

How a Bacillus “Sees” the World: Information Needs and Signaling Resources of *Mycobacterium tuberculosis*

Jorge Navarro & Pedro C. Marijuán*

Bioinformatics and Systems Biology Group, Aragon Institute of Health Sciences (I+CS), Avda. Gómez Laguna 25, Pl. 11, Zaragoza 50009, Spain
pcmarijuan.iacs@aragon.es

Abstract: Any living cell parasitizing a host organism is immersed into a molecular environment of unfathomable complexity. For the advancement of its life cycle in such “hostile” a territory, the cell has to carefully sense its environment, “see” the ongoing physiological processes taking place, and guide subsequently its own network of self-construction processes, pathological responses included. We will discuss how this informational matching occurs in the bacillus *Mycobacterium tuberculosis*, and how transcriptional programs within the global transcriptional regulatory network are deployed in response to specific signals from the environment and from within the cell itself. In the era of the bioinformatic revolution and of systems biology, it is perhaps surprising that the functional interconnection between the transcription network and the signaling system is far from clarified yet. In the extent to which the living cell can be considered as one of the central paradigms of the nascent information science, this discussion also becomes one about the essential cluster of concepts which should potentially apply to the analysis of other information-based entities.

Keywords: *Mycobacterium tuberculosis*, transcriptional regulatory network, signaling system, information and meaning, knowledge recombination

Acknowledgements: to Zong-Rong Li and his Chinese colleagues, for taking the challenge of organizing FIS 2010. The authors also acknowledge useful discussions and information exchanges on prokaryotic signaling with Carlos Martín and José Antonio Ainsa (Grupo GENMICO, University of Zaragoza), and with Yamir Moreno and Joaquín Sanz (Instituto BIFI). This paper is included within the Research Project DGA *Biología de Sistemas de Mycobacterium tuberculosis phoP: Señalización, Ciclo Celular e Interacciones Epidemiológicas*.

1. The bio-information approach

From the information science point of view, the living cell incorporates the highest trove of informational phenomena that one can think of at the molecular scale. It is a micro-world teeming up with millions of specific molecular recognition events, genetic codes, transcription and translation processes, molecular machines and self-assembling complexes, signalling systems, messengers, transducers, second messengers, regulators, effectors, connectivity networks, interferences, etc. Conspicuously, the information metaphor has become the “natural” way of talking about biomolecular phenomena, almost from the very beginning, and even more along the current *omic* and *bioinformatic revolution*. Rigorously used, however, as “information theory” it turns out to be an essential tool for

the conceptualization of generative codes, communication channels, and structural descriptions *inter alia*; but perhaps not much more. It has been a rather vague, almost indefinable, generalized information concept (or a “computer” derived one) which has been liberally used as a surrogate for any kind of anthropocentric-like interactions, autonomy, or unexplained complexities observed.

The need for more consistent a bio-information approach has been argued somewhere else (Marijuán, 2004; Marijuán & del Moral, 2007). Three basic ideas were introduced. Fundamentally, it was argued that *molecular recognition* varieties should be taken as the starting point, the very basic level to initiate the bio-information analysis (following Lin, 2001). Further it was argued that the *multidomain embodiment* of function

in enzymes and proteins allows for configuration of a new type of collective computation within an evolvable quasi-universal problem-solving “engine”, or a quasi-universal “constructor” more properly. And finally, the advancement of the cellular cycle in its adaptive coupling with the environment –signaling system mediated– was to be taken as the semiotic *reference frame* (whith the appearance of meaning, value, fitness, and intelligence), bringing thus closure to the informational “cellular theory”.

Herein we will attempt an application of some of these theoretical ideas in the practical realm, the “taste” of the pudding, precisely by means of bioinformatic and systems biology tools. In what extent can we refine the bio-information tenets by analyzing the way a well-known (and very successful) pathogen explores its environment and ekes out a living amidst the host tissues of highly advanced eukaryotes?

