of freedom are reduced from 2 to near unity within the hydrous interlayers through the enforcement of vectorial control of electron tunneling, thus filling the divide between the material and the living world. We may even think of it and its more, but reversibly oxidized affine sister, trébeurdenite $[Fe^{II}_2Fe^{III}_4O_2(OH)_{10}CO_3.3H_2O]$ – as a "mineral mitochondria"!. Yet while this putative Chapter 1 can offer the initial conditions for DYNALIFE's Chapter 2, speculations beyond the complementarities of molecular shapes based on Schrödinger's aperiodic crystal as to how the genetic code arose from this metabolic foundation are limited alas!

Role of mineral surfaces into the emergence of Life

Sainz-Díaz, C.Ignacio[1], Colín-García, M.[2], Lerin-Morales, K. M.[3], Soriano-Correa, C[4]., Martínez-Pabello, P. [2], Borrego-Sánchez, Ana[5]

[1]*Instituto Andaluz de Ciencias de la Tierra (IACT), CSIC- Universidad de Granada, Spain.*

[2]*Instituto de Geología, Departamento de Dinámica Terrestre Superficial, Universidad Nacional Autónoma de México, Ciudad Universitaria, 04510 Cd. Mx, México.*

[3]*Posgrado en Ciencias de la Tierra, Universidad Nacional Autónoma de México, Ciudad Universitaria, 04510 Cd. Mx, México*

[4] *Área de Química Computacional y Modelado Molecular, Facultad de Estudios Superiores Zaragoza, UNAM. México.*

[5] *Istituto Italiano di Tecnologia, Genova, Italy.*

e-mail: ci.sainz@csic.es

Keywords: Mineral surface, origin of the Life, organic-inorganic interactions, molecular modeling

The emergency of Life is a consequence of electron exchanges between minerals surfaces and organic molecules. This emergency needs a scenario for overcoming the entropy challenge. Minerals provide confined spaces for developing the first prebiotic reactions. Minerals provides selective inorganic membranes for prebiotic evolution. Preliminary studies on the hydrothermal vents in Iceland have detected the formation of magnesium silicate and clay minerals, where there is a high biological activity and possible prebiotic processes could have happened. Besides, aminoacids have been detected in some extraterrestrial meteorites along with silicates. These molecules can have been also precursors of prebiotic chemistry in early Earth. These reactions were likely happened in the protected confined spaces on the porous surface of chemical gardens and in the interlayer nanospaces of clay minerals. This work describes experimental and theoretical research on the sorption of bioactive molecules onto surfaces of silicate minerals, olivine and montmorillonite at different pH values. This sorption has been also studied at atomic scale by means of quantum mechanical calculations finding that this sorption is energetically favorable. The experimental crystallographic behavior has been reproduced by theoretical calculations. A favorable adsorption on the mineral surfaces is a first step for selective prebiotic processes. This multidisciplinary work is a clear example of cooperation between experimental and theoretical research.

Codon-anticodon interaction in the crystal basis model

Paul Sorba

LAPTH, CNRS, Annecy-le Vieux, France

e-amil : paul.sorba@lapth.cnrs.fr

We have proposed a few years ago a model based on continuous symmetries to consider the DNA structure as well as the mechanism of polypeptide fixation from codons (Phys.Lett. A250, 214 (1998)). Among other applications, sum rules for codon usage probabilities had been obtained in this so-called *Crystal Basis Model* (for a review see: arXiv1704.00940 [qbio.OT], published in BIOMAT 2016, 326-362, 201).

More recently imposing, still in the framework of our model, a minimum principle for the codon-anticodon interaction, the structure of the minimum set of anticodons allowing the translation-transcription for animal mitochondrial code has been obtained. Such an approach is also well adapted for an analysis of the evolution of the genetic code. Inequalities between codon usage probabilities for quartets of codons are also derived, in good agreement with the experimental situation (BioSystems 107 (2012) 113; ibid 111 (2013) 175; 141 (2016) 20). Finally, still in the context of our codon-anticodon interaction we have been able to reproduce in a recent preprint the hierarchy of codon usage frequencies observed in the observed data of twenty plants (preprint submitted to BioSystems). These works are made in collaboration with **A.Sciarrino**, Università di Napoli, Italy.

