Trends in Cell & Molecular Biology
Vol. 6, 2011

Original Communication

Biochemical lattices and networks as models of living systems: A problem of Artificial Life

Jacques Ricard*

ANBioPhy (Acides Nucléiques et Biophotonique), Université Pierre-et-Marie Curie, Paris 6, 4, Place Jussieu, 75252 Paris, France

ABSTRACT

The aim of the present paper is to discuss, in physical terms, some important features of living systems, namely their identity, their ability to communicate, to generate information, to behave as coherent wholes and as evolving systems. These properties can be defined in mathematical terms and studied on very simple systems such as biochemical lattices and enzyme networks.

KEYWORDS: communication, emergence, information, lattice, network

1. INTRODUCTION

Research programs aimed at studying biological phenomena that can be mimicked through physical systems usually belong to what is called today Systems Biology or Artificial Life. As a matter of fact, living organisms possess a number of features that, in the past, have often been considered specific of these systems. In the present paper, we are going to discuss, in the context of physical models, some features of living systems, namely their identity, their ability to communicate, their behaviour as coherent wholes and as evolving systems. Let us discuss briefly each of these properties.

Any living system constitutes an entity that possesses an identity, viz. it is different from other similar entities. In the case of present day living systems, this identity is in a way represented by the sequence of DNA base pairs, or by the bases of RNA. Hence a feature that expresses the identity of a material entity should, of necessity, display a low probability of occurrence otherwise it could not be considered a marker of this identity. Moreover many communication processes occur in living systems. Perhaps the most celebrated of these processes is the one that takes place between DNA and proteins via RNA, and corresponds in fact to a genetic information transfer between DNA and proteins. Last but not least, any living organism, even the simplest one, is a system made up of many connected elements, organelles or molecules, in such a way that the global properties of the system are, in general, different from the properties of its elements. If we could compare the features of the global system with those of its constitutive elements one could realize the system to possess either less, or more, properties than the set of its elements. In the first case, the system could be considered integrated and in the second case it could be considered emergent. There is no doubt that today's living organisms are emergent systems for they possess properties that are not borne by any of their constitutive elements but are generated through the interactions that exist between the elements of a system. Last, living organisms are spontaneously able to evolve. Moreover one can often detect some kind of progress in the process of evolution. In the case of living systems, this progress is exerted through the selection of advantageous mutations. In the case of artificial

*jkricard@aol.com

physical systems that do not possess any genetic material, mutations and selection cannot take place. However, one may wonder whether spontaneous processes of self-organization cannot take place that would lead to some kind of an evolutionary process.

As living organisms are in fact systems, one can wonder whether some biological functions, such as those presented above, could not be modelled with physical models such as biochemical lattices and networks. A giant macromolecule that possesses two classes of sites able to bind specifically ligand *x* or ligand *y* would generate such a lattice. Alternatively, any enzyme-catalysed chemical reaction is per se a network and more complex networks could be formed by associating several enzyme-catalysed chemical reactions.

2. Identity of a material entity and its information

The concept of identity of a material entity, its *essence*, has been formulated long ago by Aristotle [1]. It can be defined as the ontological principle that gives a material entity the ability to be discriminated from other similar entities. This ontological principle, as well as the ability of discriminating *one* material entity among many others, is called information [2-8]. Hence, from a practical viewpoint, information is both what makes a material entity different from its neighbours and the ability we have to identify this entity. It is evident that this concept of information of an event should be related to the probability of occurrence of that event. The smaller this probability of occurrence and the larger is the information associated with the supervening of this event.

If an event x_i has a probability of occurrence, $p(x_i)$, the information, $h(x_i)$, associated with the supervening of this event is

$$h(x_i) = f\left\{\frac{1}{p(x_i)}\right\} \tag{1}$$

where *f* is an increasing function. In order to determine the nature of this function, one can consider an event involving the independent occurrence of two other events x_i and y_j. It is evident that the information, $h(x_i, y_j)$, brought about by these two independent events, should be equal to the sum of the two informations, $h(x_i)$ and $h(y_j)$, of these events. Hence one has

$$h(x_i, y_j) = h(x_i) + h(y_j) \tag{2}$$

It then follows that the simplest expression of the *f* function is a logarithmic one for one has

$$h(x_i, y_j) = -\log p(x_i, y_j) = -\log p(x_i) - \log p(y_j) \tag{3}$$

If now the events x_i and y_j interact, the Bayes theorem requires that

$$p(x_i, y_j) = p(x_i)p(y_j|x_i) = p(y_j)p(x_i|y_j) \tag{4}$$

In this expression $p(y_j|x_i)$ is the conditional probability of occurrence of the event y_j given that x_i has already occurred. Alternatively, $p(x_i|y_j)$ is the conditional probability of occurrence of x_i given that y_j has already occurred. It then follows that

$$h(x_i, y_j) = h(x_i) + h(y_j|x_i) = h(y_j) + h(x_i|y_j) \tag{5}$$

It appears from this equation that the interaction between x_i and y_j may generate an increase, or a decrease, of the value of $h(x_i, y_j)$. In order to express the amount of joint information generated, or consumed, by the interaction between x_i and y_j one may define a new function, $i(x_i, y_j)$, called mutual information of interaction, as

$$i(x_i : y_j) = h(x_i) + h(y_j) - h(x_i, y_j) \tag{6}$$

Taking advantage of expressions (5) equation (6) becomes

$$i(x_i : y_j) = h(x_i) - h(x_i|y_j) = h(y_j) - h(y_j|x_i) \tag{7}$$

Under this form it becomes evident that if

$$h(x_i|y_j) > h(x_i) \text{ and } h(y_j|x_i) > h(y_j) \quad (8)$$

interaction between x_i and y_j generates an additional information and the system is defined as emergent. Alternatively, if

$$h(x_i|y_j) < h(x_i) \text{ and } h(y_j|x_i) < h(y_j) \quad (9)$$

the interaction between x_i and y_j produces a consumption of information and the system is integrated. In other words the function $i(x_i : y_j)$ measures the information taken up, or generated, by the interaction occurring between x_i and y_j . One can easily conceive that the identity of a system, a lattice or a network for instance, can be defined by the set of the $i(x_i : y_j)$ values associated with its nodes.

3. Identity, integration, emergence and communication in a protein lattice

Let us consider a macromolecule, for instance a protein that possesses two classes of sites each able to bind specifically ligand x or ligand y. For simplicity, we assume that the two classes of sites can bind the same number, n, of molecules of ligand x and of ligand y. One can then obtain the square lattice shown in Figure 1. This lattice possesses different features: its identity, the ability to display integration or emergence as well as to communicate a message within the lattice [8].

This lattice can be represented by a set of nodes defined by the following relationship

$$\Omega_N = \{p(N_{\kappa,\lambda}); \kappa, \lambda \in Z^+, \kappa, \lambda \le n\} \quad (10)$$

In this expression $p(N_{\kappa,\lambda})$ is the probability of occurrence of a node that has bound κ molecules of x and λ molecules of y . As κ and λ can take the successive values $0,1,2,...,n$, Ω_N collects all the nodes of the lattice (Figure 1). Setting

$$h(N_{\kappa,\lambda}) = -\log p(N_{\kappa,\lambda}) \quad (11)$$

one can define the identity of the lattice through the expression

$$\Omega_H = \{h(N_{\kappa,\lambda}); \kappa, \lambda \in Z^+, \kappa, \lambda \le n\} \quad (12)$$

One can distinguish, in the Ω_N set three subsets: $\Omega_0, \Omega_{Nx}, \Omega_{Ny}$. Ω_0 collects the probabilities of occurrence of the nodes that have bound neither *x* nor *y*. Ω_{Nx} assembles the probabilities of occurrence of the nodes that associate *x* and possibly *y*. Last but not least Ω_{Ny} brings together the nodes associated with *y* and possibly *x*. Hence the subsets Ω_{Nx} and Ω_{Ny} are defined as

$$\Omega_{Nx} = \{p(N_{i,\lambda}); i \in N, \lambda \in Z^+, i, \lambda \le n\} \quad (13a)$$

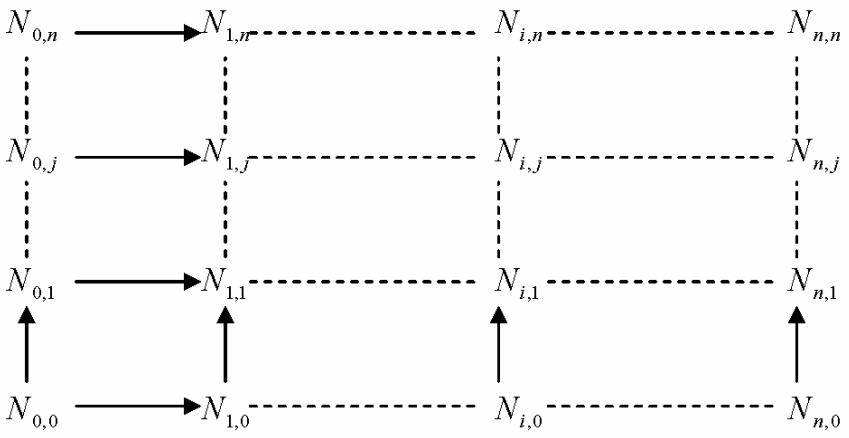


Figure 1. An ideal biochemical lattice.
The $N_{i,j}$ values represent the various states of the system that binds two different types of ligands.

$$\Omega_{Ny} = \left\{p(N_{\kappa,j}); \kappa \in Z^+, j \in N, \kappa, j \le n\right\} \quad (13b)$$