We will briefly review, first, the peculiar life cycle of *Mycobacterium* and will continue with the organization of its auto-construction processes (analyzing the transcriptional regulatory network) and the communications with the environment (throughout the cellular signalling system). At last we will put together a conceptual cluster to tentatively interrelate the systems biology approach with basic bio-information concepts: meaning, value, fitness, knowledge, and above all the *life-cycle* as the fundamental instance of reference.

2. The life cycle of *Mycobacterium tuberculosis*

Mycobacterium tuberculosis is an extraordinarily successful pathogen that currently infects approximately one-third of the global population and causes 8 million new cases of tuberculosis annually (Corbett et al., 2003). Its success is based on the singularities of its life cycle within the host organism, as it is specialized in the pathogenic attack to the central agents of the immune system: macrophages. As was pointed by Cole et al. (1998), the characteristic features of the tubercle bacillus life-cycle include its slow growth, the capability to enter and exit into dormancy, a very complex cell envelope, and a special metabolism adapted to intracellular

pathogenesis, together with a robust genetic homogeneity.

As an intracellular pathogen specialized in macrophages of the pulmonary tissue (but occasionally also within the brain, kidney, and other organs), the tubercle bacillus must be able to gain entry into macrophages. They achieve that passively, by letting themselves be engulfed into a vacuole or phagosome, which they are able to disorganize in its maturation process and subsequent fusion with the digestive lysosome. Then either they multiply intracellularly, or they may switch into dormancy and survive within the lung “granulomas” (defensive nodules organized by the immune system to contain the bacillus spread) for quite a long time; or depending on the immune competence of the host, the bacilli population within each granuloma may take a “collective decision” to implode the containing structure and subsequently be dispersed towards a new host via aerosols (Dannenberg & Rook, 1994).

The ability to persist for long periods in the host depends on the capacity of *M. tuberculosis* to acquire and utilize nutrients from the very interior of the macrophage phagosome. After detection of the specific environment, the bacillus will switch its metabolic pathways to utilize fatty acids rather than carbohydrates during infection (Boshoff & Barry, 2005). It is therefore likely that the coupling between signaling pathways and expression of different sets of genes at various stages of infection is crucial for the microbe’s survival and spread strategies.

Concretely, one of the most important signaling pathways, which includes the *phoR* and *phoP* genes, regulates a large number of genes of the TB bacilli within the phagosome, controlling important metabolic functions: hypoxia response, respiratory metabolism, response to stress, lipid synthesis, etc. (Gonzalo-Asensio et al., 2008). Many other signaling components and pathways are in charge of switching many other different genetic circuits.

3. The Transcriptional Regulatory Network

Most studies about *M. tuberculosis* have focused in the expression of individual genes and not in the internal / external instances of control concerning the guidance of the overall network. To confront this problem, we have compiled a large-scale *M. tuberculosis* transcriptional regulatory network, which has been built upon a previously published TR network (Balázsi et al., 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 et al., forthcoming). 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 which takes place in the different environments, as well as gauging the impact of the SO2 strain on physiological and immune systems of the body.

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.

