On the dynamics of gene regulatory networks

Roberto Serra
Modena and Reggio Emilia University
European Centre for Living Technology, Venice
roberto.serra@unimore.it

Abstract— **It is shown that a simplified model of genetic regulatory networks, aimed at the study of their generic properties, can shed light on some important biological phenomena. Two cases are analyzed, namely perturbations in gene expression induced in an organism by the knock-out of selected genes, and cell differentiation. The role of simplified models in biology is discussed.**

I. INTRODUCTION

The wealth of data nowadays available in genomics and other -omics has superseded our understanding of the key biological processes which they refer to, so there is a strong need for theories and models in order to make sense of the data themselves.

The movement of systems biology has contributed to shift the focus of research from a naive genocentric viewpoint to a more sophisticated approach, which takes properly into account system-level interactions. In this field, most models are quite specific, as they refer to the behaviour of a particular organ (like e.g. the heart in mammals) or to a particular genetic-metabolic subsystem.

A useful complementary approach (which has been called complex systems biology by Kaneko [1]) is centered instead on the search for generic properties, common to many different biological systems. An example of this kind of properties is the widespread fat-tailed distributions of several biological variables, and an even more striking example is provided by the scaling law which relates power consumption to body mass of different living species, irrespective of the differences among their anatomical and physiological properties [2]. Data indicate that the power consumption grows as $mass^{3/4}$. It seems indeed that the scaling law applies to very different sizes, from blue whales to small mammals and birds to mithocondria, thus spanning an impressive range of 20 orders of magnitude. Interestingly enough, the smartest explanation of this regularity and of the value of the exponent is related to the number of (relevant) spatial dimensions, i.e. 3, and to the (generic) hypothesis that biological evolution has tuned the features of living organisms in order to optimize power efficiency (under suitable constraints).

The search for generic properties of living beings had been pioneered by Stuart Kauffman, who introduced his model of random boolean networks (RBNs) more than 40 years ago in an attempt at exploring the properties of genetic regulatory networks. The model later became popular, in particular in the complex systems community, but it has only recently been shown that it can actually describe some quantitative features of real genetic regulatory networks. The possibility to verify the appropriateness of such a model has been opened by the availability of DNA microarrays, which allow genome-wide monitoring of the changes in gene expression levels. In this paper, after reviewing the RBN model in section 2, I will briefly discuss in section 3 its application to the study of perturbations induced by single gene knock-out in the yeast S. Cerevisiae, which was the first application of RBNs to the simulation of quantitative properties of real genes in a cell. Interestingly enough, the same study opens a way to test one of the strongest claims which have been put forth in complex systems biology, i.e. that evolution has led organisms to critical dynamical states, intermediate between order and chaos [3], which are sometimes referred to as "the edge of chaos". This aspect will also be discussed in section 3.

In section 4 I will describe another application of RBNs to real biological systems. In this case the quantitative data which can validate or disprove the model are not yet available, however the model is interesting in principle, as it shows that it is possible to describe several different phenomena involved in cell differentiation in a unified way by supposing that differentiation is an emergent phenomenon of a genetic regulatory network, without postulating particular gene circuits. It is also possible to devise experiments to test the hypotheses which lie at the basis of the model, as well as some of their consequences.

A final remark concerns the reference list, which might have become very long, had all the relevant original papers been mentioned. I prefer to keep the list short, quoting only a few papers where all the other relevant references can easily be found.

II. RANDOM BOOLEAN NETWORKS

The model is fairly well-known and the reader is referred to [4,5] for a more detailed description. A RBN is a dynamical system whose N variables take values {1,0} which can change in time according to a well-defined function (called transition function) of their inputs. It is convenient to think of it as a directed graph, with nodes associated to variables. If variable *i* depends upon variable *j* then there is a link from node *j* to node *i*. In the case considered here, all the nodes have exactly *k* input links, and the updating is synchronous. The network is random in that both the connections are drawn at random (choosing the *k* inputs to a node with uniform probability among the other *N-1*) and the Boolean function associated to a node is chosen at random. Usually one either chooses the Boolean function with

uniform probability from a predefined set, or generates each Boolean function by associating with a certain probability the output 1 or 0 to each of the 2^k input possible input vectors.

