Chapter 9
On the Search for Design Principles in Biological Systems

Juan F. Poyatos

Abstract The search for basic concepts and underlying principles was at the core of the systems approach to science and technology. This approach was somehow abandoned in mainstream biology after its initial proposal, due to the rise and success of molecular biology. This situation has changed. The accumulated knowledge of decades of molecular studies in combination with new technological advances, while further highlighting the intricacies of natural systems, is also bringing back the quest-for-principles research program. Here, I present two lessons that I derived from my own quest: the importance of studying biological information processing to identify common principles in seemingly unrelated contexts and the adequacy of using known design principles at one level of biological organization as a valuable tool to help recognizing principles at an alternative one. These and additional lessons should contribute to the ultimate goal of establishing principles able to integrate the many scales of biological complexity.

1 A New Discipline with Deep Roots

> There appear to exist general system laws which apply to any system of a certain type, irrespective of the particular properties of the system and of the elements involved.
> *Ludwig von Bertalanffy, 1969* [1].

The search for basic concepts and underlying principles was at the core of the systems approach to science and technology. This approach was argued to be necessary to "deal with complexities, with wholes or systems in all fields of knowledge" [1]. Its application to biology was then similarly advocated with an

J.F. Poyatos (✉)
Logic of Genomic Systems Laboratory, Spanish National Biotechnology Centre, Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain
e-mail: jpoyatos@cnb.csic.es

O.S. Soyer (ed.), *Evolutionary Systems Biology*, Advances in Experimental Medicine and Biology 751, DOI 10.1007/978-1-4614-3567-9_9,

emphasis in the need to discover principles of organization at all levels. However, this research program was somehow abandoned in later years, as mainstream biology benefited of the technical revolution associated to the rise of molecular biology and to the discovery of the structure of the DNA [2]. Interestingly, further technical revolutions (omics, large-scale computing, bio-imaging tools, to name a few) and the knowledge accumulated over decades of molecular approaches are bringing back the quest-for-principles program to the biological domain [3]. This new attempt not only promises to provide a more comprehensive view of the cell and its evolutionary and ecological constraints [4] but also hopes to bring different disciplines together for the construction of novel biomolecular components [5].

Here, I discuss some of the lessons that we are already learning in the recent search for "design principles." To this aim, I present two lessons that I derived from my own search (broader perspectives on this area can be found elsewhere, e.g., [6, 7]): (1) the importance of studying biological information processing to identify common principles in seemingly unrelated contexts and (2) the adequacy of using known design principles at one level of biological organization as a valuable tool to help recognizing principles at an alternative one. These and additional lessons should contribute to the development of a better organismic understanding of biology and to the ultimate goal of identifying its principles of organization at various levels [1].

2 Two Fundamentals on the Search for Principles

2.1 Same Principles, Different Biological Contexts

Could we identify design principles at work in unrelated biological contexts? The study of how information-processing tasks are achieved in biological systems could provide us with such broad principles, as similar tasks are found in different contexts, whether we talk of the brain sensing visual stimuli or of a bacterium anticipating a particular metabolite. Indeed, the study of bacterial computations—and bacterial protein networks—was proposed as a model system to appreciate the functioning and evolution of more complex information-processing mechanisms [8].