One can also define the probability that a node of the lattice has bound *i* molecules of *x* whether or not it has also bound molecules of *y* as

$$p(x_i) = \sum_{\lambda=0}^{n} p(N_{i,\lambda}) \quad (14)$$

Similarly, the probability that a node of the protein lattice has bound *j* molecules of *y* whether or not it has also bound molecules of *x* can be expressed as

$$p(y_j) = \sum_{\kappa=0}^{n} p(N_{\kappa,j}) \quad (15)$$

From the $p(x_i)$ and $p(y_j)$ values one can define two probability spaces Ω_X and Ω_Y as

$$\Omega_X = \left\{p(x_i); i \in N\right\} \quad (16a)$$

$$\Omega_Y = \left\{p(y_j); j \in N\right\} \quad (16b)$$

The states x_i and y_j allow to define two sets *X* and *Y* whose Cartesian product is *XY*. Its corresponding probability space is then (Figure 2)

$$\Omega_{XY} = \left\{p(x_i, y_j); i, j \in N\right\} \quad (17)$$

One can define from relations (16) and (17), *h* functions as

$$h(x_i) = -\log p(x_i) \quad (18a)$$

$$h(y_j) = -\log p(y_j) \quad (18b)$$

$$h(x_i, y_j) = -\log p(x_i, y_j) \quad (18c)$$

and from the values of these functions one can define two sets

$$\Theta_X = \left\{h(x_i); i \in N\right\} \quad (19a)$$

$$\Theta_Y = \left\{h(y_j); j \in N\right\} \quad (19b)$$

that allow to define in turn two functions $H(X)_N$ and $H(Y)_N$ as

$$H(X)_N = \sum_i \sum_j p(x_i, y_j) h(x_i) \quad (20a)$$

$$H(Y)_N = \sum_i \sum_j p(x_i, y_j) h(y_j) \quad (20b)$$

These functions are generalizations, for lattices, of the *h* functions defined above for x_i and y_j. One can also define conditional *H* functions as

$$H(X|Y)_N = \sum_i \sum_j p(x_i, y_j) h\left(x_i \middle| y_j\right) \quad (21a)$$

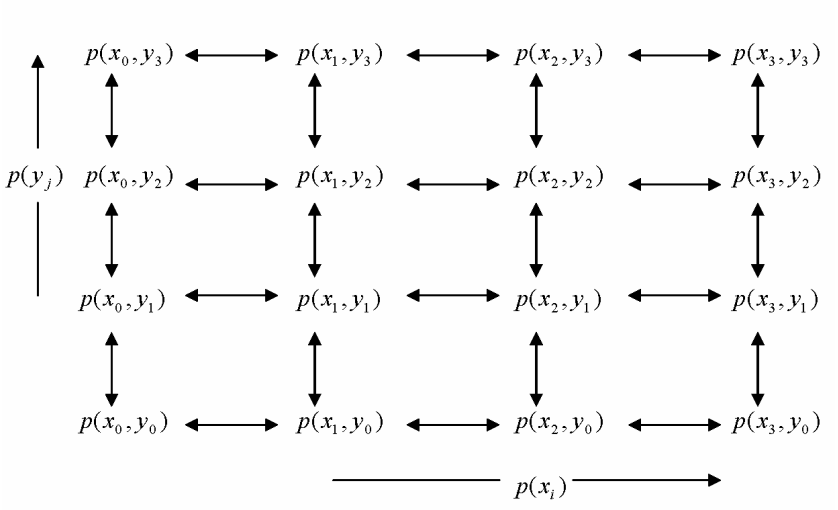


Figure 2. A biochemical square lattice.
Different types of ligands *x* and *y* are bound to the lattice. The states of the system are described by their probability of occurrence. Different $p(x_i)$ values are defined along the horizontal x axis. Similarly, different $p(y_j)$ values are defined along the vertical *y* axis (see main text).

$$H(Y|X)_N = \sum_i \sum_j p(x_i, y_j) h(y_j|x_i) \qquad (21b)$$

Functions (20) and (21) are conventionally expressed per node bearing both *x* and *y*.

It is then possible to generalize, to a population of nodes, equation (7) that was initially formulated for one node. One finds

$$I(X:Y)_N = H(X)_N - H(X|Y)_N = H(Y)_N - H(Y|X)_N \qquad (22)$$

In the general case, this expression can adopt positive, negative or zero values. In the first situation the system is defined as integrated. This means that the interaction between x_i and y_j results in the consumption of information. In the second case, the interactions between x_i and y_j generate information and the system is defined as emergent. Last, if expression (22) adopts zero values, the lattice is not a system but just a collection of independent reaction processes. The interesting idea that comes out from these results is that integration and emergence are not individual properties of the elements of the system but properties of the interactions between these elements. In the case of emergence, the whole is more than the sum of its parts. In the case of integration, it is less.

Now let us assume that *x*, for instance, has a strong affinity for the macromolecule and that *x* binding increases the affinity of *y* for the same macromolecule. Under these conditions, no macromolecule will occur in a free, unbound, state, or with one type of ligand, either *x* or *y*, bound to its surface. The probabilities of occurrence of the various states of the macromolecular lattice can then be depicted as shown in Figure 3.

Under these conditions, expressions (14) and (15) become

$$p(x_i) = \sum_{j=1}^{n} p(x_i, y_j) \qquad (23)$$

and

$$p(y_j) = \sum_{i=1}^{n} p(x_i, y_j) \qquad (24)$$

Thus, for instance, in the case of Figure 3, one has

$$p(x_1) = p(x_1, y_1) + p(x_1, y_2) + p(x_1, y_3) \qquad (25a)$$

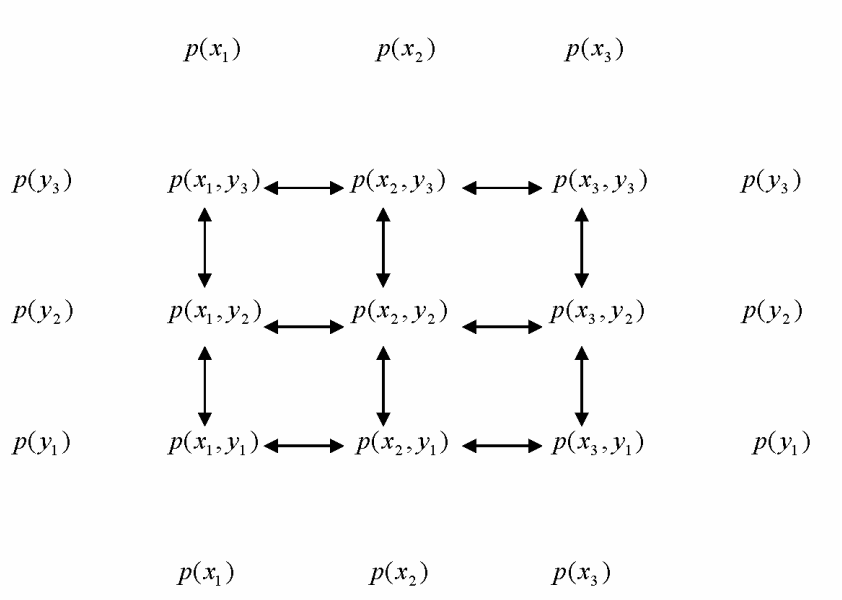


Figure 3. Connexion between $p(x_i)$ and $p(y_j)$ probabilities.
As described in the main text, the connexion between the two types of probabilities is effected thanks to the joint probabilities of the $p(x_i, y_i)$ type.

$$p(y_1) = p(x_1, y_1) + p(x_2, y_1) + p(x_3, y_1) \quad (25b)$$

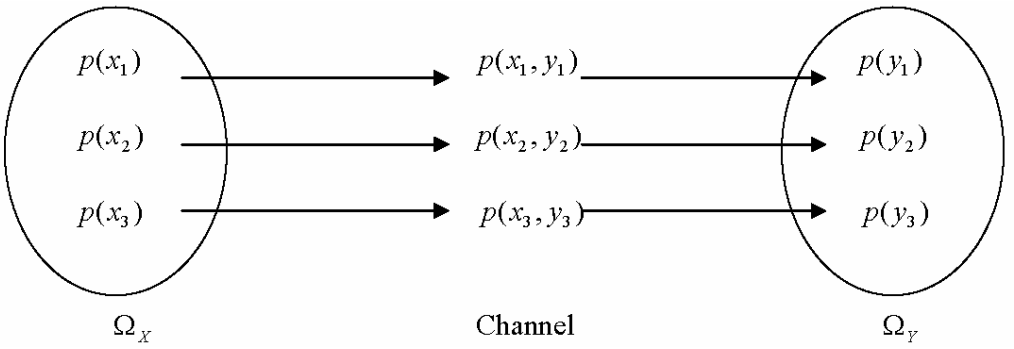
These relationships show there is a communication process between $p(x_1)$ and $p(y_1)$ through $p(x_1, y_1)$. Hence information can flow from x_1 to y_1 owing to the existence of the joint probability $p(x_1, y_1)$. One has

$$\begin{matrix} p(x_1, y_1) \\ p(x_1, y_2) \\ p(x_1, y_3) \end{matrix} \longrightarrow p(x_1, y_1) \longrightarrow \begin{matrix} p(y_1, x_1) \\ p(y_1, x_2) \\ p(y_1, x_3) \end{matrix} \quad (26)$$

Taking advantage of expressions (25) this is equivalent to

$$p(x_1) \longrightarrow p(x_1, y_1) \longrightarrow p(y_1) \quad (27)$$

This reasoning can be easily extended to all the $p(x_i)$ and the $p(y_i)$ of Figure 3. One has



(28)

This means that a given "message" is expressed in a different alphabet. From a mathematical viewpoint, the communication through a channel, of the elements of the set Ω_X up to the elements of the set Ω_Y is a one-to-one mapping of Ω_X onto Ω_Y.

