Fifth International Conference in Code Biology

Granada (Spain), 5–9 June 2018

ABSTRACTS

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Biological codes and the major evolutionary transitions

Ádám Kun[a,b,c]

[a)] Evolutionary Systems Research Group, MTA, Centre for Ecological Research, Hungarian Academy of Sciences, Klebelsberg Kuno u. 3., 8237 Tihany, Hungary [b)]Parmenides Centre for the Conceptual Foundation of Science, Kirchplatz 1, D-82049 Munich/Pullach, Germany; [c)]MTA-ELTE Theoretical Biology and Evolutionary Ecology Research Group, Pázmány Péter sétány 1/C, 1117 Budapest, Hungary
kunadam@elte.hu

The major evolutionary transitions include a list of significant events in the history of life on Earth that changed either individuality, i.e. level of selection, and/or how individuals store or transmit information, i.e. the hereditary system. Two of these transitions have clearly introduced a new code to biology: the genetic code and language.

Biological codes (1) establish a link between two worlds, (2) have an adaptor, and (3) are arbitrary. Evolutionary transitions change individuality by either joining two independent lineages (egalitarian transitions) or establish a stronger link between entities of the same lineage (fraternal transitions). Consequently, there always needs to be link between worlds (albeit these links could pre-exist). Adaptors are also potentially always present. The main question seems to be, if the mapping is by necessity arbitrary or could they be solely dictated by some physical and/or chemical law?

For example, can the genetic code be based on solely the stereochemical hypothesis? I think the answer is in principle yes, but then the question of evolvability remains. An arbitrary map is evolvable, while one dictated by set rules is not. But a predetermined mapping is easier to find. What if there is always a non-arbitrary mapping in the beginning, which is then changed/extended to an arbitrary one to allow for greater evolutionary flexibility.

What about the rest of the major transitions? The evolution of multicellularity –among others– required adhesion, cell-to-cell communication and extracellular matrix each of which has its own code, and each of which has evolved in the unicellular world. Similarly, the evolution of separate sexes requires some recognition mechanism, which can be an odour code or an acoustic code. However, these codes could also operate in asexual species and can be used outside of sexual recognition. Therefore, is the evolution of biological codes necessary for the major transitions in evolution?

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

The fractal dimension of ecoacoustic codes

Almo Farina
Department of Pure and Applied Sciences
Urbino University, Italy
almo.farina@uniurb.it

Sonifeorus animals have a peculiar capacity, still little known, to listen and interpret intra, interspecific and environmental sounds. It is reasonable to suppose the existence of ecoacoustic codes, that are the result of sound patterning, passively and actively used by a listener (soniferous animals, humans) to semiotically explore the environment. Their description, identification and interpretation pose serious problems due to the lack of a reference system that fixes an adapt species-specific scale for frequency, time and space domains utilized by soniferous species. To overpass this relevant constraint and to use ecoacoustic codes as proxy of environmental quality (species richness and diversity, abundance of resources, level of intra and interspecific interaction) a fractal model is applied to an acoustic ambient (soundscape). This model is based on the principle that the complexity of every natural system tends to a maximum when a pattern has a self-similarity at every arbitrary scale of resolution. The iteration of simple patterns has been proved to create complex structure like clouds, mountains or rivers systems. In the same fashion, we have supposed that a temporal sub-set of a soundscape is the result of elementary acoustic "bricks" (ecoacoustic codes) and that increasing the time scale such elementary units produce a self-affine pattern imparting complexity to the soundscape. The EEDI (Ecoacoustic Event Detection and Identification) procedure (Farina and Salutari 2016, Farina et al. 2017) has been applied to acoustic files from tropical and temperate regions across a range of temporal scale and fractal dimensions obtined using the box counting method.

References

Farina, A., Salutari, P., 2016. Applying the ecoacoustic event detection and identification (EEDI) model to the analysis of acoustic complexity. J. Med. Ecol. 14, 13–42.

Farina, A., Gage, S.H., Salutari, P., 2017. Testing the Ecoacoustics Event Detection and Identification (EEDI) Model on Mediterranean Soundscapes. Ecol. Indicators 85: 698-715.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

The Embodied Origins of Linguistic Codes and Psychoanalysis

Anna Aragno
140 West End Avenue, New York, NY, 10023, USA
annaragno@earthlink.net
and
Tullio Rizzini
Università degli studi di Modena e Reggio Emilia
Dipartimento di Scienze Biomediche Metaboliche e Neuroscienze
trizzin@gmail.com

This shared presentation between a neurosurgeon/psychoanalyst/psychiatrist and a psychoanalytic theoretical revisionist, each independent researcher working on different continents, converges on a single point: methodology. Freud's decoding of the language of the Primary Process through his analysis of the structure, mechanisms, and processes of signification in dreams, opened the door to an inclusive epistemology focused predominantly on the study of the Unconscious. Psychoanalysis is a Theory of Mind, a Therapeutic Modality, and a Research Methodology. Dr Rizzini's medical neurological background and Dr. Aragno's in the Arts, Humanities, and Languages, are exemplars of the breadth and vast bed of interdisciplinarity that has typified psychanalytic scholarship from its inception.

The first part of this presentation by Dr. Aragno consists of bullet points addressing pertinent information from the philosophy of science, the basic distinctions between Primary and Secondary Processes of thought, and the epigenetic developmental model of semiotic forms as these all pertain to the study of the somatic/embodied phylogenetic origins and ontogenetic recapitulation of learned cultural and linguistic codes. This will segway into Dr. Rizzini's forty years of empirical research into the instinctual somatic /embodied origins of language derived from observable gestural/oral coding processes that yielded linguistic phonemes, the extensive neuro-anatomical and vast etymological implications of which, given the extreme time limitation, cannot be detailed but may be found fully developed in his books.

The essential convergence in the study, respectively, of the developmental line of logical semiotic forms (Dr. Aragno) and the empirical evidence of the somatic/embodied origins of linguistic codes (Dr. Rizzini), rests on profound theoretical and clinical knowledge of tried and true psychoanalytic principles. Ours is a young science, of barely 100 years, but one that has precipitated seismic changes in its contribution to universal knowledge in the understanding of human nature and the unconscious, where instinctual drives and motivations originate, and most cognition takes place.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Basic neural anticipation: The problem of biological specificity

Andres Kurismaa
Department of Philosophy and History of Science
Charles University, Czechia
andres.kurismaa@gmail.com

One of the most fundamental problems in code biology concerns the biological specificity and reliability of organic codes. While these problems have found profound treatment for genetic coding (Barbieri 2015), the types of codes involved in neural information processing remain controversial and poorly understood, and this particularly with regard to their anticipatory and semantic aspects. Here, it is considered how new types of experiments on the mechanisms underlying action potential generation in single neurons (Sardi et al 2017) may be particularly relevant in this case, and complement earlier studies on the mechanisms of neural anticipation in single brain cells (Anokhin 1984, Andrianov 2006). Although in different aspects, it is argued that both of these analyses reveal unexpected complexity and selectivity in the basic processes of signal "summation" in single cells, and possibly, may allow to consider these key processes from a code biological perspective. Thus, based on early evidence from electron microscopy, as well as their own pharmacological and physiological studies, P.K. Anokhin and colleagues (Anokhin 1984; cf. Andrianov 2006) suggested that the biophysical propagation models of signal transmission (prevalent at the time, and even now) may contain serious fallacies when applied from the effector side of a neuron (axon) to its receptor side (dendrites and soma), as doing so would obscure the whole problem of information specificity and its transmission – how a particular set of inputs can be mapped to correspondingly specific outputs in individual neurons and functional systems? In addition to outlining the functional systems analysis of this problem, recent evidence revealing new types of integrative and selective processes at the basis of spike propagation in neurons is considered (Sardi et al 2017). In particular, by showing negative evidence for isotropic summation, and lack of summation and subtraction effects in combined intra- and extracellular stimulations (Sardi et al 2017), these findings may call for current models of spike generation to be significantly revised – with a view on multiple independent threshold elements within each cell that are anisotropically activated (Sardi 2017), and may express differential transmitter sensitivity and metabolism (Anokhin 1984; Andrianov 2006). If supported by further evidence, these could be important steps towards revealing the diversity of codes in single brain cells, and more generally, for understanding their reliable, yet flexible functioning in the complex anticipatory systems of the brain (Tse 2013).

References:

Andrianov, V.V. (2006). *Functional Neurochemistry of Behavioral Systemoquanta.* Moscow. Russkji Vrach.

Anokhin, P. K. (1984). Systems analysis of the integrative activity of the neuron. *The Pavlovian journal of biological science*, *19*(2), 43-101.

Barbieri, M. (2015). *Code Biology: A New Science of Life*. Springer International Publishing.

Hofmeyr, J. H. S. (2017). Basic biological anticipation. In: *Handbook of Anticipation: Theoretical and Applied Aspects of the Use of Future in Decision Making*, pp. 1-15. Springer International Publishing.

Sardi, S., Vardi, R., Sheinin, A., Goldental, A., & Kanter, I. (2017). New Types of Experiments Reveal that a Neuron Functions as Multiple Independent Threshold Units. *Scientific Reports*, *7*(1), 18036.