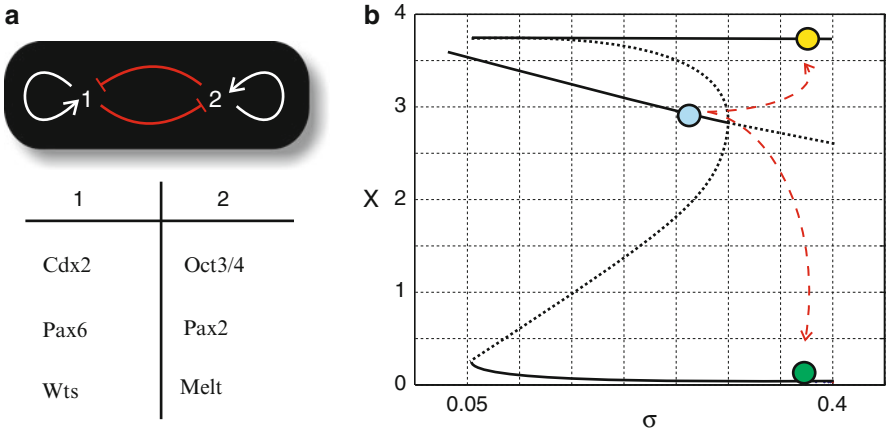
By focusing then on how information processing is especially implemented by biological circuits, we recently characterized the presence of equivalent computational tasks in circuits found in different contexts [9, 10]. We first showed how two-component genetic oscillators could act as "integrators" or "resonators" of external stimuli in a similar way to neural systems. This classification is derived from the early works of Alan Hodgkin [11] in which he proposed three broad groups of cell membrane excitability (abrupt changes in the electrical potential) applicable to different neurons. Class 1 and class 2 neurons were later associated to two different bifurcations of the resting potential (saddle-node bifurcations, class

1, and Andronov–Hopf bifurcations, class 2) that eventually determine different computational attributes, e.g., integrator (class 1) or resonator (class 2) of incoming spikes [12]. The proposed genetic circuits can show these two classes of bifurcations leading to excitability (oscillations) that relate accordingly to their computational differences [9].

Two-component genetic modules could also act as multistable switches, and in [10] we discussed how signal-processing features of such "decision" switches resemble those found in cortical circuits in monkeys, where the circuit logical units represent populations of neurons [13, 14]. Genetic decision switches are specifically constituted by two proteins, i.e., two transcriptional factors, that are autoregulated and mutually inhibit each other (mutual activation can also be considered [10]). We identified these topologies in several biological scenarios (Fig. 9.1a) and showed how the presence of relatively strong autoregulation appears as a necessary condition for the coexistence of distinct expression patterns of the circuit components (multistability). Transitions between these states can be induced by external factors, e.g., a characteristic stimulus, which effectively modify the parameters of the circuit. One of these transitions is driven by what we termed a decision switch: a genetic circuit exhibiting the coexistence of three expression states and that can transit from this regime to a bistable (two states) one. This general class of transition is linked to what is known as pitchfork bifurcation in the language of dynamical systems (Fig. 9.1b) [18].

We demonstrated the potential of decision switches for multifaceted stimulus processing, this including strength, duration, and flexible discrimination [10]. How does, for instance, strength discrimination work? And how could it be potentially tuned by selective forces? Imagine a situation in which a population of cells are all expressing a symmetric high-expression state, i.e., a steady state of the corresponding decision switch within each cell. Imagine now that a signal is acting in both components of the circuit with the same duration. This signal could modify the binding affinities of the associated transcriptional factors which is in turn reflected in the change of a distinctive parameter of the system, i.e., σ.

In Fig. 9.2a, I plotted the (HIGH, HIGH) expression states of an initial population of cells before any stimulus (cyan circles) in an X–Y concentration space. The population expresses certain variability around the expected (deterministic) value due to the presence of biochemical noise [19]. This expected value corresponds to a stable steady state of the dynamical system, one of those points where the response curves, or nullclines, of the system intersect (dashed curves in Fig. 9.2, one could also find unstable steady states [18], white dots in the same figure). A stimulus is now acting in both components of the circuit with the same strength and duration (modifying the binding affinities, denoted by a parameter σ). This change alters the available steady states of the system (Fig. 9.2b). Specifically, the previous (HIGH, HIGH) state becomes unstable (see also Fig. 9.1), and the circuit evolves with approximately the same probability toward the (HIGH, LOW) or the (LOW, HIGH) expression states, the only two steady states available in this situation (yellow/green dots, Fig. 9.2b).



