

Will systems biology offer new holistic paradigms to life sciences?

Filippo Conti · Maria Cristina Valerio ·
Joseph P. Zbilut · Alessandro Giuliani

Received: 1 April 2008 / Revised: 5 May 2008 / Accepted: 6 May 2008 / Published online: 30 May 2008
© The Author(s) 2008

Abstract A biological system, like any complex system, blends stochastic and deterministic features, displaying properties of both. In a certain sense, this blend is exactly what we perceive as the “essence of complexity” given we tend to consider as non-complex both an ideal gas (fully stochastic and understandable at the statistical level in the thermodynamic limit of a huge number of particles) and a frictionless pendulum (fully deterministic relative to its motion). In this commentary we make the statement that systems biology will have a relevant impact on nowadays biology if (and only if) will be able to capture the essential character of this blend that in our opinion is the generation of globally ordered collective modes supported by locally stochastic atomisms.

Keywords Systems biology · Bioinformatics ·
Microarray · Theoretical biology

In economics, Dee Hock, the founder of the VISA credit card association, coined the term “chaordic” referring to a system that simultaneously possess characteristics of both chaos and order. In the above context, the use of the word “chaos” is purely conversational and does not refer to the

specific “deterministic chaos” paradigm (being fully deterministic exhibits only a parody of complexity), but simply to the presence of unpredictable behaviour even in presence of some known principles of functioning.

At odds with Descartes’ dictum “Particularity and separability are infirmities of the mind, not characteristics of the Universe” (Hock 1999), the chaordic paradigm affirms that any reliable picture of the whole system must be a bottom-up one, in which general principles arise as “correlative properties” of the system contingencies (the “particularities” of the Descartes dictum) and not as consequences of top-down laws. This vision, at least *in nuce*, encompasses a holistic appreciation of the studied systems. The word holistic, in our opinion, has a too strong esoteric connotation and is decidedly too vague (the web is full of centres of holistic medicine, massage, thinking, etc.) in its present formulation to be fruitfully used in science. The aim of this paper is to try to derive a directly operational meaning to this term connecting it both to the clearly stated concept of emergence and to a set of already established experimental and data analysis tools routinely used in biological sciences, accomplishing this task we will try and explain what we perceive as the most fruitful ‘research avenue’ for systems biology. It is worth noting the emphasis on “correlative properties” as the key for system understanding is at the basis of the time honoured multi-dimensional statistics approach looking at systems as an intermingled mix of signal and noise in which signal is defined as the ‘correlated portion of information’ (Benigni and Giuliani 1994). The application of multidimensional techniques like principal component analysis or clustering techniques with a closer look to the physical implications of the obtained result is, in our opinion, the main avenue to give an operational meaning to the holistic perspective (Giuliani et al. 2004). To acquire an emergence paradigm

F. Conti · M. C. Valerio
Chemistry Department, University of Rome, ‘La Sapienza’,
Rome, Italy

J. P. Zbilut
Department of Molecular Biophysics and Physiology,
Rush University, Chicago, USA

A. Giuliani (✉)
Environment and Health Department, Istituto Superiore di
Sanita’, Rome, Italy
e-mail: alessandro.giuliani@iss.it

implies the re-consideration of the respective role of cause and effect in the observed phenomena. In the reductionist approach, the ultimate causes of observed behaviour of systems must be examined at the most fundamental level so that the collective phenomena are thought of as consequences of the action of laws posited in the microscopic world (and thus more clearly understandable at that level). On the contrary, in the emergence approach causes have the form of order parameters arising from the correlation properties of the ensembles of elements. In this last approach, principles are thus nothing different from the statistical parameters arising from such an organization. A clear example of emergentist approach can be found in the work by Klaus Von Klitzing, for which he won the Nobel prize for physics in 1983, dealing with the so called Quantum Hall effect (Von Klitzing et al. 1980). Von Klitzing and colleagues discovered that some of the most honoured “fundamental constants of physics” (i.e., Planck’s constant (\hbar), electron charge (e) and the speed of light (c)) could be derived as consequences (not causes) of the collective behaviour of semiconductors when exposed to magnetic fields. That these “principles” were consequences and not causes was clearly demonstrated by the fact they could be observed only after a given minimal dimension of the system (sufficient for a relevant statistics) was reached, i.e., they were “emergent” properties of the system not already present at the microscopic level (Laughlin 2005). This result (as many others in condensed matter physics) implies that the optimal vantage point for discovering such principles, instead of being located at the microscopic scale, is posited at the scale of the system as a whole, i.e., acquiring a holistic perspective.