The system is deterministic and synchronous, therefore if N is finite its asymptotic states are cycles (a fixed point being a cycle with period one). Depending upon its structural features (i.e. topology and boolean functions), a family of networks shows a typical time behaviour, although single network realizations can behave in a way different from that typical of their family. Indeed, some combinations of structural parameters give rise to behaviours which have been termed ordered, other combinations lead to disordered or "chaotic" states. In the case of ordered systems the typical length of the asymptotic cycles increases slowly with the system size; moreover, it often happens that two nearby initial conditions evolve to the same final attractor. In the case of disordered networks the length of the cycles increases sharply with N, and it often happens that close initial conditions lead to different attractors. Critical networks have been defined as those whose structural parameters take values which are intermediate between those of ordered and those of disordered networks.

A bold theoretical ansatz which has been proposed [3] is that critical RBNs are endowed with features which make them particularly well-suited to perform complex tasks in a changing environment; therefore it has been argued that biological evolution should have driven biological organisms in, or close to this region in parameter space.

III. Perturbations in Gene Regulatory Networks

It has recently been possible to study the expression levels of all the genes of an organism, and to compare their global properties with those of genetic network models. A more detailed description of the results summarized below can be found in [6-8] and further references quoted therein.

In an important series of experiments a single gene of *S. cerevisiae* was knocked-out, and the expression levels of all the genes, in cells with a knocked-out gene, was compared with those in normal, wild type cells. In order to make precise statements about the number of genes perturbed in a given experiment, and to compare them with Boolean models, it is required that a threshold be defined, such that the difference is regarded as "meaningful" if the ratio of the expression of gene *i* in experiment *j* to the expression of gene *i* in the wild type cell is greater than the threshold (or smaller than its reciprocal). In order to describe the global features of these experiments it is convenient to introduce the notion of avalanche, which is the number of genes affected by the perturbation induced by a particular knock-out experiment.

The knock-out experiment can be simulated *in silico* by comparing the evolution of two identical RBNs which start from the same state of an attractor, the only difference being that one gene is clamped permanently to the value 0 in the network which simulates knock-out. A gene belongs to the avalanche associated to a particular knock-out if it differs in the final states of the two networks at least once in the attractor cycle. The initial simulations were performed using a classical RBN with 2 input connections per node, restricting the set of Boolean functions to the so-called canalyzing ones. The data set concerns 6325 genes and 227 experiments and the comparison with the experimental distribution of avalanches turned out to be good.

The reason why such a simple model worked so well has been uncovered by analytical methods which have proven that the distribution of avalanches depends only upon the outdegree distribution, while the indegree distribution plays no role [7]. Moreover, in the case of classical random Boolean networks, where the distribution of outgoing connections is Poissonian, it can be also proven that the distribution of small avalanches depends only upon a single parameter, the so-called Derrida exponent which is given by the equation:

$$\lambda \equiv (1-q)A$$

where A is the average connectivity of the network and q is the probability that a chosen node does not change its value when one (and only one) of its inputs has changed (note that q depends on the choice of the set of Boolean functions). λ had been introduced in the past in order to distinguish between ordered and disordered dynamical regimes (1 being the critical value), and it turns out that it also rules the distribution of avalanches. Therefore it is possible to estimate its value from the distribution of avalanches, so these analyses provide a general way to test the criticality hypothesis and, within the limitations of the data set presently available, they also provide support to it.

IV. Cell Differentiation

One of the major challenges in complex systems biology is that of providing a general theoretical framework to describe the phenomena involved in cell differentiation, i.e. the process whereby stem cells, which can develop into different types, become progressively more specialized. The model described below (for more details see 9-10 and further references quoted therein) is an abstract one (it does not refer to a specific organism or cell type) and it aims at reproducing the most relevant features of the process: (i) the existence of different degrees of differentiation, that span from totipotent stem cells to fully differentiated cells; (ii) stochastic differentiation, where populations of identical multipotent cells stochastically generate different cell types; (iii) deterministic differentiation, where signals trigger the progress of multipotent cells into more differentiated types, in well defined lineages; (iv) limited reversibility: differentiation is almost always irreversible, but there are limited exceptions under the action of appropriate signals; (v) induced pluripotency: fully differentiated cells can come back to a pluripotent state by modifying the expression of some genes and (vi) induced change of cell type: modification of the expression of few genes can directly convert one differentiated cell type into another.

