VARIETIES OF BIOLOGICAL INFORMATION: A MOLECULAR RECOGNITION APPROACH TO SYSTEMS BIOLOGY AND BIOINFORMATICS

Jorge Navarro, Ángel Goñi-Moreno & Pedro C. Marijuán

Abstract: *Bioinformatic and systems biology developments should be accompanied not only by a plethora of computer tools, but also by an in-depth reflection on the distinctive nature of biological information. In this work we attempt a consistent approach to the multiple varieties of information in the living cell by starting out from the conceptualization of molecular recognition phenomena. Subsequently, an elementary approach to the "informational architectures" behind cellular complexity may be chartered. In the interplay of the different informational architectures two cellular subsystems should be highlighted: on the one side the transcriptional regulatory network, and on the other, the cellular signaling system that is in charge of the interrelationship with the environment. The embodiment of functional agents and the peculiar handling of DNA sequences along the evolutionary process will suggest a parallel with the von Neumann scheme of modern computers, including the cellular capability to "rewrite the DNA rules" along ontogenetic development.*

Keywords*: Molecular recognition, Informational architectures, DNA addresses, Transcriptional regulatory network, Cellular signaling system, von Neumann scheme.*

ACM Classification Keywords*: D. Software. D.1 Programming Techniques*

Introduction: molecular recognition as a key to biological organization

The approach followed here is based on a bottom-up strategy, linking informational architectures, and in general the structures of cellular organization, to molecular recognition modalities [Conrad, 1996], [Marijuán, 2003, 2009].

Molecular recognition is but one of the fundamental territories of chemistry. Paraphrasing Shaik's words [Weinhold and Landis, 2007], it is the central element from which an entire bio-chemical and evolutionary universe is constructed. Actually, molecular *specificity* and molecular *affinity*, which provide the ground for any molecular recognition phenomena are amongst the most essential concepts of classical chemistry and molecular sciences —as all chemical reactions are based on the relative specificity of the intervening molecular partners and on their mutual affinity or free energy availability. It means, in other words, that the "making and breaking of bonds" is what makes possible the mutual recognition and the formation of complexes between biomolecular partners.

By far, it is in the highly heterogeneous molecules that constitute living matter where the chemical phenomenon of molecular recognition reaches its maximal universality, ubiquity, and combinatory capabilities. The myriad of molecular recognition encounters at the cytoplasm of a bacterium are taking place in a highly organized and systematic way: no "insulating wires" are needed. And this is one of the most remarkable information processing resources of the living cell (a wired cell would be unthinkable of at the molecular scale!).

Apparently, all the multitude of specific molecular matching events in the cell are occurring on a case by base basis, beyond any useful molecular taxonomy, but this is not the case, as we are going to discuss. Precisely we will distinguish classes of informational architectures based on molecular recognition considerations. The most important functional interrelationship in the living cell concerns the population of protein biomolecular agents which are built by means of transcriptional and translational processes performed upon the sequential

arrangement of nucleic acids. This means the class of "diluted" architectures based on networks of enzymes and proteins versus the class of "sequential" architectures.

We will particularly discuss how a network of gene expression relationships is organized within a concrete living cell (*Mycobacterium tuberculosis*) and how the topological governing of this network is deployed by the cellular signaling system in charge of the interrelationship with the environment. The embodiment of functional agents and the peculiar handling of DNA sequences along the evolutionary process will suggest a parallel with the von Neumann scheme of modern computers, including the cellular capability to "rewrite the DNA rules" along ontogenetic development.

Looking for a unitary molecular recognition background

Regarding the question of how many specific recognition encounters may be distinguished within the biomolecular "soup" of any living cell [Goodsell, 1991], it is surprising that in spite of the ubiquity and universality of biomolecular recognition phenomena, they are not well focused in their general categorization yet.

Molecular recognition, like any other specific chemical reaction, simply implies the "making and breaking of bonds". The problem with biomolecular recognition instances is that they involve an amazing variety and combinatorics of almost any type of chemical bond (and particularly of Coulombian motifs), which together provide specificity and affinity to the intermolecular encounters: covalent bonds, hydrogen bonds, hydrophobic / hydrophilic forces, dipole forces, van der Waals forces, ionic Coulombian forces, etc. Dozens or even hundreds of weak bonds may participate, for instance, in the formation of a protein-protein specific complex.

Quite probably, measuring molecular recognition and establishing its crucial parameters and variables can only be realized biologically on a case-by-case basis. At least this is the current trend in most molecular biological and molecular dynamic approaches.