1	2
Cdx2	Oct3/4
Pax6	Pax2
Wts	Melt

Fig. 9.1 Structure and dynamics of the decision switch. The switch is constituted by two autoregulated logical units (**a**) that mutually inhibit each other. These units are commonly transcriptional factors like those found in the context of mammalian embryogenesis (Cdx2, Oct3/4) [15], visual system specification (Pax6, Pax2) [16], and *Drosophila* eye development (Wts, Melt) [17]. (**b**) Bifurcation diagram of the steady-state level of one of the elements (X, adimensional) of a decision switch as a function of the constituent proteins' binding affinities (quantified by σ). This switch corresponds to a transition in which the initial symmetric expression state (HIGH, HIGH) (cyan dot) becomes unstable as σ changes (continuous to *dotted line*). Only two expression states, (HIGH, LOW) and (LOW, HIGH)—*yellow/green dots*—remain. The circuit goes then from having three to two coexisting expression states by means of a pitchfork bifurcation [18]. These states could correspond to the following fates in the examples of the table in (**a**): (Cdx2, Oct3/4), (HIGH, HIGH) → precursor cells, (HIGH, LOW) → trophectoderm, (LOW, HIGH)→ inner cell mass, (Pax6, Pax2), (HIGH, HIGH) → early eye epithelium, (HIGH, LOW) → optic cup, (LOW, HIGH) → optic stalk; (Wts, Melt), (HIGH, LOW) → "yellow" photoreceptor, (LOW, HIGH) → "pale" photoreceptor

The previous scenario leads to a balance distribution of the initial population into two new subpopulations, after the stimulus, expressing two complementary phenotypes. This is associated to the symmetric partition of the X–Y space into two basins of attraction (white/gray areas in Fig. 9.2b; a typical stochastic trajectory, due to biochemical noise, to one of these attractors is shown in black). How could this distribution be tuned? One possibility is that the stimulus strength acting on each circuit components is different. In this way, the circuit computes differences in stimulus strength and codes the result of this computation in the phenotypic distribution of the population [10]. Moreover, unbalanced distributions as response to symmetric signals could be obtained if the circuit presents some relative asymmetry of its components. This could be related, for instance, to different binding affinities of the two transcriptional factors. In this situation, even a signal acting equivalently in both proteins would originate a biased decision-making process. In a limiting case, asymmetries could turn the decision completely deterministic as in Fig. 9.2c, d. In this case, the population expressing the (HIGH, HIGH) state univocally goes to a single expression state, as the initial state fully resides in the basin of attraction of the (LOW, HIGH) state, when the stimulus is present, and stays trapped in that state as the signal is gone.

