

Chapter 2

Why Systems Biology Can Promote a New Way of Thinking

Alessandro Giuliani

Abstract This chapter deals with the effect Systems Biology had on the Nature of what we consider ‘an explanation’ in Biological Science. I try and demonstrate how the most relevant change carried out by Systems Biology approach was the shift from the molecular layer as the definitive place where causative process start to the elucidation of the among elements (at any level of biological organization they are located) interaction network as the main goal of scientific explanations. This change of perspective allows to dissipate a widespread idealistic nightmare looking at the single molecules as Maxwell-demon-like intelligent agents. The recognition that genes work in networks has as consequence the existence of discrete ‘allowed global modes’ of gene expression. This theoretical expectation was verified by the incredibly narrow space of different tissues (each corresponding to a largely invariant gene expression profile)—around 200 tissue types for all the metazoans emerging from the transfinite number of possible combinations of the expression values of around 30,000 genes. This is a crucial step for generating a scientifically sound framework to address global biological regulation.

Systems Biology approach makes obsolete the debate between ‘reductionist’ and ‘holistic’ approach in favor of a ‘middle-out’ paradigm formally identical to the time honored chemical thought. This is probably the brightest promise of Systems Biology to scientific knowledge.

Keywords Attractor in systems biology · Maxwell’s demons · Levinthal paradox · Network · Protein contact network (PCN) · Metabolic network

2.1 Introduction

The classical form in which biological systems are described (being they metabolic charts, gene expression regulation pathways, protein-protein interaction maps, food webs and so forth) corresponds to a set of nodes linked by edges in which the nodes

A. Giuliani (✉)

Environment and Health Department, Istituto Superiore di Sanità,
Viale Regina Elena 299, 00161 Roma, Italy
e-mail: alessandro.giuliani@iss.it

are the basic elements of the described system (genes, proteins, metabolites and so forth) and the edges connecting them some rules of the kind ‘is transformed into’ or ‘is increased by’.

The figures normally present in books and scientific papers implicitly consider these pathways as linear causative chains in which a signal starting from a molecular perturbation, after a sequence of if-then events, emerge as a biological end-point (Tun et al. 2011). Normally these processes are referred as ‘cascades’ provoking a progressive amplification of the initial stimulus (MacFarlane 1964). On the contrary, biological effectors (being them genes, proteins, hormones.), with only few exceptions, work in networks, and this fact implies a completely different form of biological regulation with respect to the ‘cascade’ model: the *entire* network has, thanks to its wiring structure, few preferred modes corresponding to the stable configuration of the network itself (Tun et al. 2011; Kauffman 1993; Huang et al. 2005), any perturbation, being it pharmacologically induced or coming from a mutation in a crucial gene, ends up into one or the other of these allowed states without any simple relation with the features of the applied perturbation (Tun et al. 2011).

Figure 2.1, taken from (Huang 2009), depicts the change in perspective shifting from pathway to network paradigm.

Without entering in the physical processes instantiating such intermingled (and largely invariant) networks, Systems Biology scholars can make use of a purely phenomenological view on biological regulation adopting some general concepts of dynamics. This is a necessity, if we consider that, thanks to the development of high throughput methodologies the graphs corresponding to the ‘perceived’ regulation networks became larger and larger and ask for some form of global analysis in order to get rid of their wild multiplicity.

The approach considering the graph as a system of differential equations in which an entering stimulus, correspondent to a modification of a peripheral node of the network, is progressively processed according to the wiring architecture and kinetics constraints of the network itself, while being the most potentially exhaustive avenue of research is severely hampered, in the case of biological systems, by a lot of problems. First of all the practical impossibility to attach to the whole set of edges reliable kinetic-like weights for quantifying the entity of the between elements correlation. Only in the case of very small networks this can be done by means of the statistical estimation of the parameters from experimental data, but it is well known that in physiological settings these weights can vary of orders of magnitude (Laughlin et al. 2000). Moreover in many cases we cannot rely on the complete knowledge of the wiring diagram of the network. For these (and other) reasons many authors preferred a purely topological approach to the analysis of biological networks, considering the presence of a link between two nodes as a pure yes/no binary relation and limiting themselves to statistical descriptions making use of the so called graph-invariants, i.e. a collection of indexes that, relying on the simple count of nodes and edges, enable the analyst to identify crucial elements of the network (like the so-called hubs, nodes engaged in a very large amount of relations) or to highlight specific features of the entire network architecture responsible for some aspects of the studied system