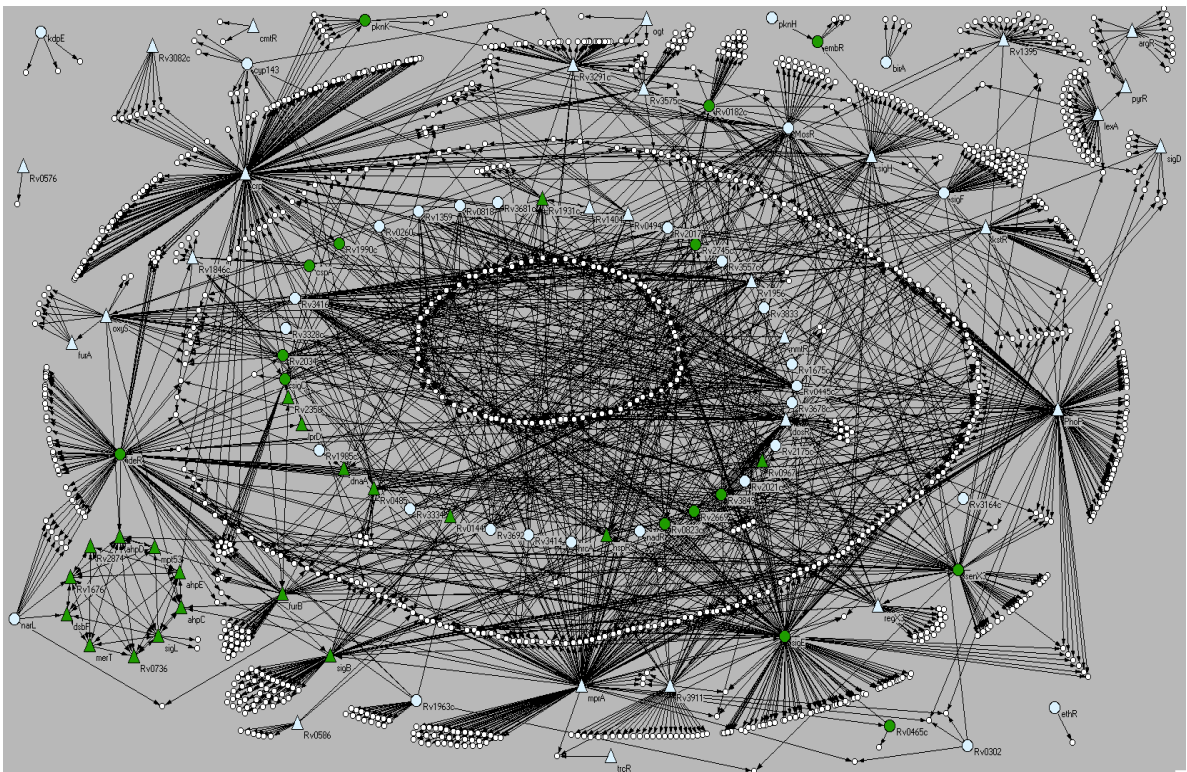


Figure 1. The Transcriptional-Regulatory (ETR) Network of *M. tuberculosis*. Nodes represent Mt's genes, and links represent their regulatory interactions. Transcription factors appear either green or blue, depending on whether they have known transcriptional regulator or not. The white nodes represent 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 expanded network.

Table 1 shows a comparison between the new network ETR herein proposed and the TR network proposed by Balázsi (the largest to date). In this table, it can be seen that the new network has substantially increased the number of genes, the number of links, and the number of transcription factors. This obviously means new possibilities for an improved structural and functional knowledge of *M. tuberculosis* (Navarro et al., forthcoming).

The network shows a clear organization in structural levels that correspond with the complex functions and life-cycle stages of the bacillus. 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 cl*, not changing sequence in these genes (Guet et al., 2002).

In the mutant *M. tuberculosis* SO2 (*phoP* mutation), it is a mere nucleotide change in *phoP* what causes a substantial change in its transcription network, life cycle and pathogenic capacity, implying thus that *phoP* is in control of a vast array of complex

functions: hypoxia response, respiratory metabolism, stress response, synthesis of membrane lipids, and virulence factors (Gonzalo-Asensio et al., 2008). This high density of functional links out from *phoP* can be clearly appreciated in Figure 1.

In fact, *phoP* appears as the most connected “hub” of the network, second only to *crp* transcription factor. Indeed, we can establish that *phoP* regulates key functions required for the intracellular survival and persistence of *M. tuberculosis*. Therefore, inactivation of *phoP* results in downregulation of genes required to successfully survive within macrophages and consequently in *M. tuberculosis* attenuation.

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 micro-organism structures and life-cycle (Levine & Tjian, 2003).