The importance of such a discovery can be hardly overrated. Here we do not see “different” principles explaining the system organization with respect to already known physical laws—here we see the same principles as “spontaneously arising” from system organization and strictly dependent on the scaling of the system in study.

In theoretical physics there is nowadays a strong battle between the emergentist (collective first) and reductionistic (microscopic first) approaches (Laughlin 2005). In biology things could be much clearer so to immediately shift in the direction of an emergence based approach if scientists were not hindered by an assumed ideological paradigm according to which explanations must be pursued at the microscopic level.

It is informative to view complex behaviour at different scales: the reliability of observations varies significantly. While we can be sure that if we shout at a rabbit (complex, macroscopic system), it will run away, thus giving us the impression of a fully ordered deterministic system. On the other hand, the results of a molecular genetic experiment

on the regulation of a specific gene of the same rabbit will be much more noisy and strictly dependent on myriads of boundary conditions and experimental recipes going from the rabbit strain, to the specific organ from which the cells are harvested, to the temperature and pH at which the cells are stored (Laughlin 2005).

The same kind of reasoning holds true even if we simply think of the fact that medical diagnosis (involving the analysis of the emergent properties of an incredibly complex system) is much more reliable than the results of the research involving the single enzymes, genes, receptors, or metabolites involved in the corresponding disease.

The presence of order parameters giving rise to strongly reproducible emergent behaviours at the whole system level can thus be accepted as obvious by anyone as well as the fact that the most fundamental level could not be the most promising level from which to look at biological systems.

Clearly the program to try to explain the properties of phase transitions solid–liquid–ice from atomic level properties of water is perfectly legitimate as well as to study protein folding starting from protein sequences, simply we must stress that not necessarily the most basic level is the place where all the definitive explanations live (and this is the reason why we do not look at particle physics when dealing with phase transitions).

The mythology of the “single gene level” as the privileged locus of the “ultimate and definitive” explanation of anything still persists. In our opinion this mythology has its roots in the heredity concept: the gene is what remains unchanged generation after generation so it must encompass all the relevant information for explaining and predicting whole system behaviour. This very naïve concept forgets (together with a myriad of other aspects that we do not mention here), the learning ability of single systems, the developmental processes, the continuous exchange with the environment, the functional inter-relations among different genes, the degeneracy of the genotype-phenotype mapping, and the presence of other forms of heredity not carried by genes. The single gene concept has demonstrated such an appeal that only recently have a number of scientists started expressing a need to change direction taking into consideration the collective behaviour of large ensembles of genes (Holter et al. 2000; Stern et al. 2007; Wilkins 2007; Tsuchiya et al. 2007; Ahn et al. 2006). Let’s then try and sketch how a completely different approach can be envisaged and what systems biology has to do with it.

Biological systems, by the exploitation of suitable energy sources, achieve spontaneous self-organization (order) allowing them to reach high levels of diversity and complexity by means of adaptive processes. From the thermodynamic point of view, the actual decrease of

entropy of the system, relative to its organization, is balanced by the entropy increase of the surrounding environment. The whole level emergent properties (the most basic of all being: the organism can perform a metabolism sufficient to sustain its life) impose the constraints to the molecular organization, but these constraints can be managed in a relatively flexible way by the microscopic level atomisms due to their extreme redundancy and richness of interaction patterns. This allows for the display of a huge repertoire of possible solutions that appear as equivalent in terms of the perceived result (the organisms can live in a myriad of different environments by the use of very different energy sources and passing through many diverse intermediate states).