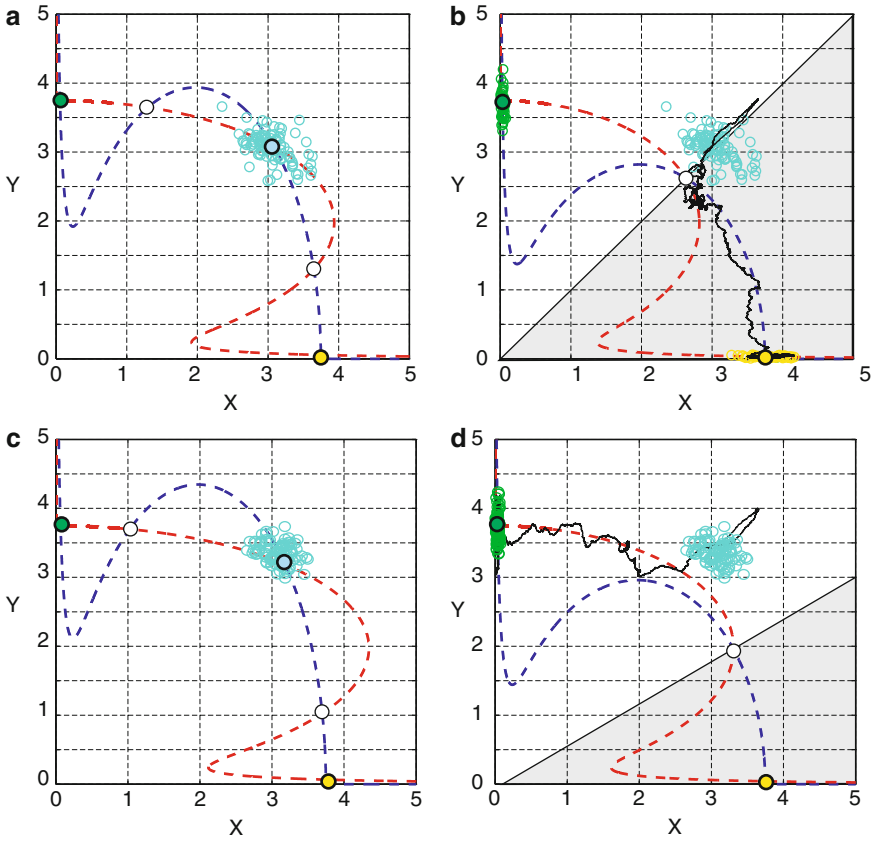


Fig. 9.2 Decision-making of an initial (HIGH, HIGH) population (*cyan circles* in (**a**)) in concentration space (adimensional). After experiencing a stimulus affecting the binding affinities of both proteins (X,Y) equivalently, ∼50% of cells evolve stochastically toward the (HIGH, LOW) or (LOW, HIGH) state. However, a circuit with a slight asymmetry in binding affinities evolves deterministically after the same stimulus to a single expression state (LOW, HIGH). Here, unstable steady states are represented by *white dots*. The available basins of attractions of each circuit when the signals are present are highlighted in *white/gray areas*. The stochastic change (due to biochemical noise) in concentration space of two typical circuit expression states are represented in *black*

2.2 Known Principles as Tools to Find New Principles

The search for principles in biology is not restricted of course to biological circuits. A different domain in which this approach was followed focuses on understanding how specific nucleotide sequences are recognized by regulatory proteins [20]. In this case, even the potential presence of principles was under consideration [21] (a debate commonly found in discussions about principles, see conclusions).

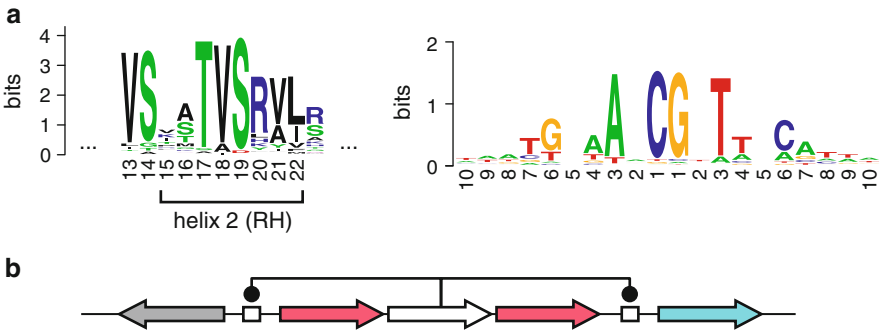


Fig. 9.3 Section of the logo for the alignment of ∼2,700 HTH-LacI domains (**a**), including the (AA-15, AA-16) recognition AAs (coordinates denote position in the alignment [22]). The helix-2 (or recognition helix; RH) is embedded in the full HTH motif. The logo for the alignment of the set of BSs associated to 370 LacI family members (BS sequences from RegTransBase [27]) is also shown. (**b**) Local regulation at the core of phylogenetic footprinting includes both autoregulation which can be linked to the regulation of an upstream divergent operon and downstream unidirectional adjacent regulation (BSs, *white boxes*) [23]. *White arrow* denotes the transcriptional factor within a given operon

How do transcriptional factors recognize their cognate binding sites? Several aspects can influence this process. Direct readouts are linked to how selected amino acids (AAs) and nucleotides (NTs) partners do determine specificity, while indirect readouts are mostly related to how structural features around the contacting AAs could modify the recognition (this could also include aspects of DNA bending, etc). These last features indeed limit the existence of a universal AA/NT recognition code. Could we nevertheless find recognition principles of broad applicability? We asked recently to what extent such a wide-coverage code might actually be found [22]. Notably, we looked for recognition rules by introducing a new approach to phylogenetic footprinting based on *a known principle of biological circuits*: the pervasive presence of local regulation in prokaryotic transcriptional networks [23].