Tse, P. (2013). *The Neural Basis of Free Will: Criterial Causation*. Mit Press.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

p-Adic View of the Genetic Code

Branko Dragovich[1,2] and Nataša Ž.Mišić[3]
[1]Institute of Physics, University of Belgrade, Belgrade
[2]Mathematical Institute, Serbian Academy of Sciences and Arts, Belgrade, Serbia
[3]Research and Development Institute Lola Ltd, Belgrade, Serbia
dragovich@ipb.ac.rs

The genetic code is a map from the set of 64 codons onto the set of 20 amino acids and one stop signal. It is almost unique in all living organisms with respect to a huge number of mathematical possibilities. The vertebrate mitochondrial (VM) code is relatively simple and other dozens genetic codes can be considered as its slight variations. In the VM code, an amino acid is coded by one, two or three codon doublets. When two codons code the same amino acid,one can say that they are close, or similar,in the informational sense. It is shown (see e.g. [1,2,3]) that the p-adic distance is an adequate mathematical tool to describe codon closeness (similarity).

p-Adic distance (p is a prime number) is the most employed example of ultrametrics and plays a central role in p-adic analysis and its applications in modeling physical and biological systems with hierarchical structure (for a recent review see e.g. [4]). In p-adic approach to modeling genetic code it is enough to use p-adic distance between some integers. In fact, p-adic distance is related to divisibility of difference between two integers with respect to prime number p: larger divisibility – smaller distance. p-Adic distance between any two numbers is smaller or equalto 1. So, as p-adic distance is smaller, the related numbers are p-adic closer.

In our approach to modeling the VM code we assign 64 natural numbers in the form $a_0 + a_1 5 + a_2 5^2$ to the 64 codons appropriately identifying nucleotides in codons with digits a_i in these numbers. In particular, we take the following identification: **C (Cytosine) = 1**, **A (Adenine) = 2**, **U (Uracil) = T (Tymine) = 3**, and **G (Guanine) = 4**. With respect to the smallest 5-adic distance between codons one obtains 16 quadruplets. Each of these quadruplets splits into two doublets under 2-adic distance. Then each of these 32 doublets, which contain two p-adically closest codons, codes one amino acid or stop signal. This p-adic approach has been extended to many other aspects of the genetic code.

In our talk we will review some main published results and also present some new developments.

[1] B. Dragovich and A. Dragovich, A p-adic model of DNA sequence and genetic code, *p-Adic Numbers Ultrametric Anal. Appl.* **1** (1), 34-41 (2009), arXiv:q-bio.GN/0607018v1.
[2] B. Dragovich and A. Dragovich ,p-Adic modelling of the genome and the genetic code, *Computer J* .**53** (4), 432-442 (2010), arXiv:0707.3043v1 [q-bio.OT].
[3] B. Dragovich, A. Yu. Khrennikov and N. Ž. Mišić, Ultrametrics in the genetic code and the genome, *Appl.Math.Comput.***309**, 350-358 (2017), arXiv:1704.04194 [q-bio.OT].
[4] B. Dragovich et al., p-Adic mathematical physics: the first 30 years, *p-Adic Numbers Ultrametric Anal. Appl* .**9** (2), 87-121 (2017), arXiv:1705.04758 [math-ph].

The authors acknowledge support from the Ministry of Education, Science and Technological Development of Serbia (Projects: OI 173052, OI 174012, TR 32040 and TR 35023).

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Getting a Handle on Events Leading to the Universal Genetic Code

Charles W. Carter, Jr[1], and Peter R. Wills[2]
[1]Department of Biochemistry and Biophysics,
University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7260;
carter@med.unc.edu
[2.] Department of Physics, Auckland University, Auckland New Zealand

We recently summarized both experimental and theoretical arguments against the notion that genetic coding was ever present in the context of an RNA World (Carter and Wills 2017; Wills and Carter 2017). An important component of these arguments was the combination of a bidirectional gene encoding the ancestor to Class I aminoacyl-tRNA synthetases (aaRS) on one strand and that for Class II aaRS on the other strand with nano environmental sensing evident in the identity elements for recognition of cognate tRNAs by aaRS (Carter and Wolfenden 2016, 2015; Wolfenden et al. 2015). In other words, we have now experimentally characterized molecular species with appropriate properties to have implemented genetic coding. An important corollary is that it has now become possible to examine these events in more detail, perhaps to the extent of reconstructing other intermediates in the elaboration of the code itself. We will discuss four areas where we may expect progress in the near future:
i) It should be possible to decipher the pattern of tRNA acceptor stem bases that dictate that differentiate tRNAs into those recognized by the two Classes.
ii) It may be possible to identify architectural bases for the discrimination between amino acids with large and small side chains in the aaRS structures of the two Classes, in keeping with the statistically significant difference between Class I and II aaRS substrates.
iii) Structure-based multiple sequence alignments should allow construction of more accurate phylogenetic trees constrained by the coupling implicit in the bidirectional coding ancestry.
iv) It may then become possible to identify reduced alphabets constrained by robust phylogenetic analysis and to reconstruct and experimentally characterize bidirectional genes based on them. Supported by NIGMS 78227.

References

Carter, C. W., Jr., and Richard Wolfenden. 2015. 'tRNA Acceptor-Stem and Anticodon Bases Form Independent Codes Related to Protein Folding', *Proc. Nat. Acad. Sci. USA*, 112 7489-94.

———. 2016. 'Acceptor-stem and anticodon bases embed amino acid chemistry into tRNA', *RNA Biology*, 13: 145–51.

Carter, Charles W, Jr, and Peter R Wills. 2017. 'Interdependence, Reflexivity, Fidelity, and Impedance Matching, and the Evolution of Genetic Coding', *Molecular Biology and Evolution*, In Press.

Wills, Peter R., and Charles W. Carter, Jr. 2017. 'Insuperable problems of an initial genetic code emerging from an RNA World', *Biosystems*, In Press.

Wolfenden, Richard, Charles A. Lewis, Yang Yuan, and Charles W. Carter, Jr. 2015. 'Temperature dependence of amino acid hydrophobicities', *Proc. Nat. Acad. Sci. USA*, 112 7484-88.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Metabolic iterations and futile cycles

Chris Ottolenghi
Biochimie Métabolomique et Protéomique
Hôpital Necker Enfants Malades – Inserm / Institut Imagine
Université Paris Descartes, France
chris.ottolenghi@parisdescartes.fr

Metabolic cycles are traditionally subdivided according to whether looping applies to cofactors or substrates. Cofactor-regenerating cycles are essential for homeostasis (1), while substrate-regenerating cycles are less easy to understand, except when they are associated with metabolite shuttling between cell compartments (2). In fact, the discoverer of the first metabolic cycles had long emphasized that distinct co-substrates behave in different ways : in the same cycle, some substrates or reactants are regenerated, while others undergo chemical transformations (such as acetyl-CoA combustion). The substrates being regenerated were termed "catalytic metabolites" by Krebs himself, admittedly in full astonishment, and it is unfortunate that this term is no longer in use.

In addition, it is useful to consider that the Krebs cycle and other complex metabolic cycles are related to metabolic shuttles, differing in essentially only one feature : they modify their cargo in addition to carrying it somewhere. Consistently, some of these "metabolite transforming" cycles are (partly) associated with intercompartmental shuttling (e.g. the urea cycle), and others are obviously associated with transport activity when *they are duplicated* on both sides of a boundary, like the Krebs cycle. Indeed, under the control of distinct isoenzymes (often coded by distinct genes), the Krebs cycle catalytic metabolites embody *very extensive shuttling* between cytosol and mitochondria (2). It is tempting to propose that carrier/catalytic metabolites may define modular units of function for metabolite-(co)determined mechanisms, well beyond mere shuttling. As a striking example, *the glycerophospholipid cycle can drive sequential, alternating storage and delivery of progressively modified signal metabolites* (3).

Similar iterative mechanisms might well assist other complex functions, such as cell memory formation (4), and perhaps even classical cycles which are still only partly understood, such as the serine-folate source of nucleotides (5). In addition, as yet unrecognized iterative mechanisms might be associated with the growing list of new "futile" cycles identified in stem cells and cancer.

Identifying transport codes, perhaps extended to liquid-liquid phase separations (6), even more so because metabolite transporter proteins remain a highly neglected area of study (7), will provide a key to further progress in exploring metabolism.

1. Hofmeyr JHS et al. Metabolic control analysis of moiety-conserved cycles. *Eur. J. Biochem* 1986; 155: 631-641.
2. Sridharan GV et al. Discovery of substrate cycles in large scale metabolic networks using hierarchical modularity. *BMC Syst Biol.* 2015; 9:5.
3. Dennis EA & Norris PC. Eicosanoid storm in infection and inflammation. *Nat Rev Immunol.* 2015;15(8):511-23.
4. Barbieri M. A New Model for Biology. In Barbieri M, *The Organic Codes,* 2003, Cambridge University Press, UK, pp. 67-91.
5. Ducker GS et al. Reversal of Cytosolic One-Carbon Flux Compensates for Loss of the Mitochondrial Folate Pathway. *Cell Metab*. 2016;24(4):640-641.
6. Hyman AA et al. Liquid-Liquid Phase Separation in Biology. *Annu. Rev. Cell Dev. Biol.* 2014 ; 30:39–58.
7. César-Razquin A et al. A Call for Systematic Research on Solute Carriers. *Cell* 2015;162(3):478-87.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Symmetry and Evolution of the Genetic Code

Diego Gonzalez and Simone Giannerini
University of Bologna, Department of Statistical Sciences
CNR/IMM Bologna Section
gonzalez@bo.imm.cnr.it

The origin and early evolution of the genetic code is one of the most challenging problems in evolutionary molecular biology. In this contribution we review a number of results on this problem that derive from a mathematical model of the genetic code [1,2,3]. Following this approach, we are able to explain the origin of degeneracy in the early code, a genetic code ancestor of LUCA's code that is supposed to have the same degeneracy distribution of the extant Vertebrate Mitochondrial Genetic Code. Progresses along this line rely heavily on the use of conservation and symmetry features that parallels their role on fundamental and particle physics. The main new results here presented regard Rumer's anti-symmetry and its conservation in the evolution of the genetic code [5].