The key hypotheses are that the differentiation process is an emerging property due to the interactions of very many genes (so its main features should be shared by a variety of different organisms) and that cellular noise plays a crucial role. To check these hypotheses a noisy version of the RBN model can be used (briefly refereed to as NRBN).

Noise is modelled as a transient flip of a single node, chosen at random. Attractors of deterministic RBNs are

unstable with respect to noise even at these low levels, and if a node is flipped for a single time step in an attactor state one sometimes observes transitions from that attractor to another one. Therefore, by flipping all the states belonging to the attractors of a RBN, it is possible to create a complete map of the transitions among the attractors. In these conditions single attractors can no longer be associated to cell types, as it is usually assumed [4]. Ribeiro and Kauffman [11] observed that it is possible to identify in the attractors' landscape subsets of attractors, which they called Ergodic Sets, which entrap the system in the long time limit, so the system continues to jump between attractors which belong to the set. Unfortunately it turns out that most NRBNs have just one such set: this observation rules out the possibility to associate them to cell types.

A possible solution to this problem is based on the observation that flips are a kind of noise fairly intense, as they amount to silencing an expressed gene or to express a gene which would otherwise be inactive: a particular transition may well be an event too rare to happen with significant probability in the cell lifetime, if it can happen only by perturbing a specific gene, or very few ones. It is possible therefore to introduce a threshold θ, and to neglect all the transitions having an occurrence probability lower than that. In such a way, the notion of Ergodic Set has to be modified in that of Threshold Ergodic Set (briefly, TES), a set of attractors linked only by jumps having a probability higher than θ, that entrap the system in the long time limit. A TES is therefore a subset of attractors which are directly or indirectly θ-reachable (reachable by means of transition whose probability exceeds the threshold θ) from at least another member of the set, and from which no transition can allow escaping. The threshold clearly is related to the level of noise in the cell, and scales with the reciprocal of the frequency of flips [9].

An ergodic set can be described therefore as a TES with θ=0; by increasing the threshold, one usually observes the birth of more and more TESs until, above a certain level, all the attractors of the deterministic model are also independent TESs. It is therefore possible to associate cell types to TESs, that represent coherent stable ways of functioning of the same genome even in the presence of noise. Several authors, on theoretical and experimental bases, associate different levels of noise to different levels of differentiation, the noise being higher the less differentiated the cell is. So the degree of differentiation appears to be related to the possibility for an undifferentiated cell to wander in a portion of phase space greater than the corresponding portions covered by more differentiated cells. In the NRBN model a convenient proxy for the available portion of phase space could be the number of different attractors belonging to the TES associated to that cell. A 0-threshold TES could therefore be associated to a totipotent cell, while as the threshold is increased smaller TESs appear, corresponding to more differentiated biological forms, until at high enough threshold values all the attractors are TESs, thus representing the fully differentiated cells. The increase of the threshold would correspond to a decrease of noise level, that could be related to an improvement in the mechanisms whereby fluctuations are kept under control. This association of differentiation to changes in the noise level represents the most stringent outcome of the model, and could be amenable to experimental test.

This hypothesis explains in a straightforward way the fact that there are different degrees of differentiation (i.e. property *i*), corresponding to different threshold values. It is also straightforward to describe stochastic differentiation (i.e. property *ii*): in this vision the fate of a cell depends on the particular attractor where the system is found when the noise level changes. The new cell type will be that corresponding to the new TES to which the attractor belongs at the new threshold level.

