

Self-organized living systems: conjunction of a stable organization with chaotic fluctuations in biological space-time

BY CHARLES AUFFRAY¹, SANDRINE IMBEAUD¹,
MAGALI ROUX-ROUQUIÉ² AND LEROY HOOD³

¹*Genexpress, Functional Genomics and Systemic Biology for Health,
CNRS FRE 2571, 7 rue Guy Moquet, BP8,
94801 Villejuif Cedex, France (auffray@vjf.cnrs.fr)*

²*Biosystémique Modélisation Ingénierie, Institut Pasteur,
25–28 rue du Dr Roux, 75724 Paris Cedex 15, France*

³*Institute for Systems Biology, 1441 North 34th Street,
Seattle, WA 98103-8904, USA*

Published online 2 May 2003

Living systems have paradoxical thermodynamic stability, the intrinsic property of self-organization, fluctuation and adaptation to their changing environment. Knowledge accumulated in the analytical reductionist framework has provided useful systematic descriptions of biological systems which appear to be insufficient to gain deep understanding of their behaviour in physiologic conditions and diseases. A state-of-the-art functional genomics study in yeast points to the current inability to appraise ‘biological noise’, leading to focus on few genes, transcripts and proteins subject to major detectable changes, while currently inaccessible small fluctuations may be major determinants of the behaviour of biological systems. We conjecture that biological systems self-organize because they operate as a conjunction between the relatively variable part of a stable organization and the relatively stable part of a chaotic network of fluctuations, and in a space with a changing number of dimensions: biological space-time. We propose to complement the precepts of the analytical reductionist framework with those of the biosystemic paradigm, in order to explore these conjectures for systems biology, combining in an iterative mode systemic modelling of biological systems, to generate hypotheses, with a high level of standardization of high-throughput experimental platforms, enabling detection of small changes of low-intensity signals, to test them.

Keywords: biosystemic paradigm; functional genomics; reductionism;
self-organization; standardization; systems biology

1. Introduction: a self-organizing definition of self-organization

Once upon a time
An assembly of noble gentlemen

One contribution of 18 to a Theme ‘Self-organization: the quest for the origin and evolution of structure’.

Ventured to self-organize
 In the quest of a definition
 Of self-organization
 What is self-organization?
 That is the question ...

The German physicist (Konrad Kaufmann)

A tautology?

Well, the theory of everything says that everything is self-organization. Whether that will stand with time is another matter, but it makes it comfortable for us, because everything we are going to talk about within the frame of that theory is self-organization, even if we do not use the word.

The Canadian cosmologist (Lee Smolin)

If Lee Smolin is right, then, like Monsieur Jourdain (see Poquelin 1671), who is 'using prose without knowing it', we are in fact all working on self-organization, even if we do not use the words. Or, in other words, as Monsieur de La Palice would say: 'if things self-organize, it is because they organize themselves'.

A scientific modus operandi?

The way I understand this is as follows: self-assembly or organization is the metamorphosis of a physical or chemical state to a different form. The alterations require a catalyst with the appropriate qualities and sufficient energy that can trigger a chain of reactions together with a feedback mechanism to capture and stabilize the new state.

The Canadian immunologist (Tak Mak)

I would like to limit my comment to the world of viruses where the viral components self-organize to functional infective complex viral particles. This is achieved using purified protein and nucleic acid constituents. The only system where the entire pathway for a complex virus is described is that of the assembly of bacteriophage phi6 nucleocapsids.

The Finnish virologist (Dennis Bamford)

A generic property of inert and living systems?

There have not existed good definitions for self-organization. The very concept emerged in the 1950s and 1960s in the automata theory, when mainly discrete systems were considered. For instance, Mesarovic (1962) defined a self-organizing system as a system which starts from an initial state as defined by certain relations, and then changes its structure by using other relations: 'self-organization means emergence of new functions and systems of functions, whereby the set of these functions is often

optimized and eventually ordered spatially and hierarchically'. Klir & Valach (1967) stated that a self-organizing system can improve itself, and acquire properties which were not even thought of when it was created. I very firmly think that a mere structure is not what results in self-organization, at least in the sense that we use this concept in biology.

The Finnish technologist (Teuvo Kohonen)

Self-organization is a process in which pattern at the global level of a system emerges solely from numerous interactions among the lower-level components of the system. Moreover, the rules specifying the interactions among the system's components are executed using only local information, that is, without reference to the global pattern. In short, the pattern is an emergent property of the system, rather than a property imposed upon the system by an external ordering influence. The meaning of pattern in this context is simply a definite arrangement of objects in space, or in time, or both. Biological examples of pattern include such diverse entities as a school of fish, a raiding column of army ants, the synchronous flashing of fireflies and the complex architecture of a termite mound.