References

[1] D.L. Gonzalez and Marcello Zanna, *Una nuova descrizione matematica del codice genetico*, Systema Naturae 2003, Vol. 5, pp. 219-236.
[2] D.L. Gonzalez, Can the genetic code be mathematically modelled?, Medical Science Monitor, 10(04), HY 11-17, 2004.
[3] D.L. Gonzalez, S. Giannerini, R. Rosa, *The non-power model of the genetic code: a paradigm for interpreting genomic information*. Philosophical Transactions of the RoyalSociety A: Vol 374(2063), 2015.0062, 2016.
[4] E. Fimmel, L. Strüngmann, *Yury Borisovich Rumer and his "biological papers" on the genetic code*. Philosophical Transactions of the Royal Society A, Vol 374(2063), 2015.0228, 2016.
[5] D.L. Gonzalez, S. Giannerini, R. Rosa, *On the origin of degeneracy in the genetic code*, Technical Report, 2017.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Artibiotics: An Artificial Biology Musical Experiment

Eduardo Reck Miranda
Interdisciplinary Centre for Computer Music Research (ICCMR)
Plymouth University, United Kingdom
eduardo.miranda@plymouth.ac.uk

Artibiotics is an experimental musical composition for percussion ensemble and electronics. It articulates the development of Artificial Biology, which is a new concept that I conceived in 2017, during a residency at Wagner Lab, in the Institute of Medical Microbiology and Hygiene, Regensburg, in collaboration with Biofaction, Vienna. Essentially, Wagner Lab is unravelling the structure and function of naturally occurring antibiotics with a view on engineering new kinds of synthetic antibiotics. In a nutshell, their research involves shuffling the DNA code of known antibiotics, synthesise the new molecules and test them *in vivo* against specific kinds of bacteria. The more I learned about Wagner Lab's synthetic biology work, the more fascinated I became by the idea of synthesising new chimeric proteins. I wondered if it would be possible to develop the notion of synthetic biology further, to the point where the very biochemistry of Biology would become chimeric, blending musical and biological processes. In many ways, the scientific modus operandi that I witnessed at Wagner Lab reminded me of my own creative modus operandi: I felt that the process of synthesising new proteins is not so different from the process of creating new pieces of music. This gave me the idea of inventing a new concept: *Artificial Biology*, or *A-Biology*, which is the surreal biology of a parallel universe of musical and biological molecules. The A-Biology framework consists of a number of pieces of software that processes strands of DNA, including: (a) Miranda machine, (b) rhythmator, (c) pitch-folding and (d) post-translational musifications.

At the core of A-Biology is the *Miranda machine*: an abstract Turing machine-like processor that manipulates a sequence of DNA symbols according to a set of rules. Given a DNA strand, the Miranda machine derives a program that processes the DNA strand itself. That is, a DNA strand encodes instructions to modify itself and generate offspring, which in turn modify themselves and so on.

The *rhythmator* translates a given DNA sequence into a rhythmic sequence. In A-Biology, amino acids are represented in terms of rhythms, referred to as *nucleo-rhythms*. This enables the rhythmator to parse a DNA strand and translate its codons into *rhythmic codons*. Vocabularies of nucleo-rhythms are built either manually or automatically. The latter works by extracting basic rhythmic figures from given musical scores. For *Artibiotics* I generated a number of vocabularies from Brazilian samba. *Pitch-folding* is inspired by the phenomenon of protein folding. It takes into account the orientation information of DNA commands yielded by the Miranda machine to 'fold' the derived program. The DNA commands activate a number of algorithms for generating pitches in function of the orientation of the respective DNA commands.

Finally, *post-translational musifications* are modifications made on the resulting musical sequences generated by the previous processes. The results from these modifications are referred to as *musical molecules*. Post-translational musifications involve processes that might be applied to the sequences automatically (e.g., the A-Biology equivalent of dehydration in standard Biology) and manually. Post-translational musifications give me opportunity to exercise my aesthetic judgement and adapt the sequences to produce musical molecules for specific musical contexts during the compositional process.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Genetic Code Analysis Toolkit (GCAT) – New Tool to Develop Hypotheses on Origin and Evolution of the Genetic Code

(Parts 1&2)

Elena Fimmel (Part I)
Mannheim University of Applied Science
e.fimmel@hs-mannheim.de
and
Markus Gumbel (Part II)
Mannheim University of Applied Science
m.gumbel@hs-mannheim.de

(Joint work with Lutz Strüngmann)

The genetic code can be seen as the major key to biological self-organisation. In fact, all living organisms regardless of whether they are plants, bacteria or mammals have the same molecular bases: adenine, cytosine, guanine, and thymine. Unidimensional sequences of these bases contain the genetic information for the synthesis of proteins in all forms of life. Thus, one of the most pivotal questions is to explain why evolution has produced the current genetic code and why it is as it is. Motivated by these fundamental questions, a new software tool Genetic Code Analysis Toolkit (GCAT) was developed which can be used to investigate the properties of the genetic code in order to develop hypotheses on its origin and evolution. The main focus of the tool has been put on the graphical visualisation of the data. In Part I of the talk some basic information about historical background and selected questions that led to the development of GCAT will be given. In Part II of the talk we will describe in short the tool GCAT and give a couple of applications presenting new results on circular codes and some ancient code.

How to crack the sugar code

Hans-Joachim Gabius
Institute of Physiological Chemistry, Faculty of Veterinary Medicine,
Ludwig-Maximilians-University, Veterinärstr. 13,
80539 Munich, Germany
gabius@lectins.de

Carbohydrates are chemically predestined to form biochemical messages (glycans) of a maximum of structural diversity in a minimum of space. Coding by sugars explains the wide occurrence of glycan chains as part of cellular glycoconjugates (1,2). These signals are read and translated into effects by tissue receptors (lectins), their functional pairing being highly specific and often intimately co-regulated for optimal (patho)physiological significance (2,3).
In addition to structural and functional analysis of individual components such as a certain lectin, two lines of research are being pursued to delineate rules of the sugar code: i) fully programmable design of glycan surface presentation on supramolecular cell models for interaction studies (bottom-up) (4) and ii) targeted modulation of cell surface glycome (top-down). Generation of lectin variants by modular transplantation for functional assays is a means to give a meaning to the natural characteristics of protein architecture (5).
The strategic combination of these approaches will lead to an understanding of structural and topological parameters that govern the flow of information via glycans.

(1) H.-J. Gabius (ed) The Sugar Code. Fundamentals of glycosciences. Wiley-VCH, Weinheim, 2009;
(2) H.-J. Gabius, *Biosystems*, in press;
(3) H.-J. Gabius et al., Cell. Mol. Life Sci. 2016; 73: 1989-2016;
(4) Q. Xiao et al., PNAS, in press;
(5) J. Kopitz et al., Angew. Chem. Int. Ed. 2017; 56: 14677-81.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

From a world without codes to the codes of life: from self assembly to directed self assembly at the dawn of life

Julyan Cartwright
Instituto Carlos I de Física Teórica y Computacional
Universidad de Granada, 18071 Granada, Spain
julyan@onsager.ugr.es

I shall present an overview of where we are today in understanding how material self-organized and self-assembled at the beginnings of life on Earth, and how it may have begun to be directed through information storage; that the biology of codes was the beginning of biology. I shall compare and contrast early life with how biological development uses and direct self-organization and self-assembly processes in ontogenesis.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Histone code and higher-order chromatin folding: a perspective

Kirti Prakash [1,2] and David Fournier [3]
[1]Physico-Chimie Curie, Institut Curie, CNRS UMR 168, 75005 Paris, France
[2]Micron Advanced Bioimaging Unit, Department of Biochemistry,
University of Oxford, Oxford, UK
[3]Faculty of Biology and Center for Computational Sciences,
Johannes Gutenberg University Mainz, 55128 Mainz, Germany
kirtiprakash1.61@gmail.com

Histones are punctuated with small chemical modifications that alter their interaction with DNA (1, 2). One attractive hypothesis stipulates that certain combinations of these histone modifications may function, alone or together, as a part of a predictive histone code to provide ground rules for chromatin folding (3, 4). We consider four features that relate histone modifications to chromatin folding: charge neutralisation, molecular specificity, robustness and evolvability. Next, we present evidence for the association among different histone modifications at various levels of chromatin organisation and show how these relationships relate to function such as transcription, replication and cell division. Finally, we propose a model based on biosemiotics (5, 6), where the histone code can set critical checkpoints for chromatin to fold reversibly between different orders of the organisation in response to a biological stimulus.