We consider the extensive LacI family of transcriptional regulators [24] as a model system to address the previous question. In these proteins, their helix-turn-helix (HTH) domain interacts with a group of cognate binding sites (BSs) [25]. We aligned ∼2,700 nonredundant HTH-LacI domain sequences using a database of prokaryotic genomes [26] and identified potential BS sequences associated to 370 LacI family members (using [27]). The binding modes associated to these comparative analysis corroborated patterns previously identified with a few structural studies and let us hypothesize that a large subfamily of LacI could present a recognition code between specific NTs of their binding sequence (nucleotides NT-5, NT-6) and specific AAs of their HTH domain (AA-15, AA-16; see Fig. 9.3a).

To search for wide-coverage rules between these NT/AA pairs, we needed to identify—at a large scale—native BSs for each considered TF, independently of

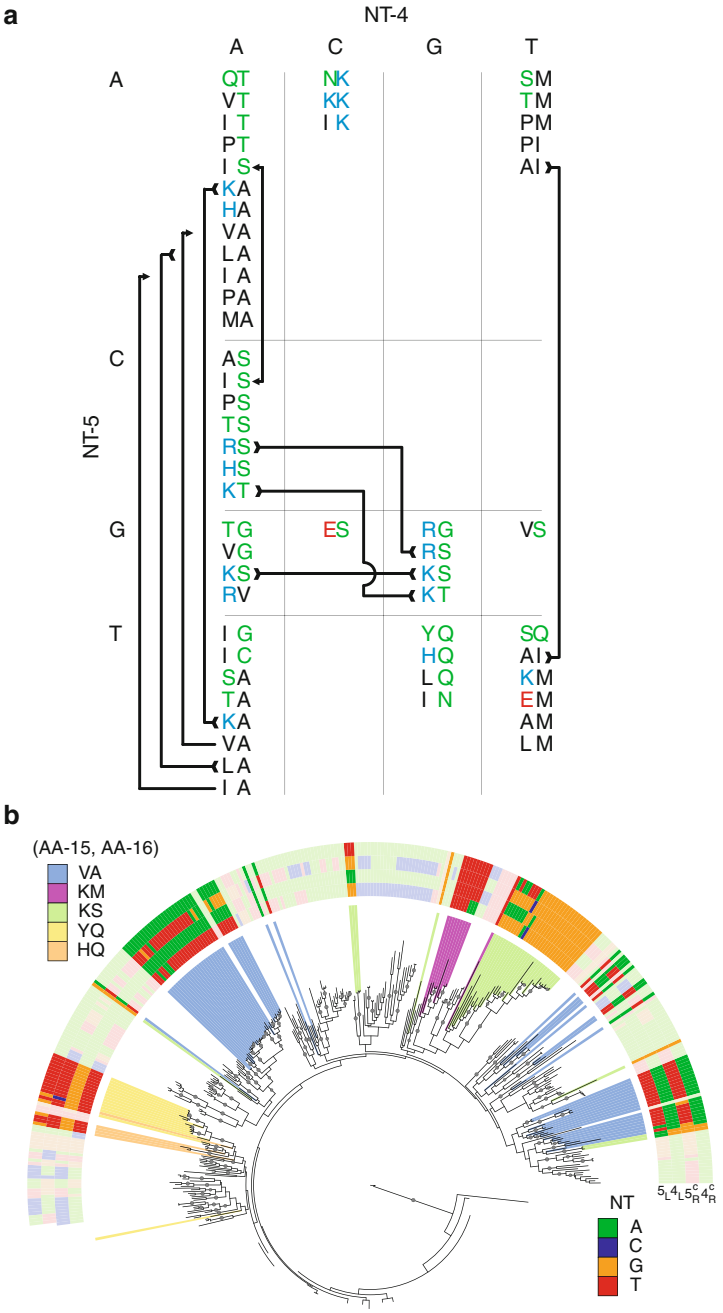


Fig. 9.4 (AA-15, AA-16) sequences recognizing the same (NT-5, NT-4) pair were grouped in (**a**). Here, we only considered significant palindromic NT sequences. Recognition degeneracies are represented by divergent (intrinsic) or convergent (extrinsic, see main text) *arrows*. Colors denote polar (*green*), basic (*blue*), acidic (*red*) and hydrophobic (*black*) AAs. In (**b**) a fraction of the