The French and Belgian ethologists
(Guy Theraulaz and Jean-Louis Deneubourg)

Wandering through the Web in search of definitions

Self-organization: A spontaneously formed higher-level pattern of structure or function that is emergent through the interactions of lower-level objects.

(Flake 2000)

Self-organization is a process where the organization (constraint, redundancy) of a system spontaneously increases, i.e. without this increase being controlled by the environment or an encompassing or otherwise external system.

(Heylighen 1997)

Self-organization may be defined as a spontaneous (i.e. not steered or directed by an external system) process of organization, i.e. of the development of an organized structure. The spontaneous creation of an 'organized whole' out of a 'disordered' collection of interacting parts, as witnessed in self-organizing systems in physics, chemistry, biology, sociology, ... is a basic part of dynamical emergence.

(Heylighen 1997)

To be self-organized or not to be self-organized
That is the question ...

The German physicist again (Konrad Kaufmann)

Concerned grumbling voices in the background (the Swedish self-organizers John Skår and Ingemar Ernberg): 'Well, Mr Chairman, it is time to go to the next speaker'.

2. The analytical reductionist framework has enabled identification of the relatively stable organization of living systems

In the autumn of 1649, René Descartes travelled to Stockholm in response to the invitation of Queen Kristina, who wanted him to provide insights into his ‘method’ to rightly conduct reasoning and search for the truth in the sciences, as published in his already famous book *The discourse of the method* (Descartes 1637). The scientists who gathered at the Nobel Forum of the Karolinska Institute during the exceptionally beautiful summer of 2002 have all in some way or another inherited and worked in the framework of the four precepts of the method outlined by Descartes 365 years ago.

This conceptual framework of thinking has enabled great advances in understanding nature, particularly in physics and the sciences developed from it, such as chemistry. In biology, however, while it has allowed substantial progress in our understanding of what living systems are made of, there is a growing recognition that it has failed to provide the basis for equivalent deep understanding of biological functioning. As a result, biologists and physicians have not gained a similar understanding of their subject of interest as physicists and chemists have. Consequently, the promised applications of biological knowledge to improve human health and control of environmental matters seem to be further out of reach.

In some ways, we are facing a situation similar to the children chasing the rainbow to dig at its foot in search of hidden gold, encouraged by a ray of Sun in the midst of heavy rain. Before our journey in the quest for knowledge, triggered by curiosity as well as our sense of duty to bring useful applications for the benefit of mankind, is stalled, many of us feel that we need to redefine our strategy. In order to do so, it is time to revisit the four precepts of the method, identify possible limitations embedded in them, and lay the foundations for a renovated conceptual framework for biology. In doing so, we need to take into account the various attempts developed over the past half century to develop other conceptual frameworks, including cybernetics and systemic paradigms.

We will first revisit Descartes’s analytical framework, and indicate in broad terms how the community of biologists has used it to identify what can be viewed as the relatively stable organization of living systems. Moving towards a very practical implementation of systems biology, we will then discuss one particular example of functional-genomics study, arguing that, contrary to what people have said so far, they reveal not only the relatively stable part of the system, but also the chaotic fluctuations underlying biological function. We will then propose, based on the work of a number of people, that we should use conjunctive reasoning in order to help the development and implementation of system-level understanding in biology and its use for practical biomedical applications.

(a) *The four Cartesian precepts of the method and biology*

René Descartes, 365 years ago, in his *Discours de la méthode pour bien conduire sa raison et rechercher la vérité dans les sciences*, or ‘Discourse on the method to conduct reasoning and search for truth in the sciences’, contributed to the foundation of modern science by proposing a method based on four precepts.

Le premier était de ne recevoir jamais aucune chose pour vraie, que je ne la connusse évidemment être telle: c’est à dire d’éviter soigneusement

la précipitation et la prévention et de ne comprendre rien de plus en mes jugements, que ce qui se présenterait si clairement et si distinctement à mon esprit, que je n'eusse aucune occasion de le mettre en doute.

The first was never to accept anything for true which I did not obviously know to be such; that is to say, carefully to avoid precipitancy and prejudice, and to include nothing more in my judgements than what was presented to my mind so clearly and distinctly that I could have no occasion to doubt it.

The underlying principles are objectivity and determinism, which we use every day in biology without questioning them. In order to develop a deeper understanding of living systems, we need to be very careful that we keep an open mind and to question our practices in the conduct of science, not rushing into dead ends, and refraining from making claims of very significant advances that appear in the media, when we have in fact made only small steps forward.

Le second, de diviser chacune des difficultés que j'examinerais, en autant de parcelles qu'il se pourrait, et qu'il serait requis pour les mieux résoudre.