As a future program in the framework of COST, we plan with **Claire Lesieur** (member of MC) to investigate the codon usage bias from genomics to proteomics still using the Crystal Basis Model. More precisely to wish to test the hypothesis that codon usage bias responds to protein usage. For this purpose, we could compare the codon bias usage and the codon-anticodon interactions of the whole genome, the genes of proteins involved in misfolding diseases and the genes of proteins highly expressed within a unique species. As a second project, we plan also to consider protein ancestors to investigate the role of the parameter q in the Crystal Basis Model along the ages.

Inference of coupling functions between interacting dynamical systems

Tomislav Stankovski[1,2]

[1]*Faculty of Medicine, Ss Cyril and Methodius University, Skopje 1000, North Macedonia*
[2]*Department of Physics, Lancaster University, Lancaster LA1 4YB, United Kingdom*

e-mail: t.stankovski@ukim.edu.mk

Keywords: coupling functions, interactions, dynamical systems, dynamical inference

Interacting dynamical systems abound in nature and often the interest is not only to understand if, but also how they interact i.e. to reveal the functions and mechanisms that define and connect them. Coupling functions contain detailed information about the functional mechanisms underlying the interactions and prescribe the physical rule specifying how an interaction occurs. We used a method based on dynamical Bayesian inference in order to model and reconstruct the coupling functions from the data of interacting dynamical systems. The method accounts also for potential time-varying dynamics and noise interferences, as for example those encountered in biological systems. The effectiveness of the method is demonstrated on two cases: coupled chaotic systems in state space, and neural cross-frequency coupling observed through phase dynamics. Special attention is spent on considerations if this methodological framework can be applied for modelling also dynamical models of genetic components and networks, as it lays down also open questions for future developments in this direction.

Circular Codes in the Genetic Information

Lutz Strüngmann[1]

[1]*Mannheim University of Applied Sciences,*
Paul-Wittsack Str. 10,
68163 Mannheim,
Germany

e-mail: l.struengmann@hs-mannheim.de

Keywords: Circular Codes, Genetic Code, Frame-Shift, Translation

Circular codes can be seen as a mathematical concept to detect and possibly correct frame-shift errors during signal transmission. In the middle 90's such a circular code *X* was identified in the genes of bacteria, eukaryotes, plasmids, and viruses by a comprehensive statistical investigation. The code *X* contained the 20 trinucleotides that appeared to be the codons that had the highest preference for the correct reading frame compared to frames 1 and 2. Since then intensive research on circular codes in the genetic information and their potential role in maintaining the correct reading frame during the translation process in the ribosome has been done by various authors. In particular, *X*-motifs were identified in (i) genes "universally" (ii) tRNAs of prokaryotes and eukaryotes; (iii) rRNAs of prokaryotes (16S) and eukaryotes (18S), in particular in the ribosome decoding center where the universally conserved nucleotides G530, A1492, and A1493 are included in the X-motif; and (iv) genomes (non-coding regions of eukaryotes). Circular codes have a highly complex structure and the ones found in genes possess additional properties like e.g. self-complementarity that reflect their biological nature.

In our talk we give a short introduction to the theory of circular codes and an overview on the concepts from mathematics, statistics and bioinformatics that have been used to study these creatures. An evolutionary model of the genetic code will be presented and we give some recent results on circular codes and their biological applications.

Fast search for associations in genetic datasets

Valeriy Titarenko[1], Sofya Titarenko[2]

[1]*School of Biological Sciences, University of Manchester, United Kingdom*
[2]*School of Mathematics, University of Leeds, United Kingdom*

e-mail: S.Titarenko@leeds.ac.uk

Keywords: genetic markers, pattern mining, optimisation

Many genetic traits are associated with specific nucleotides at known positions. One of the ways to identify them is to collect genetic data for case and control groups, i.e. people with or without a studied phenotype. Then for predefined positions within the human genome, we check which single nucleotide polymorphisms (SNP) are observed for each specimen.

Some genetic diseases are associated with single SNPs, while others are only observed when multiple genetic markers are present simultaneously. One of the approaches to searching for multiple genetic markers is using pattern mining methods combined with an appropriate statistical test (e.g. chi-squared test). But, unfortunately, the computational cost is one of the main problems met in such methods.

Modern hardware provides millions of variants (positions within a genome), and we may use thousands of specimens. Each variant has four options: homozygous for the first/second allele, heterozygous, and missing genotype. For two variants, there are nine meaningful combinations. For the whole dataset, we may need to consider trillions of combinations.

