

Communication

Networks as a Privileged Way to Develop Mesoscopic Level Approaches in Systems Biology

Alessandro Giuliani

Environment and Health Department, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Roma, Italy; E-Mail: alessandro.giuliani@iss.it; Tel.: +39-6-4990-2579; Fax: +39-6-4990-2999

Received: 29 April 2014; in revised form: 21 May 2014 / Accepted: 28 May 2014 /

Published: 30 May 2014

Abstract: The methodologies advocated in computational biology are in many cases proper system-level approaches. These methodologies are variously connected to the notion of “mesosystem” and thus on the focus on relational structures that are at the basis of biological regulation. Here, I describe how the formalization of biological systems by means of graph theory constitutes an extremely fruitful approach to biology. I suggest the epistemological relevance of the notion of graph resides in its multilevel character allowing for a natural “middle-out” causation making largely obsolete the traditional opposition between “top-down” and “bottom-up” styles of reasoning, so fulfilling the foundation dream of systems science of a direct link between systems analysis and the underlying physical reality.

Keywords: bioinformatics; graph theory; chemistry; molecular biology; systems analysis

1. Introduction

The impressive development of high throughput screening procedures allowing thousands of different measures on the same sample generated in the last two decades unexpected problems to biological research.

The technical advancements made biologists to face a puzzling paradox: they fulfilled the dream of measuring the entire ensemble of players at a specific layer of biological regulation in parallel, without knowing how to manage this information burden. The huge disproportion between variables (in the order of thousands) and independent statistical units (samples, in the order of dozens) made irrelevant (for obvious over-fitting problems) the classical inferential statistics approaches used since almost a century [1,2]. On the other hand, the classical “box-and-arrow” style used so far for sketching a biological

mechanism in which the boxes are the relevant players (genes, proteins, metabolites) and the arrows correspond to a relation between the nodes at the extremities (e.g., “stimulates” “increases” “modifies”) becomes out-of-scope when the elements go from the order of 9–10 to the order of thousands. The emerging field of “systems biology” attained momentum in this last decade to face the above problems [3]. Sui Huang in his review [4] indicated two main approaches (and consequently groups of system biologists) to face the above-sketched “curse of dimensionality”. The first (and still more relevant in terms of number of followers) is reductionist (Prof. Huang refers to this view as “the localist view (those who see the trees first)”), the second approach is instead fully systemic (Prof. Huang refers to this view as “the generalist view (those who see the forest first)” [4]. Both the approaches build upon the construction of networks, *i.e.*, complex graphs having as nodes the players (genes, proteins, metabolites) linked by edges (functional relations to be discovered by data analysis applied to high-throughput techniques output). Localist view applies network approaches to single out “relevant genes and pathways” out of a plethora of false hits, on the contrary globalist approach considers networks as the image in light of a system working as a whole and network invariants as mesoscopic relevant variables for understanding the system at hand. To make a long story short, localists think systems biology main task is to develop efficient pattern recognition methods, globalists are persuaded the real task is to establish a physically motivated view of biological regulation. The localist view considers network approaches as a smart data analysis method to provide biologists with the “relevant genes” to consider, in the usual deterministic bottom-up way, as the primary causes of phenomena. Globalist view has the ambition of deriving an explanation at the mesoscopic level of among-genes relation structure. In the following, I will demonstrate how the second (truly systemic) approach is motivated by the reality of biological systems and not only by our epistemological preferences.

2. Microscopic Complexity Generates Macroscopic Simplicity

Let us start with some largely known (even if often misunderstood) facts:

- (1) Human DNA molecule is a giant polymer approximately two meters long confined into a space of the scale of microns [5]. This implies the logical information stored in the DNA sequence is not freely accessible like in the RAM of a computer or in an abstract Turing machine. The most crucial step allowing a specific transcription pattern of the stored information (*i.e.*, the correct genes at the correct expression level) is the unfolding of finely selected patches of the polymer correspondent to the genes of interest. Excluding the presence of any Maxwell’s demon, we must imagine the presence of relatively few “allowed configurations” of the DNA molecule exposing the “gene patterns” correspondent to different phenotypic states [6].

This state of affairs has two immediate consequences: (a) regulation does not happen only on a gene-by-gene basis but in terms of “global modes” spanning the entire genome and (b) only a discrete (and relatively small) number of gene expression patterns do exist.