The second, to divide each of the difficulties which I would examine into as many parts as it would be possible, and as might be required to resolve them best.

This second Cartesian precept is related to the underlying principle of reductionism, and contains something which is often overlooked in some of the discussions about its limits: 'and as might be required to solve them', which relates to the precept of pertinence that we will discuss later. Currently, biology appears mostly as a descriptive science that accumulates data, with the idea in the background that if we divide living systems into very small parts, and if we stumble upon some difficulty in understanding, we should divide it more, and maybe at some point we will have enough resolution to address our questions. As we continue to work in this way in biology, we face the same hurdles in system reconstruction and extraction of basic principles previously encountered in physics, for example.

Le troisième, de conduire par ordre mes pensées, en commençant par les objets les plus simples et les plus aisés à connaître, pour monter peu à peu, comme par degrés, jusques à la connaissance des plus composés; et supposant même de l'ordre entre ceux qui ne se précèdent point naturellement les uns les autres.

The third, to conduct my thoughts in order, beginning with the simplest and easiest objects to know, to rise little by little, as it were by steps, up to the knowledge of the most complex; and assuming even order between those which do not precede each other naturally.

This precept deals with complexity and combinations of elements in living systems. If we take the human genome, for example, which is made of linear arrangements of a total of three billion nucleotides, each one of four possible types, the number of possible human genomes is four to the power of three billion, or ten to the power of

about one billion. This number is incommensurate with the number of elementary particles in the Universe, estimated by the physicists at about 10^{70} . So there is no way nature or we, even by using the Universe as a computer, could explore completely such a huge combinatorial, pointing to an intrinsic limitation of the power of analytical reductionism in the life sciences.

Et le dernier, de faire partout des dénombrements si entiers, et des revues si générales, que je fusse assuré de ne rien omettre.

And the last, in every case to make enumerations so complete, and reviews so general, that I should be assured not to omit anything.

In this final precept, Descartes supported the idea that exhaustiveness, comprehensiveness, coverage, is essential. This translates today into the idea that knowing all the nucleotides of the human genome, or all species living on Earth, for example, should enable us, at last, to do something that we would not be able to do if our knowledge was incomplete: uncover hidden underlying principles or develop new medical treatments.

(b) Biology: a science of information in hierarchical flux

The way we have organized our thinking and practice in biology is that of a science of information organized and flowing in hierarchical and directional flux from genes, to transcripts, proteins, interactions, metabolism, cells, tissues, organisms, populations and ecosystems. This is following Descartes's third and fourth precepts of organizing things in chains of reasoning and systematically collecting comprehensive information at each level. Proceeding this way is expected to enable identification of key biological mechanisms and objects, allow engineering and manipulation of the biological systems that we study, with the goal for example of understanding the origin of life and evolution, and to prevent or cure diseases.

We have built the ability to collect diverse types of biological data for the identification and characterization of genes, transcripts and proteins of an organism, including coding regions, functional domains, regulatory sequences, polymorphisms and structure-function relationships. These data are then used to unravel the organization and regulation of metabolic pathways and networks, taking into account the role of small molecules and of the environment to describe normal physiological and perturbed pathological conditions. The comparison between organisms, highlighting similarities and differences, makes it possible to document evolution and development. Tools have been developed for knocking out specific genes and replace them, modify them and shuttle them between organisms.

In the past decade, stimulated by (and enabling) the Human Genome Program, experimental methods in biology have evolved towards high-capacity quantitative data collection, combining advances in instrumentation, automation and computation. Sequencing and mapping of a variety of animal, plant and microbial genomes was made possible by the automated DNA sequencer and synthesizer; macro- and microarrays of cDNAs and oligonucleotides are used to assess changes in gene-expression profiles by monitoring global transcriptomes; mass spectrometry, liquid chromatography and electrophoresis are combined and integrated to monitor proteomes; and advances in imaging and high-speed multiparameter cell sorting are transforming cellular biology.

Table 1. *Survey of the biomedical literature*

(PubMed (12 million references, December 2002, <http://www.ncbi.nlm.nih.gov/entrez>) was used to assess how many publications contain specific key words (given below) of the biological information flux in their title or abstract, with queries such as DNA, RNA, protein, cell and function, and of the ‘omics’ sciences, with queries such as genome, transcriptome, proteome, metabolome and physiome. Informatics and the usual term bioinformatics (although computational biology would be more appropriate, see text) were used to assess biology as an information science in the biomedical literature. The percentages on the left are the number of papers for each word given as a fraction over all PubMed, and the percentages on the right are the fractions of the papers for each word, as compared with the number for the word on the same line in the left column (genome versus DNA, transcriptome versus RNA, etc.)