A few references, however, could provide some interesting insights about molecular-recognition generalities. First, [Meggs,1998] about "biological homing", contemplated particularly from a Coulombian "lock and key" combinatory point of view; then [Lin, 2001] about the changes in thermodynamic entropy and entropy of mixing derived from molecular similarity changes; and finally [Carlton, 2002], with original proposals for measuring the information content of any complex molecular system.

Symmetry considerations

The usefulness and depth of symmetry considerations in molecular recognition phenomena, as emphasized by [Lin, 2001], are self-evident. Symmetry allows a direct classification of biomolecular recognition occurrences by means of three ordering categories: *identity, complementarity,* and *supplementarity.* They respectively mean: recognition by sharing identical molecular properties (e.g., self-organization of phospholipids in membranes), recognition by means of complementary properties of the molecular partners (e.g., moieties, or the nucleic acids' double helix), plus recognition through a quasi-universal capability to wrap or envelop any molecular shape by building a complex molecular scaffold of weak bonds around the target (e.g., enzymic active sites, protein complexes).

In the *supplementarity* case (not contemplated by Lin's approach), the partial surfaces involved are inherently sloppy in their specificity and have a very variable affinity. Possibly we could keep calling *complementarity* to this facultative and highly variable interrelationship, but at the cost of leaving in the dark a very interesting distinction with respect to the very clean and holistic matching between complementary moieties (molecular fractions). Let us illustrate the additional difference introduced by means of a literary metaphor. Daniel Defoe's character, Mr. Gulliver, could be matched by *identity* with his seaman fellows (dressing the same uniform, for instance), or by

complementarity in the relationships with his loving wife Mrs. Gulliver; but throughout *supplementarity* he was matched by a motley crew of Lilliputians who had built many kinds of small 'bonds' around his bodily parts.

Informational architectures

From an organizational point of view, the previous categories based on symmetry considerations would be reflecting the global distribution of molecular functions within the cell, the different classes of informational architectures (see Table 1):

- *identity* in the structural self-organization of membrane and cytoskeleton support systems (the structural & support architecture),

- *complementarity* in the informational memory-banks of nucleic acids (the sequential architecture),

- *supplementarity* in the active sites and recognition-surfaces of enzymic molecular machinery (the diluted, processing architecture).

Identity	Complementarity	Supplementarity
nucleotides/RNA	RNA/RNA pairing	enzymes/substrates
nucleotides/DNA	RNA/DNA pairing	enzymes/effectors
amino acids/protein chains	RNA/ribozymes	enzymes/cofactors
phospholipids/membranes	RNA/ribosomes,	enzymes/proteins
tubulins/microtubules	RNA/amino acids	antibodies/antigens
actins/microfilaments	RNA/ribonucleoproteins	receptors/peptides
clathryn/vesicles	DNA/DNA pairing	receptors/transmitters
carbohidrates/glycoproteins	DNA/polimerases	receptors/ligands
lipids/lipoproteins	DNA/promoters	receptors/hormones
	DNA/histones	channels/ions
	DNA/transcription factors	channels/nucleotides
	DNA/repressors	channels/ligands
		proteins/chaperons
		proteins/protein kinases
		proteins/proteinphosphatases
		proteins/proteases
		proteins/proteasomes
		proteins/converter enzymes
		proteins/protein multimers
		proteins/protein complexes
		proteins/protein machines

Table 1. Basic categories of molecular recognition in the living cell.

Together these three architectural classes integrate a "universal processing and constructing system" capable of exploiting an endless variety of boundary conditions at the molecular scale. In the astonishing matching games performed by the molecular crews of the living cell, so to speak, the "Lilliputian populations" made out from amino acids are coded into a sequential genome, are altered systematically, and are evolutionarily selected. This cellular "society" becomes organized in a very sophisticate way so that both the *functions* and the circumstantial *addresses* of the multifarious molecular crews —as we will discuss— are put together into the same information bank.

Embodiment of the functional agents: the diluted architecture

Enzymes and proteins, the agential stuff of the "diluted architecture" coded onto the DNA, appear as flexi-molecular machines with a life cycle of their own [Ho, 1995]. Their constitutive structure of linked amino acids is permanently caught into a state of flow, from birth at ribosomes to final degradation at proteasomes. In actuality, it is in the enigmatic folding process taking place at chaperons (in itself a computational NP-problem) where enzymes and proteins acquire their machine-like characteristics, which enable them to perform a regular function within the cell.