The physical bases of a communication process can be found in a molecular lattice of the type shown in Figure 4. As shown in Figure 4, the succession of x_1, x_2, x_3 is reproduced, or "translated", in the sequence y_1, y_2, y_3 only if the terms $p(x_i, y_j)$ are negligible when $i \neq j$,. In that case, information is transferred in both directions, from the sequence x_1, x_2, x_3 to the sequence y_1, y_2, y_3 , and conversely. If this condition is not fulfilled, the expression of a sequence is scrambled during a communication process.

One can find out, as an example of biological reversible information transfer, the information transfer between DNA and mRNA. If we consider, for instance, a DNA region that can be transcribed into a mRNA segment, it is not surprising to observe that this process is reversible for the number of the "coding terms" is the same for the DNA and the RNA segments. This prediction could have been made before the experimental results proving the reversibility of the transcription process. If, alternatively, Ω_X is the "codons space" and Ω_Y the "aminoacids space" the information flow can follow one direction only, namely from Ω_X to Ω_Y. Contrary to what had been initially thought, this could not be considered the "central dogma of molecular biology" [9, 10] for it is in fact the property of any code, biological or not, in which the two probability spaces are not isomorphic. As shown in Figure 3, the transfer of information between *x*

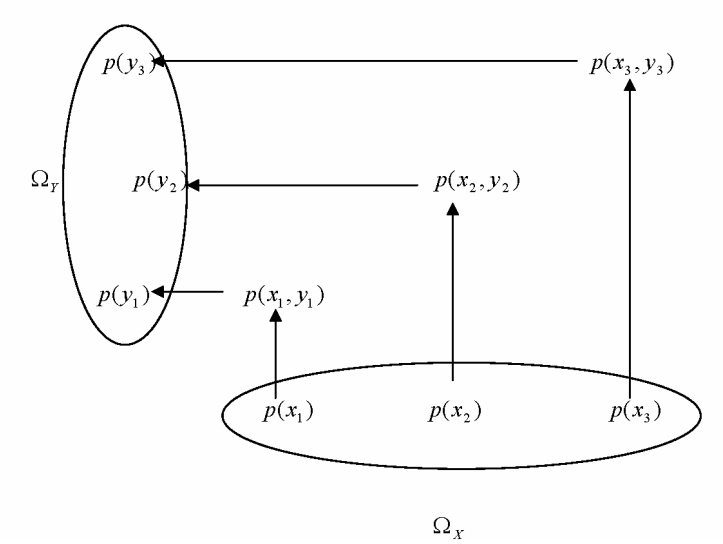


Figure 4. The principle of the communication channel.
The communication between the two sets Ω_X and Ω_Y is effected through joint probabilities of the $p(x_i, y_i)$ type.

and *y* is possible only through the term that is common to both $p(x)$ and $p(y)$.

It thus appears that a molecular lattice, and more generally any kind of network, can display three fundamental properties of living systems, namely: an identity that can be expressed through a specific amount of information borne by the various nodes of the lattice; the possibility of the system to generate its own information; and the ability of the lattice to communicate a message in a channel through a different language or alphabet. It is intuitively obvious that the concept of emergence of information by a system is antagonistic to that of communication of a message by the same system for a communication process requires integration of that system. Indeed it is quite possible that different regions of the same system could possess different "functions", viz. some regions could communicate whereas others could generate information. Whatever that may be, there is a switch that allows the same region of the system to communicate, or to generate information.

The model of Figure 3 predicts that if $p(x_1, y_1) \approx p(x_1) \approx p(y_1)$ the communication channel between $p(x_1)$ and $p(y_1)$ works both ways, viz. from $p(x_1)$ to $p(y_1)$ and from $p(y_1)$ to $p(x_1)$. Alternatively, if $p(x_1) > p(x_1, y_1) > p(y_1)$ the information transfer will occur from $p(x_1)$ to $p(y_1)$ (Figures 3 and 5). One can then notice that the simple model of Figure 3 can take account of the fact that information in biological systems can be transferred in one, or in two directions.

4. The switch between emergence and communication of information

It is therefore important to develop a mathematical rule that defines the conditions generating a communication or, alternatively, an emergence process. This mathematical rule, which is the very basis of any communication process, is called the subadditivity principle in communication theory [7, 11, 12]. It constitutes a switch between two different functions, namely the communication of a message and the emergence of information.

Equation (22) above can be rewritten as

$$I(X:Y)_N = \sum_i \sum_j p(x_i, y_j) \log \frac{p\left(x_i \middle| y_j\right)}{p(x_i)} \quad (29)$$

and this function can adopt positive or negative values. In order to find out the conditions that

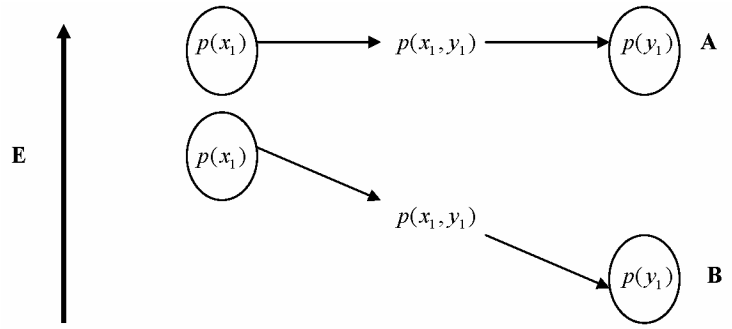


Figure 5. A bidirectional and monodirectional communication process.
In **A**, the communication between the $p(x_1)$ and the $p(y_1)$ is bidirectional for $p(x_1)$ and $p(y_1)$ have about the same order of magnitude, viz. the corresponding material entities possess about the same energy. In **B**, the communication process is unidirectional for it is assumed that $p(x_1) > p(y_1)$.

generate communication or, alternatively, emergence of information, one defines a new function, $I^{\bullet}(X:Y)_N$, that possesses values smaller than, or equal to, $I(X:Y)_N$. Moreover, if subadditivity condition applies, $I^{\bullet}(X:Y)_N$ adopts a zero value. In order to find out this function, one can take advantage that, for all $x > 0$, one has

$$x - 1 \geq \ln x \tag{30}$$

that can be rewritten as

$$x - 1 \geq \frac{1}{M} \log x \tag{31}$$

with $M = \log e$. It follows that

$$\log e(x - 1) \geq \log x \tag{32}$$

Let us set

$$x = \frac{p(x_i)}{p\left(x_i \middle| y_j\right)} \tag{33}$$

It follows from expression (32) that

$$\log e\left\{\frac{p(x_i)}{p\left(x_i \middle| y_j\right)} - 1\right\} \geq \log \frac{p(x_i)}{p\left(x_i \middle| y_j\right)} \tag{34}$$

and the function $I^{\bullet}(X:Y)_N$ is then defined as

$$I^{\bullet}(X:Y)_N = -\log e \sum_i \sum_j p(x_i, y_j)\left\{\frac{p(x_i)}{p\left(x_i \middle| y_j\right)} - 1\right\} \tag{35}$$

that should be compared to equation (29), viz.

$$I(X:Y)_N = -\sum_i \sum_j p(x_i, y_j) \log \frac{p(x_i)}{p\left(x_i \middle| y_j\right)} \tag{36}$$

According to expression (34), it appears that $I(X:Y)_N \geq I^{\bullet}(X:Y)_N$. Equation (35) can be rearranged to

$$I^{\bullet}(X:Y)_N = -\log e \sum_i \sum_j \left\{\frac{p(x_i, y_j)\, p(x_i)}{p\left(x_i \middle| y_j\right)} - p(x_i, y_j)\right\} \tag{37}$$

As the conditional probability $p\left(x_i \middle| y_j\right)$ can be rewritten as

$$p\left(x_i \middle| y_j\right) = \frac{p(x_i, y_j)}{p(y_j)} \tag{38}$$

it follows that

$$I^{\bullet}(X:Y)_N = -\log e \sum_i \sum_j \left\{p(x_i)\, p(y_j) - p(x_i, y_j)\right\} \tag{39}$$

or

$$I^{\bullet}(X:Y)_N = \log e\left\{\sum_i \sum_j p(x_i, y_j) - \sum_i p(x_i) \sum_j p(y_j)\right\} \tag{40}$$

If $I^{\bullet}(X:Y)_N = 0$, one has, of necessity, $I(X:Y)_N \geq 0$. The conditions that generates the

relation $I^{\bullet}(X:Y)_N = 0$ defines the so-called subadditivity principle, namely

$$\sum_i \sum_j p(x_i, y_j) - \sum_i p(x_i) \sum_j p(y_j) = 0 \quad (41)$$

An interesting condition that generates subadditivity is

$$\sum_i \sum_j p(x_i, y_j) = \sum_i p(x_i) = \sum_j p(y_j) = 1 \quad (42)$$

The situation described in expressions (42) implies that every node of the lattice associates both x and y. In such a situation, there cannot exist a node bearing no ligand, or only one of them. Then the system can be defined as a communication channel functioning in one or two directions. This is the situation depicted in Figure 3.

It appears, however, from expression (42) that, if various states of the system are associated with one ligand only, or with none of them, expression (41) does not hold anymore. Conversely, if relationships (42) apply, the subadditivity condition holds and information is transferred in the system from place to place. It is then evident that the concept of subadditivity implies that a communication process is taking place in the lattice.