References:

1. Jenuwein, Thomas, and C. David Allis. "Translating the histone code." *Science* 293.5532 (2001): 1074-1080.
2. Turner, Bryan M. "Cellular memory and the histone code." *Cell*111.3 (2002): 285-291.
3. Prakash, Kirti, and David Fournier. "Evidence for the implication of the histone code in building the genome structure." Biosystems (2017).
4. Prakash, Kirti, and David Fournier. "Histone code and higher-order chromatin folding: A hypothesis." Genomics and Computational Biology 3.2 (2017): 41.
5. Barbieri, Marcello, ed. *The codes of life: the rules of macroevolution*. Vol. 1. Springer Science & Business Media, 2007.
6. Barbieri, Marcello. "From biosemiotics to code biology." *Biological Theory* 9.2 (2014): 239-249.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Mixed circular codes and their role in evolution

Lutz H. Strüngmann
Institute for Mathematical Biology, Faculty of Computer Science,
Mannheim University of Applied Sciences, Paul-Wittsack-Str. 10
68163 Mannheim, Germany
l.struengmann@hs-mannheim.de

(joint work with Elena Fimmel, Christian Michel, Francois Pirot, and Jean-Sébastien Sereni)

Self-complementary trinucleotide circular codes have been found in large populations of genes and are assumed to play an essential role in maintaining the reading frame during the translational process in the ribosome. In spite of theories speculating about a dinucleotide or tetranucleotide ancestor code of the current genetic code, also dinucleotide and tetranucleotide circular codes have been of great interest. In this talk we consider mixed circular codes, i.e. circular codes that consist of n-nucleotide words over the alphabet of nucleic bases {A,C,G,T} for various n, e.g. a mixture of dinucleotides and trinucleotides. Based on [1] we discuss basic properties of such mixed circular codes and give a construction method for maximal mixed circular codes. Moreover, we will show how the trinucleotide circular code found in genes can be extended to a mixed circular code which could have been an intermediate code when passing from the dinucelotide world to the trinucleotide world. Finally, we put our work in an evolutionary context.

[1] Fimmel E., Michel Ch. J., Francois Pirot, Jean-Sebastian Sereni, and Strüngmann L.: Mixed circular codes in the genetic information, work in progress.

Representationalism Revised: The Impact of Code Biology

Majid Davoody Beni
Assistant professor at the Amirkabir University of Technology
Department of Management, Science and Technology
No. 424, Hafez Street,Tehran, PO Box 3313-15875
m.davoody@aut.ac.ir

Representationalism had been the orthodox view in the philosophy of perception and cognitive science for a while. The classical theories of John Locke and Immanuel Kant could be understood as pioneering versions of representationalism. And the appearance of classical computationalism (i.e., Turing computation) provided further reasons for understanding cognition in terms of symbol processing in a computing system (Fodor 1975; Fodor 2008). Also, Hermann von Helmholtz' theory of perception provided a venue for channelling Kantian insights of representationalism into contemporary computational neuroscience (Swanson 2016). However, representationalism has been notoriously vague about the biological mechanisms that underpin perceptual and cognitive representations. This flaw made representationalism vulnerable to the objections of anactivists and embodied theorists (Chemero 2009; Hutto and Myin 2013; Beni 2017). This is because neuroanatomy of the brain provides little support for the representationalism and its conception of information processing.

In this talk, I embark on presenting a revised version of representationalism which can be reconciled to ecological and embodied theories of cognition. To do so, I will draw on the resources of the code-model of biosemiotics, and more specifically, on Marcello Barbieri's notion of the "the universal neural code" as well as Grabe and Sachse's (2017) recent account of "olfactory code". I shall argue that the code model allows for using the notions of symbol processing and perceptual/cognitive representation. But it is important to notice that it provides a base for accounting for representationl mechanisms in biologically viable terms. That is to say, the code model provides the necessary biological basis for accounting for the capacities of organic systems with species-specific representations of the world. The emerging account of representation goes beyond the classical theories of computation and could be unpacked along the lines of the ecological and embodied theories. I conclude the talk by arguing that it is be possible to reconcile two diverging philosophical accounts of perception and cognition. Representationalism and embodied account could be unified if we replace classical computationalism with the code model of organic symbol processing.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Robustness of the standard genetic code to mutations and mistranslations regarding various amino acid properties

Małgorzata Wnętrzak, Paweł Błażej, Paweł Mackiewicz
Faculty of Biotechnology, University of Wrocław,
Wrocław, Poland
malgorzata.wnetrzak@uwr.edu.pl

The distinct structure and universality of the standard genetic code (SGC) have fascinated the scientists ever since the first amino acid assignments were discovered. Therefore, there are several hypotheses trying to explain the origin and evolution of this code. One of them postulates that the SGC evolved to minimize harmful effects of amino acid replacements in proteins, caused by mutations and translational errors. Many investigations concerning this hypothesis have already been carried out but they were focused mainly on the genetic code robustness regarding the polar properties of amino acids. However, if the SGC has indeed evolved to minimize the errors in proteins, many different amino acid properties, often competing with each other, must have influenced this process. Therefore, to study the possible adaptability of the genetic code, we applied the Strength Pareto Evolutionary Algorithm to perform eight-objective optimization based on the costs of changes in eight amino acid properties, resulting from all possible single point mutations in codons. These eight properties are the representatives of clusters grouping many similar or related amino acid features. By using such approach we avoided an arbitrary choice of physicochemical properties of amino acids to consider in this study. Our results show that it is possible to find theoretical genetic codes which are better optimized for every objective than the standard genetic code, but the latter is still definitely closer to the best optimized codes than to the worst ones. It implies that there is a tendency to minimize the costs of amino acids replacements in the standard genetic code but it is not fully optimized

The Evolution of Cell Adhesion Codes

Marcella Faria
Dactyl Foundation for the Arts & Humanities
almeidapradomarcella@gmail.com

Adhesion is a central mechanism in the development of multicellular life. It underlies cell movement, communication and differentiation. Remarkably, the presence of only three classes of interconnected molecules is sufficient to establish functional adhesive interactions between cells: Extracellular Matrix components (ECMs); Cell Adhesion Molecules (CAMs); and cytoskeleton microfilaments. In the present work, we shall dissect the nature and the dynamical interactions between these three classes of molecules as they appear in evolution, to postulate that their organization into organic codes parallels the development of multicellularity. Adhesion codes are based on the use of CAMs as adaptors that establish arbitrary, yet specific, correspondences between the extracellular world presented by the ECM, and intracellular world, as shaped by the cytoskeleton. We review the literature to show that these three classes of molecules have very remote evolutionary origins, and were progressively co-opted for adhesive functions, being eventually articulated in the triadic structure that is a signature feature of organic codes. We draw on the published data for the description of molecular and structural novelties that allow for the elaboration of increasingly complex: molecular structures at cell level, *i.e.* the ECM, the cell junctions, the synapses; cell structures at tissue level, *i.e.* biofilms and filaments, the epithelium, the endothelium; and cell behaviors at cell, tissue and systems level, *i.e.* aggregating, polarizing, and networking. We additionally correlate the come into being of new structural and functional patterns of adhesion with some macro evolutionary transitions. To do so, we separately case study prokaryotes, metazoans and vertebrates for their specificities. Finally, in line with the mechanism we have previously proposed for the complexification of Signal Transduction codes, we emphasize that Adhesion Codes also had evolved by becoming coupled to an increasing number of phenotypical transitions along the phylogenetic history of living organisms, namely: aggregation, migration, differentiation and somatic activation. So that changes in physiological conditions as reflected by ECM composition and organization become signs for cytoskeleton remodeling akin to each of the aforementioned processes.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Seven major types of codes

Marcello Barbieri
Dipartimento di Morfologia ed Embriologia
Via Fossato di Mortara 64a, 44121 Ferrara, Italy
brr@unife.it

The existence of many organic codes in living systems has not yet become part of Modern Biology and this means that their role in the great events of evolution is still largely ignored. The first major transition was the appearance of a population of primitive systems, collectively known as the *common ancestor*, that gave origin to the genetic code, but so far no consensus has been reached about the underlying mechanism.
The second major transition was the origin of the first cells, more precisely of three distinct cell lineages that are known as Archaea, Bacteria and Eukarya. The descendants of the common ancestor could produce specific proteins but not specific responses to the environment because they had not yet acquired a modern signal transduction system. As the genetic code was necessary to produce specific proteins, the signal transduction codes were necessary to produce specific cell behaviours, and this is why the origin of these codes was instrumental to the origin of the first cells.
The third major transition was the evolution of the common ancestor of all eukaryotes (LECA), a process that went on for about 2,000 million years, roughly from 3.5 to 1.5 billion years ago. After the genetic code and the signal transduction codes, the prokaryotes evolved no other major code, whereas the eukaryotic progenitors acquired splicing codes, histone code, cytoskeleton codes, compartment codes and others. This suggests that the common eukaryotic ancestor evolved in stages by a multiplicity of organic codes.
The fourth major transition was the origin of embryonic development, and in this case too we encounter a plurality of codes such as the *Hox* code, the adhesive code, the transcriptional codes, the apoptosis code and the bioelectric code.
The fifth major transition was the origin of instincts and feelings, the processes that brought consciousness into existence. They appeared when some organisms became capable of producing mental representations of the signals from the sense organs with mechanisms that are most likely universal because all animals experience hunger, thirst, sexual drive and the like. It is also likely that instincts and feelings are based on codes because there is no necessary connections between sense organs and mental representations.
The sixth major transition was the origin of the 'faculty' of language, a faculty that is probably based on one or more universal codes because all human beings have the potential to learn a language.
The seventh major transition was the origin of culture and in this case there is little doubt on the underlying mechanism because all human artifacts are based on codes.
There are, in conclusion, at least seven major types of codes in the world, but in all cases we face the same basic question: how does a system give origin to a code? This is one of the greatest problems of science and it is precisely this problem that lies at the heart of the new research field of Code Biology.