The presence of multiple solutions to the same problem (and thus the basic degeneracy of the structure/function problem) arises very early in biological organization: a single protein (an object in the twilight zone between chemistry and biology) presents a multiplicity of almost equally energetically available configurations, and this multiplicity of possible states allows the protein to display a rich dynamics that is necessary for playing its physiological role (Finkelstein and Galzitskaya 2004), moreover the same basic ‘average structure’ can be obtained by completely different sequences (Branden and Tooze 1991) or different ‘structures’ generated on demand by the same sequence (Dunker et al. 2002). The same degeneracy holds at all the levels of biological organization from genetic regulation networks (Krishnan et al. 2007) up to ecological communities (Guill and Drossel 2008).

This implies that simple energetic considerations are not endowed with a sufficiently discriminant power to guide our research toward a unique and satisfying solution.

In order to understand the organization of a biological system we need both classical energy constraints and topological (energetically neutral?) invariants emerging as bottom up (not necessarily induced by superimposed energy minimization principles) organizational principles. Even without advocating a vitalistic principle that we consider outside the range of science, we must in any case think of still neglected dimensions of optimisation that could be “energetic” but outside the reach of what we nowadays call energy balances.

The discovery of such new ‘optimisation principles’ should be in our opinion the main topic of systems biology agenda instead of the generation of more or less sophisticated mathematical formalization of ‘already established’ biological pathways.

These organizational principles, in order to be discovered, ask for a synthesis of top-down and bottom-up approaches continuously exchanging the perspective from where to look at a system. In some sense it is a continuously changing parallax view. This need to go back and

forth between the two top-down and bottom-up views was very well described by Dhar (2007). The basic material for this enterprise (sadly enough because this was not the aim this kind of research was developed for and this provokes many problems) comes from the so called “-omics” sciences.

Genomics and Proteomics describe cellular behaviour in the space of genetic regulation and protein expression, respectively. Metabolomics instead locates the system in the space of the relative abundance of the small organic molecules constituting the metabolite pool. All these -omics, with the only partial exception of Metabolomics which, being born into (and still largely confined to) a chemically oriented world is largely devoid of the ideological idiosyncracies typical of biology, were developed in a strictly reductionistic, fundamental first, important laws living in the microscopic layer, scientific environment. The general idea common to the -omics approach was: the reductionistic approach fails not because is intrinsically flawed when dealing with systems in which the integration of many different elements is the most important aspect (this is where the term organism comes from) but simply because we still do not know all the actors of the play. When we will eventually know every tiny element concurring to the scene, the entire picture will (more or less automatically) become clear.

Pretty soon, the very first results of differential gene expression experiments, demonstrated the state of affairs was completely different from what the initial proponents of genomic science expected: the functionalities they expected to be in play in the various analysed situations either were not there or were present together with many hundreds that were completely unexpected (Stern et al. 2007). The reproducibility of the single gene level results, while very high in technical terms (PCR based single gene replicas of microarray results invariably confirm the microarray datum) was practically null at the biological level (e.g. in a patient/control discrimination for a specific disease—the most discriminating genes change abruptly from one study to another) so inhibiting the initial hopes for an efficient and ready to use diagnostic tool. It was in this crisis situation that biologists ask hard sciences specialists for some help.

This help was initially of a pure technical nature, just to evaluate the most macroscopic statistical paradox of microarray experiments. One important question was how to locate dozens of “significant” genes out of a collection of 20,000, which is an operation with a high risk of chance correlations. With the passage of time the questions were more refined and involved a certain appreciation of the “actual content” of the study, such as the development of gene interaction networks consistent with the microarray experimental results. These more refined questions gave

rise to the actual emphasis on “Systems Biology” in which styles of reasoning borrowed from hard sciences like physics and engineering officially entered biology (Ahn et al. 2006; Kitano 2004).

The point is that, in the great majority of cases, these styles of reasoning are borrowed in a “defensive” fashion. To make a classical, old-fashioned, actors = single genes picture, mathematicians (in a broad sense, the actual persons can be physicists, statisticians, chemists, quantitatively sophisticated biologists—here we simply indicate with this term someone who is not scared by numbers) have the task to cleanse the messy material coming from -omics experiments and translate it into pretty graph-like structures with genes (protein, metabolites...) as nodes and arrows as edges connecting them. The majority of scientists are in general very fond of these Mandala-like pictures they use as a base for meditation and for explaining (in general in a post-hoc way) many phenomena. Again we are in a “Particularly and separability as infirmities...” Cartesian paradigm.