Enzyme (and protein) function is but a continuation of the folding process. Apparently it implies a clear and regular succession of enzymic states: specific molecular recognition of the substrate, mutual coupling, lowering of the activation energy, interconversion between forms of energy, exit of the substrate transformed into product, and culmination of a regular work cycle [Marijuán and Westley, 1992], [Urry, 1995]. As a matter of fact, classical biochemical approaches have described this regular functioning of the enzyme through deterministic rate equations, non-linear ones that are often analyzed in a linear simplified way by means of control theory.

Nevertheless, this functioning may also be approached probabilistically. A stochastic dynamics –*molecular automata*– where enzymes "fire" their state transitions according to probabilities derived from the free energy differences in between states, can be more realistic than classical equations of control theory [Marijuán, 1994]. Moreover, such probabilistic dynamics would be closer to the stochastic nature of transitions in the "post-folding" energy landscape from which the different states of the enzyme cycle are derived [Frauenfelder, 1991], [Shimizu and Bray, 2001].

Apart from the classical discussion about determinism versus stochasticity in the enzyme's function, we have to pay attention to the role that the *embodiment* of the function plays in the way such functionality is deployed cellularly. For instance, the whole organization of degradation processes or *degradomics* (traditionally forgotten) nowadays appears almost as complex as the transcription process itself [Marijuán, 1996, 2002]. The very duration of the biomolecular agent depends on this planned process of degradation. Besides, some of the striking ecological regularities found in living organisms —so to speak, depending on their "efficient" biomass— might be related to the commonality of processes or stages in the life cycle of each molecular agent. Concretely, animals in their metabolic rates and life spans [Atasanov, 2005], and plants in their photosynthetic surfaces and life spans [Wright, 2004], deploy an amazing constancy that can be explained only by taking into account the strict coupling at the molecular level between stochastic dynamics and degradation process, for any enzyme or protein functionally active.

Primary versus secondary addresses

A parsimonious approach to the function of the biomolecular agent has to pay attention not only to the functional *"what"* dictated in the active site of the enzyme, and to its global *duration*, but also to a series of accompanying processes distributed over different parts of the molecular structure, which may include: modulation by effectors,

intracellular transportation, permanent (post-translational) modification, formation of complexes, time-frames derived from transcription and translation, and as already said the final (or partial) degradation.

Thus, the *"what"* of the functional clause should be accompanied by many other circumstances such as: *how fast, where, which way, with whom, when,* and *how long*. In general, the functionalities of the active site and the retinue of accompanying processes are independently defined onto the DNA sequences, constituting *addresses* which are separately coding for function (*"primary address"* coding the active site), and also for the other operation of control, transportation, splicing, modification, complexes, transcription-translation, degradation, etc. (each one implying some specific *"secondary address"* in the DNA coding, irrespective that they may be functionally operative in the DNA, RNA, or in the protein stages).

In prokaryotes, the global arrangement of embodiment processes is simpler than in eukaryotes, in correspondence their protein components are smaller and contain fewer domains comparatively. The possibility of systematic tinkering upon multiple modules and domains becomes one of the most distinctive evolutionary strategies of eukaryotes, the tool-box of their multicellularity. A serial-combinatoric arrangement of exons and introns (which usually constitute folding domains), tailored for each tissue by differential splicing, allows eukaryotes a far bigger proteome than prokaryotes (around one or two orders of magnitude) without multiplying the number of genes involved [Claverie, 2001].

By tinkering and playing combinatory games upon exons and introns containing a vast array of secondary addresses, eukaryotic cells may systematically explore and change the whole boundary conditions surrounding the triggering of each biomolecular function —mastering all those circumstances of *when, where, how fast, which way, for how long, with whom,* etc., which together co-determine the functional action of any eukaryotic enzyme or protein [Marijuán, 2003].

The generation of variety within biological genetic algorithms is surprisingly complex in most eukaryotic genomes, potentially involving occurrences such as: SNPs, repetitive DNA, mobile elements, transposons, retrotransposons, telomere shortening, gene and segmental duplications, chromosome fissions and fusions, whole genome duplications, symbiosis, etc. The striking complexity of eukaryotic bauplans and organismic physiologies has been achieved only by the combined action of all these engines of variation impinging upon the set of different addresses involved in the embodiment of the biomolecular functional agents.

Subsequently, in the evolutionary problem-solving strategies of prokaryotic and eukaryotic cells, their very different DNA grammars would imply crucial differences. Their respective universality in evolutionary problem solving is addressed directly towards the solution of *molecular "assimilation"* phenomena in one case, while in the other case it is addressed towards harnessing *molecular "organization"* phenomena (morphology and differentiation).