| term | number | percentage | term | number | percentage |
|-------------|-----------|------------|----------------|--------|------------|
| informatics | 8 276 | 0.07 | bioinformatics | 3 438 | 41 |
| DNA | 703 031 | 6 | genome | 78 081 | 11 |
| RNA | 359 501 | 3 | transcriptome | 264 | 0.07 |
| protein | 2 606 196 | 22 | proteome | 1 836 | 0.07 |
| cell | 2 538 254 | 21 | metabolome | 33 | 0.001 |
| function | 5 643 642 | 47 | physiome | 19 | 0.000 3 |

Powerful methods have been developed for the systematic perturbation of biological systems, including targeted inactivation, replacement and modification of genes in cells and organisms using polymerase chain reaction, RNA interference, anti-sense oligonucleotides and shuttle expression vectors for integration/excision based on homologous recombination.

This has resulted in a growing need for analysis, annotation and integration of diverse datasets, with the goal of not only dealing with information, but extracting some knowledge from it. With the emergence of a systems-biology paradigm, more emphasis is now placed on data and knowledge modelling.

Computational biology, or bioinformatics, methods have changed the landscape of biology, dealing with storage and distribution of large amounts of the diverse types of data registered in a variety of electronic databases about DNA, RNA, protein sequences and, more recently, protein interactions, transcription factors, expression profiles and metabolic networks. The tools of computer science are now widely used not only for data analysis, annotation and integration using simple, complex and global queries, but also for modelling, prediction and engineering of deterministic and stochastic models of biological systems, with the goal of developing the ability to identify emerging properties, formulate new hypotheses and test them. All these efforts are facing the problems of information completeness, exactness and updating. Community-wide efforts are ongoing for distributed annotation and text-based analysis of the literature, as represented by PubMed, a repository of over 12 million publications in the biological sciences of the last 35 years.

Fifty years ago, when Jim Watson started his work that led to the discovery of the structure of DNA, he reported in his famous book *The double helix* (Watson 1968), that he went to the library and he had to read less than 10 papers on the important question of that time, whether the genetic material was DNA or protein. Reading those few papers, he could make up his mind that the hypothesis that was the most likely to be true in his opinion was that DNA was the substance of heredity, and

therefore he should work on the structure of DNA. Today, a beginner who wants to work on DNA will have to choose among over 700 000 papers registered in PubMed (table 1), on RNA a little less, both representing 3–6% of the total literature. The situation is even worse for proteins and cells, with some 2.5 million papers (there are 3.4 million papers containing the word ‘metabolism’). And the worst situation is for ‘function’, since about half of all the biomedical literature contains this word.

We suggest that this is a strong indication that, whereas we now know a great deal about the elements of living systems and the basic principles of their organization, and have a wealth of information on biological function, the fact that there are thousands of definitions of ‘function’ points to our limited understanding of how living systems really function.

During the last decade, in conjunction with the development of large-scale and high-throughput biology, new words ending with the suffix ‘-ome’ have emerged in the literature. There are already over 78 000 papers with the word ‘genome’, representing 11% of the number of those papers that contain the word ‘DNA’. In contrast, there are only 264 with the word ‘transcriptome’ (a global set of RNA transcripts), and 1836 with the word ‘proteome’ (global set of proteins), which are 0.07% of those that contain ‘RNA’ and ‘protein’, respectively, and these numbers are growing very rapidly, much faster than PubMed as a whole. The words ‘metabolome’, with 33 papers, and ‘physiome’ with 19, represent less than one out of 100 000 of the papers related to ‘cell’ and ‘function’. We suggest that those papers may represent the emerging new trends in the biological sciences and it could be interesting to look at them to start with.

Although biology is appearing more and more as an information science (303 306 papers contain the word ‘information’), the number of papers with the word ‘informatics’, at 8276, is surprisingly small and represents less than 0.07% of the biomedical literature. Yet 41% of PubMed’s papers contain the word ‘bioinformatics’, reflecting the great deal of attention paid to the issues of data collection, storage, analysis, data modelling and knowledge extraction, which is encouraging. There are 12 105 articles with the adjective ‘computational’, of which 2863 also contain the term ‘computational biology’ and 508 contain the term ‘computational genomics’, reflecting the same trends.

3. Functional-genomics studies reveal chaotic fluctuations of gene-expression levels

(a) *Expression profiling in yeast: the state of the art*

In order to illustrate the power and limits of functional-genomics studies, let us take an example of a functional-genomics study in yeast, considered as the state of the art (Hughes *et al.* 2000).