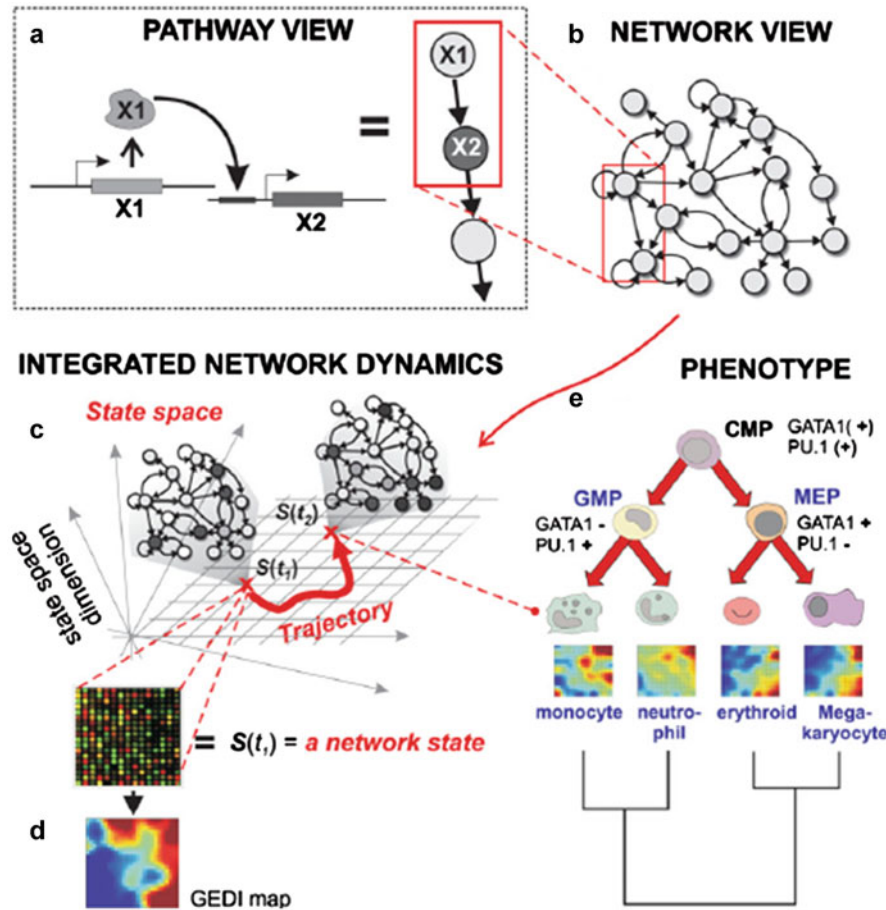


Fig. 2.1 Panel **a** reports the usual pathway view: a molecule produced by gene $X1$ acts as modulator (positive or negative) of gene $X2$ that in turn acts on another gene and so forth. If we consider these linear pathways are part of a network (panel **b**) we understand how the only allowable states are those corresponding to the network configurations that occupy energy minima in the state space having as dimensions the actual values of the nodes (panel **c**). The presence of a strong correlation structure among the nodes (in panel **d** the gene expression network is reported in which the node values correspond to the expression values of different ORFs) creates a 'rugged energy landscape' over the state space with only few valleys correspondent to differentiated tissues having a strongly invariant gene expression profile expressed as GEDI (Gene Expression Dynamics Inspector) map in which each pixel corresponds to a gene whose expression value is paralleled by a different color. These maps, collectively correspond to observed phenotypes (panel **e**)

behaviour (this is the case of the so called 'scale-free' architecture that was demonstrated to be at the basis of the huge resilience of biological systems) (Watts and Strogatz 2004).

The consideration of biological systems at the coarse-grain level of the graph topological approach is, in my opinion, a very important first step for the development of a sort of biological statistical mechanics in which the actual behaviour of the global system can be predicted by a convenient statistics over its constituent parts (Giuliani 2010).