Controling gene expression: the sequential architecture

The elements of the diluted architecture are all of them coded into the sequential architecture, thus the control of gene expression by the elements of the former will give an overall picture of the self-modification capabilities of the cellular system. Traditionally most studies have focused in the expression of individual genes and not in the texture of the overall network or in the internal/external instances of control concerning the guidance of gene expression. Currently, however, transcriptional regulatory networks are built for different prokaryotic microorganisms and eukaryotic specialized cell-types or cellular functions.

As an instance of such networks, the authors have compiled a large-scale *M. tuberculosis* transcriptional regulatory network, which has been built upon a previously published TR network [Balázsi, 2008] the largest to date, with further addition of different kinds of resources pertaining to publicly available sources: DNA

microarrays, operons, orthology approaches, and synthetic biology experiments [Navarro & Marijuán, 2010]. See Figure 1. Our compilation forms part of ongoing studies tending to the development of a *new vaccine* based on the mutant strain SO2 [Martín, 2006]. The objective is to contribute to a better understanding of both the transcriptional control by the system and the re-organization of the cell cycle that takes place in the different environments, as well as gauging the impact of the SO2 strain on physiological and immune systems of the body.

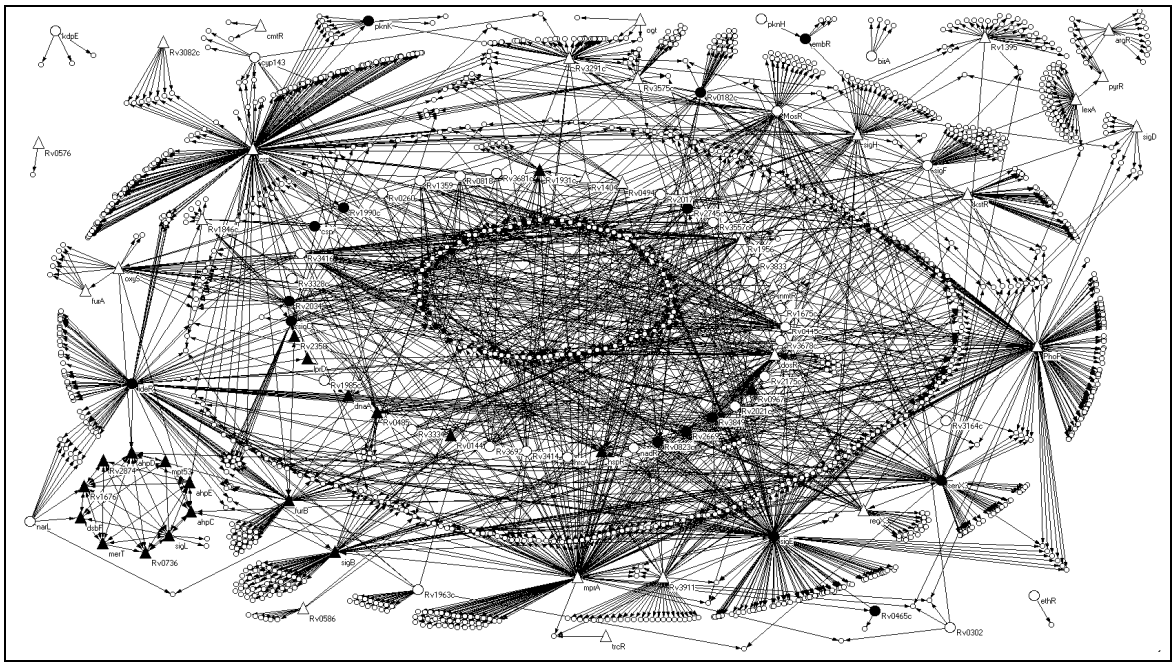

Figure 1. The Transcriptional-Regulatory (ETR) Network of *M. tuberculosis*. Nodes represent *Mt*'s genes, and links represent their regulatory interactions. Transcription factors appear either green or blue, depending on whether they have known transcriptional regulator or not. The white nodes represent output elements without transcriptional activity. The triangle nodes represent protein transcription factors that auto-regulate their own expression. Approximately 35% of the genome is covered by this network. (Modified from: Navarro, 2010).

The 1,400 network nodes represented in Figure 1 correspond all of them to specific genes of *M. tuberculosis* and their protein products, while the 2,304 links correspond to gene expression regulatory interactions by 94 transcription factors.

The network shows a clear organization in structural levels that correspond with the complex functions and life-cycle stages of this highly sophisticate pathogen. Although the functions are relatively well defined in modules or communities, they can change dramatically by simply "rewiring" some connections of the genetic network. This has already been made in other bacteria: it has been demonstrated experimentally that it is possible to make that a bacterium synthesizes (or not) a green fluorescent protein simply by exchanging the regulatory regions of genes, *lacR*, *tetR* y *lambda cI*, not changing sequence in these genes [Guet et al., 2002].