Towards Synthetic Codes in Biology

Markus Schmidt
Biofaction KG
Kundmanngasse 39/12
1030 Vienna, Austria
www.biofaction.com
www.markusschmidt.eu
schmidt@biofaction.com

Science and engineering usually follow three stages that go from (1) description to (2) analysis to (3) synthesis. Taking genetic code as an example, the description of the DNA molecule by Crick and Watson was followed by the deciphering of the genetic code and subsequent analysis of its evolutionary robustness, possible origins and relative performance compared to alternative codes. With the advent of synthetic biology and especially xenobiology, we are now at the brink of entering the third phase: the design and synthesis of non-canonical genetic codes. This paper will (1) discuss the different motivations for designing alternative genetic codes (e.g. efficiency in industrial biotech, genetic firewall); (2) provide an overview of existing promising concepts, processes and metrics (e.g. from IT and cryptography); and (3) suggest options, opportunities and pitfalls on the pathway towards synthetic genetic codes.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Nuclear inositide and lamin codes: Implications for epigenetic regulation of multicellular organisms

Nadir M. Maraldi
University of Bologna, Bologna, Italy
nmaraldi@area.bo.cnr.it

At the cellular level crucial cross-roads allow the communication between diverse worlds. Environmental chemico-physical signals which cannot enter the cell must be transduced into a cellular language through second messengers. The mechanism of signal transduction takes place at different levels, which are not always represented by physical boundaries. In fact, receptors that act as adaptors able to recognize the agonist and to bind the second messenger, are either inserted into the plasma membrane, or linked to cytoskeletal elements (Toker, 2002), or within the nucleoplasm (Maraldi, 2008). A continuous flux of environmental information reaches the genome, affecting its structural arrangement and causing a variety of responses, including DNA replication, apoptotic death, and selective gene sequence transcription. This control system should be intuitively located at the nuclear envelope (NE); however this is not the case, because the NE undergoes disassembly at each cell cycle and the genome is exposed to the cytoplasmic environment during mitosis. The molecular component that maintains its relationship with the chromatin even during mitosis is represented by nuclear lamins, that during the interphase self-assemble giving rise to the nuclear lamina, and which interact with chromatin-associated proteins during the entire cell cycle. A crucial step of the cell evolution is represented by the appearance of cells with a nucleus in which at least a nuclear lamin is expressed.

Multicellular organisms that undergo an embryonic development require not only differentiation but also cell memory, that allows the cells to indefinitely reproduce their state of differentiation, and cell pattern, that is the ability of cells to arrange themselves in three-dimensional forms with respect to the surrounding cells. This crucial event could involve a complex epigenetic mechanism in which lamins, by interacting with histones and other chromatin associated proteins, may play a central role. Lamins, and particularly type A lamins, have been thus identified as responsible for the phenotype maintenance and referred to as “guardians of the soma” (Hutchinson and Worman, 2004).

Several posttranslational modifications characterizes lamin A expression; in differentiated normal cells the lamin A maturation steps are very efficient and prelamin A species are almost undetectable. Impairment in the process result into altered tissue-specific cell differentiation programs, giving rise to the impressive phenotypic alterations of the diseases collectively referred to as laminopathies, in which the an anticipated or delayed maturation of lamin A, result in perturbation of tissue-specific differentiation programs. Some experimental evidence support a link between differential lamin A variants expression and tissue-specific cell commitment. It is thus conceivable that the lamin posttranslational modifications constitute a *lamin code*, utilized by metazoans cells to induce tissue-specific cell differentiation. Although the rules of this code are at the moment not deciphered, the presence of a set of *adaptors*, represented by NE transmembrane proteins (NETs), and of *effectors*, constituted by epigenetic repressors that modulate the chromatin arrangement and the gene expression, strongly support the possibility that the system could be included in the increasing number of organic codes that characterize cell evolution (Maraldi, 2018).

Toker A. (2002), Phosphoinositides and signal transduction. *Cell. Mol. Life Sci.* 59: 761-779.
Hutchison C.J. and Worman H.J. (2004), A-type lamins: guardians of the soma? *Nat. Cell Biol*. 6: 1062-1067
Maraldi N.M. (2018) , The lamin code. *BioSystem Special issue in Code Biology.* M. Barbieri Ed. *In press*

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Some Scaling Properties of the Nucleon Numbers in the Standard Genetic Code

Nataša Ž. Mišić
Research and Development Institute Lola Ltd, Belgrade, Serbia
nmisic@rcub.bg.ac.rs

Numerous arithmetical regularities obtained for the nucleon numbers of the genetic code constituents (the canonical amino acids and the canonical nucleobases) [1,2,3], indicate the need for a research aimed on determination of the nature of this phenomenon, primarily on *the demarcation weather these regularities are the accidental occurrences or the results of a globally acting ordering process*. One of the approaches to this task is the extension of the initial approach of the revealing of arithmetic regularities, predominantly in the form of the perfect balances of nucleon numbers which are the multiplies of the decimal number 37 [1], by more global and approximate analysis that would include some *scaling properties*.

The motivation for such approach lies in the fact that a global ordering inherently is some multiscale phenomena, and that the number 37 is a unique number in the decimal system characterized by *the discrete scale invariance* [3]. A supporting evidence follows from the relation of the number 37 to the golden mean Φ and its multiplicative inversion ψ. Namely, a simple modification of Φ, ϕ-polynomials gives the irrationals, Ψ and ψ, which integrate *the self-similarity and the scaling by powers of 10*, i.e. $\Psi\psi = 10$ and $\Psi - \psi = 1$, where $\Psi^2\psi = 10\Psi = 37.015 \ldots \approx 37$. The values Ψ and ψ also satisfy *the two basic relations* $[10\Psi] + [10\psi] = 37 + 27 = 64 = [10(\Psi + \psi)]$ and $[10\Psi^2] - [10\psi^2] = 137 - 73 = 64$, where the function $[\cdot]$ rounds to the nearest integer [3].

This approach revealed that a nucleon number distribution in the Standard Genetic Code (SGC) is such that the average nucleon numbers are very close to $10\Psi^2 \approx 137$ for the total amino acid set, $100\Psi \approx 370$ for the total DNA codon set and consequently $10(\Psi^2 + 10\Psi) = 10\Psi^3 \approx 507$ for both sets (valid only for the reduced codon spaces of the SGC by 3 termination codons) [3]. Additionally, the degeneracy pattern, in the form of Rumer's principal division on the *four*fold degenerate and *two*fold degenerate codon halves together with their associated amino acids [4], has a nucleon distribution of the total amino acid sets which satisfy *the second basic* Ψ, ψ*-relation*, i.e. the ratio $64 : 73 : 137$, respectively (valid by the round-nearest integer function $[\cdot]$) [3].

Motivated by the researching related to the circular codes in the genome and the genetic code [5], as wells as by the existence of nucleon balances for the Gamow's division of the SGC [1], here we further investigated a nucleon distribution related to the mentioned SGC properties and the Rumer's canonical nucleobase order $C, G, U/T, A$ [4]. It can be shown that it is possible to execute the bipartition of the SGC so that this two disjoint sets in relation to the total set satisfy *the first basic* Ψ, ψ*-relation*, i.e. the ratio $27 : 37 : 64$, for the next different levels of the reduced SGC by 3 termination codons: the number of codons and the average frequency of single nucleobase, the total nucleon numbers of the amino acid sets, the total nucleon numbers of the RNA and DNA codon sets and consequently the total aggregate nucleon numbers of amino acid set and cognate codon set (valid by the round-nearest integer function $[\cdot]$).

Although the main presented results, along with some other ones, cannot be accepted as evidence of our hypothesis that the arithmetical regularities of nucleon numbers in the SGC are the result of a global ordering principle, they can be considered as the supporting facts for some other theories and hypotheses on the origin and evolution of the SGC. Before all, a nucleon fine tuning in the SGC on the level of molecular *sets* supports the hypothesis that the ancestral genetic code was ambiguous [6,7,8]. It suggests an

existence of some non-local mechanism involved in direct codon/amino acid (group) interactions and the emergence of SGC resulted of top-down and bottom-up causation intertwining [3], what is also compatible with some phylogenomic studies [9] and supports a mineral-mediated origin of the SGC [10]. It supports the hypothesis that the selective driving forces acting during the emergence (ancient phase) and evolution (modern phase) of the genetic code are different [11].

The author acknowledges support from the Ministry of Education, Science and Technological Development of Serbia (Projects: TR 32040 and TR 35023).