We make a different proposal that could give an holistic flavour to systems biology (and altogether give a non-zero contribution to the advancement of science) suggesting computational (i.e., not-scared-by-numbers) scientists, instead of simply giving mainstream biologists material to confirm on ‘solid mathematical bases’ what they already know and peacefully contemplate their arrows and nodes Mandala, of looking at what those -omics data propose per se careless of old fashioned gene-centric explanations. We propose to scientists to concentrate on the robust portion of high throughput technologies like the strong correlations observed in gene expressions or in metabolomic data, where they will find astonishing “collective organizations” urgently asking for a new thinking more than intermingled networks usually confined in very minor components of the data.

The most evident and macroscopic collective phenomena asking for consideration is with no doubt the existence of an extremely reproducible characteristic level of expression for all the many thousands of genes of a cell line. Figure 1 reports the correlations of two different strains of the same cell line (mouse macrophages wild type and Myd88 knock out, respectively) with the vector points representing the expression level of approximately 23,000 gene products (Hirota et al. 2005). The correlation between the two cell samples, spanning the whole genome expression is remarkable, and this kind of behaviour is encountered every time, in any microarray experiment, whenever two different populations of the same cell line (notwithstanding which stressor, drug or mutation is inserted) are plotted. This invariance is what constitutes the individuality of a given cell line and we are far from the understanding of the bases of such an ordered and

repeatable behaviour that is practically unique in biology. What is for sure is that this is a “scalable” behaviour, reproducible with random extractions of genes up to a certain minimum number and with no relation to the specific functions of the involved gene products. This extremely ordered behaviour (that has its counterpart in time constituted by the presence of whole genome rhythms spanning billions of cells in a colony, thus falsifying the ergodic hypothesis (each cell in a plate makes its own game) at the basis of “molecular first” hypotheses) (Tsuchiya et al. 2007; Klevecz et al. 2004) is evident when reaching a minimum number of considered genes. When we look inside the single gene behaviours we observe erratic variability not consistent with the large scale ordering. An examination of Fig. 1 immediately explains this conundrum: looking at exceptional behaviour of single genes corresponds to picking up (and considering it as the relevant information) the points escaping the linear relation at a larger extent (the genes significantly affected by treatment), but these points are very few and, still more important, these erratic points are where the influence of noise is maximal so that it is perfectly sound that we cannot derive from them a reliable information.

Nevertheless the legacy of ‘single gene reductionist paradigm’ forces the analysis in this highly non-rational direction, while the basic point to explain is the robust and repeatable ordering of thousands of different gene expression that instead does not receive any particular attention and is given ‘for granted’ by biologists in the total absence of any rational explanation for it.

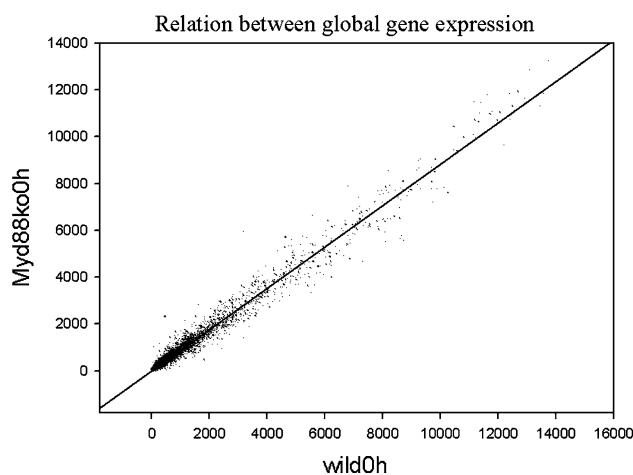


Fig. 1 The Figure reports the correlation between the values of expression of around 23,000 genes (the vector points of the figure) relative to two different populations of blood cells (macrophages) bearing a mutation as for a very important gene involved in innate immunity (Myd88ko) and wild type respectively. Pearson r (product moment correlation coefficient) between the gene expression vectors of the two populations is near the maximum attainable ($r = 0.998$)

The above sketched case story highlights the classical signature of the order/stochastic blend we started with in the beginning of this paper: a fully ordered pattern on a large, population (23,000 genes) based level, supported by a disordered behaviour of its constituent elements (single genes variations). Only if Systems Biology will dare to tackle the analysis of these still unknown large scale order process will it become a powerful tool in the opening of new scientific horizons.