The genome of the bacillus contains more than 4,000 genes, and close to 190 transcription factors. Of this entire repertoire, the new ERT network represents 94 transcription factors and 1,400 genes. So there is plenty of room for future improvement of the network, as new laboratory works will describe new links derived from other transcription factors not worked out yet. In general, the number of transcription factors per genome translates into greater genetic network connectivity, which is correlated with increased complexity of the microorganism structures and life cycle [Levine and Tjian, 2003].

Cellular signaling systems: topological governance

By itself the transcription network is "blind". In other words, the coupling between the sequential and the diluted architectures needs adaptive capability to respond to environmental demands. This is done by means of signaling guidance, so to partially deploy the genetic circuits in response to relevant happenstances of the environment or from within the cell. The *topological governance* of the transcription regulatory network, the decision of what parts should be activated or what particular circuits should be inhibited, is achieved thus by the cellular signaling system or *signalome.*

Prokaryotic signaling systems

In prokaryotes, a variety of molecular systems are involved in the *signalome*, ranging from simple transcription-sensory regulators (a single protein comprising two domains), such as the well-known *embR, alkA* or *furB*, to those systems of multiple components and interconnected pathways that regulate key stages of the cell cycle, such as latency, pathogenesis, replication, and dispersion. A basic taxonomy of bacterial signaling systems was proposed by the authors somewhere else [Marijuán, 2010], which was centered on "the 1-2-3 scheme" (see Figure 2):

The first level of signaling complexity corresponds to simple regulators, "the one-component systems (OCS)." In fact, most cellular proteins involved in cellular adaptation to changing environments, in a general sense, could be included as participants in this primary category [Galperin, 2005]. Around one hundred OCS elements may be present in a moderately complex prokaryotic cell.

Increasing the scale of complexity, the "two-component systems (TCS)" appear, which include a histidine kinase protein receptor and an independent regulatory response; conventionally they are considered as the central paradigm in prokaryotic signaling systems, and in fact, a number of intercellular communication processes among different species are carried out by these specialized systems. A few dozen TCS pathways may be present in the prokaryote.

To maintain conceptual coherence, an additional category, the "three-component system (ThCS)" should apply to two-component systems that incorporate additional non-kinase receptor for activating the protein kinase (eg, methylated receptors described for the chemotaxis.) Very few pathways are showing the ThCS arrangement but they are very important ones (e.g., chemotactic guidance).

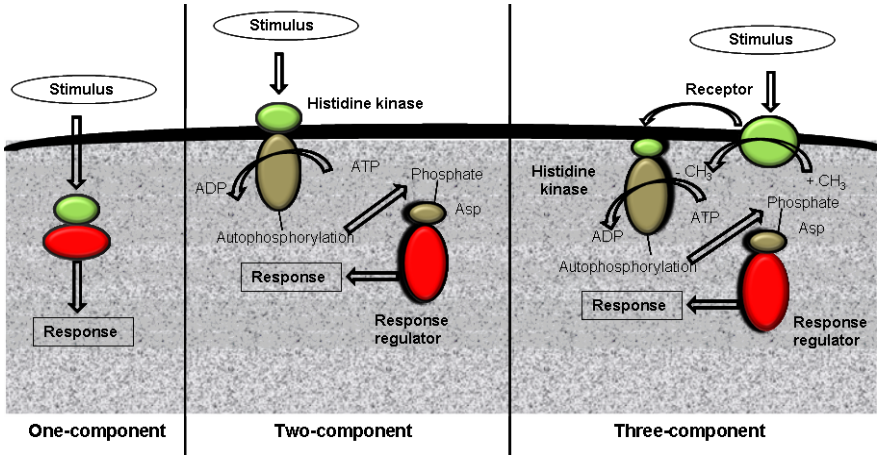

Figure 2. The three characteristic signaling pathways developed by prokaryotes. The external stimulus is perceived either by an internal receptor–transducer (left), or by a transmembrane histidine kinase that connects with a response regulator (center), or by an independent receptor associated to the histidine kinase (right). (Modified from: [Marijuán, 2010]).

Eukaryotic signaling systems

In eukaryotes the signaling system comprises many hundreds of different classes of dedicated molecular agents (receptors, ion channels, transducers, amplification cascades, second messengers, intermediate effectors, final effectors) that can be arranged differently in each tissue. See Fig. 3 for a very simplified scheme. It is very important that, in multicellular organisms, every cell-type has tailored its *specialized signalome* along its developmental trajectory, in dependence of its own history of received signals and self-modifying processes [Marijuán, 2002].