The authors studied 300 yeast mutants and chemical treatments under a single growth condition, monitoring the expression profiles of 5835 genes by comparing the hybridization of microarrays or DNA chips with fluorescently labelled cDNAs transcribed from RNAs of the mutants versus a common control. A total of 1 650 500 hybridization data points were generated, including 63 negative control experiments; inverse fluorophore labelling was used to compensate for possible differences in incorporation during labelling, and half of the experiments were performed in duplicate.

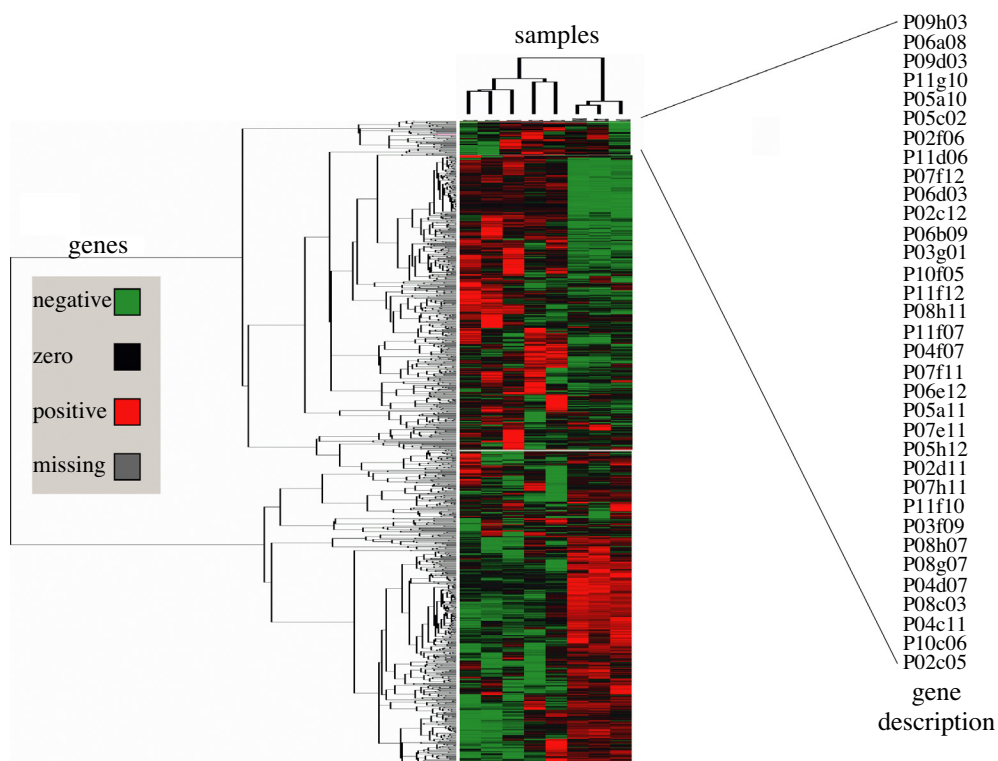


Figure 1. Cluster analysis: a typical dendrogram.

The results were then subjected to two-dimensional hierarchical clustering, a method used to group together genes or samples associated with similar hybridization patterns.

Such results are usually displayed in an image map similar to the one represented in figure 1, in which all the genes are associated with a colour code, indicating those which are strongly activated in green, and those which are strongly repressed in red, when they are compared in different conditions. They are then typically grouped into clusters that reflect changes in their activity or modules that have some biological meaning. In the end, a list of genes selected for further functional studies is established.

The discovery of eight gene functions and a previously unknown drug target, reported by Hughes *et al.*, as significant as it might be, does not seem to represent a very large output compared with the massive accumulation of data. In fact, most of the data produced are ignored because of a lack of conceptual and practical tools to deal with them. The phrasing by the authors of their own analysis of the results points to the current limitations in such studies.

They found that 'nearly all the experiments resulted in a two-fold or greater alteration in the abundance of at least one transcript' and that in the 'set of 63 controls: simultaneous growth of isogenic wild-type cultures, the vast majority also include

at least one gene with greater than two-fold induction or repression ... with up to 20-fold changes for some genes'. They 'reasoned that these fluctuations represent a form of *biological noise*' (our italics) and concluded that 'a single growth protocol was sufficient to generate functional data for roughly half of the mutants, and to evoke responses from a large majority of genes.'

This already tells us that the background on which they analyse the data is subject to important fluctuations, which they have to take into account. The 'biological noise' they refer to might in fact be normal biological fluctuation that should be assessed by comparing different growth conditions, and which we should be able to analyse with the mathematics of chaos.