the TF location in the LacI family phylogenetic tree. This search relied on the assumption of the conservation of binding mode regardless of the evolutionary distance. How could we find BSs for each of the TF considered? We made use of a fundamental circuit-based principle found in prokaryotes: the widespread presence of auto- and neighbor regulation [23]. We thus grouped regulators sharing the same sequence of recognition residues, and within each of these "recognition classes," we looked for potential BSs in the intergenic regions around the operon encoding the TF (see Fig. 9.3b). By applying phylogenetic footprinting techniques [28] on these sequences, we obtained a nucleotide logo from each alignment of BSs associated to a recognition class.

One could then naively accept that only BS logos with high information content in both NT-4 and NT-5 would confirm the hypothesis of a recognition code. However, the presence of low-information positions—in these two NTs—might not necessarily reject our assumption but rather be explained by degeneracies in the recognition process, an expected feature of extant codes [29]. Ambiguities elucidated by intrinsic degeneracies [a particular (AA-15, AA-16) recognition class shows some degeneracy in recognition of several BS sequences] are compatible with our starting hypothesis. The code hypothesis must be revised, or even rejected, when extrinsic degeneracies (TFs of same recognition class binding opposite BSs) are common. Indeed, we only found a few cases of TFs with the same sequence in the specificity pair but recognizing incompatible BSs. This absence of extrinsic degeneracies suggests the presence of an AA/NT recognition code (Fig. 9.4a) [22].

How could we validate the predictions of the proposed rules? We followed three complementary strategies. First, we confirmed some of the theoretical predictions with experimental data of LacI mutants (e.g., [30]). These mutational studies supported code predictions. Second, we used several observations of our analysis to certify the natural counterpart of a binding mode previously considered only to be a laboratory construct (the binding of LacI to the synthetic site SymL [31]). Finally, we identified convergence events in the recognition process (same AAs associated to the same NTs throughout the gene tree, Fig. 9.4b), i.e., as a consequence of the stability of the binding mode, evolution finds the same recognition solution repeatedly. Thus, by using a known principle in one level of organization (local regulation at the circuit level), we were able to identify new principles at a different one (an AA/NT recognition code at the protein/DNA level).

Fig. 9.4 full gene tree involving all TFs with BSs is shown. External color code displays the specificity-associated positions—to help visualization of palindromic combinations right positions are read in the complementary (*c*) strand: (NT-5_L, NT-4_L; NT-5^c_R, NT-4^c_R). The color background in several branches corresponds to different recognition AAs (a few classes were only enhanced). *Dots* in branches denote bootstrap values larger than 80 (for 100 trees total, see [22] for full tree and further details)

3 Conclusions

The possibility of finding simplifying principles in biology was contemplated since the early days of General System Theory [1]. This was part of a broader interest to search for principles in many diverse complex systems, a task that seems particularly timely nowadays [39]. Is this search correct in biology? People with different backgrounds (molecular biologists, physicists, engineers, etc.) would probably answer differently—some opinions being particularly adverse—but they would all most likely agree that some "pragmatic guidelines" [32] are at least needed if we are to understand biological systems.