Let us consider the lattice of Figure 1. In this scheme we have three different types of nodes: a node, $N_{0,0}$, which is not associated with a ligand *x* or *y* ; several nodes, $N_{i,0}$, that are associated with ligand *x* only; several nodes, $N_{0,j}$, associated with ligand *y* only; and nodes, $N_{i,j}$, associated with both ligands x_i and y_j. One has then

$$p(N_{0,0}) + \sum_i p(N_{i,0}) + \sum_j p(N_{0,j}) + \sum_i \sum_j p(N_{i,j}) = 1 \quad (43)$$

It follows that

$$\sum_i p(x_i) = 1 - p(N_{0,0}) - \sum_j p(N_{0,j}) \quad (44a)$$

$$\sum_j p(y_j) = 1 - p(N_{0,0}) - \sum_i p(N_{i,0}) \quad (44b)$$

$$\sum_i \sum_j p(x_i, y_j) = 1 - p(N_{0,0}) - \sum_i p(N_{i,0}) - \sum_j p(N_{0,j}) \quad (44c)$$

and one has

$$\sum_i \sum_j p(x_i, y_j) - \sum_i p(x_i) \sum_j p(y_j) = \sigma_\rho =$$
$$= p(N_{0,0}) \left\{ 1 - p(N_{0,0}) - \sum_i p(N_{i,0}) - \sum_j p(N_{0,j}) \right\} - \sum_i p(N_{i,0}) \sum_j p(N_{0,j}) \quad (45)$$

which can be rearranged to

$$\sigma_\rho = p(N_{0,0}) \sum_i \sum_j p(N_{i,j}) - \sum_i p(N_{i,0}) \sum_j p(N_{0,j}) \quad (46)$$

Hence it appears that the subadditivity rule is not followed if expression (46) above assumes negative values, that is if

$$\sum_i p(N_{i,0}) \sum_j p(N_{0,j}) > p(N_{0,0}) \sum_i \sum_j p(N_{i,j}) \quad (47)$$

This condition implies that many nodes of the lattice are associated with *x* or *y* but not with both of them.

The first important conclusion that can be derived is that subadditivity condition σ_ρ can be nil or positive and the lattice is an integrated system that communicates information from place to place. The second interesting conclusion that can be formulated, when σ_ρ assumes negative values, is the emergence of information generated by the system.

5. Reduction or emergence in a macromolecular lattice

Classical molecular biology, which has developed since 1950, is based, in most cases, on what could be called an ontological reductionism, viz. the idea that a system can be studied and understood from its simple decomposition into its elements and the independent study of these elements. The concept of reduction is not restricted to the study of the problems of complexity. It is in fact the philosophical doctrine that aims at defining the predicates of a theory T_h in terms of the predicates of another theory T_l, more general and embracing [13-16]. T_h is called a “high-level” and T_l a “low-level” theory. Let us represent the set of the predicates of T_h by C_h and that of T_l by C_l. The fact that C_h is included in C_l, viz.

$$C_h \subset C_l \tag{48}$$

implies that it is possible to reduce the high-level to the low-level theory. This type of analysis can be applied to the reduction of a system to its component sub-systems. If such a reduction is not feasible this means that the sub-systems interact and from this interaction emergence of information and novel properties take place.

Let us consider, as an example, a macromolecular lattice *XY* originating from the interaction between two linear binding sequences, *X* and *Y*, of ligand *x* and ligand *y* on their respective protein sites. If these sites do not interact, that is if the binding of *x* does not affect that of *y*, and conversely, the probability that the lattice *XY* has bound *i* molecules of *x* is

$$p(x_i) = \frac{\binom{n}{i} K_x^i x^i}{(1 + K_x x)^n} \tag{49}$$

In this expression, *n* is the largest number of ligand molecules *x* that can be bound to the protein. K_x is the corresponding binding constant of *x* to the macromolecule. Similarly, the probability that the lattice has bound *j* molecules of *y* is

$$p(y_j) = \frac{\binom{n}{j} K_y^j y^j}{(1 + K_y y)^n} \tag{50}$$

In this expression K_y is the binding constant of *y* to the macromolecule. Last but not least, the probability that the lattice has bound both i molecules of x and j molecules of y is

$$p(x_i, y_j) = \frac{\binom{n}{i}\binom{n}{j} K_x^i x^i K_y^j y^j}{(1 + K_x x)^n (1 + K_y y)^n} \tag{51}$$

One can observe that expression (51) is the product of equations (49) and (50) viz.

$$p(x_i, y_j) = p(x_i)\, p(y_j) \tag{52}$$

This relationship is valid only because the binding of x does not affect that of y and conversely. One can then conclude that the binding properties of *x* and *y* by the system *XY* can be correctly described from the independent studies of the binding of *x* and *y* alone on the system (Figure 6).

This situation is feasible only because the binding processes of x and y to their respective sites do not interact. The lack of interaction implies that the emergence of information is not to be expected under these conditions and that the binding properties of x in the presence of y, or of *y* in the presence of *x*, could be reduced to the binding properties of *x* or *y* alone. If, however, the binding sites for *x* and *y* were in interaction it would have been impossible to reduce the binding properties of the global system to the individual properties of *x* binding, or of *y* binding, considered in isolation.

6. Non-equilibrium as a source of emergence in simple enzyme networks

Enzyme reactions can be considered simple biochemical networks as they describe the successive transformations of states that take place during the enzyme conversion of substrates into products. Most enzyme reactions involve two substrates and two products. They can be described by the simple model

$$AX + B \leftrightarrow A + XB$$

where *AX* and *B* are the substrates, *A* and *XB* the products. The reaction is catalyzed by enzyme *E*. We are going to discuss, in the following, the situation where the enzyme binds the two substrates before releasing the two products. With two substrates, *AX* and *B*, one can postulate there exists either two different sequences, or one sequence, leading to the ternary enzyme-substrates complex *EAXB*. These sequences are shown in Figure 7. The first sequence corresponds to a random process where the enzyme binds *AX* or *B* first and *B* or *AX* afterwards (Figure 7A). If, alternatively, the substrate binding process follows a compulsory sequence, this means there exists a first and a second substrate that bind to the enzyme (Figure 7B). The kinetics of product appearance can be followed under steady state conditions and the overall system is a network of chemical processes taking place under both non-equilibrium and open conditions.

From such a network one can derive the probabilities of occurrence of states *EAX*, *EB* and *EAXB* as well as the shifts between the standard quasi-equilibrium and the actual non-equilibrium of the system (the *u*'s). One finds

$$p(AX) = \frac{K_1[AX](1+K_3[B]) + u_{EAX}}{1+K_1[AX]+K_3[B]+K_1K_2[AX][B]+u_{EAX}+u_{EB}+u_E} \quad (53a)$$

$$p(B) = \frac{K_3[B](1+K_4[AX]) + u_{EB}}{1+K_1[AX]+K_3[B]+K_1K_2[AX][B]+u_{EAX}+u_{EB}+u_E} \quad (53b)$$

$$p(AX,B) = \frac{K_1K_2[AX][B]}{1+K_1[AX]+K_3[B]+K_1K_2[AX][B]+u_{EAX}+u_{EB}+u_E} \quad (53c)$$

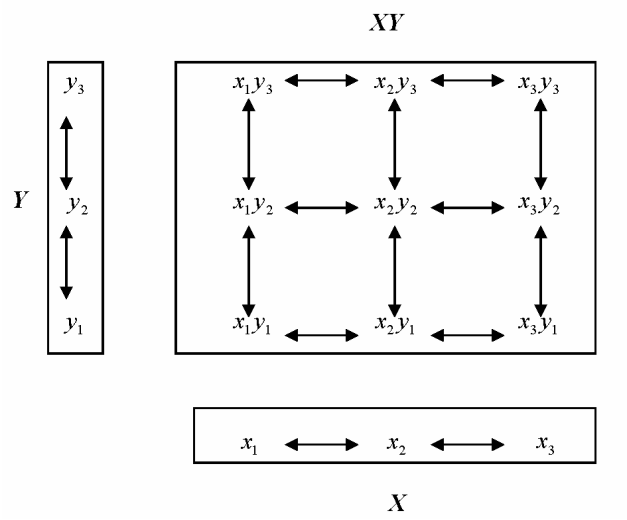


Figure 6. Illustration of the concepts of complexity and reduction.
In general, the properties of the ***XY*** system cannot be reduced to the properties of ***X*** and ***Y*** considered in isolation. Such a reduction is possible, however, if the properties of *x* and *y* remain unaltered after they have interacted.

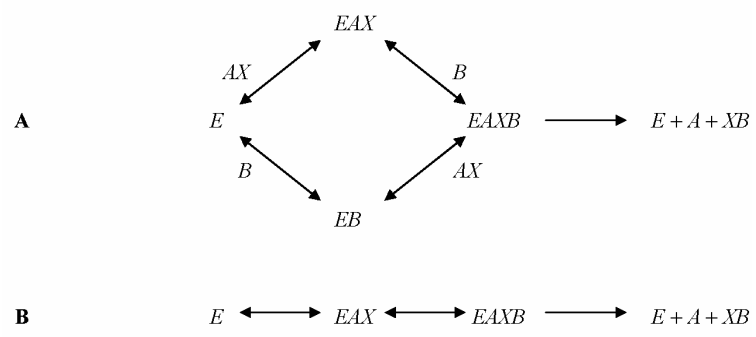


Figure 7. Simple random and ordered enzyme reactions as networks.
A, Random binding of substrates, AX and B, to the enzyme.
B, Ordered binding of substrates, AX and B, to the enzyme.