Network	TR Network	ETR Network	Increase (%)
Genes	783	1400	617 (79%)
Links	937	2304	1367 (146%)
Transcriptional factor	45	94	49 (109%)

Table 1. Comparison between the network proposed by Balázsi et al. (TR), and the new expanded network (ETR). The increase column refers to different properties that have been augmented in this study, showing in brackets the % increased, all of them are compared with the network proposed by Balázsi et al., 2008.

4. Signaling System components

By itself the transcription network is “blind”. It needs signaling guidance to deploy its genetic circuits in response to relevant happenstances of the environment or from within the cell. A variety of molecular systems are involved in signal detection, 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.

As happens in many other prokaryotes, the signaling system of *Mycobacterium tuberculosis* is not very well-known yet. But it participates of the basic taxonomy of bacterial signaling systems proposed by the authors somewhere else (Marijuán et al., 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).

- 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.
- 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.)

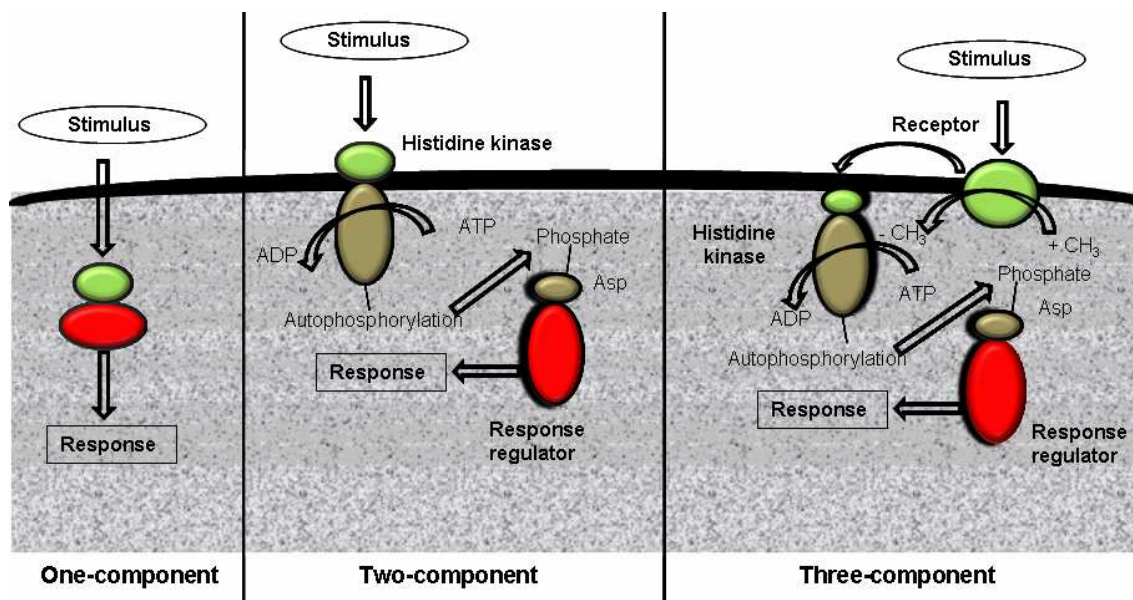


Figure 2. The three characteristic signaling pathways developed by prokaryotes. The external stimulus is perceived either by an internal receptor–transducer (left), or by a transmembrane histidine kinase that connects with a response regulator (center), or by an independent receptor associated to the histidine kinase (right). These three different options imply very different information processing capabilities and metabolic costs.

It is interesting to see how *Mycobacterium tuberculosis* fares with respect those signaling categories, particularly in relation to the average signaling components found in other prokaryotes. Analyzing the information encoded in 145 prokaryotic genomes, Ulrich et al. (2005) found that in many prokaryotic species there is a robust correlation between the total number of signaling elements encoded in the genome, and the total number of genes in the same genome, or simply the square of the genome size (Ulrich et al. 2005). This correlation also occurs for OCS and TCS independently taken. See Figure 3, where OCS and TCS have been represented separately, and where the corresponding figures of *Mycobacterium* have been singled out. Conspicuously, *Mycobacterium* stands out as having a very high level of OCS and a moderate level of TCS.