In the case of statistical mechanics of inanimate systems this was the case with the Boltzmann microscopic definition of entropy as a statistical index computed over the microstates frequency distribution of the studied system (Giuliani 2010). This deliberate coarse-grain approach that abandoned the dream of following the trajectories of the single elements for a population level view, enabled scientists to get a link between microscopic and macroscopic physical descriptions (Laughlin et al. 2000; Watts and Strogatz 2004; Giuliani 2010; Karsenti 2008).

In the following I will try and describe the search for a Boltzmann-like approach to biology by the critical analysis of different regulation network-like systems, in the same time I hope it will be clear how this effort is strictly consistent with very fertile lines of epistemological lines of thought, mainly chemical research tradition and multidimensional statistics (Di Paola et al. 2012; Benigni and Giuliani 1994).

2.1.1 The Concept of Attractor in Systems Biology

The concept of attractor was developed in dynamical systems theory, where the whole system is thought as evolving towards a preferred (minimal energy) state called an attractor set, and represented such as a point, a curve, and a manifold in the state space. The study of folding process in proteins, where the impossibility for the linear chain of amino-acids to randomly explore all the possible configurations in biologically plausible times before settling down in the native 3D structure (the so called Levinthal paradox taking its name by the crucial observation made by Cyrus Levinthal that in his 1969 paper (Levinthal 1969) computes in the order of millions of years the duration of protein folding process as compared by the seconds to minutes effective actual time) is the field of biomolecular science where a ‘goal-oriented’ trajectory driven by the existence of a preferred configuration ‘attracting’ the system trajectories in the state space was studied more in depth.

Figure 2.2 reports a simplified view of the folding process of a protein in energetic terms: this representation is named ‘folding funnel’ (Dill and Chan 1997) and stresses the fact that different initial states corresponding to different positions in the upper part of the funnel converge on the same potential well at the bottom of the funnel.

The fact the potential well bottom (equilibrium state) does not correspond to a fixed configuration but to a set of possible states makes it possible protein dynamics that is crucial for exerting its physiological role. The fact protein molecule behavior can be fully interpreted as the dynamics of a network in which the amino-acid residues are the nodes and the between-residues contacts the edges (Di Paola et al. 2012) makes the folding funnel metaphor perfectly suited for gene expression network, the only difference being the knowledge we have of the physical forces shaping

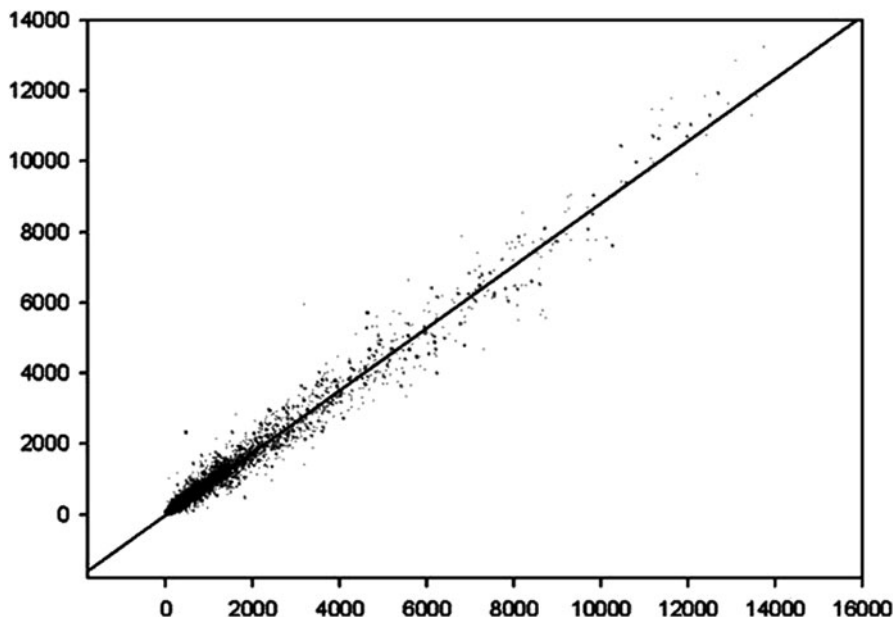


Fig. 2.3 The X and Y axes of the figure correspond to two independent samples of the same tissue type (in this case macrophages). The approximately 23,000 vector points correspond to different gene expressions coming from a microarray experiment. Notwithstanding the fact the graph spans four order of magnitudes there is a remarkable order of gene expression level corresponding to a Pearson correlation coefficient $r = 0.99$ between the profiles. This invariance comes from the fact each tissue is an attractor in the gene expression space, the (relatively small) scattering around the identity line corresponds to the motions ‘inside the attractor’, these motions are analogous to the dynamics of a protein molecule around its native state and are the only ones eventually affected by disease states or pharmacological perturbations. (Giuliani 2010)