They analysed the data by 'two-dimensional hierarchical clustering of the most prominent gene behaviours ... followed by statistical significance measurement of clusters' and identified '15 highly significant clusters', four of which are 'potentially *misleading* experiment clusters ... identified by the groups of genes induced or repressed, or by the fact that their composition makes *little biological sense*'. Whereas this might be true in the framework of past knowledge, we have to make sense of such clusters, not to dismiss them because their contribution contradicts some established relationships or it is surprising to us. To put it in another way, we have to get back to the first precept of Descartes, to be open-minded, and not to make *a priori* judgements on those datasets. In fact, what they have done is to take only hints on essentially previous knowledge that points to that part of the living systems that is mostly stable, highlighting rare large variations, thus defining major genes, transcripts and proteins that behave as stable attractors.

They continue by pointing out that 'the unexpected identification of very low-magnitude transcriptional regulation of the mitochondrial ribosome suggests that such low-amplitude but meaningful regulatory patterns might be common ... and can be shown to be biologically significant'. Our assumption is that most of the relevant biological information is embedded in this part of the data: the small variations of small signals, which collectively could contain more information than the large variations of a handful of genes that are listed as the result of hierarchical clustering. This, as they point out, 'underscores the importance of generally high-quality internally consistent data, and the necessity of developing means to compensate for noise or other biases.' In order to do so, we have to be extremely rigorous and implement a mode of data collection that is compatible with an ability to dig into the data pool in a meaningful manner.

(b) *Quality assurance in functional genomics*

For that purpose, we have developed a quality-assurance process for micro-array analysis including 39 steps, including experimental design, gene collection, sample collection, array preparation, probe synthesis, hybridization, data transformation and knowledge extraction (Imbeaud *et al.* 2003). Emphasis is put initially on experimental design, an issue which has been poorly addressed, if not entirely ignored in many expression profiling studies (Yang & Speed 2002). It is very important to assess the power of the data-capture scheme, and to establish beforehand how many samples are needed to address specific biological questions, what is the biological variation expected, compared with that coming from the experiment, and how many replicas are needed to ensure statistical significance of the results (limiting false negatives and false positives).

The process then has to be performed step by step very carefully, with sequential quality controls, otherwise the data produced in vast quantity may be flawed because the experiment design is not appropriate to the biological question addressed, the experimental noise embedded into it is not assessed, and therefore we cannot take full advantage of it. We are engaged in partnerships with instrument companies to develop platforms that are capable of generating very reliable data. This will empower us to go deeper and deeper in the comparison of many samples, which is really required for this type of study. We have been able so far to look only at the top of the iceberg of the data pool, whereas in fact most biology related to complex genetic regulatory networks is likely to reside in the bottom part of that iceberg.

New principles of conducting science are needed, not only for reasoning, or extracting basic principles, but also for design and reduction to practice, based on real data. Therefore, quality assurance of the experimental platforms is essential. It is not yet in place in genomics, except for DNA sequencing. It took 21 years from the discovery of the sequencing chemistry in 1977 to the implementation of a high-throughput quality control pipeline for DNA sequencing in 1998, which made it possible to generate the vast amount of data now assembled in almost complete descriptions of the human genome sequence. For the transcriptome and proteome, the relevant technologies are only in their infancy, and we are far from controlling them with the same degree of accuracy as DNA sequencing. As a result, the power of those technologies is to generate a vast amount of data, but our ability to take advantage of it is extremely limited. It is going to take another few years for transcriptome studies to reach maturity, and this is going to be even harder for proteome technologies, which deal with other levels of complexity. So we have to be very careful in both combining very fundamental principles and revising our experimental practice.

4. Systemic modelling using conjunctive reasoning will help the development of system-level understanding in biology

An exhaustive knowledge of the structure, function and relation of the components of biological systems is necessary but insufficient to understand phenotypes. We need to approach biological systems as systems in which feedback, redundancy and modularity ensure system optimization, stability and robustness. Functional genomics and systems biology require an integrated software platform for data collection, simulation, visualization, analysis and hypothesis generation. In our community we are actively promoting the implementation of a framework for functional genomics and systems biology which starts by defining the components of a biological system and collecting the relevant biochemical and genetic data on a global scale with high-throughput platforms, using them to formulate an initial model of the system, then systematically perturb the components of the system with the methods alluded to above, and study the results. By comparing the observed responses to those predicted by the model, it is then possible to refine the model so that its prediction fit best to the experimental observations. Finally, new experimental perturbations are conceived and tested in order to distinguish between the multiple competing hypotheses (Ideker *et al.* 2001a).

However, as pointed out by Morin & Le Moigne (1999) and Simon (1969, 1996), complex systems cannot be apprehended globally, only through partial but intelligible models. ‘When one decomposes a system, it decomposes itself’, it loses its properties,

and connections to some of the fundamental properties that emerge only when the parts get together are lost.