In eukaryotes, rather than a simple taxonomy like the 1-2-3 seen in prokaryotes, the scheme of signaling pathways becomes an interconnected network. Some of the main pathways are shown in a linear way in Figure 2. Here, the general "detection, measurement, and intervention" character of the signalome has to be emphasized. The second messengers (cAMP, cGMP, Ca, InsP3, diacylglicerol, ceramide...) are dramatically modified in their concentrations by the different signaling paths that have been transiently activated, within a generalized cross-talking among all activated paths. Particularly throughout the very fast changes in second messenger concentrations, an integrated perspective (measurement) of the different internal and external influences at play is obtained within the cell, and is subsequently passed towards intermediate chains and the final effectors.

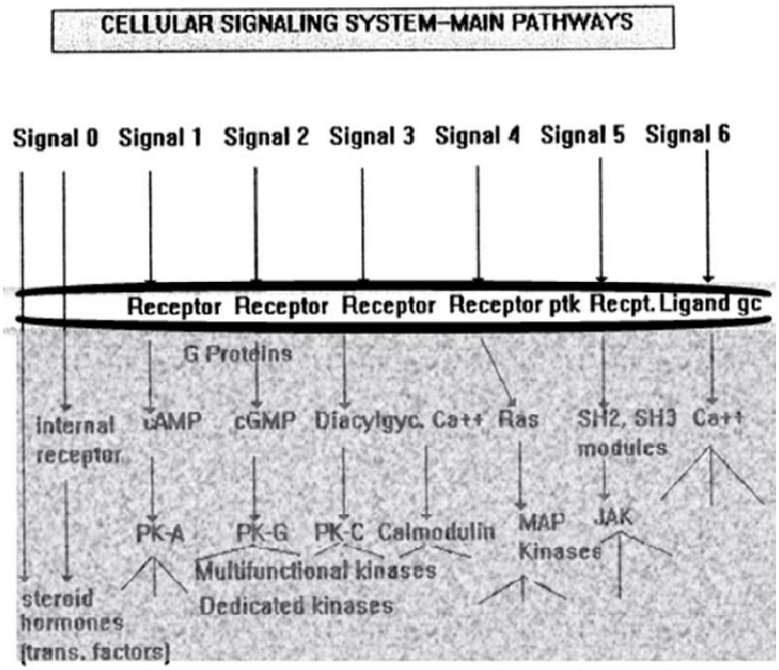

Figure 3. Representation of the principal classes of signaling pathways in eukaryotic cells. The signaling paths in the left (steroids) are the slowest ones, usually associated with cell fate and hormonal effects. Paths 1 to 3, mediated by G proteins, are faster and have a great amplification (ideal ones for sensory receptors), counting with numerous variants. Path 4 corresponds to control of development and cell cycle. Path 5 represents the customary access for neuropeptide action. Path 6, ligand-gated channels, is the genuine cortical path for fast neurotransmitters (GABA, Glutamate). The representation is highly simplified and does not include further effector cascades and vertical-lateral cross talking between paths. (Modified from: [Marijuán, 2003].

At the end of the signaling command-chain, the nuclear machinery is waiting to be fed with a combination of *ad hoc* signals in order to change the transcriptional status of the genome [Janes, 2005]. This nuclear part of the whole signalome apparatus has already been implementing the *histone code,* in order to allow a tight grip upon the euchromatin-heterochromatin states which regulate access to transcription –so that the well measured signals from the cytoplasmic signalome may be finally enacted as a new transcription program in relation with the advancement of the cell cycle or with the specialized function of the cell [Janes, 2005]. As already said, a comparison with the "direct" topological governance of the transcriptional regulatory network we have discussed in the prokaryotic case is not viable, except in the most simplified cases (e.g., steroid paths, degradome).

Everything has to converge factually on the cell cycle: signaling system, transcription, metabolism, protein synthesis, protein degradation, network organization, and control of the cell cycle itself [Marijuán, 1996]. As successive rounds of cell replication are accomplished, there occurs a functional rewriting of "DNA rules" by the signaling system. The transcriptional status of most DNA regions is systematically altered as cells advance along the totipotent, pluripotent and stem cell path, until completion of the tissular differentiation and specialization is achieved [Gasser, 2002]. The euchromatine / heterochromatine state of a number of genomic regions is irreversibly altered by signaled implementation of the previously mentioned "histone code".