of fact in two very important papers (Overington et al. 2006; Hopkins 2008) Overington, Hopkins and colleagues gave a very crude (but statistically clear) picture of the state of pharmacology research and development: the number of new drugs arriving at the market stage dramatically decreased starting from the 80’s of the last century and the classes of receptors they are supposed to bind were already known since 50 years, the concept of a ‘druggable genome’ with myriads of new drug targets supposed to be revealed by genome project simply does not exist or, better, the targets are not ‘druggable’. The same basic idea of network stable states allows to understand what happened: the only ‘simple targets’ whose modification is expected to give rise to a macroscopic, organism-scale observable effect are those located at the extreme periphery of the interaction network, while the modification of a node located in the internal position of the network is immediately buffered by the feedback relations so that the system cannot be modified by pharmacological intervention (Tun et al. 2011).

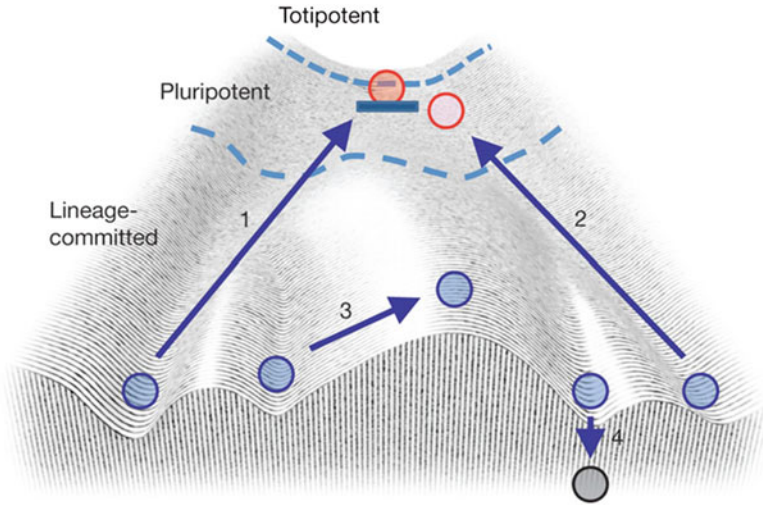


Fig. 2.4 At the cell population level, the ensemble of cells' potency in iPS reprogramming process can change in a probabilistic manner like rolling up and down on the epigenetic energy landscape towards a specific valley having definite potency (attractor state). A lower potency can be pushed up by a competent stimulus to another allowed potency level on the landscape corresponding to another (less stable) equilibrium endowed with a higher differentiation potential energy. Figure cited from (Yamanaka 2009)

On the contrary, the recent Nobel prize to Yamanaka and Gurdon tells us a completely different story (Yamanaka 2009; Yamanaka and Blau 2010): a bunch of effectors, in a still largely unknown way, is able to transmit to the 'system as a whole' an effective stimulus able to push the entire network along a 'counter-gradient' trajectory going back to an higher energy state corresponding to an undifferentiated, totipotent state. It is remarkable that Yamanaka explains this effect using the Conrad Waddington epigenetic landscape (Waddington 1957) a precursor of energy landscapes, in which unstable, 'high energy' states (and thus states in which the system can be modified by an external modification) are represented as ridges and stable states (attractors) by valleys, as depicted in Fig. 2.4, coming from a Yamanaka paper (Yamanaka 2009).

The attractor view allows scientists to eliminate the need to impose the presence of 'intelligent agents' (Maxwell's demons) in order to get rid of specific and finely tuned behaviors: the system 'lives' in a non-uniform state space (the ensemble of all the possible system configurations) characterized by a so called 'rugged landscape' (Frauenfelder 1991) where the energy minima (valleys of the landscape, quasi-equilibrium configurations) correspond to attractor states (Frauenfelder 1991).