1. V.I. Shcherbak, *J. Theor. Biol.* **166** (1994) 475–477; V.I. Shcherbak, *Biosystems* **70** (2003) 187–209; In *The Codes of Life*, M. Barbieri (Ed.), Springer, 2008, 153–188.
2. A.B. Verkhovod, *J. Theor. Biol.* 170 (1994) 327-330; A.M. Downes, B.J. Richardson, *J. Mol. Evol.* **55** (2002) 476–490; M.M. Rakočević, *J. Theor. Biol.* **229** (2004) 221–234.
3. N.Ž. Mišić, *Proc. NEUREL* (2010) 97–102; *Neuroquantology* **9** (2011) 702–715; *Proc. TABIS* (2014) 101–123; *FU Phys. Chem. Tech.* **14** (2016) 275–298.
4. Y.B. Rumer, *Proc. Acad. Sci. USSR* **167** (1966) 1393–1394; *Proc. Acad. Sci. USSR* **183** (1968) 225–226.
5. D.G. Arques, C.J. Michel, *J. Theor. Biol.* **182** (1996) 45–58; C.J. Michel, M. Pellegrini, G. Pirillo, *J. Theor. Biol.* **389** (2016) 40–46; E. Fimmel, A. Danielli, L. Strüngmann, *J. Theor. Biol.* **336** (2013) 221–230; E. Fimmel et al. *J. Math. Biol.* **70** (2015) 1623–1644.
6. C.R. Woese, *Proc. Natl. Acad. Sci. USA* **54** (1965) 1546–1552.
7. J. Maynard Smith, E. Szathmáry, *The Major Transitions in Evolution*, Oxford: Freeman, 1995.
8. M. Barbieri, *Biol. Theory* **10** (2015) 301–310; *Biol. Theory* **11** (2016) 102–112.
9. G. Caetano-Anollés, M. Wang, D. Caetano-Anollés, *PLoS ONE* **8** (2013) e72225; D. Caetano-Anollés, G. Caetano-Anollés, *Life* **6** (2016) 43; In *Evolution of the Protein Synthesis Machinery and its Regulation*, G. Hernandez, R. Jagus (Eds.), Springer, 2016, 9–30.
10. J.P. Ferris, *Elements* **1** (2005) 145–149; H.J. Cleaves, A.M. Scott, F.C. Hill, J. Leszczynski, N. Sahai, R. Hazen, *Chem. Soc. Rev.* **41** (2012) 5502–5525.
11. E. Szathmáry, *Trends Genet.* **15** (1999) 223–229; R.D. Knight, S.J. Freeland, L.F. Landweber, *Trends Biochem. Sci.* **24** (1999) 241–247; *Nat. Rev.* **2** (2001) 49–58.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Horizontal Gene Transfer Promotes the Universality of the Genetic Code

Ned Budisa
Berlin Institute of Technology/TU Berlin, Department of Chemistry, Biocatalysis Group
Müller-Breslau-Straße 10, D-10623 Berlin, Germany,
budisa@chem.tu-berlin.de

The complete ribosomal protein synthesis cycle and codon-amino acids associations are universally preserved in all life taxa on Earth. This process is accompanied by a set of hierarchically organized recognition and controlling events at different complexity levels. It starts with amino acid activation by aminoacyl tRNA synthetases (aaRS) followed by cellular matching with the acceptor units of their cognitive tRNAs ("operational RNA code") and ribosomal codon-anticodon pairing of messenger RNA ("triplet code"). However, this codon-anticodon matching is possible only when protein translation machinery (translation factors, ribosome) accepts an esterified amino acid. This capacity ("charge code") correlates mainly with the amino acid nature and the identity elements in the tRNA 3D structure. A fourth potential "folding code" (also referred as "stereochemical code") between the translation dynamics, sequence composition and folding of the resulting protein can also be defined in the frame of the 'Anfinsen dogma' followed by post-translational modifications. All these coding events as well as the basic chemistry of life are deemed invariant across biological taxa due to the horizontal gene transfer (HGT) making the 'universal genetic code' the 'lingua franca' of life of earth. When cells (or organelles) are prevented from transmitting genetic information (i.e., HGT) the deviations in the above-mentioned coding events become inevitable. A better understanding of these codes, in particular the mechanisms of their conservation in the context of HGT could provide a guide for the experimental engineering of the ribosomal protein biosynthesis machinery. This is highly relevant, among others, in attempts to create synthetic life forms in genetic isolation by using tailored "minimal genomes" and may explain the necessity for multiple coding evens in nature.

References

[1] Kubyshkin, V., Acevedo-Rocha, C. and Budisa N. On universal coding events in protein biogenesis. *Biosystems*. **2017**, *S*0303-2647_30278-2.doi: 10.1016/j.biosystems.2017.10.004.

[2] Kubyshkin, V. and Budisa N. Synthetic alienation of microbial organisms by using genetic code engineering: why and how? *Biotechnol. J*. **2017**, *12* (8), doi: 10.1002/biot.201600097.

[3] Aggarwal, N., Bandhu, A.V. and Sengupta, S. Finite population analysis of the effect of horizontal gene transfer on the origin of a universal and optimal genetic code. *Phys Biol.* **2016**, 13(3):036007. doi: 10.1088/1478-3975/13/3/036007.

[4] Acevedo-Rocha, C. and Budisa, N. Xenomicrobiology: a roadmap for genetic code engineering. *Microb. Biotechnol*. **2016**, *9*, 666-676, doi: 10.1111/1751-7915.12398.

[5] Budisa, N. Xenobiology, New-to-Nature Synthetic Cells and Genetic Firewall. *Curr Org Chem* **2014**, *18*, 936–943, doi: 10.2174/1385272818081406161543O1.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Genetic coding of sense and antisense peptide interactions: *The complementarity code*

Nikola Štambuk and Paško Konjevoda
Ruđer Bošković Institute, Bijenička cesta 54, HR-10000 Zagreb, Croatia
stambuk@irb.hr , pkonjev@irb.hr

During the last four decades a growing body of evidence has been accumulated on the specific interaction of the complementary peptides, i.e. sense and antisense peptides specified by the complementary DNA and RNA sequences. The recognition rules of sense–antisense peptide interactions have been investigated by a number of research groups and authors, e.g. Biro, Blalock, Mekler, Miller, Root-Bernstein, Siemion and Štambuk [1]. Today, the efficiency of this approach, based on complementary coding in 3'→5' and 5'→3' directions, has been experimentally verified for more than 50 ligand–acceptor (receptor) systems, and it represents a promising field of research in biology and medicine [1]. *The complementarity code* of sense and antisense peptide interactions combines the physico-chemical properties of amino acids with bidirectional coding and stereochemical interaction. It is a sequence code using the elements of the Standard Genetic Code (SGC) and the recently-described Carter-Wolfenden tRNA acceptor-stem code (tRNA ASC) [1,2]. As such, it possesses all the criteria for an organic code [3].

Important characteristics of *the complementarity code* are [1]:

1. the translation of complementary amino acid codon pairs extracts 27 sense–antisense amino acid pairs in the 3'→5' direction, 52 pairs in the 5'→3' direction, setting up *the complementarity code* table [1];
2. interactions of complementary amino acid pairs, in terms of hydrophobic (SGC) and lipophilic properties (tRNA ASC), strongly depend on the central purine base of the mRNA codon and its pyrimidine complement of the tRNA anticodon—in both translation directions;
3. complementary translation extracts 3 identical clusters of sense–antisense amino acids that pair according to the character of the second SGC base columns: polar (2^{nd} A) with nonpolar (2^{nd} U) and neutral (2^{nd} G) with neutral (2^{nd} C)—irrespective of the hydrophobic and lipophilic amino acid properties, and translation direction (3'→5' and 5'→3');
4. the pairing of polar–nonpolar and neutral–neutral amino acids is identical for the hydrophobic and lipophilic amino acid properties, despite the fact that hydrophobic and lipophilic amino acid scales are uncorrelated;
5. both mRNA (SGC) and tRNA (tRNA ASC) carry the same information with regard to amino acid hydrophobicity and lipophilicity.

The complementarity code for the peptide-peptide interactions will be discussed in the context of virtual screening, antisense peptide design and *de novo* protein construction [1,4]. The presented algorithms could be used for the analysis of the structure, function and evolution of protein and nucleotide sequences.

References

[1] Štambuk, N., Konjevoda, P., Turčić, P., et al., 2017. Biosystems S0303-2647(17)30279-4; doi: 10.1016/j.biosystems.2017.10.009. [Epub ahead of print]
[2] Carter, C.W., Wolfenden, R., 2015. Proc. Natl. Acad. Sci. USA. 112, 7489–7494.
[3] Barbieri, M., 2015. Code Biology—A New Science of Life, Springer, Cham.
[4] Štambuk, N., Konjevoda, P., 2017. Information 8, 29; doi:10.3390/info8010029.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Sparse DNA representation to identify genomic codes in the human sequence

O. Paredes, R. Romo-Vázquez, J. A. Morales
Departamento de Ciencias Computacionales, CUCEI,
Universidad de Guadalajara, Mexico
omar.paredes@alumnos.udg.mx

The genome is all the information that is encoded inside the DNA molecules, information that is encoded in genomic codes. A genomic code is a set of specific rules (mapping) that correlates the set of signs, i.e. DNA sequences, and the set of significance, i.e. biological functions as coding, splicing sites, regulatory function, and others. Examples of genomic codes includes genetic code, biosequence codes, splicing code and regulatory codes. Trifonov proposed that the codes in the biosequences are sequence patterns of variable sequences [1]. However, in the symbolic DNA representation these patterns may not be easy to find due to the unknown size of the patterns and the overlapped property of them. A numerical DNA representation of the signs or genomic signal may be a suitable representation because there are techniques in signal processing that look for patterns or periods that are the numerical representation of the genomic codes. In this work we proposed the implementation of signal processing techniques (i.e. Fourier and wavelet transform) to map the symbolic DNA to a sparse representation that is suitable to identify some genomic codes. We found that human regulatory regions with different significance can be distinguishable between them by its genomic codes in the domain of the signal transforms sparse representation. These results suggest that the genomic codes can be identified and characterized in the human and other organisms in representations more suitable that the deep-rooted symbolic DNA representation. Furthermore, the identification of other genomic codes in the daily increasing number of sequence genomes gives a strong support of the codepoietic nature of the genomes.