This change of perspective asks for a sudden leap in the relevance of creativity and invention of scientists with respect to the adhesion of already established knowledge: large scale collective phenomena ask for the development of totally new constructs (such as the eigengene (Holter et al. 2000), a mode of expression involving simultaneously the expression variability of thousands of genes) for which a biologically accepted counterpart still does not exist. In our opinion going along the still non-explored avenues of the arise of collective organization from intrinsically stochastic elements could be an extremely fascinating and fruitful agenda for systems biology scientists.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Ahn AC, Tewari M, Poon CS, Phillips RS (2006) The limits of reductionism in medicine: could systems biology offer an alternative? *PloS Med* 3(6):e208
- Benigni R, Giuliani A (1994) Quantitative modeling and biology: the multivariate approach. *Am J Physiol* 266(35):R1697–R1704
- Branden CL, Tooze J (1991) *Introduction in protein structure*. Garland, New York
- Dhar PK (2007) The next Step in biology: a periodic table. *J Biosci* 32:1005–1008
- Dunker KA, Brown CJ, Lawason DJ, Iakoucheva LM, Obradovic Z (2002) Intrinsic disorder and protein function. *Biochemistry* 41:6573–6582
- Finkelstein AV, Galzitskaya OV (2004) Physics of protein folding. *Phys Life Rev* 1:23–56
- Giuliani A, Zbilut JP, Conti F, Manetti C, Miccheli A (2004) Invariant features of metabolic networks: a data analysis application on scaling properties of biochemical pathways. *Physica A* 337(1–2):157–170
- Guill C, Drossel B (2008) Emergence of complexity in evolving niche-model food webs. *J Theor Biol* 251:108–120
- Hirotsani T, Yamamoto M, Kumagai Y, Uematsu S, Kawase I, Takeuchi O, Akira S (2005) Regulation of lipopolysaccharide-inducible genes by MyD88 and Toll/IL-1 domain containing adaptor inducing IFN-beta. *Biochem Biophys Res Commun* 328:383–392
- Hock D (1999) *Birth of the chaordic age*, ISBN 1-57675-074-4, Berret-Kochler Inc
- Holter NS, Mitra M, Maritan A, Cieplak M, Banavar JR, Fedoroff N (2000) Fundamental patterns underlying gene expression profiles: simplicity from complexity. *Proc Natl Acad Sci USA* 97(15):8409–8414
- Kitano H (2004) Biological robustness. *Nat Rev Genet* 5:826–837
- Klevecz RR, Bolen J, Forrest G, Murray DB (2004) A genome wide oscillation in transcription gates DNA replication and cell cycle. *Proc Natl Acad Sci USA* 101:1200–1206
- Krishnan A, Giuliani A, Tomita M (2007) Indeterminacy of reverse engineering of gene regulatory networks: the curse of gene elasticity. *Plos ONE* 6:e562
- Laughlin RB (2005) *A different universe. Reinventing physics from the bottom down*. Basic Books
- Stern S, Dror T, Stolovici E, Brenner N, Braun E (2007) Genome-wide transcriptional plasticity underlies cellular adaptation to novel challenge. *Mol Syst Biol* 3:106–112
- Tsuehchia M, Wong ST, Yeo ZX, Colosimo A, Palumbo MC, Farina L, Crescenzi M, Mazzola A, Negri R, Bianchi MM, Selvarajoo K, Tomita M, Giuliani A (2007) Gene expression waves: cell cycle independent collective dynamics in cultured cells. *FEBS J* 274:2874–2886
- Von Klitzing K, Dorda G, Pepper H (1980) New method for high accuracy determination of fine structure constant based on quantized Hall resistance. *Phys Rev Lett* 45(6):494–497
- Wilkins AS (2007) For the biotechnology industry, the penny drops (at last): genes are not autonomous agents but function within networks. *BioEssays* 29:1179–1181