Lessons on such quest for principles can also be observed, and here, I discussed two. The first one reproduces early discussions of Systems Theory on the presence of isomorphisms in science. I showed how specific ideas discussed in (computational) neuroscience (integrator/resonator oscillators, flexible discrimination, etc.) could also be applicable to genetic modules [9, 10]. Further interactions between these fields are expected [33]. Genetic circuits could be thus genuinely implementing a number of dynamical principles (e.g., relaxation oscillators) found in many unrelated areas. Moreover, knowledge of a particular principle at one level of biological organization could help us to understand/identify patterns at a distinct level. This is my second proposal that I illustrated with an example of how to identify TF-DNA recognition rules in bacteria [22]. To this aim, we made use of a widespread principle of bacterial regulation, i.e., local control of expression [23], to modify standard phylogenetic footprinting techniques. Additional examples include the use of principles of metabolic [34] or signaling [35] networks to better understand the forces behind gene dispensability and duplication.

Could one anticipate future research areas to develop? Let me suggest two that balance both fundamental and applied research. We can broadly ask first, thanks to current technology, which principle is at use in a particular biological context, how is it genetically implemented (e.g., which type of circuit and molecular agents), and what type of evolutionary forces could have originated such genetic architecture as compared to other accessible alternatives (e.g., [36–38]). These studies could also contribute to our understanding of how adaptive forces eventually work at different levels of biological organization [40], a crucial question indeed. A second attractive area is that of combining those principles directly observed in biological systems to motivate the construction of novel molecular (e.g., synthetic biology [5]) or engineering (e.g., bio-inspired robotics [41]) systems with many possible applications. The successful, and unsuccessful, assembly of these artificial systems would in the end help us to better delineate the evolutionary processes and ecological scenarios contributing to the emergence of the principles tested [42].

Given the relevance of these practical and fundamental questions, the pursuit of identifying a number of guiding rules in biology emerges as a decisive one.

Acknowledgments I thank Raúl Guantes for discussions over the years, Francisco M. Camas for discussions and comments on an earlier draft, and Ministerio de Ciencia, Tecnología e Innovación (Spain) Grant BFU2008-03632/BMC for funding.

References

1. von Bertalanffy L (1969) General systems theory. George Braziller, New York
2. Watson JD, Baker TA, Bell SP et al (2008) Molecular biology of the gene. Pearson/Benjamin Cummings, San Francisco
3. Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999) From molecular to modular cell biology. Nature 402(Suppl 6761):C47–52
4. Nurse P, Hayles J (2011) The cell in an era of systems biology. Cell 146:850–854
5. Khalil AS, Collins JJ (2010) Synthetic biology: applications come of age. Nat Rev Genet 15:367–377
6. Savageau MA (1976) Biochemical systems analysis: a study of function and design in molecular biology. Addison Wesley, Boston
7. Alon U (2007) Introduction to systems biology: design principles of biological circuits. Chapman and Hall/CRC, Boca Raton
8. Bray D (2009) Wetware a computer in every living cell. Yale University Press, New Haven
9. Guantes R, Poyatos JF (2006) Dynamical principles of two-component genetic oscillators. PLoS Comput Biol 2:e30
10. Guantes R, Poyatos JF (2008) Multistable decision switches for flexible control of epigenetic differentiation. PLoS Comput Biol 4:e1000235
11. Hodgkin AL (1948) The local electric changes associated with repetitive action in a non-medulated axon. J Physiol 107:165–181
12. Izhikevich EM (2000) Neural excitability, spiking, and bursting. Int J Bifurcat Chaos Appl Sci Eng 10:1171–1266
13. Machens CK, Romo R, Brody CD (2005) Flexible control of mutual inhibition: a neural model of two-interval discrimination. Science 307:1121–24
14. Wong KF, Wang XJ (2006) A recurrent network mechanism of time integration in perceptual decisions. J Neurosci 26:1314–1328
15. Niwa H, Toyooka Y, Shimosato D, Strumpf D, Takahashi K et al (2005) Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell 123:917–929
16. Schwarz M, Cecconi F, Bernier G, Andrejewski N, Kammandel B et al (2000) Spatial specification of mammalian eye territories by reciprocal transcriptional repression of Pax2 and Pax6. Development 127:4325–4334
17. Mikeladze-Dvali T, Wernet MF, Pistillo D, Mazzoni EO, Teleman AA et al (2005) The growth regulators warts/lats and melted interact in a bistable loop to specify opposite fates in Drosophila R8 photoreceptors. Cell 122:775–787
18. Strogatz SH (2000) Nonlinear dynamics and Chaos: with applications in physics, biology, chemistry and engineering. Perseus, Cambridge
19. Kaern M, Elston TC, Blake WJ, Collins JJ (2005) Stochasticity in gene expression: from theories to phenotypes. Nat Rev Genet 6:451–464
20. Seeman NC, Rosenberg JM, Rich A (1976) Sequence-specific recognition of double helical nucleic acids by proteins. Proc Natl Acad Sci USA 73:804–08
21. Matthews BW (1988) Protein-DNA interaction. No code for recognition. Nature 335:294–295
22. Camas FM, Alm EJ, Poyatos JF (2010) Local gene regulation details a recognition code within the LacI transcriptional factor family. PLoS Comput Biol 6:e1000989
23. Camas FM, Poyatos JF (2008) What determines the assembly of transcriptional network motifs in *Escherichia coli*? PLoS One 3:e3657