[1] Trifonov EN: The multiple codes of nucleotide sequences. *Bull Math Biol* 1989, 51:417–432.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Coding for randomness in evolutionary models of bet hedging

Omri Tal
Max Planck Institute for Mathematics in the Sciences
Inselstr. 22-26, 04103, Leipzig, Germany
omrit1248@gmail.com , omri.tal@mis.mpg.de

Kelly's famous work on optimal gambling strategies and the financial value of information was arguably the first convincing attempt at applying concepts from information theory for analysis in a different field – and this in light of Shannon's contemporary famous caution against jumping on "the bandwagon" of information theory. Evolutionary bet hedging models under fluctuating environments in the literature have followed rather directly from this earlier financial framework, utilizing notions such as growth rate maximization and proportional betting, while providing an operational interpretation of an information channel. I will highlight the analytic and conceptual limitations of such models, propose further generalizations to the information-theoretic framework to better account for evolutionary processes, and in particular focus on a serious gap in research, a missing link -- the need for a (genetic) code or mechanism for adaptive developmental phenotypic *randomization*.

On Nullomers and Circular Codes

Oreste Piro
University of Balearic Islands, Mallorca, Department of Physics
oreste.piro@gmail.com
Simone Giannerini
University of Bologna, Department of Statistical Sciences
simone.giannerini@unibo.it
Diego L. Gonzalez
CNR/IMM Bologna Section
gonzalez@bo.imm.cnr.it

A nullomer is a sequence of nucleotides that cannot be found in a given DNA sequence (or chromosome, or complete genome, etc.). In short, it is a forbidden sequence. Some studies have shown that nullomers are overrepresented and highly conserved [3]. It has been hypothesized that forbidden sequences are associated to deleterious effects on organisms. Thus, research in this area might be useful, for example, for pesticide and antibiotic development, for drug target identification, for species classification and evolution, etc [1,2]. On the other hand, circular codes have been demonstrated to be present in coding sequences and they might be related to the mechanisms of retrieving and maintenance of the correct reading frame by the ribosome in protein synthesis [4,5]. Circular motifs, i.e. sequences of codons that belong to circular codes, are over-represented in coding sequences [6]. In addition, in the theory of symbolic dynamical systems, the recognition of absent sequences is a fundamental tool to construct dynamical rules from data [7]. Several methods have also been developed to construct grammars associated to the underlying symbolic dynamics when the generating system is known. In addition, in the theory of symbolic dynamical systems, the recognition of absent sequences is a fundamental tool to construct dynamical rules from data. In this contribution we study possible connections between these two important aspects of coding sequences from the point of view of dynamical systems theory.

References

[1] A.J. Pinho, P.J.S.G. Ferreira, S.P. Garcia and J. Rodrigues, *On Finding Minimal Absent Words*, BMC Bioinformatics, 10, 137 (2009).
[2] G. Hampikian and T. Andersen, Pac. Symp. Biocomput., 355-66, *Absent Sequences: Nullomers and Primes* (2007).
[3] D. Vergni and D. Santoni, *Nullomers and High Order Nullomers in Genomic Sequences*, PLOS One, 0164540, (2016)
[4] D.G. Arquès and C.J. Michel, *A complementary circular code in the protein coding genes*, Journal of Theoretical Biology, 182, 45-48 (1996).
[5] D.L. Gonzalez, S. Giannerini and R. Rosa, *Circular codes revisited: A statistical approach*, Journal of Theoretical Biology, 275, 21-28 (2011).
[6] K. El Soufi , C.J. Michel, *Circular code motifs in genomes of eukaryotes*, J Theor Biol. 2016 Nov 7;408:198-212.
[7] J. B. Pierrehumbert, *On attributing grammars to dynamical systems*, Journal of Phonetics 18, 465-477 (1990).

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

The structure of the genetic code as an optimal graph clustering problem

Paweł Błażej
Department of Genomics, Faculty of Biotechnology
University of Wrocław, Poland
pawel.blazej@uwr.edu.pl

Various methodologies have been used to study the features of the standard genetic code (SGC) and assess the level of its potential optimality. Here, we introduced a new general approach to evaluate the quality of the genetic code structure. This methodology comes from graph theory and allows us to describe new properties of the genetic code in terms of conductance. This parameter measures the robustness of codon groups against the potential changes in translation of the protein-coding sequences generated by single nucleotide substitutions. We described the genetic code as a partition of an undirected and unweighted graph, which makes the model general and universal. Using this approach, we showed that the structure of the genetic code is a solution to the graph clustering problem. We presented and discussed the structure of the codes that are optimal according to the conductance. Despite the fact that the standard genetic code is far from being optimal according to the conductance, its structure is characterized by many codon groups reaching the minimum conductance for their size. The SGC represents most likely a local minimum in terms of errors occurring in protein-coding sequences and their translation.

The spectrum of biomolecular codes: an informational perspective

Pedro Marijuán, Jorge Navarro, Raquel del Moral
Bioinformation Group
Aragon Institute of Health Science
(IACS)/Aragon Health Research Institute (IIS Aragón). 50009 Zaragoza, Spain.
pcmarijuan.iacs@aragon.es

We argue that the living system is composed by elements with different *informational architectures*. It involves the distinction of different modalities of molecular recognition. Clearly three main classes of informational architectures in the living cell can be distinguished depending on the specific molecular recognition processes they perform: a) the "sequential" (DNAs, RNAs) based on the mutual recognition between *complementary* motifs in molecular chains that can be joined and separated; b) the "processing network" based on specific *shape recognition* performed by enzymes and proteins upon their substrates or targets; and c) the "structural" (membranes, organelles, cytoskeleton, etc.) based on the mutual recognition of *identity* in the relationships between components. These three ways of molecular recognition can be easily expressed in symmetry terms as well. Thereafter, the self-construction of the cellular system via the central genetic code actually implies a number of *microcodes*, whenever and wherever information has to travel with specificity from a particular class of architecture to another—i.e., when sequential motifs have to be copied or be recognized by enzyme processors, when proteins have to recognize each other, or when organelle structures have to be built or modified. Whatever functional operation has to be performed by the synthesized enzymes and proteins, it implies the generation of new recognition codes among the heterogeneous architectural partners involved (e.g., for protein modification, transportation, splicing, complexes, etc.), including the final degradation in the proteasome as well. All complex functions would partake of similar basic designs based on an assemblage of low-level and middle-level codes for the matching between the different architectures involved.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Self-maintaining Molecular Codes

Peter Dittrich
Bio Systems Analysis Group, Institute of Computer Science
Friedrich-Schiller-Universität Jena
07745 Jena, Germany
peter.dittrich@uni-jena.de

Previously, a formal method to assess the semantic capacity of a chemical reaction network to process "meaningful" information has been suggested (Görlich/Dittrich, PLoS ONE, 8(1), e54694, 2013). The basic idea is to measure how easy it is to implement with this network a molecular code, which is an arbitrary (contingent) mapping between species, that is, a mapping that cannot be inferred from knowing the species alone. A preliminary computational analysis of various chemical systems revealed a quite large spectrum of different semantic capacities. The hypothesis has been derived that life over the course of evolution is gaining access to chemistries with increasing semantic capacity. A recent study revealed molecular codes in metabolic network models; however basically each code contained an ubiquitous species like water, so that the contingency of the codes found is questionable. As a consequence, the theory has to be refined along the following two questions: (1) Which component of a (molecular) system should be regarded as a sign, meaning, or code-maker, respectively? This would be important for deriving experimental procedures and algorithmic tools for finding and characterizing molecular/organic codes. In the current molecular code approach a sign (and meaning) is represented by the presence of a particular molecular species. There a countless alternatives, like a particular concentration interval of a particular species. (2) A second closely related question is: Can the molecular code be actually been used and how useful is it? If an ubiquitous species, like water, is a sign (or meaning), the code is obviously not usable. In another case, if the context (code-maker) cannot be generated and maintained by a system the code cannot be realized (implemented) and thus not been used. As a potential solution, this contribution explores an extension of the molecular code concept by including the formal concepts of self-maintenance and chemical organizations and the less formal concepts of autopoiesis and unity. The idea is to say that a code is usable, if a code-maker (context) exists that is contained in a chemical organization. Thus this requires a code-maker to be contained in a self-maintaining set Furthermore, we can demand that signs and meanings are not in the closure of this self-maintaining set. The concept is presented along various examples and its implications for analyzing molecular codes in real systems is discussed.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Does genetic coding need quantum biology?