Each system accommodates towards the energy minimum nearest to it, consistently with the marked 'context dependence' (e.g. sensitivity to microenvironment) of biological regulation. Metaphorically, C. H. Waddington (Waddington 1957) suggested that cell fate would be determined by a trajectory toward a local minimum

(attractor) on epigenetic energy landscape, where a series of “valleys” and “ridges” describe stable cellular states (local minima) and barriers (local maxima) between those states, respectively. The epigenetic landscape is a proposal for the existence of global molecular regulation in cell fate decision. The word ‘global’ underlines the fact ‘energy’ is computed over the entire state space (in the case of transcription dynamics, the genome-wide expression) and not over few specific genes.

Clearly, as above stated, this is a pure phenomenological proposal that does not enter the molecular mechanisms supporting it even if cytoskeleton organization (Ingber 1999) or confinement by phase transitions (Hyman and Simons 2012) allowing the selection of specific pathway in complex microenvironments are very plausible candidates. Limiting ourselves to data analysis coming from actual biological experimentation, it is sufficient to imagine these discrete states corresponding to ‘allowed positions in the transcriptome space’, coming from the presence of a still unknown origin field sensed by the entire genome and driving its collective behavior (Huang et al. 2005). It is worth noting that a very basic ‘toolbox’ made of principal component analysis (principal components being the coordinated fluxes of variations of many different genes), network invariant descriptors (with the assignment to each node a set of measures related to its role in the network wiring), cluster analysis (very dense clusters in the phase space correspond to attractors) are sufficient to undergo such ‘biological dynamics’ avenue of research (Huang et al. 2005; Huang 2009; Benigni and Giuliani 1994).

In (Huang et al. 2005), the authors offer a very thorough proof-of-concept of the relevance of considering a differentiated state of a cell population as an attractor in the proper dynamical sense. They demonstrate that, after perturbation induced by two completely different chemical stimuli (atRA and DMSO respectively) initially inducing a completely different response in terms of gene expression, the system returns back to the same attractor point in the genome expression phase space (see Fig. 2.5).

The above behavior corresponds exactly to the basic definition of an attractor as a state ‘attracting’ the perturbation trajectories of the system. This stems from the fact attractor states are stable and, if the perturbation is not sufficiently strong to push the system outside its ‘basin of attraction’, soon or later the system will come back to its original attractor state losing memory of the nature of the initial perturbation. Thus the specific differences in mechanism of action of the two effectors are not so relevant in terms of the resulting effect that is largely dependent on the affected system modes.

A very important consequence of the presence of an attractor-like regulation is the impossibility to maintain the classical discrimination of house-keeping vs. specifically regulated genes and the importance of low-variance genes with the consequent need to re-cast the idea of what a ‘pathway’ (or a ‘gene signature’) is (Tsuchiya et al. 2010; Venet et al. 2011).

Clearly it is for sure that different genes have different discrimination ability for specific diseases, and again these specific genes could be more useful than other for diagnostic purposes, but this has only to do with our specific discrimination goals: the system as it works in a self-coherent way on the whole-genome scale.

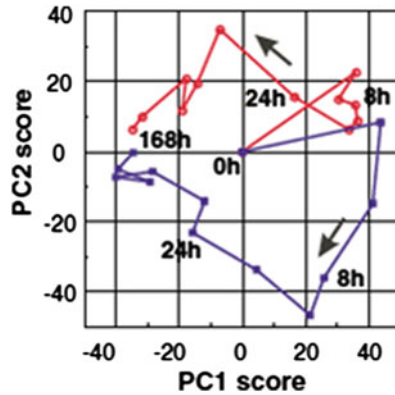


Fig. 2.5 Attractor approaching coming from two different trajectories: principal component analysis for highly expressed 2773 genes following atRA and DMSO stimulus shows two different trajectories on the space spanned by the first two principal components (PC1 and PC2; *red circles* for atRA; *blue squares* for DMSO). Figure comes from (Huang et al. 2005) and allows to appreciate how two different stimuli initially make the system to go away from the same initial state (0 h) toward two different directions, when the transients settle down and the system reaches a new stable configuration, the two trajectories converge to the same attractor state (168 h) losing memory of their different paths. The strong between genes correlation allows to collapse the entire multidimensional expression space into a bi-dimensional component space

Especially relevant is the fact that the genome dynamics involve both highly and lowly expressed genes, which are generally considered noisy and insignificant in microarray experiments.