In these expressions u_{EAX}, u_{EB} and u_E are the shifts between standard quasi-equilibrium and the actual non-equilibrium of the system [17]. Hence the *u*'s express how the system departs from standard quasi-equilibrium state up to the actual non-equilibrium conditions. In expressions (53) the constants $K_1, K_2, ...$ are the ratios $k_1 / k_{-1}, k_2 / k_{-2}, \ldots$ of rate constants. The shifts between the actual non-equilibrium conditions of the system and its quasi-equilibrium state is expressed by the values of the perturbation terms viz.

$$u_{EAX} = \frac{kk_1[AX](k_{-3} + k_4[AX])}{k_{-3}k_{-4}(k_{-1} + k_2[B]) + k_{-1}k_{-2}(k_{-3} + k_4[AX])} \quad (54a)$$

$$u_{EB} = \frac{kk_3[B](k_{-1} + k_2[B])}{k_{-3}k_{-4}(k_{-1} + k_2[B]) + k_{-1}k_{-2}(k_{-3} + k_4[AX])} \quad (54b)$$

$$u_E = \frac{k(k_{-1} + k_2[B])(k_{-3} + k_4[AX])}{k_{-3}k_{-4}(k_{-1} + k_2[B]) + k_{-1}k_{-2}(k_{-3} + k_4(AX])} \quad (54c)$$

Moreover simple inspection of the enzyme network suggests that the binding of the substrates, *AX* and *B*, to the enzyme may interact, thus leading to an increase, or a decrease, of the corresponding information. As already outlined, one can express in quantitative terms the degree of emergence, or of integration, of the network as

$$i(AX : B) = \log \frac{p(AX|B)}{p(AX)} = \log \frac{p(B|AX)}{p(B)} \quad (55)$$

As expected, $i(AX : B)$ will be negative and the system will be emergent if $p(AX|B) < p(AX)$ and $p(B|AX) < p(B)$. One can derive, for instance, the expression of $p(AX|B)$ and one finds

$$p(AX|B) = \frac{K_4[AX]}{1 + K_4[AX] + u^{\bullet}_{EB}} \quad (56)$$

where $u^{\bullet}_{EB} = u_{EB} / K_3[B]$. Hence, $u^{\bullet}_{EB}$ assumes the form

$$u^{\bullet}_{EB} = \frac{kk_{-3}(k_{-1} + k_2[B])}{k_{-3}k_{-4}(k_{-1} + k_2[B]) + k_{-1}k_{-2}(k_{-3} + k_4[AX])} \quad (57)$$

and the ratio $p(AX|B) / p(AX)$ can be expressed as

$$\frac{p(AX|B)}{p(AX)} = \frac{K_4[AX] + K_1K_4[AX]^2 + K_3K_4[AX][B] + K_1K_2K_4[AX]^2[B] + N(u)}{K_1[AX] + K_1K_4[AX]^2 + K_3K_4[AX][B] + K_1K_2K_4[AX]^2[B] + D(u)} \quad (58)$$

In this relationship $N(u)$ and $D(u)$ express how non-equilibrium conditions play a part in the appearance of emergence, or of integration, of the system. If the system is close to equilibrium, which implies that $N(u) = D(u) = 0$, it will be emergent if $K_1 > K_4$ and $K_3 > K_2$. Alternatively, the system will be integrated if $K_1 < K_4$ and $K_3 < K_2$. However, in agreement with relationship (58), one should expect the conditions for integration or emergence to be altered if the system departs from quasi-equilibrium. The difference $D(u) - N(u)$ is a measure of the role played by non-equilibrium conditions on the degree of emergence of the system. The larger this difference and the greater is the degree of emergence of the system. The difference $D(u) - N(u)$ can be expressed under two equivalent forms

$$D(u) - N(u) = \left(1 - \frac{K_4}{K_1}\frac{k_{-1} + k_2[B]}{k_{-1}}\right)u_{EA} + K_1[AX]u^{\bullet}_{EB} + u_{EA}u^{\bullet}_{EB} \quad (59)$$

and

$$D(u) - N(u) = u_{EA} + K_1[AX]\left(1 - \frac{K_2}{K_3}\frac{k_{-3} + k_4[AX]}{k_{-3}}\right)u^{\bullet}_{EB} + u_{EA}u^{\bullet}_{EB} \qquad (60)$$

Expression (59) reaches maximum positive values if

$$K_1 >> K_4 \quad \text{and} \quad k_{-1} >> k_2[B] \qquad (61)$$

Similarly expression (60) reaches large positive values if

$$K_3 >> K_2 \quad \text{and} \quad k_{-3} >> k_4[A] \qquad (62)$$

Under either of these conditions expressions (59) and (60) reduce to

$$D(u) - N(u) = u_{EA} + K_1[AX]u^{\bullet}_{EB} + u_{EA}u^{\bullet}_{EB} \qquad (63)$$

that can possess positive values only. It then follows from this reasoning that conditions (61) and (62) that drift the system away from quasi-equilibrium tend to generate emergence of information in the enzyme network.

If, alternatively, conditions

$$K_4 >> K_1 \quad \text{and} \quad k_2[B] >> k_{-1} \qquad (64a)$$

$$K_2 >> K_3 \quad \text{and} \quad k_4[A] >> k_{-3} \qquad (64b)$$

are fulfilled, the expression of $D(u) - N(u)$ becomes equal to

$$D(u) - N(u) = u_{EA}\left(1 + u^{\bullet}_{EB} - \frac{K_4}{K_1}\frac{k_2[B]}{k_{-1}}\right) \qquad (65)$$

which is likely to adopt negative values for K_1 is much smaller than K_4 and $k_2[B]$ much larger than k_{-1}. Hence, under the conditions (64a and b), if the system drifts away from quasi-equilibrium it tends to increase its integrated character.

In the case of the situation depicted in Figure 7B the two substrates bind to the protein following a compulsory order. Under these conditions one has

$$p(AX) = \frac{K_1[AX] + K_1K_2[AX][B] + u_{EAX}}{1 + K_1[AX] + K_1K_2[AX][B] + u_{EAX} + u_E} \qquad (66a)$$

$$p(B) = p(AX, B) = \frac{K_1K_2[AX][B]}{1 + K_1[AX] + K_1K_2[AX][B] + u_{EAX} + u_E} \qquad (66b)$$

$$p\left(B\middle|AX\right) = \frac{K_2[B]}{1 + K_2[B] + u^{\bullet}_{EAX}} \qquad (66c)$$

with

$$u_E = \frac{k}{k_{-2}} + \frac{kK_2[B]}{k_{-1}} \qquad (67a)$$

$$u_{EAX} = \frac{k}{k_{-2}}K_1[AX] \qquad (67b)$$

$$u^{\bullet}_{EAX} = \frac{k}{k_{-2}} \qquad (67c)$$

It then follows that

$$\frac{p\left(B\middle|AX\right)}{p(B)} = \frac{1 + K_1[AX] + K_1K_2[AX][B] + u_{EAX} + u_E}{K_1[AX] + K_1K_2[AX][B] + K_1[AX]u^{\bullet}_{EAX}} \qquad (68)$$

If the system is close to equilibrium, viz. if the *u*'s are close to zero

$$\frac{p\left(B\middle|AX\right)}{p(B)} = \frac{1 + K_1[AX] + K_1K_2[AX][B]}{K_1[AX] + K_1K_2[AX][B]} > 1 \qquad (69)$$

and far from quasi-equilibrium one has

$$D(u) - N(u) = -\frac{k}{k_{-2}} - \frac{kK_2[B]}{k_{-1}} < 0 \qquad (70)$$

It then follows that the sequential ordered system can only be integrated.

In the case of a sequential random binding of substrates to the enzyme, emergence of information implies that the catalytic constant of the enzyme system increases, thus making the catalyst more efficient. The classical transition state theory [18, 19] implies that the expression of

any rate constant in a chemical system is related to the corresponding free energy of activation. Thus, for instance, the expression of the catalytic constant k is

$$k = \frac{k_B T}{h} \exp(\Delta G^{\neq} / RT) \qquad (71)$$

where k_B is the Boltzmann constant, T the absolute temperature, h the Planck constant, R the gas constant, T the absolute temperature and $\Delta G^{\neq}$ the free energy of activation. The smaller the $\Delta G^{\neq}$ value and the higher is the corresponding catalytic constant.

We have seen that a situation of emergence implies that

$K_1 > K_4$ and $k_{-1} > k_2[B]$

$K_3 > K_2$ and $k_{-3} > k_4[AX]$

These thermodynamic conditions are illustrated in Figure 8. The two diagrams imply that the ternary *EAXB* state is close to the transition state $EAXB^{\neq}$.

It is clear that the conditions that generate emergence in such a simple network increase the efficiency of the catalyst. In fact "emergence" means "emergence of catalytic power". Under these conditions, the ternary *EAXB* state becomes closer to the transition state, $E[A..X..B]^{\neq}$, of the catalytic reaction (Figure 8), thus increasing the catalytic constant of the process. To a large extent, emergence is due to the lack of equilibrium of the system. In fact, non-equilibrium is a source of catalytic power. It is because the $EAXB$ state has become close to the $E[A..X..B]^{\neq}$ state that the catalytic power of the whole system has increased (Figure 8).

The situation is symmetrical for an integrated system. In that case one has

$K_4 > K_1$ and $k_2[B] > k_{-1}$

$K_2 > K_3$ and $k_4[AX] > k_{-3}$

and the free energy level of the *EAXB* state is smaller than that of the initial state (Figure 9). Under these conditions the *EAXB* state has a low level of energy and is well "below" the $E[A..X..B]^{\neq}$ transition state (Figure 9). This implies that the catalytic constant is very small.

If the binding of the substrates to the enzyme follows a compulsory order the system is, of necessity, integrated. The difference $D(u) - N(u)$ is then negative. Depending on the final ground state *EAXB* has a low, or a high, energy level the system will be close, or far away, from equilibrium but, in any case, it will be integrated.

7. Multienzyme networks

As we have already mentioned, enzyme reactions are usually not isolated processes. They are associated as to form meta-networks that play a fundamental role in life processes.

7.1. Definition and main properties of meta-networks: micro- and macro-nodes

A network of biochemical reactions can be considered a network of networks, or a meta-network. Such a situation takes place because any enzyme-catalysed reaction is itself a network. The ensemble of these connected elementary networks can therefore be considered a network of networks, or a meta-network. Within this structure, there exists a time-hierarchy as the events that take place within an enzyme process are usually much faster than those involved in the connection between different enzyme reactions. As a matter of fact, such processes require that a product is released from an enzyme and diffuses to another enzyme that will be involved in its further transformation. As diffusion, or transfer processes, are usually much slower than the catalysed chemical transformation of a substance into another one, it is logical to consider as macro-nodes of the meta-network the "fast" chemical events involved in any of the catalysed chemical reactions and as links of the meta-network the "slow" transport processes of the chemical substances from macro-node to macro-node. Such meta-network should possess some important properties that are described below.