From the point of view of formal systems, this unusual characteristic of "rewriting the own rules" could be significant concerning the cellular automata field [Wolfram, 2002], perhaps opening new paths towards new types of biologically inspired cellular automata capable of negotiating complex morphological / differentiation spaces.

Concluding comments: on evolutionary problem-solving strategies

With the spiraling and multiplication of cellular cycles along the eukaryotic developmental path, the hallmark for a new type of biocomplexity is set. The prokaryotic cell cycle has been loaded with a formidable complexity in eukaryotes, setting the stage for a number of *emergences* discussed in computational and philosophical fields. For instance: transition from molecular stochasticity to systemic robustness and quasi-determinism, organization of cellular signaling systems, establishment of an "informational" cell cycle (based upon molecular recognition dynamics), interplay of cellular bottom up causality with organismic top-down causality, exhibition of behavior endowed with autonomy and agency, etc.

In the extent to which this complexity growth of eukaryotes has been built by tinkering upon the scheme of *functional or primary addresses* and *secondary addresses* put together onto the same DNA memory, the parallel with the von Neumann scheme of modern computers seems unavoidable –for in computers, logical functions and memory addresses are also put together into the CPU memory. See some other recent interpretations in [Danchin, 2009], [Yan, 2010].

Because of this DNA scheme in eukaryotic cells, the evolutionary genetic algorithms for physiological problem-solving are largely parallelized in eukaryotes. The different components of the biomolecular solutions may be tinkered with separately, and linked together later on [Peisajovich, 2010]. Besides, every molecular stage (transcription, folding, transportation, modification, complexes, degradation), specifically coded onto DNA addresses, may be used as a new functional element of control. Solutions may be chosen, then, from an augmented set of molecular building blocks.

The so called "Central Dogma" of classical molecular biology should not be taken as a closed black-box; rather the successive stages and intermediate transcripts could participate as legitimate molecular partners, each one endowed with endogenous recognition capabilities, within a whole *transmolecular matrix* of controlling interactions [Marijuán, 2002, 2003]. As an instance, in the recently discovered phenomenon of RNA interference,

scores of micro RNAs are transcribed for the only purpose of using up their molecular recognition capabilities within the context of other DNA transcription and RNA translation events, collectively known as "gene silencing."

In actuality, the evolutionary coupling between the two informational architectures of life, the sequential and the amorphous, has explored almost every conceivable cellular bauplan and organism physiology [Mojica, 2009]. Life has thrived throughout the deployment of an organization with amazing informational capabilities and systemic emergences. We might argue that prokaryotes have used those very capabilities mostly towards the direct solution of *molecular assimilation* problems (in their encounter with environmental substances), while eukaryotes have tamed a fascinating developmental complexity by evolving towards the general solution of *molecular organization* problems.

Based on molecular recognition phenomena, the dynamics of life is informational, and non-mechanical. In this regard, the following three principles may tentatively summarize various directions discussed in this paper:

1. The living cell is an open system continuously engaged in the advancement of a manifold trajectory: the life cycle of self-re-production.

2. The cellular advancement across the life cycle may be propelled (or nullified) not only by the availability of environmental affordances but also —and mostly— by signaling events.

3. The effects of signaling events in the living cell are irrespective of their material underpinning; biological information is decoupled from its mater and energy counterparts and becomes symbolic, "semiotic", though it always relates to self-production processes of the life cycle.

Bibliography

[Atanasov, 2005] Atanasov A.T. The linear alometric relationship between total metabolic energy per life span and body mass of poikilothermic animals. BioSystems, 82, 137-142 (2005).

[Carlton, 2002] Carlton M. The information paradigm. Posthumous compilation available at: http://conway.cat.org.au/~predator/paradigm.txt (2002.)

[Claverie, 2001] Claverie J.M. What If There Are Only 30,000 Human Genes? Science 291pp. 1255-1257 (2001)

[Conrad M, 1996]. Cross-scale information processing in evolution, development and intelligence. BioSystems 38 pp. 97-109. (1996)

[Danchin , 2009]. Bacteria as computers making computers. FEMS Microbiol Rev 33: 3-26 (2009).

[Frauenfelder, 2009] Sligar S.G. and Wolynes P.G. (1991) The Energy Landscapes and Motions of Proteins. Science 254 pp. 1598-1603 (2009).

[Galperin, 2005]. A census of membrane-bound and intracellular signal transduction proteins in bacteria: Bacterial IQ, extroverts and introverts. BMC Microbiol. 5: 1–19 (2005).

[Gasser, 2002]. Visualizing Chromatin Dynamics in Interphase Nuclei. Science 296, 1412-1416 (2002).