Following the avenue opened by these authors, we endeavoured to revisit the systemic theoretical framework developed over the past decades, which intends to describe an active object in its context, dealing with what it does, not only what it is, as has been the case for the most part during the flourishing era of cellular and molecular biology. Systemic modelling considers the active object as a processor which modifies itself and its environment. Since we always reason on models, representations, retaining only what is important, it is not essential to know all the elements to know or understand biological function (Roux-Rouquié & Le Moigne 2002).

New precepts are needed in order to deal with the challenges of systems biology, and they require that we go beyond classical analysis and reductionism. This is not to say that we have to get rid of Descartes's precepts, because their use has proved very powerful and will remain so as far as we are concerned with the question of knowing what biological systems are made. But when we try to deal with the question of function and phenotype, we recognize that these precepts, based on disjunctive reasoning, have proved insufficient. Therefore, we have started to explore system modelling based on conjunctive reasoning, with the goal of helping the development of system-level understanding of biological systems. This will require the integration of expertise from physics, mathematics and informatics for data collection, simulation, and visualization.

(a) *Four precepts for systems biology beyond analytical reductionism*

We propose adding to the four Cartesian precepts new precepts to operate in systems biology beyond reductionism (Auffray *et al.* 2003).

The first is *contextualization*, complementing the basic principle of objectivity, which identifies objects within the environment in which they function and transform themselves, and associates them with a function, rather than just identifying them in isolation with clearly distinct characters and properties, independently of their environment. In doing so, we avoid losing all the connections and the interactions which are essential for biological function and self-organization, as already pointed out by von Foerster (1960), who proposed 'to use the term self-organizing system, while being aware of the fact that this term becomes meaningless, unless the system is in close contact with an environment, which possesses available energy and order, and with which our system is in a state of perpetual interaction ...'.

The second is *relatedness*, which complements (or perhaps contradicts) the precept of decomposition or division 'in as many parts as it may be possible'. It consists of identifying interactions that modify the nature or behaviour of interacting objects, pointing to the fact that what is important are the interactions and not the objects themselves. An extreme view is that, in fact, there is no such thing as an object, because it is changing all the time, and that therefore everything is interaction. Pointing to the limitations of the reductionist approach, the use of the precept of relatedness should allow much more dynamic interplay between objects and experiments.

The third is the precept of *conditionality* complementing the precept of causality. It is intended to identify the rules that determine the behaviour of interacting objects

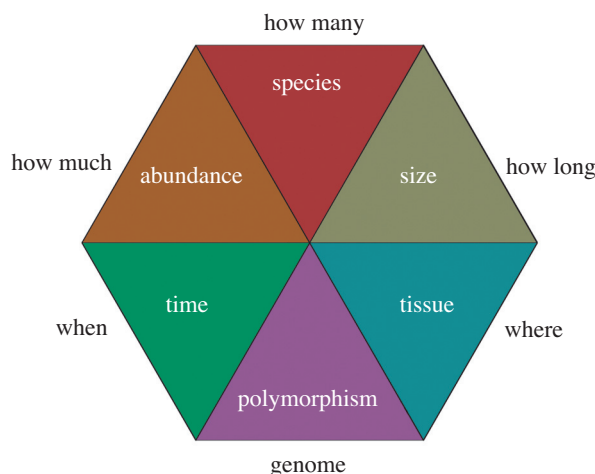


Figure 2. The multiple dimensions of the transcriptome (proteome). Each dimension of the transcriptome (proteome) is represented by a coloured sector.

leading to their organization, as pointed out by Edgar Morin around a quarter of a century ago (Morin 1977, 1980). That has to be compared with the idea of just reordering the parts by linking them in long chains of simple reasoning, because nature is not organized only through those types of chains, as discussed earlier. Living systems appear to be made of an organization which is mostly stable, with limited variation, coupled with a universe of fluctuations which we assume to be chaotic in nature. If that view is correct, it will be a central theme of biology for the coming century.

Finally, more emphasis should be placed on the precept of *pertinence* than on exhaustivity. What we have been doing so far is trying to be comprehensive, exhaustive, ‘to make enumerations so complete to be assured not to omit anything’, and now what we have to do is to identify modules that are considered relevant to model the proper dimensionality of relations within the environment.

(b) *Two conjectures for systems biology*

A first conjecture is that living systems have the ability to organize themselves as the result of a conjunction occurring through an interface between the variable part of a mostly stable organization, and the stable part of a chaotic network of small fluctuations.

These small fluctuations, which are inaccessible to the tools currently available, may be the major determinants of the behaviour of biological systems, because they convey collectively the most information. Detection of small changes in low-intensity signals will require the development of a new conceptual and practical framework combining in an iterative mode systemic modelling of biological systems, to generate hypotheses, together with a high level of standardization of high-throughput experimental platforms enabling reliable cross comparisons, to test them.