A meta-network is an ensemble of connected catalysed chemical reactions that constitute the macro-nodes of the system. Each macro-node is made up of several connected micro-nodes, viz. the various states of the corresponding catalysed chemical reaction. The links that associate different macro-nodes are the "slow" transport processes of chemical substances from enzyme to enzyme.

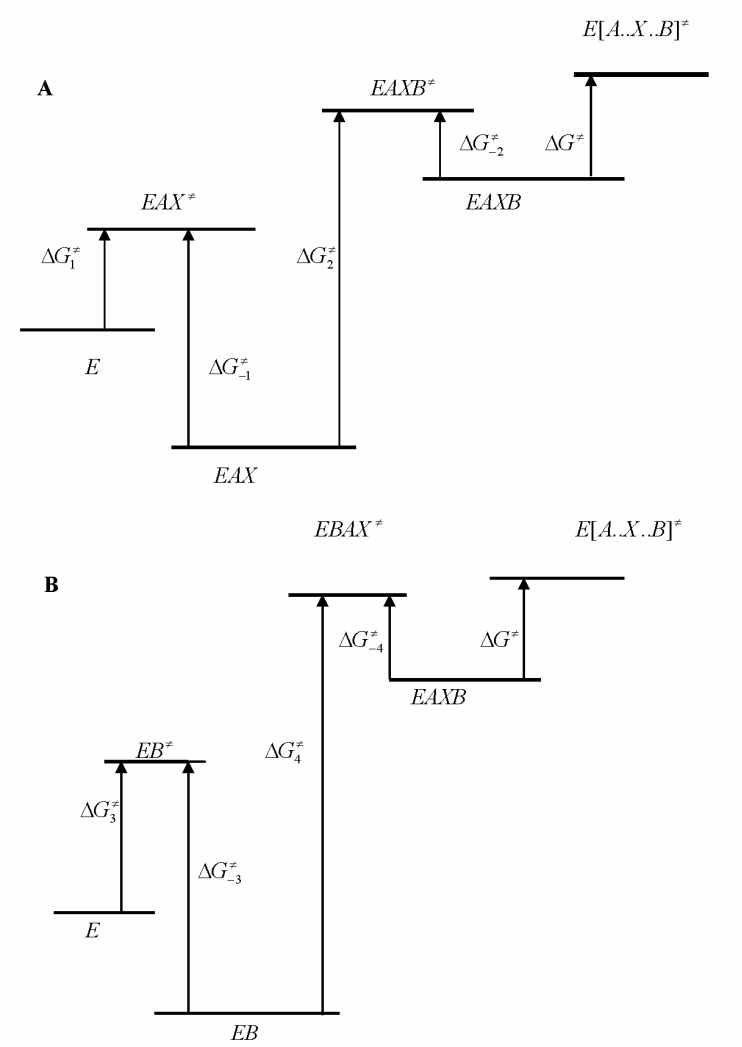


Figure 8. Free energy profiles for an enzyme process that fulfils the conditions of emergence.
A, Free energy profile of the process involving AX as the first substrate.
B, Free energy profile of the process involving B as the first substrate.
In either case the energy level of the ternary complex $EAXB$ becomes close to that of the transition state $E[A..X..B]^{\neq}$. The immediate consequence of this situation is a decrease of $\Delta G^{\cong}$ and an increase of the catalytic constant. Under these conditions, the enzyme system is far away from thermodynamic equilibrium.

The links associating macro-nodes are therefore directed processes that are slow relative to substrate binding, catalysis and product release of individual catalysed chemical reactions. Such events should therefore be modelled with directed graphs.

In a meta-network a given macro-node can possibly play either a major, or a minor, role. The importance of this role is expressed by the probability of occurrence of the corresponding macro-node, viz. the probability of occurrence of the corresponding enzyme reaction. It will be shown latter that this probability of occurrence is directly related to the degree of connection of this macro-node. It follows from this statement that a poorly connected macro-node cannot possess a high probability of occurrence. In fact, the probability of occurrence of a node is an expression of the overall network topology.

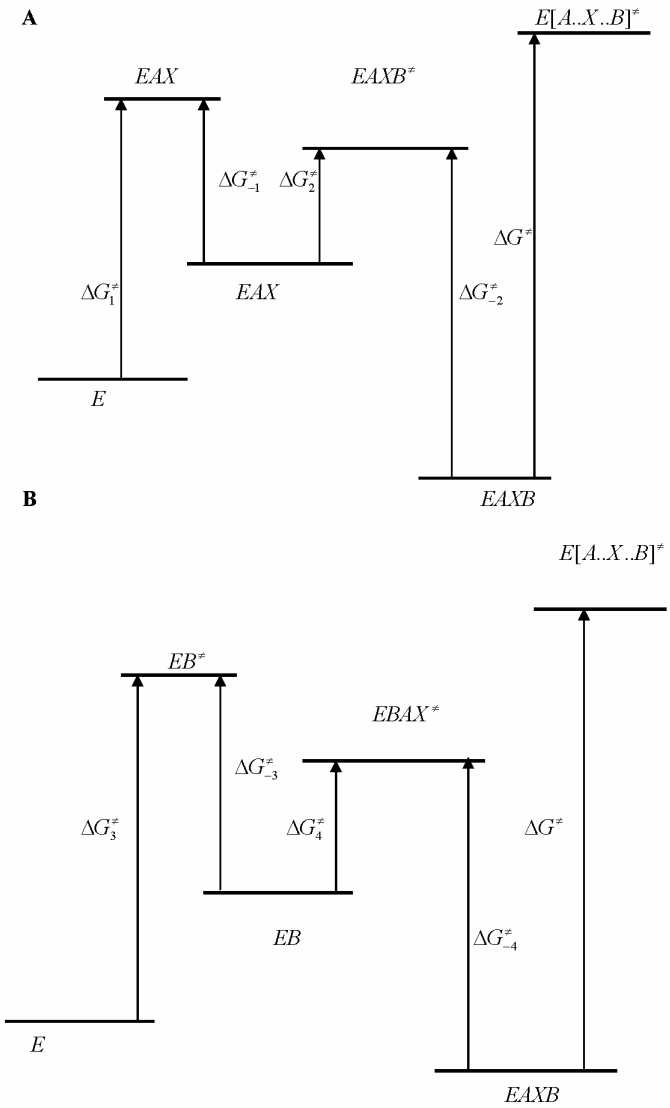


Figure 9. Free energy profiles for an enzyme process that fulfils the conditions of integration.
A, Free energy profile for the process involving AX as the first substrate.
B, Free energy profile for the process involving B as the first substrate.
In either case the energty level of the ternary complex $EAXB$ is far below that of the transition state complex $E[A..X..B]^{\neq}$. This implies a large increase of $\Delta G^{\neq}$ and a decrease of the catalytic constant. Under these conditions the enzyme system is close to thermodynamic equilibrium.

Meta-networks should be open thermodynamic systems with an input and an output of matter. As a consequence they should comply with the laws that govern such open systems and there cannot exist a "general science of networks" [20, 21] that could apply equally well to metabolic processes and to networks of social relationships, for instance, that do not have to meet the laws of thermodynamics of open systems.

7.2. Identity and information of metabolic networks

As we have seen previously, the identity of a functioning biological structure relies upon both its content of Aristotelian information and its local distribution. This statement can be applied to metabolic networks. Let us consider a macro-node of a metabolic network, the probability that enzyme E_i has bound enzyme substrate A_i is

$$p(A_i)_{Ei} = \frac{[E_i A_i] + [E_i A_i B_i]}{Y_i} \quad (72)$$

where Y_i , defined for the random binding of substrates A_i and B_i to enzyme E_i , is equal to

$$Y_i = [E_i] + [E_i A_i] + [E_i B_i] + [E_i A_i B_i] \quad (73)$$

Then the probability that the meta-network, *N*, has bound substrate A_i is

$$p(A_i)_N = \frac{[E_i A_i] + [E_i A_i B_i]}{Y_T} \quad (74)$$

where

$$Y_T = \sum_{i=1}^{n} Y_i \quad (75)$$

As

$$p(Y_i) = \frac{Y_i}{Y_T} \quad (76)$$

it follows that

$$p(A_i)_N = \frac{[E_i A_i] + [E_i A_i B_i]}{Y_i} \frac{Y_i}{Y_T} = p(A_i)_{Ei}\, p(Y_i) \quad (77)$$

where $p(Y_i)$ is the probability of occurrence of node Y_i in the meta-network. If substrate A_i binds to enzyme E_i , and only to this enzyme, the conditional probability that E_i binds A_i , given it has already bound B_i , is the same whether E_i is isolated or included in a meta-network. One has then

$$p\left(A_i \middle| B_i\right)_{Ei} = p\left(A_i \middle| B_i\right)_N \quad (78)$$

It is then possible to define functions $h(A_i)_N$ and $h\left(A_i \middle| B_i\right)_N$ as

$$h(A_i)_N = -\log\left[p(Y_i)\, p(A_i)_{Ei}\right] \quad (79a)$$

$$h\left(A_i \middle| B_i\right)_N = -\log p\left(A_i \middle| B_i\right)_N = -\log p(A_i \middle| B_i)_{Ei} \quad (79b)$$

It follows that the amount of information consumed, or released, at the level of node Y_i of the meta-network is

$$I(A_i : B_i)_N = h(A_i)_N - h\left(A_i \middle| B_i\right)_N \quad (80)$$

equal to

$$I(A_i : B_i)_N = I(A_i : B_i)_{Ei} - \log p(Y_i) \quad (81)$$

It follows from this expression that the Aristotelian information consumed, or produced, by a node of the meta-network is equal to the information of the same node considered in isolation affected by a term expressing the probability of occurrence of this node. The smaller this probability of occurrence and the larger is the importance of this term. This implies that if the number of nodes of the network is very large the probability of occurrence of Y_i , $p(Y_i)$, is very small and $-\log p(Y_i)$ is large and positive. Hence the fact that an enzyme is part of a network gives this enzyme additional information that expresses the topology of the global network.