24. Weickert MJ, Adhya S (1992) A family of bacterial regulators homologous to Gal and Lac repressors. J Biol Chem 267:15869–15874
25. Lewis M (2005) The *lac* repressor. Compt Rendus Biol 328:521–548
26. Alm EJ, Huang KH, Price MN, Koche RP, Keller K et al (2005) The MicrobesOnline web site for comparative genomics. Genome Res 15:1015–1022
27. Kazakov AE, Cipriano MJ, Novichkov PS, Minovitsky S, Vinogradov DV et al (2007) RegTransBase – a database of regulatory sequences and interactions in a wide range of prokaryotic genomes. Nucleic Acids Res 35:D407–412
28. Ureta-Vidal A, Ettwiller L, Birney E (2003) Comparative genomics: genome-wide analysis in metazoan eukaryotes. Nat Rev Genet 4:251–262
29. Desjarlais JR, Berg JM (1992) Toward rules relating zinc finger protein sequences and DNA binding site preferences. Proc Natl Acad Sci USA 89:7345–7349
30. Sartorius J, Lehming N, Kisters B, von Wilcken-Bergmann B, Müller-Hill B (1989) *lac* repressor mutants with double or triple exchanges in the recognition helix bind specifically to lac operator variants with multiple exchanges. EMBO J 8:1265–1270
31. Perros M, Steitz T (1996) DNA looping and Lac repressor-CAP interaction [comment on "Crystal structure of the lactose operon repressor and its complexes with DNA and inducer"]. Science 274:1929–1930 [author response 1931–1932]
32. Keller EF (2002) Making sense of life: explaining biological development with models, metaphors, and machines. Harvard University Press, Cambridge
33. De Schutter E (2008) Why are computational neuroscience and systems biology so separate? PLoS Comput Biol 4:e1000078
34. Papp B, Pál C, Hurst LD (2004) Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. Nature 429:661–664
35. Soyer OS, Creevey CJ (2010) Duplicate retention in signalling proteins and constraints from network dynamics. J Evol Biol 23:2410–2421
36. Acar M, Becskei A, van Oudenaarden A (2005) Enhancement of cellular memory by reducing stochastic transitions. Nature 435:228–232
37. Camas FM, Blázquez J, Poyatos JF (2006) Autogenous and nonautogenous control of response in a genetic network. Proc Natl Acad Sci USA 103:12718–12723
38. Cağatay T, Turcotte M, Elowitz MB, Garcia-Ojalvo J, Süel GM (2009) Architecture-dependent noise discriminates functionally analogous differentiation circuits. Cell 139:512–522
39. Ostrom E (2005) Understanding institutional diversity. Princeton University Press, Princeton
40. Okasha S (2006) Evolution and the levels of selection. Oxford University Press, Oxford
41. Pfeifer R, Lungarella M, Iida F (2007) Self-organization, embodiment, and biologically inspired robotics. Science 318:1088–1093
42. Floreano D, Keller L (2010) Evolution of adaptive behaviour in robots by means of Darwinian selection. PLoS Biol 8:e100029