There exist also several processes, e.g. during the embryogenesis, in which cell differentiation is not stochastic but it is driven towards precise, repeatable types by specific chemical signals, which activate or silence some genes. These signals can be simulated by permanently fixing to 1 or 0 the state of some nodes. However this single action doesn't influence the level of noise, and therefore doesn't enable differentiation: in order to have deterministic differentiation it is necessary that so-called "switch" nodes exist, whose permanent perturbation coupled with a change in noise level always leads the system to the same TES. The existence of switch nodes has actually been verified to be a common property (found in about 1/3 of the nets), thereby proving the effectiveness of the model (i.e. property *iii*).

Moreover, by simulating the overexpression of a few genes, it has been possible to simulate also the other properties summarized above, and in particular the important processes of induced pluripotency and transitions among different cell types.

V. Conclusions

The above examples show that relatively simple generic models of gene regulatory networks are able to describe the quantitative features of the perturbations induced by gene knock-out, and to form the basis of an interesting model of cell differentiation. It goes without saying that more sophisticated models might be necessary to describe other important properties.

But the point which is worth stressing is that even models which are based on crude approximations may well provide insights on complex phenomena. This is well known in physics, where the aim is often that of finding very general properties, and simple models which display these properties are considered very useful. On the contrary, researchers in biology and social sciences often overstate the need for detailed models, which entrap all the features and the interactions which they suppose might be important - a requirement which, if taken too seriously, might even prevent the development of dynamical medelling in those fields. What might be envisaged is a hierarchy of models, where the simpler ones, which however capture some key properties, are used to understand some of the most relevant aspects, and to suggest further experiments. They may well be complemented by more detailed models able to provide more accurate quantitative (and sometimes also qualitative) behaviours.

ACKNOWLEDGMENT

Useful discussions with Stuart Kauffman, Annamaria Colacci, Marco Villani, Alex Graudenzi, Alessia Barbieri and Chiara Damiani are gratefully acknowledged. This work has been supported by the Italian MIUR-FISRproject nr. 2982/Ric(Mitica) and by the Dice project of the Fondazione di Venezia.

REFERENCES

[1] Kaneko, K. (2006) Life: An Introduction to Complex System Biology. Springer-Verlag, Berlin, New York

[2] West, G.B., Brown, J.H.(2005) The origin of allometric scaling laws in biology from genomes to ecosystems: towards a quantitative unifying theory of biological structure and organization. J. Exp. Biol. 208: 1575-1592

[3] Kauffman, S.A. (1995) At Home in the Universe. New York, Oxford University Press

[4] Kauffman S.A. (1993) The Origins of Order. New York, Oxford University Press

[5] Aldana M., Coppersmith S., Kadanoff L.P. (2003) Boolean dynamics with random couplings. In: Kaplan, E., Marsden, J.E., Sreenivasan, K. R. (Eds), Perspectives and Problems in Nonlinear Science, 23–89. Springer Applied Mathematical Sciences Series

[6] Serra, R. Villani, M., Semeria, A. (2004) Genetic network models and statistical properties of gene expression data in knock-out experiments. J. Theor. Biol. 227: 149-157, 2004

[7] Serra, R. Villani, M., Graudenzi, A., Kauffman, S.A. (2007) Why a simple model of genetic regulatory networks describes the distribution of avalanches in gene expression data. J. Theor. Biol. 249: 449-460

[8] Serra, R., Villani, M., Graudenzi, A., Colacci, A., Kauffman, S.A. (2008) The simulation of gene knock-out in scale-free random boolean models of genetic networks. Networks and heterogeneous media 3 (2), 333-343

[9] Serra, R. Villani, M., Barbieri, A., Kauffman, S.A., Colacci, A. (2010) On the dynamics of random boolean networks subject to noise: attractors, ergodic sets and cell types. J. Theor. Biol. 265: 185-193

[10] Villani, M., Barbieri, A., Serra, R. (2011) A Dynamical Model of Genetic Networks for Cell Differentiation. PLoS ONE 6 (3): e17703

[11] Ribeiro, A.S., Kauffman, S.A. (2007) Noisy attractors and ergodic sets in models of gene regulatory networks. J. Theor. Biol. 247: 743-755