These mathematical considerations allow one to define and express on quantitative grounds the identity of such a network. This identity could be defined by the following expression

$$Identity = \sum_i I(A_i : B_i)_N \quad (82)$$

and the mean Aristotelian information per node is

$$< I(A:B) >= \sum_i p(Y_i) I(A_i : B_i)_{Ei} - \sum_i p(Y_i) \log\, p(Y_i) \quad (83)$$

The first term of the right-hand side member of this expression represents the mean contribution of the micro-states. The second term, called topological information, expresses how the connections of the macro-states contribute to the mean information of the whole system. This contribution relies upon the topology of the network made up of these macro-states, each of them having a probability $p(Y_i)$ dependant upon the network topology.

If our aim is to illustrate how a meta-network is organized, the best way is to consider as an example the ideally simple situation of an open

system made up of four enzyme reactions (Figure 10). The transition constant, τ_i , between two macro-states is in fact the product of different contributions: $p(A_i, B_i)$, the probability of occurrence of the ternary complex, $E_i A_i B_i$, k_i the corresponding catalytic constant, and $k_i^{D\bullet}$ the apparent diffusion constant of this reaction intermediate from enzyme to enzyme. One has then

$$\tau_i = k_i k_i^{D\bullet} p(A_i, B_i) \tag{84}$$

with

$$k_i^{D\bullet} = k_i^D \frac{\partial [S_i]}{\partial x} \tag{85}$$

where k_i^D is the true diffusion constant of substrate S_i .

One can describe the simple system of Figure 10 by four differential equations plus a conservation equation. One has

$$\frac{dp(Y_1)}{dt} = \frac{v_i}{Y_T} + \tau_4 p(Y_4) - \tau_1 p(Y_1) \tag{86a}$$

$$\frac{dp(Y_2)}{dt} = \tau_1 p(Y_1) - \tau_2 p(Y_2) \tag{86b}$$

$$\frac{dp(Y_3)}{dt} = \tau_2 p(Y_2) - (\tau_3 + \tau_0) p(Y_3) \tag{86c}$$

$$\frac{dp(Y_4)}{dt} = \tau_3 p(Y_3) - \tau_4 p(Y_4) \tag{86d}$$

and

$$Y_T = Y_1 + Y_2 + Y_3 + Y_4 \tag{87a}$$

$$p(Y_1) + p(Y_2) + p(Y_3) + p(Y_4) = 1 \tag{87b}$$

$$\frac{v_i}{Y_T} = \tau_0 p(Y_3) \tag{87c}$$

Under steady state conditions the time derivative vanish and equations (86) can be rewritten as

$$\begin{bmatrix} \tau_1 & -\tau_2 & 0 & 0 \\ 0 & \tau_2 & -(\tau_0 + \tau_3) & 0 \\ 0 & 0 & \tau_3 & -\tau_4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} p(Y_1) \\ p(Y_2) \\ p(Y_3) \\ p(Y_4) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{88}$$

Solving this Cramer system allows to derive the expressions of the probabilities $p(Y_i)$. One finds

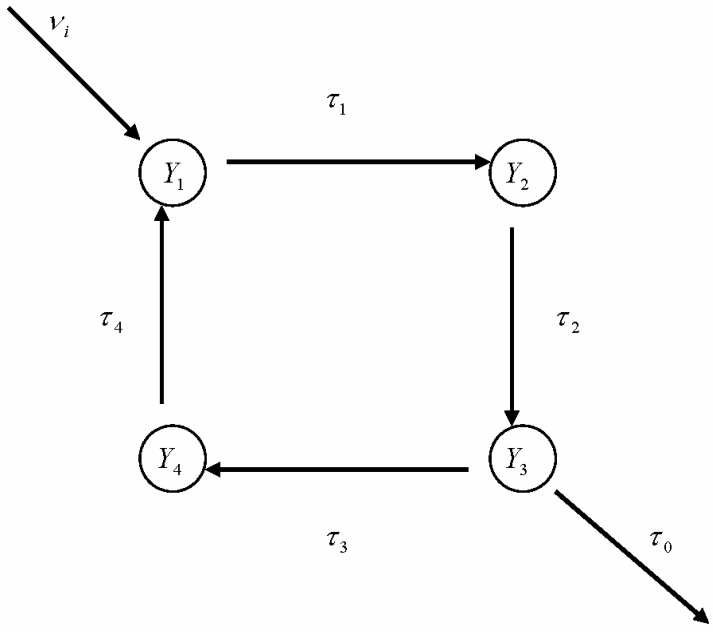


Figure 10. An ideal regular metabolic network.
The nodes Y_i are the individual enzyme reactions, the τ 's are the links between the nodes. The system is open, viz. it possesses an input and output of matter.

$$p(Y_1) = \frac{\tau_2\tau_3\tau_4 + \tau_0\tau_2\tau_4}{\tau_1\tau_2\tau_3 + \tau_1\tau_2\tau_4 + \tau_1\tau_3\tau_4 + \tau_0\tau_2\tau_4 + \tau_0\tau_1\tau_4 + \tau_2\tau_3\tau_4} \quad (89a)$$

$$p(Y_2) = \frac{\tau_1\tau_3\tau_4 + \tau_0\tau_1\tau_4}{\tau_1\tau_2\tau_3 + \tau_1\tau_2\tau_4 + \tau_1\tau_3\tau_4 + \tau_0\tau_2\tau_4 + \tau_0\tau_1\tau_4 + \tau_2\tau_3\tau_4} \quad (89b)$$

$$p(Y_3) = \frac{\tau_1\tau_2\tau_4}{\tau_1\tau_2\tau_3 + \tau_1\tau_2\tau_4 + \tau_1\tau_3\tau_4 + \tau_0\tau_2\tau_4 + \tau_0\tau_1\tau_4 + \tau_2\tau_3\tau_4} \quad (89c)$$

$$p(Y_4) = \frac{\tau_1\tau_2\tau_3}{\tau_1\tau_2\tau_3 + \tau_1\tau_2\tau_4 + \tau_1\tau_3\tau_4 + \tau_0\tau_2\tau_4 + \tau_0\tau_1\tau_4 + \tau_2\tau_3\tau_4} \quad (89d)$$

It appears immediately from these equations that the transition constant of the output τ_0 plays a part in the expression of the probability of occurrence of the nodes. In particular, the rate constant of output τ_0 contributes to increase the probability of occurrence of the nodes located up-stream the output of the network. It thus appears that it would be incorrect to consider the network as a closed, isolated system. In fact its interactions with the outside play a significant role in the probabilities of occurrence of the nodes. This conclusion can be generalized to any type of open biochemical network. If one considers, for instance, the fuzzy-organized network shown in Figure 11, one can derive the probabilities of occurrence of the nodes and one finds

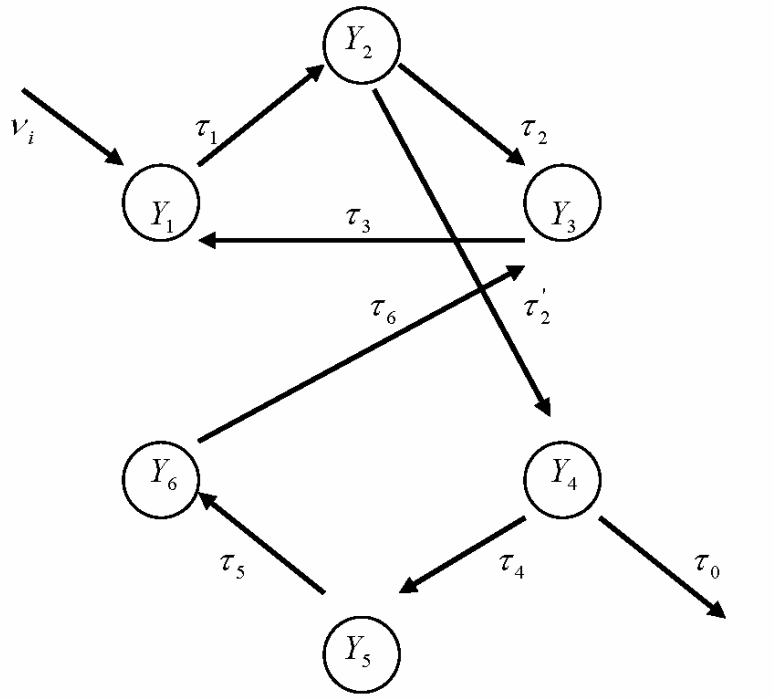


Figure 11. A fuzzy-organized metabolic network.
As previously, the nodes are enzyme reactions and the τ 's the links between the nodes.
As previously the system is open.

$$p(Y_1) = \frac{\tau_3\tau_5\tau_6(\tau_4+\tau_0)(\tau_2+\tau_2^{'})}{\Delta} \quad (90a)$$

$$p(Y_2) = \frac{\tau_1\tau_3\tau_5\tau_6(\tau_4+\tau_0)}{\Delta} \quad (90b)$$

$$p(Y_3) = \frac{\tau_1\tau_2\tau_5\tau_6(\tau_4+\tau_0)}{\Delta} + \frac{\tau_1\tau_2^{'}\tau_4\tau_5\tau_6}{\Delta} \quad (90c)$$

$$p(Y_4) = \frac{\tau_1\tau_2^{'}\tau_3\tau_5\tau_6}{\Delta} \quad (90d)$$

$$p(Y_5) = \frac{\tau_1\tau_2^{'}\tau_3\tau_4\tau_6}{\Delta} \quad (90e)$$

$$p(Y_6) = \frac{\tau_1\tau_2^{'}\tau_3\tau_4\tau_5}{\Delta} \quad (90f)$$

with

$$\Delta = \tau_5\tau_6(\tau_4+\tau_0)(\tau_1\tau_3+\tau_1\tau_2+\tau_2\tau_3+\tau_2^{'}\tau_3) + \tau_1\tau_2^{'}(\tau_3\tau_4\tau_5+\tau_3\tau_4\tau_6+\tau_3\tau_5\tau_6+\tau_4\tau_5\tau_6) \quad (91)$$

It then appears that the probability of occurrence of any node cannot be arbitrarily defined. It depends in fact upon the network topology considered as a whole. We shall see in the next Section that these conclusions imply a dramatic change of our views about the mathematical description of metabolic networks.