That *M. tuberculosis* features such a high number of One Component Systems (OCS) is rather unexpected. By far they are the most abundant signaling resource, and yet

they have not been subject to specific compilations for *Mycobacterium*. The authors' educated guess, based on their ongoing work on transcription factors, and the compilation about prokaryotic signaling by Ulrich et al. (2010), allow for an estimated of 140-160 OCS, which in turn can be further categorized into several functional classes (enzymatic, transducers, ECFs, etc.).

Two Component Systems (TCS) are present in the genome of *Mycobacterium tuberculosis* at least in 11 complete TCS pathways, plus two orphan histidine-kinase genes, and six orphan response regulator genes (Haydel & Curtis, 2006). The existence of these eight additional orphan response regulator and histidine-kinase genes indicates that coordinate regulation could be very complex, implying interconnection between signalling pathways. Table 3 shows the 11 TCS pathways and their corresponding signaling function (that in several cases is not yet well-established).

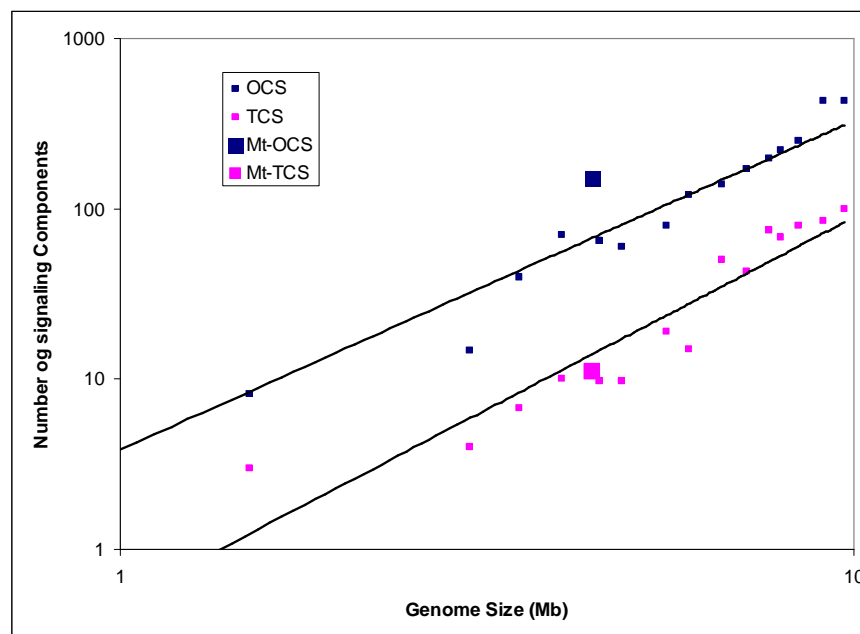


Figure 3. Relationship between the number of one- and two-component signaling systems and the genome size. The graphics is in a double logarithmic scale. Each point indicates the average number of OCS or TCS components for each

Name	Characteristic Function
SenX3-RegX3	Survival during phosphate limitation
TcrA-TcrB	Intracellular growth (?)
PhoP-PhoR	Virulence, lipids & metabolism
NarL-NarS	O ₂ , CO or NO levels
PrrA-PrrB	Macrophage infection (Mg ²⁺)
MprA- MprB	Persistence (CO ₂ Krebs)
KdpE-KdpD	K ⁺ sensing & transporting
TrcR-TrcS	Metabolic switch (repressor)
DosR-DosS	Hypoxia response (O ₂)
MtrA-MtrB	Intracellular growth (Iron)
TcrX-TcrY	Intracellular growth (?)

Table 2. Catalogue of the *M. tuberculosis* two-component regulatory genes.

Three Component Systems (ThCS) are also present in *Mycobacterium*, at least through two signaling pathways endowed with *seronin-threonine* kinase receptors, which have been explicitly claimed as such (Shrivastava et al., 2009). Presumably these eukaryotic-like kinases (11 in total), similar to those of eukaryotic signaling systems, and very rare in prokaryotes, have been obtained through horizontal gene transfer, and they are deployed in order to “sense” how the host cell is organizing its defense responses –and to manipulate them.