Six dimensions of the transcriptome (or proteome) are represented in figure 2: how many species of them there are, how long they are, how many of them there are, where they are located in the organism, what is the influence of genome variation,

and when they are expressed during development, physiology, disease and evolution. Many have argued that we need only two dimensions to cluster our data and find the five genes that are important for a given disease of interest. In fact, in highlighting the six dimensions mentioned above, we are just beginning to address the central problem of dimensionality in biology.

A second conjecture is that living systems operate in a space with a changing number of dimensions (biological space-time), and that it is this very ability that makes them able to self-organize. The kind of modules that system approaches start to reveal (Davidson *et al.* 2002; Ideker *et al.* 2001b) are embedded structures with their own number of dimensions, and the challenge for the future is to find, in each biological system, which is the correct number of dimensions that have to be considered in a given context.

5. Conclusion

The central goal for the coming decades will be to build the ability to measure small changes of small signals in biological systems. If this can be achieved, through cross-disciplinary contributions, we may be able to gain a better understanding of living systems, but also to design and perform subtle chaotic piloting of biological systems with complex mixtures of biomolecules to prevent or treat diseases, with limited adverse effects.

We thank the members of our teams and the speakers and participants of the Nobel Symposium for their contributions, particularly Dennis Bamford, Konrad Kaufmann, Teuvo Kohonen, Tak Mak, Lee Smolin and Guy Theraulaz, and we thank Odile Brasier, Christine Couillaud and Lena Norlund for manuscript editing.

References

- Auffray, C., Imbeaud, S., Roux-Rouquié, M. & Hood, L. 2003 From functional genomics to systems biology: conceptual and practical issues. (Submitted.)
- Davidson, E. H. (and 24 others) 2002 A genomic regulatory network for development. *Science* **295**, 1669–1678.
- Descartes, R. 1637 *Discours de la méthode pour bien conduire sa raison et rechercher la vérité dans les sciences* (transl. C. Auffray). Paris: C. Angot.
- Flake, G. W. 2000 The computational beauty of nature. In *Computer explorations of fractals, chaos, complex systems, and adaptation*. Cambridge, MA: MIT Press.
- Heylighen, F. 1997 Self-organization, emergence and the architecture of complexity. In *Principia cybernetica*. (Available at ftp://ftp.vub.ac.be/pub/projects/Principia_Cybernetica/Papers_Heylighen/Self-Organization_Complexity.txt.)
- Hughes, T. R. (and 21 others) 2000 Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126.
- Ideker, T., Galitski, T. & Hood, L. 2001a A new approach to decoding life: systems biology. *A. Rev. Genomics Hum. Genet.* **2**, 343–372.
- Ideker, T. *et al.* 2001b Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934.
- Imbeaud, S., Graudens, E. & Auffray, C. 2003 ‘The 39 steps’ in expression profiling with microarrays. (In preparation.)
- Klir, J. & Valach, M. 1967 *Cybernetic modeling*. Princeton, NJ: Iliffe Books.
- Le Moigne, J. L. 1990 *La modélisation des systèmes complexes*. Paris: Dunod.

- Mesarovic, M. D. 1962 On self-organizational systems. In *Self-organizing systems* (ed. M. C. Yovits *et al.*). Washington, DC: Spartan Books.
- Morin, E. 1977 *La méthode. La nature de la nature*. Paris: Le Seuil.
- Morin, E. 1980 *La méthode. La vie de la vie*. Paris: Le Seuil.
- Morin, E. & Le Moigne, J. L. 1999 *L'intelligence de la complexité*. Paris: L'Harmattan.
- Poquelin, J. B. (Molière) 1671 *Le gentil bourgeois*. Paris: Le Monnier.
- Roux-Rouquié, M. & Le Moigne, J. L. 2002 The systemic paradigm and its relevance to the modeling of biological functions. *C. R. Acad. Sci. Paris Ser. III* **325**, 419–430.
- Simon, H. A. 1969 *The sciences of the artificial*, 1st edn. Cambridge, MA: MIT Press.
- Simon, H. A. 1996 *The sciences of the artificial*, 3rd edn. Cambridge, MA: MIT Press.
- von Foerster, H. 1960 On self-organizing systems and their environments. In *Self organizing systems* (ed. M. C. Yovits & S. Cameron), pp. 31–50. New York: Pergamon Press.
- Watson, J. D. 1968 *The double helix: a personal account of the discovery of the structure of DNA*. New York: Atheneum.
- Yang, Y. H. & Speed, T. 2002 Design issues for cDNA microarray experiments. *Nat. Rev. Genet.* **3**, 579–588.