GENERAL DISCUSSION

The identity of present day living organisms relies upon a specific sequence of RNA bases or of DNA base pairs. The concept of identity, however, is far more general and versatile than the one offered by the primary structure of a macromolecule. It relies upon that of information, which was initially proposed by Aristotle [1] and considerably developed by Shannon [4] in such a way there is no a priori difficulty in assuming the existence of prebiotic systems that could be devoid of any nucleic acid [10]. Information is both what makes a material entity different from its neighbours and the ability we have to identify this entity. It is then evident that information should be related to the concept of probability. The smaller the probability of occurrence of an event and the larger is the information associated with the supervening of this event. As our aim is to explain, on physical bases, biological events by the concept of system, there is no a priori obligation to refer to the classical concept of macromolecule.

The concept of information is quite general in the sense it can be used to define the identity of a living system, to express how the system can communicate with others, how the formation of a system requires the consumption of information and conversely how a system can spontaneously generate its own information. The consumption of information is associated with the idea that the system forms an integrated whole that can communicate with others whereas spontaneous production of information by a system means emergence of novel properties related to the component sub-systems. In this case one can state that “the whole is more than the sum of its parts”. It is remarkable that simple biochemical systems such as lattices and networks of catalysed chemical reactions possess these properties that are present at a higher degree of complexity in living organisms.

Reversible communication between two sets of nodes of a protein lattice requires that the energy levels of these nodes be similar. This implies, for instance, that the joint probability $p(x_i, y_j)$ is about the same order of magnitude as $p(x_i)$ and $p(y_j)$. This is probably what is occurring during the reversible process of information transfer between DNA and mRNA. As a matter of fact, one can obtain synthesis of mRNA form DNA and conversely synthesis of DNA from mRNA. This is a straightforward consequence of physical chemistry. If, alternatively, the energy level of x_i is much higher than that of y_j and if the energy level of $x_i y_j$ is located midway between that of x_i and that of y_j then the communication process tends to be unidirectional, from x_i to y_j. More precisely, this means that $p(x_i)$ is larger than $p(y_j)$ and such that one has $p(x_i) > p(x_i, y_j) > p(y_j)$.

It appears that, in a protein lattice, one can observe the existence of either a communication process within the protein edifice, or the emergence of information. As these two processes cannot coexist at the same place, it is important to know the conditions that facilitate either a communication process, or the emergence of information. The switch between these two types of processes depends upon a physical principle called the principle of subadditivity that states that if

$$\sigma_\rho = \sum_i \sum_j p(x_i, y_j) - \sum_i p(x_i) \sum_j p(y_j) = 0$$

then the communication processes take place within the lattice in the absence of any process of emergence. Alternatively, if σ_ρ <0, then emergence of information spontaneously takes place. The first situation requires that all the nodes of the lattice are occupied by both ligands x and *y*. Conversely, emergence of information implies that some nodes of the lattice are not occupied, or occupied by only one type of ligand, either *x* or *y*.

Classical molecular biology is based on some kind of ontological reductionism, more precisely the idea that a system can be studied and understood through its decomposition into its elements. The concept of reduction, we are referring to, is the idea that predicates of a theory can be expressed in terms of another theory more general and embracing. This type of analysis can be applied to the reduction of a system to its component sub-systems with the belief that the independent studies of these sub-systems will be sufficient to understand and explain the global properties of the system itself. The possibility of deducing and explaining the global properties of a system from those of its component sub-systems is, however, problematic. Thus, for instance, the properties of a simple rectangular lattice *XY* can be deduced from those of *X* and *Y* only if the binding of *x* to its sites does not affect that of *y*, and conversely. One has to recognize that, in most cases, the properties of the whole cannot be deduced from the properties of the parts.

An important point is to understand the physical origin of emergence in a biochemical network. This point can be studied even with relatively simple systems such as enzyme reactions involving two substrates that bind randomly to an enzyme. Such apparently simple process can be close to equilibrium or far away from this equilibrium [22]. What the present study has shown is that, increasing the non-equilibrium character of the system, increases its tendency to generate information. Emergence for non-equilibrium systems means emergence, or enhancement, of catalytic activity for the ternary enzyme-substrate state becomes closer to the enzyme-transition state of the reaction, thus increasing the efficiency of catalysis. It thus appears that non-equilibrium is a source of catalytic power. One can speculate that during the process of evolution non-equilibrium may have played an important role, not only for explaining the emergence of novel properties, but also for explaining the emergence of new shapes and forms in living systems [23-25].

Networks we have been referring to in the present paper are not only isolated enzyme reactions but also networks of enzyme reactions. Such networks can be defined as meta-networks. This type of organization implies that any node is an enzyme reaction connected to other enzyme reactions according to a certain topology. Such a situation implies some kind of time hierarchy within a meta-network, viz. the events taking place within the various nodes, or macro-nodes, are much faster that the transfer processes from macro-node to macro-node. Moreover such meta-networks should be open structures with an input and output of matter. Then the probability of occurrence of a macro-node, viz. an enzyme reaction, is a certain mathematical function of the rates of connexion of the other macro-nodes, including the output rate from the meta-network. Metabolic networks are thus open dynamic structures. This mode of description of metabolic networks is quite different from the classical one in which the nodes are the metabolites connected as a closed structure [21]. In some of these networks, defined as scale-free, a small number of nodes called “hubs” are highly connected. This means that the “hubs” have a low probability of occurrence but are highly connected. In the perspective developed in

the present paper a meta-network should possess a number of properties that are listed below.

A metabolic meta-network is an open structure with an input and an output of matter. This view is at variance with the idea that there could exist a general science of networks that would include, for instance, networks of social relationships and metabolic networks [21]. The first type of network does not have to comply with the laws of thermodynamics of open systems whereas the second one cannot violate its principles and laws.

In a metabolic meta-network, the free energy change upon going from a macro-node to another one should be independent upon the pathway followed. This implies the existence of some constraints between the corresponding rate constants.

It is not possible to define the degree of connexion of a node independently of its probability of occurrence. These thermodynamic requirements have not been taken into account so far in current literature [21]. The very fact there could exist highly connected nodes with a low probability of occurrence is hardly compatible with the idea that the probability of occurrence of a node is directly related to its degree of connexion. This conclusion appears hardly compatible with the existence of "hubs", viz. of highly connected nodes of poor probability of occurrence. Put in other words, one should not forget that metabolic networks are open systems and should be considered as such.

REFERENCES

1. Aristotle. Métaphysique (French translation),. Livre Z, Librairie philosophique J. Vrin, Paris 2000.
2. Shannon, C. E. 1948, Bell System Technical Journal, 27, 379-423.
3. Shannon, C. E. 1948, Bell System Technical Journal, 27, 623-656.
4. Shannon, C. E. 1949, The Mathematical Theory of Communication, University of Illinois Press, Urbana.
5. Kullback, S. 1959, Information Theory and Statistics, Wiley and Sons, New York.
6. Cover, T. M. and Thomas, J. A. 1991, Elements of Information Theory, Wiley and Sons New York.
7. Callager, R. G. 1964, Information Theory, In Margenau H. and Murphy H. (Eds.) The Mathematics of Physics and Chemistry, Vol II, 190-248.
8. Ricard, J. 2006, Emergent Collective Properties, Networks and Information in Biology. Elsevier, Amsterdam, Boston.
9. Crick, F. H. C., Barnett, L., Brenner, S., and Watts-Tobin, R. J. 1961, Nature, 192, 1227-1232.
10. Crick, F. H. C. 1970, Nature, 227, 561-563.
11. Adami, C. 1998, Introduction to Artificial Life. Springer Telos.
12. Ricard, J. 2006, J. Non-Equilib. Thermodyn., 31, 103-152.
13. Hempel, C. G. 1970, Philosophy of Natural Science, Engelwood Cliffs, New Jersey, Prentice Hall.
14. Hull, D. L. 1988, Science as a Process. University of Chicago Press, Chicago.
15. Nagel, E. 1961, The Structure of Science. Harcourt, Brace and World, New York.
16. Robinson, J. D. 1986, Reduction, explanation and the quests of biological research. Philosophy of Science, 53, 333-353.
17. Laidler, K. J. 1958, The Chemical Kinetics of Enzyme Action. Oxford, Clarendon Press.
18. Laidler, K. J. 1969, Theories of Chemical Reactions Rates, McGraw-Hill, New York.
19. Castellan, G. W. 1973, Physical Chemistry (Second Edition), Addison-Wesley, Reading.
20. Barabasi, A. L. 2002, Linked: The New Science of Networks. Perseus Publishing, New York.
21. Albert, R. and Barabasi, A. L. 2002, Rev. Mod. Phys., 74, 47-97.
22. Ricard, J. 2010, C. R. Biologies, 333, 769-778.
23. Nicolis, G. and Prigogine, I. 1977, Self-Organization in Nonequilibrium Systems. John Wiley and Sons, New York.
24. Ricard, J. 1999, Biological Complexity and the Dynamics of Life Processes. Elsevier, Amsterdam, New York.
25. Kauffman, S. A. 1993, The Origins of Order. Oxford University Press, New York.