The eventual local fine-tuning exerted on a gene-by-gene scale can be superimposed to the global “genome scale” regulation but explains a comparatively minor effect [7].

- (2) Tissues are populations made up of billions of cells that must work in a highly coordinated way (e.g., think of the necessity for cardiac muscle cells to oscillate in synchrony in order to obtain

a normal heartbeat). This coordination implies the specific folding/refolding of chromosomes described in the previous point, must span the entire tissue, thus implying the necessity of a “communication system” unifying the entire tissue. This coordination was demonstrated to happen even in the very artificial situation of cells cultured on a Petri dish that in turn display very clear “gene expression waves” indicating a highly coordinated expression pattern spanning billions of cells. The existence of a cytoskeleton network creating a continuous communication system traversing different cells and entering the nuclei, received many experimental proofs [8,9], even if we are still very far to clarify the biophysical mechanisms of the information transmission along this network.

- (3) Basic physical chemistry tells us that an interaction involving the ordered encounters of three particles is practically impossible to achieve in a purely diffusive regime in solution. Ordinary metabolic processes like the biosynthesis of lipids [10] involve a dozen of ordered hits: this means we are forced to consider metabolism as a phenomenon occurring inside a structured solid phase. The presence of such ordered phase spanning supra-molecular scales is evident considering the dramatic effects exerted by lack of gravity in cells grown in the space shuttle [11]: thousands of genes appear to be dis-regulated. If we consider gravity has a null effect on the molecular scale, the huge biological consequences of the lack of gravity imply a top-down causation in which the maintaining of a supra-molecular order encompassing huge populations of cells comes before the molecular scale events and not the other way around as normally assumed by molecular biologists [12].

All the above facts (and many others I did not mention but again related to the emergence of a striking order from a wild combinatorial (see for example [13]) tell us of the existence of few privileged forms. Here the word form is intended in its fundamental meaning of a specific set of relation between elements and roughly corresponds to the notion of system governing biological regulation. The existence of few “allowed” forms is evident at all the levels of biological regulation. Only 200 tissues (each one corresponding to a strictly invariant profile of average expression pattern involving around 30,000 genes) are present in all the metazoans, only 1000 folds are sufficient to explain the three-dimensional arrangements of the millions of different protein molecules, only three basic body plans explain the shape of all the animals [14].

This means that the space of what is “effectively observed” at a macroscopic scale is drastically reduced with respect to the space of what can be “potentially observed” in terms of microscopic information. A wild complexity at the molecular scale results into a fundamental simplicity (few patterns) at the macroscopic scale. In other words, microscopic complexity gives rise to macroscopic simplicity through the agency of relatively few mesoscopic organization principles.

3. Networks are the Most Effective Paradigm to Look for Mesoscopic Organization Laws

From what said above it should be clear why the goal of systems biology is not to find a needle in a haystack isolating the “most relevant pathway” (or gene, or protein) embedded in a mess of irrelevant information. On the contrary, the task of a truly “systems” biology is to look for the laws of form [15] explaining the presence of few allowable patterns out of the myriads of possible alternatives. The quest

for these laws of form can only marginally rely on the usual “differential equations” style: biology deals with non-stationary systems in which the rate constants change abruptly across many orders of magnitude [16] moreover, it is practically impossible to study in isolation the single elements of a system. An enzymatic reaction has completely different properties when is studied in a test tube with respect to when is studied *in vivo* together with a set of other co-occurring reactions [17].

Graph theory allows for a very natural approach to systems biology focusing on the relation between the elements of the studied phenomenon. We can roughly describe the network approach as the answer to the question “What can we derive from the sole knowledge of the wiring diagram of a system?” This wiring diagram does not imply any deep knowledge: the links (edges) between the systems elements (nodes) could simply represent very basic relations as “is correlated with”, “is located near”, and “communicates with” without specifying any functional form of the relation. A graph G is made of a set of nodes linked by a set of edges and can be described by a set of measures located at different level of definition going from microscopic (single nodes) to macroscopic (entire network) passing by mesoscopic (architectural principles) levels [18].