5. Information and the advancement of a life cycle: meaning, value, adaptability (fitness), and knowledge

The recent boom in network studies has motivated work about transcriptional regulatory networks in very different model systems and from very different stances. For instance, prokaryotic genomes have been compared to computer operating systems in terms of the topology and evolution of their regulatory networks (Yan et al., 2010); and new terms such as “paleome”, “cenome” “minimal genome” and others have also been coined in this computational direction (Danchin, 2009). Still missing, however, is the connection between transcription networks and the signaling system. In our view, the ensemble of signaling components –OCS, TCS, and ThCS– are the “*topological*

governors” of the transcriptional regulatory network, deciding the potential connections that are going to become real, and vice versa, which of the existing ones should be discontinued. A “smart”, adaptable life-cycle is the outcome.

The cell life-cycle is advanced by the confluence of genes permanently expressed (house-keeping systems) and genes that are expressed after detection of specific environmental “affordances.” It is in the latter where the bulk of complexity lies, at least from the information-science point of view. Rather than merely energy, what the cell enacting a complex life-cycle continuously needs is “information” about its environment. And these information needs are covered by the plethora of signaling resources that we have been discussing in the case of *Mycobacterium tuberculosis*.

The living cell enacts a new way of existence, an active “informational” one that is based on the capability to keep the own structures in a permanent state of “flow”, by piling up synthesis and degradation, activation and inhibition of its network of productive processes in a way that reminds critically self-organized phenomena (Marijuán, 2009). Thus, the signaling system provides the cell with a generic capability to make *distinctions on the adjacent*. The living cell may systematically respond to signal coalitions (“affordances”) from the environment, and produce the

meaning they imply, by letting the signals themselves circulate throughout the signaling system pathways and meddle with the ongoing self-production “flow”. Meaning may be defined throughout *molecular mining*: as the (signal) induced changes in transcriptional connectivity and in constitutive enzyme-protein populations, plus associate metabolites and substrates.

The *relevance* and *value* of impinging signals can subsequently be gauged within a crescendo of occurrences: metabolic buffering, second messengers’ alteration, transcriptional rewiring, and successive crossing of the cell-cycle “checkpoints”. Completion of the cell cycle is the fundamental instance of reference. The phenomenon of *knowledge* may be

appended too, once the DNA codes of the successful signaling components (protein domains) have been evolutionarily selected, refined, and cohered within the life cycle. Domain *shuffling* –actually, *knowledge recombination* in the realm of the living cell– appears as the main evolutionary force propelling the development of signaling systems and achieving the adaptive coupling with the environment, either in prokaryotes or in eukaryotes.

Finally, the way a bacillus “sees” the world becomes a proto-phenomenon of biological information. Now the novelty is that by means of systems biology tools and network science we can start to translate the whole conceptual cluster around information in rather precise molecular terms.