These findings give an immediate explanation to the recent ‘iconoclastic’ results obtained by Venet et al. (2011) thoroughly commented in (Jordan 2012) demonstrating the practical equivalence between random collection of genes and specific signatures for breast cancer prognosis. Along the same line is the finding of the complete equivalence of different random gene selections for tracking hematopoietic differentiation demonstrated by Felli et al. (2010). In the attractor model, lowly expressed genes are effective players in global gene regulation, given they are integral part of collective expression modes elicited by the perturbation (treatment, mutation, differentiation stimulus, etc.); this implies that any sufficiently dense sampling of genetic probes gives us a relevant picture of the collective mode (Felli et al. 2010; Censi et al. 2011).

A very recent work describing the architecture of whole genome regulation as emerging from results coming from ENCODE project (Gerstein et al. 2012) is consistent with the view of a dense interconnected network working as a whole and thus asking for system-level description of gene expression dynamics.

2.1.2 *‘Bottom-up’, ‘Top-down’ or (Better) ‘Middle-out’?*

If we assume a classical molecular approach, we make the implicit assumption that the ‘effective flux’ of causation starts at the most microscopic level of the biological matter and progressively emerges at more macroscopic levels by an interaction chain.

We refer to this approach as ‘bottom-up’ and it was at the basis of molecular biology research in these last 50 years: each disease, each general condition is approached by looking for its molecular determinants.

On the other hand, a physician adopts a ‘top-down’ diagnostic approach, even if he is convinced the basic causative layer of the still unknown disease he suspects a given patient is affected is located at some fundamental level, he must orient the search for a proof of his conjectures in a top-down way by looking for objective data (biomarkers from blood or urine analysis, image analysis as NMR or X-rays, biopsies.) collected starting from the goal, and then driven by the global state of the patient that implicitly is supposed to influence the microscopic findings. The same ‘top-down’ approach is implicitly assumed in ‘goal-driven’ phenomena like development even if embryologists actively look for the way to turn development into a bottom-up explanation, being the top-down approach considered as a constraint arising from the lack of a sufficiently accurate knowledge of development.

All in all, the choice of one approach or the other is often not-decidable: from a certain point of view the ‘ultra-reductionist’ exclusively bottom-up perspective is totally unrealistic for the obvious reason any organism is subjected to a huge number of top-down constraints coming from the fact they live in a physical world (gravity (Ingber 1999), electromagnetism (Sebastian et al. 2001), thermodynamics (Shakhnovic 2006)) as well as from higher order perturbations affecting molecular targets (synaptic modifications induced by learning (Malenka and Nicoll 1993), hormonal changes due to psychological and social stress (Catalani et al. 2011)). All these phenomena ask for a top-down causation complementing the bottom-up mechanisms.

On the other hand, a purely top-down approach will end up into a pure descriptive/diagnostic and mainly tautological body of knowledge in which any knowledge element is at its best a ‘diagnostic marker’ of something else or, worst, a necessary consequence of a global principle. This happens for example in some misconceptions of evolutionary theory that virtually inhibit any fundamental research on the causes of observed phenomena by simple stating ‘if it is there it means that it must be there because it is convenient’ that is in some cases is nothing more nothing less than tautological ‘just-so-stories’.

Network (or better graph) paradigm are located half-way between these two extremes and for their very basic nature make these two opposite epistemological approaches obsolete.

The classic Königsberg bridge problem introduced graph theory in eighteenth century. The problem had the following formulation: does there exist a walk crossing each of the seven bridges of Königsberg exactly once? The solution to this problem appeared in ‘Solutio Problematis ad geometriam situs pertinentis’ in 1736 by Euler

Systems and Synthetic Biology

Singh, V.; Dhar, P.K. (Eds.)

2015, XI, 385 p. 114 illus., 75 illus. in color., Hardcover

ISBN: 978-94-017-9513-5