An example of microscopic level descriptor is the “node degree” corresponding to the number of edges ending (starting) from a given node, while the distribution of node degrees over the entire network, allowing to discriminate alternative wiring models as random and scale-free networks, is an example of macroscopic descriptor [19]. The average shortest path (*asp* or characteristic length of the graph) is a mesoscopic descriptor corresponding to the average of the “shortest paths” (*i.e.*, the paths linking two nodes of the network using the minimum number of edges) computed over all the distinct node pairs in G .

These views are each other naturally connected without the agency of any explicit model other than the wiring diagram, thus the removal of a node with an high “Betweenness” (a microscopic, node-level feature corresponding to the number of shortest paths passing by the node) will have an elevated impact on the value of the mesoscopic variable *asp*.

This “natural multilevel” character of graphs allows for the discovery of laws of form in terms of network descriptors optima. The need of maximizing the efficiency of signal transmission in allosteric transitions (global configuration changes of protein molecules following the interaction with a ligand at a specific site [20]) makes *asp* minimization a crucial mesoscopic principle driving protein structures [21].

The stress induced by a chronic disease like atrial fibrillation on heart, increases the global connectivity (average node degree) of the gene expression network (the correlation matrix between different gene expression levels) [22]. The edges whose deletion breaks the total connectivity of the metabolic network (the graph having metabolites as nodes and enzymes transforming metabolite i into j as edges) are the most probable candidate to be the target of lethal mutations [23]. In a recent work Dehmer *et al.* [24] demonstrated the efficiency of graph invariants coming from the gene expression signatures of different patient samples in predicting different disease states with no explicit reference to the specific nature of the involved nodes (genes).

4. Conclusions

The above examples (selected from a rapidly growing literature) mark the birth of a “middle-out” [18,25,26] style of reasoning, in which the focus is neither in the search for fundamental

atomisms from which to start to reconstruct the entire picture (bottom-up approach) nor in the search for consequences of macroscopic conditions (top-down approaches). Network paradigm allows to focus on the space “between-objects” in which lives the relational structure linking the different players without the need of knowing in advance neither the general laws governing the system at hand nor the fine structure of the constituent elements. This approach is particularly fruitful in biology where an incredibly rich phenomenology generates from the relations among many different heterogeneous elements. It is worth noting, as described in the first paragraph, biological networks are not only a convenient way to formalize data but do have a material counterpart in phenomena like DNA compaction, protein folding, and metabolism. This allows an immediate translation of the systemic approach into the elucidation of physical realities so allowing for a full exploitation of systems analysis potentialities.

Acknowledgments

I wish to thank my friends and colleagues Luisa Di Paola, Mariano Bizzarri, Sui Huang, Masa Tsuchiya, Katrina Erenpreisa, Kumar Selvarajoo for their continued interest and fruitful discussions on these themes.

Conflicts of Interest

The author declares no conflict of interest.

References

1. Roden, J.C.; King, B.W.; Trout, D.; Mortazavi, A.; Wold, B.J.; Hart, C.E. Mining gene expression data by interpreting principal components. *BMC Bioinform.* **2006**, *7*, doi:10.1186/1471-2105-7-194.
2. Sharov, A.A.; Dudekula, D.B.; Ko, M.S.H. A web-based tool for principal component and significance analysis of microarray data. *Bioinformatics* **2005**, *21*, 2548–2549.
3. Kitano, H. Systems biology: A brief overview. *Science* **2002**, *295*, 1662–1664.
4. Huang, S. Back to the biology in systems biology: What can we learn from biomolecular networks? *Briefings Funct. Genom. Proteom.* **2004**, *2*, 279–297.
5. Yoshikawa, K.; Takahashi, M.; Vasilevskaya, V.V.; Khokhlov, A.R. Large discrete transition in a single DNA molecule appears continuous in the ensemble. *Phys. Rev. Lett.* **1996**, *76*, 3029–3034.
6. Giuliani, A. Collective motions and specific effectors: A statistical mechanics perspective on biological regulation. *BMC Genomics* **2010**, *11*, doi:10.1186/1471-2164-11-S1-S2.
7. Tsuchiya, M.; Selvarajoo, K.; Piras, B.; Tomita, M.; Giuliani, A. Local and Global responses in complex gene regulation networks. *Phys. A* **2009**, *388*, 1738–1746.
8. Huber, F.; Schnauss, J.; Ronicke, S.; Rauch, P.; Müller, K.; Fütterer, C.; Käsel, J. Emergent complexity of the cytoskeleton: from single filaments to tissue. *Adv. Phys.* **2013**, *62*, 1–112.
9. Butcher, D.T.; Alliston, T.; Weaver, V.M. A tense situation: Forcing tumor progression. *Nat. Rev. Canc.* **2009**, *9*, 108–121.
10. Ohlrogge, J.; Browse, J. Lipid biosynthesis. *Plant Cell* **1995**, *7*, 957–970.