We propose a novel search framework to handle this issue efficiently by solving the original problem using modern computation. It consists of the following key principles: 1) efficient data organisation; 2) sorting algorithms optimised for modern computer architecture.

In recent years the storage capacity of even budget computers has vastly increased (up to 128GB), so available data require only a fraction of memory available. On the other hand, all modern computers allow optimisation using the SIMD principle (single instruction, multiple data). Therefore, we suggest rearranging genetic data in a specific way to make SIMD instructions work the fastest way. This is done by interleaving blocks of data for known genotypes but keeping them close to each other. We experiment with different sizes of blocks suggesting the optimal block size. All necessary arithmetic operations and searching algorithms are rewritten to handle the new data storage type by applying appropriate masks.

As a result, it becomes possible to run the problem previously solved using cluster machines on a standalone workstation.

One point of view on DNA-RNA transcription

Slobodan Zdravković

Institut za nuklearne nauke Vinča, Univerzitet u Beogradu, 11001 Beograd, Serbia

e-mail: szdjidji@vin.bg.ac.rs

Keywords: Helicoidal Peyrard-Bishop model, Breather, DNA-RNA transcription, Demodulated standing solitary wave

This work relies on the well-known helicoidal Peyrard-Bishop model of DNA dynamics [1]. We start with the Hamiltonian and, using Hamilton's equations of motion, we obtain two partial different equations, one of which is nonlinear. There are two approaches, or approximations, to solving this nonlinear equation. One of them is a continuum approximation yielding kink solitons moving along the DNA chain [2]. Relevant for this work is a semi-discrete approximation [3]. This approach brings about a modulated solitary wave called a breather [1,3]. The whole mathematics is explained in the article [1] and the book chapter [4]. It is known that the transcription occurs only at the transcription segments (TSs). These are the segments where DNA is surrounded by RNA polymerase molecules. It is important to stress that the transcription is possible because the DNA molecule opens locally at TSs. It was shown that the local opening could be seen as a DNA breathing mode with extremely high amplitude [5], which, otherwise, can be conceived as a resonance mode [6].

The local opening implies a significantly smaller coupling between base pairs. This is crucial for the transcription but is not enough. It was shown that demodulated breather is possible and is much more convenient [3]. This mode prolongs the interaction between nucleotides belonging to DNA and RNA polymerases. A physical meaning of the breather is that only the nucleotides within it oscillate. When the wave reaches TS, the transcription begins. It is important that the nucleotides belonging to TS have enough time for the interaction. This time is increased if the breather stops for some time. It is possible and explained in Ref. [3]. Therefore, it has been suggested that the demodulated standing soliton is crucial for the transcription.

References:

[1] S. Zdravković, J. Nonlin. Math. Phys., **18** Suppl. 2 (2011) 463–484.

[2] S. Zdravković, D. Chevizovich, A.N. Bugay, A. Maluckov, Chaos **29** (2019) 053118.

[3] S. Zdravković, M.V. Satarić, A.Yu. Parkhomenko, A.N. Bugay, Chaos **28** (2018) 113103.

[4] S. Zdravković, Nonlinear Dynamics of DNA Chain, In *Nonlinear Dynamics of Nanobiophysics*, Edited by S. Zdravković and D. Chevizovich, Springer, 2022, 29-66.

[5] S. Zdravković and M.V. Satarić, Europhys. Lett. **78** (2007) 38004.

[6] S. Zdravković and M.V. Satarić, Europhys. Lett. **80** (2007) 38003.

SCIENTIFIC COMMITTEE
Diego L. Gonzalez (CNR-IMM Bologna)
Julyan Cartwright (CSIC-IACT Granada)
Simone Giannerini (University of Bologna)
Greta Goracci (Free University of Bozen-Bolzano)
Marco S. Nobile (Ca' Foscari University)

ORGANIZING COMMITTEE
Greta Goracci (Free University of Bozen-Bolzano)
Marco S. Nobile (Ca' Foscari University)
Alessandra Volpato (Free University of Bozen-Bolzano)

COST Action
CA21169

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

Funded by
the European Union

Università
Ca'Foscari
Venezia
Department of Environmental
Sciences, Informatics and Statistics

unibz
Freie Universität Bozen
Libera Università di Bolzano
Università Liedia de Bulsan