[Goodsell, 1991]. Inside a living cell. TIBS 16, 203-206 (1991).

[Guet, 2002] Elowitz M.B., Hsing W. & Leibler S. Combinatorial Synthesis of Genetic Networks. Science 296 (5572): 1466-70 (2002).

[Ho, 1995]. Bioenergetics. The Open University, London, (1995).

[Janes, 2005] Albeck J.G., Gaudet S., Sorger P.K., Lauffenburger D.A. and Yaffe M.B. A Systems Model of Signaling Identifies a Molecular Basis set for cytokine-Induced Apoptosis. Science 310, pp. 1646-53 (2005).

[Levine & Tjian , 2003]. Transcription regulation and animal diversity. Nature 424: 147-51 (2003).

[Lin, 2001].The Nature of the Chemical Process. 1. Symmetry Evolution –Revised Information Theory, Similarity Principle and Ugly Symmetry. Int. J. Mol. Sci. 2 pp. 10-39 (2001).

[Marijuán, 1994]. Enzymes, automata and artificial cells. In Computing with biological metaphors, ed. by R.C. Paton. Chapman & Hall, London, pp. 50-68. (1994)

[Marijuán, 1996]. The Cell as a Problem-solving 'Engine'. In Computation in Cellular and Molecular Biological Systems, ed. by R. Cuthberson, M. Holcombe and R. Paton. World Scientific, Singapore, pp. 183-194 (1996).

[Marijuán, 2002]. Bioinformation: untangling the networks of life. BioSystems 64, pp. 111-118 (2002)

[Marijuán, 2003]. From inanimate molecules to living cells: the informational scaffolding of life. In Energy and Information Transfer in Biological Systems, ed. by F. Musumeci, L.S. Brizhik and M.W. Ho. World Scientific, Singapore (2003).

[Marijuán, 2009]. The advancement of information science: is a new way of thinking necessary? tripleC 7(2): 369-75 (2009).

[Marijuán, 2010]. Navarro J. & del Moral R. (2010). On prokaryotic intelligence: strategies for sensing the environment. Biosystems 99: 94-103 (2010).

[Marijuán and Westley, 1992]. Enzymes as molecular automata: a reflection on some numerical and philosophical aspects of the hypothesis. BioSystems 27: 97-113 (1992)

[Martin, 2006]. Tuberculosis vaccines: past, present and future. Curr Opin Pulm Med. 12 (3): 186-9 (2006).

[Meggs, 1998] Biological homing: hypothesis for a quantum effect that leads to the existence of life. Medical Hypothesis 51, pp. 503-506 (1998).

[Mojica, 2009] Navarro J., Marijuán P.C. & Rafael Lahoz-Beltra. (2009). Cellular "bauplans": Evolving unicellular forms by means of Julia sets and Pickover biomorphs. BioSystems 98, 19–30 (2009).

[Navarro, 2010]. Transcriptional Regulatory Network of M. tuberculosis: Functional and Signaling Aspects. Master Thesis. Universidad de Zaragoza (2010).

[Peisajovich, 2010] Garbarino J.E., Wei P. and Lim W.A. Science 328: 368-72 (2010).

[Shimizu and Bray, 2001]. Computational Cell biology –The Stochastic Approach. In Foundations of Systems Biology, ed. by H. Kitano. The MIT Press, Cambridge (2001).

[Urry, 1995]. Elastic Biomolecular Machines. Scientific American January 44-49 (1995).

[Weinhold and Landis, 2007]. High Bond Orders in Metal-Metal Bonding. Science, 316,.61-3 (2007).

[Wolfram, 2002]. A New Kind of Science. Wolfram Media Inc. (2002).

[Wright, 2004] Reich P.B. et al. --33 authors total-- (2004) The worldwide leaf economics spectrum. Nature 428, 821-7 (2004).

[Yan, 2010] Fang G., Bhardwaj N., Alexander R.P. & Gerstein M. Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks. PNAS 107 (20): 9186-91 (2010).

Authors' Information

Jorge Navarro López– *Bioinformation and Systems Biology Group, Aragon Institute of Health Sciences (I+CS,) Zaragoza, Spain.*

e-mail: jnavarro.iacs@aragon.es

Angel Goñi-Moreno– *Natural Computing Group, Universidad Politécnica de Madrid, Spain; e-mail: anandgel@gmail.com*

Pedro C. Marijuán – *Bioinformation and Systems Biology Group, Aragon Institute of Health Sciences (I+CS,) Zaragoza, Spain.*

e-mail: pcmarijuan.iacs@aragon.es