Peter R Wills
Department of Physics, University of Auckland PB 92109,
Auckland 1142, Aotearoa-New Zealand
p.wills@auckland.ac.nz

Uniquely quantum effects and phenomena such as quantum superposition, coherence, entanglement and tunnelling have been found sometimes to play a pivotal role in the molecular chemistry of a broad range of biological processes, including olfaction, direction-finding, enzymatic catalysis, photosynthesis, and electron transport. These findings have fueled speculation that the extraordinary functional order which organisms maintain and on which all biological systems depend is fundamentally a manifestation of the quantum nature of the universe, that is to say, biology cannot be explained in terms of classical physics.
This would seem to imply that biological systems process information ("compute") in a quantum fashion, dealing in qbits, rather than classical information of the sort found in the works of Shannon, Turing, von Neumann and others. If the introduction of quantum information and computing into biological theory sheds new light on the coding problem it will have revolutionary consequences. However, I will argue that Quantum Biology is something of a distraction from the conceptually more difficult theoretical problem of understanding how it is that molecular information ever acquired encoded meaning and enabled the genetic inheritance of phenotypic traits. It is better first to follow the paths of Code Biology and Biosemiotics to discover what understanding can be gleaned from thinking in classical terms before embarking on the incorporation of quantum complexities into our analysis of the genotype-phenotype relationship that underpins our current evolutionary explanations. I will illustrate this with reference to recent work on the origin of the hardware and software of the genetic code.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

From genotype and phenotype, to sociotype: the essential codes of human life

Raquel del Moral, Jorge Navarro, Pedro Marijuán
Bioinformation Group
Aragon Institute of Health Science
(IACS)/Aragon Health Research Institute (IIS Aragón). 50009 Zaragoza, Spain.
rdelmoral.iacs@aragon.es

In the framework of the "sociotype" study, a new construct developed by the authors within the conceptual triad genotype-phenotype-sociotype that allows the theoretical integration of essential qualitative and quantitative aspects of social networking, we explore the conjoint metrics of both social structures and social relationships (natural conversation) of individuals. *Each individual establishes his/her fundamental codes of the social niche within his/her cortical memories* by distinguishing four relational dimensions: family, friends, work/study and acquaintances. Thereafter, we have investigated the respective figures – structural and communicational – for these four domains. In this pursuit, the two questions of *how much do we talk* and *with whom* have been quantitatively surveyed for a total of 1,475 individuals (1,000 female, 49.79 median age), who were interrogated about their social interactions by means of the SOCQ sociotype questionnaire and also other questionnaires (GHQ12, UCLA, Eysenck) and about the time spent in conversation, as well as about the number of social contacts within the different realms. It is interesting that the results obtained about the number of social bonds differ from well-known assumptions such as the Dunbar number. The results about conversation time seem to be in agreement with some of the scarce literature available. Finally, given the non-Gaussian distribution obtained for our data concerning both the number of social contacts and the conversation times, we have explored the plausibility of a recently proposed fit: the Planckian distribution equation. It makes sense that our personal codes of the social niche are caught into the inner competition for resources of our interpersonal memories that mobilize our "attention economy".

Dichotomic Classes and Entropy Optimization in Coding Sequences

Simone Giannerini, Diego Gonzalez, Greta Goracci
University of Bologna, Department of Statistical Sciences
simone.giannerini@unibo.it
gonzalez@bo.imm.cnr.it
greta.goracci@unibo.it

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

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.

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.

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.

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.

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.

S. Giannerini, G. Goracci, (2018) "Small sample asymptotics for goodness-of-fit tests.", Technical Report, University of Bologna, Department of Statistical Sciences.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Contributions of Genetic Code Structure to the Evolution of Overlapping Genes

Stefan Wichmann, Zachary Ardern, Klaus Neuhaus, Siegfried Scherer
Technical University of Munich
stefan.wichmann@tum.de

The redundancy in the genetic code's mapping between codons and amino acids enables multiple amino acid sequences to be encoded in parallel on a DNA strand. This theoretical property of the code is known to be utilised in viruses. There is increasing evidence for overlapping encoding in other organisms, including many genes fully embedded within other coding sequences in alternate reading frame.
The structure of the standard genetic code places constraints on sequence evolution in general, and on the evolution of proteins encoded in overlapping genes (OLGs) in particular. Here we investigate some aspects of the code's structure which are not found among similar, but randomly generated codes, and which contribute to the evolution or conservation of OLGs.
The genetic code has previously been found to be 'optimal' for various properties such as robustness to point mutations. There has been some work on the contribution of the code's structure to the evolution and conservation of overlapping genes, but most of this has focused on the "minus one" frame which is the direct translation of the reverse complement of a coding sequence without offset, unlike with the other frames. But, in light of the increasing evidence for OLGs across all organisms, the other frames (i.e., +2, +3, -2, -3) also require attention, along with the relationship amongst different properties contributing to the evolution of genes in each possible reading frame.
As part of the groundwork for this larger project, we improved and quantified the method used to determine optimality. The main difficulty with the previously used methods is the choice of the 'negative control' set of alternative genetic codes. Almost every putatively optimal quantity in the literature was determined with a different negative control set. We present and discuss these aspects of the methods, and discuss their relevance to the evolution of OLGs:
1. Extent of 'optimality' depends strongly on the 'negative control' comparison code set chosen (observed for average overlapping ORF length estimate, frameshift error robustness, mutational robustness of OLGs
2. Extent of 'optimality' differs by reading frame (observed for: average overlapping ORF length estimate, frameshift error robustness, mutational robustness of OLGs).
3. A hierarchy of optimal properties can be constructed. Properties higher in rank are optimal when compared to larger comparison code sets; lower ranking properties are optimal in smaller, more restricted, code sets.

Simplexity and vicariance: a role for organic codes?

Stephen J. Cowley,
University of Southern Denmark
Cowley@sdu.dk

Cognitive psychology shows that more than one kind of process shapes human thinking (e.g. Evans & Frankish, 2009). Hughlings Jackson (1874) identified a similar disjunction in language that, before computationalism, was widely accepted. I use the concepts of *simplexity* (Berthoz, 2012) and *vicariance* (Berthoz, 2017) to link these 'dual process' views. Thus, heterarchichal control enables neurophysiology to set off processes that override their own ongoing organization.

Living organization evolved to serve bio-systems: Berthoz traces this to simplex mechanisms whose properties ensure that, in time, new kinds of simplexity evolve. The case connects the eye-ball's rolling, its part in seeing, how looking develops, and how cultural norms constrain human gaze. Heterachical control allows slower processes to stand in for faster ones; flexibility emerges with vicariance (where one process substitutes for another). Light detection can be managed by the CNS, seeing can aid movement by an organism-in-an-environment, looking can become as a social resource, and gaze can be co-opted by culture. In vicariance a slow system (A2) offers a process based in a disjunction in the operating of a faster one (A1). Having clarified the model, I offer two suggestions. First, language co-varies with attention: slow vocalizing-hearing is able to mesh with looking such that speaking and action can override seeing (one can think/speak *as* one attends). Second, the link may use organic coding (see, Cowley, 2006). Thus, when a slower system (A2) overrides a faster one(A1's ongoing responding) it sets off a semantic synthesis that can set a parameter. Thus seeing can ground brain-based learning, an organism can manage looking, its looks can be used in coordinating action and, using gaze, cultural values can be integrated with how a person draws on normative practice.

References

Berthoz, A. (2012). Simplexity: Simplifying Principles for a Complex World (G. Weiss, Trans.) Yale University Press.

Berthoz, A. (2017). *The Vicarious Brain, Creator of Worlds*. Harvard University Press.

Cowley, S. J. (2006). Language and biosemiosis: Towards unity?. *Semiotica*, *2006*(162), 417-443.

Evans, J. S. B., & Frankish, K. E. (2009). *In two minds: Dual processes and beyond*. Oxford: Oxford University Press.

Hughlings Jackson, J. (1874). On the nature of the duality of the brain. *Medical Press and Circular*, *1*, 80-86.

Fifth International Conference in Code Biology – Granada (Spain) 5-9 June 2018

Negative entropy is necessary, but not sufficient to define information

Wanderley Dantas dos Santos
State University of Maringa, Department of Biochemistry,
Av. Colombo, 5790, Zona 7, Bloco I-89, room 15, Maringa, Parana, Brazil
wdsantos@uem.br

Boltzmann defined entropy in terms of the number of possible microstates the components of a system can assume:

$$S = k_B \ln w$$

where, k_B is the Boltzmann constant expressing the amount of kinetic energy of each particle in function of the temperature.

Later, Shannon established that information *I* is a decision between at least two options and could be measured as the log of the number of options (symbols) in a message. In a binary code, each digit can assume one of two states (0/1), thus the amount of information varies with the base 2 log of the number of bits. Leon Brillouin generalized this idea to show that the amount of information in a physical system can be measured as the logarithm of the ratio between the initial number of possible structures P_0, and the number of possible structures P_1 after a gain in information (I_1>0):

$$I = \mathrm{K} \ln P_0/P_1.$$

Here, K is a proportionality constant that can be adjusted according to the desired unities. For K=$\log_2 e$, then $I=\log_2(P0/P_1)$ gives the information in bits. For K=k_B then, $I=k_b \ln(P_0/P_1)$ gives information in Joules K^{-1}, the unities of entropy. Considering the initial entropy $S_0 = k \ln P_0$ and the final entropy (after a gain in information) $S_1 = k \ln P_1$, it is possible to deduce that

$$S_1 = S_0 - I_1.$$

demonstrating that information = negative entropy or negentropy. The physical interpretation is that a chemical bound is a decision (defined state) among possibilities (other possible microstates): the very definition of information provided by Shannon. The amount of bonds is inversely proportional to the entropy (1/w); therefore, information can be expressed in reciprocal terms to those of Boltzmann entropy:

$$-S = k \ln (1/w)$$

In contrast with these quantitative treatments, more recently, Marcello Barbieri examined qualitative features of information. In informatics, semiotics and molecular biology information is transported in specific code sequences (bits, letters or monomers, respectively). These sequences are not determined by the intrinsic properties of their components, since many different sequences are equally possible. Instead, they always result from a 'coding process' in which the specific sequence of symbols (binary, alphabetic, genetic) is oriented by a 'guideline'. A piece of information (byte, word, gene), although is an objective physical entity, is irreducible and immensurable: can only be described by naming their components, in their exact order. In turn, most homogeneous phases (e.g. water) are made by 'monostable' components, i.e. components that can assume only one distinguishable stable state and, therefore, 'cannot form codes'. Therefore, a gain in negentropy (e.g. by freezing it) does not result in a gain in information.

In conclusion, while a gain in information requires a gain in negentropy, a gain in negentropy does not necessarily implies a gain in information. Because of its symbolic nature, physical information can only be defined on a system capable to produce codes as those composed by latches, genes and so on.