References

- Balázsi G., Heath A.P., Shi L. & Gennaro M.L. (2008). The temporal response of the *Mycobacterium tuberculosis* gene regulatory network during growth arrest. *Molecular Systems Biology* 4: 225.
- Boshoff H.I. & Barry C.E. 2005. Tuberculosis—metabolism and respiration in the absence of growth, *Nat. Rev. Microbiol.* 3: 70–80.
- Cole, S. T., Brosch, R., Parkhill, J. & 39 other authors. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393: 537–44.
- Corbett E. L., Watt C. J., Walker N., Maher D., Williams B. G., Raviglione M. C., Dye C. (2003). The Growing Burden of Tuberculosis: Global Trends and Interactions with the HIV Epidemic. *Archives of Internal Medicine* 163: 1009–21.
- Dannenberrg A.M. & Rook G.S.W. (1994). Pathogenesis of pulmonary tuberculosis: interplay of tissue-damaging and macrophage-activating immune responses - dual mechanisms that control bacillary multiplication. In Bloom, BR. Tuberculosis: pathogenesis, protection, and control. ASM Press, Washington, 1994.
- Galperin M.Y. (2005). A census of membrane-bound and intracellular signal transduction proteins in bacteria: Bacterial IQ, extroverts and introverts. *BMC Microbiol.* 5: 1–19.
- Gonzalo-Asensio J., Soto C.Y., Arbués A., Sancho J., Menéndez M.C., García M.J., Gicquel B. & Martín C. (2008). The *Mycobacterium tuberculosis* *phoPR* Operon Is Positively Autoregulated in the Virulent Strain H37Rv. *Journal of Bacteriology* 190 (21): 7068-78.
- Guet C.C., Elowitz M.B., Hsing W. & Leibler S. (2002). Combinatorial Synthesis of Genetic Networks. *Science* 296 (5572): 1466-70.
- Haydel S.E. & Clark-Curtiss J.E. (2006). Global expression analysis of two-component system regulator genes during *Mycobacterium tuberculosis* growth in human macrophages. *FEMS Microbiology Letters* 236 (2): 341 – 47.
- Levine M. & Tjian R. (2003). Transcription regulation and animal diversity. *Nature* 424: 147-51.
- Link S.K. (2001). The Nature of the Chemical Process. 1. Symmetry Evolution – Revised Information Theory, Similarity Principle and Ugly Symmetry. *Int. J. Mol. Sci.* 2(1), 10-39
- Marijuán P.C. & del Moral R. (2007). The Informational Architectures of Biological Complexity. In: European Conference on Computing and Philosophy ECAP-2005, Proceedings. Cambridge University Press.
- Marijuán P.C. (2004). Information and life: towards a biological understanding of informational phenomena. *tripleC* 2(1): 9-19.
- Marijuán P.C. (2009). The advancement of information science: is a new way of thinking necessary? *tripleC* 7(2): 369-75.
- Marijuán P.C., Navarro J. & del Moral R. (2010). On prokaryotic intelligence: strategies for sensing the environment. *Biosystems* 99: 94-103.
- Martin C. (2006). Tuberculosis vaccines: past, present and future. *Curr Opin Pulm Med.* 12 (3): 186-9.

- Shrivastava R., Ghosh A.K. & Das A.K. (2009). Intra- and intermolecular domain interactions among novel two-component system proteins coded by Rv0600c, Rv0601c and Rv0602c of *Mycobacterium tuberculosis*. *Microbiology* 155: 772-779
- Ulrich L.E. & Zhulin I.B. (2010). The MiST2 database: a comprehensive genomics resource on microbial signal transduction. *Nucleic Acids Res.* 38: 401-7.
- Ulrich L.E., Koonin E.V. & Zhulin I.B. (2005). One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol.* 13: 52–6.
- Yan K.K., Fang G., Bhardwaj N., Alexander R.P. & Gerstein M. (2010). Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks. *PNAS* 107 (20): 9186-91.

About the Author

Pedro C. Marijuán

Engineer and Doctor in Cognitive Neuroscience (PhD Thesis on "Natural Intelligence", Universidad de Barcelona, 1989). During more than 20 years, the author has advanced research on the nature of information & communication, and on biological intelligence, looking at both the molecular-cellular and organismic (brain) realms; he was co-founder with Michael Conrad of FIS (Foundations of Information Science); organizer and co-organizer of several international conferences in the FIS field and organizer of the International Cajal Conference (Zaragoza, 1999); he also was organizer of the regular series of conferences "Ateneo del CPS" and "Foros Tecnológicos y Empresariales" (2006-08). His publications include *The Dual Brain* (book in Spanish: "El cerebro dual"), Editorial Hacer, Barcelona (1997). Editor of two special issues on FIS: *Foundations of Information Science*, *Journal BioSystems* (1996,1998), and editor of the special issue *Cajal and Consciousness at the Annals of the New York Academy of Sciences* (2001).