11. Hammond, T.G.; Benes, E.; O'Reilly, K.C.; Wolf, D.A.; Linnehan, R.M.; Taher, A.; Kaysen, J.H.; Allen, P.L.; Goodwin, T.J. Mechanical culture conditions effect gene expression: Gravity-induced changes on the space shuttle. *Physiol. Genom.* **2000**, *3*, 163–173.
12. D'Anselmi, F.; Valerio, M.C.; Cucina, A.; Galli, L.; Proietti, S.; Dinicola, S.; Pasqualato, A.; Manetti, C.; Ricci, G.; Giuliani, A.; Bizzarri, M. Metabolism and cell shape in cancer: A fractal analysis. *Int. J. Biochem. Cell. Biol.* **2011**, *43*, 1052–1058.
13. Tompa, P.; Rose, G.D. The Levinthal paradox of the interactome. *Protein Sci.* **2011**, *20*, 2074–2079.
14. Martindale, M.Q.; Lee, P.N. The development of form: Causes and consequences of developmental reprogramming associated with rapid body plan evolution in the bilaterian radiation. *Biol. Theory* **2013**, *8*, 253–264.
15. Denton, M.; Marshall, C. Laws of form revisited. *Nature* **2001**, *410*, doi:10.1038/35068645.
16. Von Dassow, G.; Meir, G.; Munro, E.M.; Garret, M.O. The segment polarity network is a robust developmental module. *Nature* **2000**, *406*, 188–192.
17. Welch, G.R.; Keleti, T. On the “cytosociology” of enzyme action *in vivo*: A novel thermodynamic correlate of biological evolution. *J. Theor. Biol.* **1981**, *93*, 701–735.
18. Giuliani, A.; Filippi, S.; Bertolaso, M. Why network approach can promote a new way of thinking in biology. *Front. Genet.* **2014**, *5*, 1–4.
19. Emmert-Sreib, F.; Dehmer, M. Networks for systems biology: Conceptual connection of data and function. *IET Syst. Biol.* **2011**, *5*, 185–207.
20. Di Paola, L.; De Ruvo, M.; Paci, P.; Santoni, D.; Giuliani, A. Protein contact networks: An emerging paradigm in chemistry. *Chem. Rev.* **2013**, *113*, 1598–1613.
21. Tasdighian, S.; Di Paola, L.; De Ruvo, M.; Paci, P.; Santoni, D.; Palumbo, P.; Mei, G.; Di Venere, A.; Giuliani, A. Modules identification in protein structures: The topological and geometrical solution. *J. Chem. Inf. Model.* **2014**, *54*, 159–168.
22. Censi, F.; Giuliani, A.; Bartolini, P.; Calcagnini, G. Multiscale graph theoretical approach to gene regulation networks: A case study in atrial fibrillation. *IEEE Trans. Biomed. Eng.* **2011**, *99*, 1–5.
23. Palumbo, M.C.; Colosimo, A.; Giuliani, A.; Farina, L.; Essentiality is an emergent property of metabolic network wiring. *FEBS Lett.* **2007**, *581*, 2485–2489.
24. Dehmer, M.; Mueller, L.A.J.; Emmert-Streib, F. Quantitative network measures as biomarkers for classifying prostate cancer disease states: A systems approach to diagnostic biomarkers. *PLoS ONE* **2013**, *8*, doi:10.1371/journal.pone.0077602.
25. Csermely, P.; Korcsmáros, T.; Kiss, H.J.; London, G.; Nussinov, R. Structure and dynamics of molecular networks: A novel paradigm of drug discovery: A comprehensive review. *Pharmacol. Therapeut.* **2013**, *138*, 333–408.
26. Laughlin, R.B.; Pines, D.; Schmalian, J.; Stojković, B.P.; Wolynes, P. The middle way. *Proc Natl. Acad. Sci. USA* **2000**, *97*, 32–37.