



Biological modelling/Biomodélisation

Systems biology and the origins of life? Part II. Are biochemical networks possible ancestors of living systems? Networks of catalysed chemical reactions: Non-equilibrium, self-organization and evolution

Jacques Ricard

Institut Jacques-Monod, 2, place Jussieu, 75251 Paris cedex 05, France

ARTICLE INFO

Article history:

Received 29 September 2010

Accepted 8 October 2010

Keywords:

Emergence

Enzyme reactions

Integration

Networks

Non-equilibrium dynamic structure

Self-organization

Thermodynamics

ABSTRACT

The present article discusses the possibility that catalysed chemical networks can evolve. Even simple enzyme-catalysed chemical reactions can display this property. The example studied is that of a two-substrate proteinoid, or enzyme, reaction displaying random binding of its substrates A and B . The fundamental property of such a system is to display either emergence or integration depending on the respective values of the probabilities that the enzyme has bound one of its substrate regardless it has bound the other substrate, or, specifically, after it has bound the other substrate. There is emergence of information if $p(A) > p(A|B)$ and $p(B) > p(B|A)$. Conversely, if $p(A) < p(A|B)$ and $p(B) < p(B|A)$ the system is integrated. The first condition is likely to occur if the system is far from quasi-equilibrium. Moreover, in such systems, emergence results in an increase of the energy level of the ternary EAB complex that becomes closer to the transition state of the reaction, thus leading to the enhancement of catalysis. Hence a drift from quasi-equilibrium is, to a large extent, responsible for the production of information and enhancement of catalysis. Non-equilibrium of these simple systems must be an important aspect that leads to both self-organization and evolutionary processes. These conclusions can be extended to networks of catalysed chemical reactions. Such networks are, in fact, networks of networks, viz. meta-networks. In this formal representation, nodes are chemical reactions catalysed by poorly specific proteinoids, and links can be identified to the transport of metabolites from proteinoid to proteinoid. The concepts of integration and emergence can be applied to such situations and can be used to define the identity of these networks and therefore their evolution. Defined as open non-equilibrium structures, such biochemical networks possess two remarkable properties: (1) the probability of occurrence of their nodes is dependant upon the input and output of matter in, and from, the system and (2) the probability of occurrence of the nodes is strictly linked to their degree of connection. The higher this degree of connection and the higher is their probability of occurrence. These conclusions are in clear disagreement with the static descriptions of the so-called scale-free metabolic networks.

© 2010 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

1. Introduction

In the accompanying article [1], we have outlined the fact that any living organism is able to reproduce, to

display some sensitivity to a time-arrow, to possess an identity and to evolve. In the previous paper, it was shown that three of these features were already met in apparently simple physical systems, namely encapsulated catalysed chemical reactions. In this Note we intend to discuss the last point, namely to decide whether or not a network of chemical reactions can display some sort of evolution.

E-mail address: jkrCARD@aol.com.

Indeed one cannot think of evolution in classical neo-Darwinian terms for these tentative prebiotic systems were not supposed to display any real mutation followed by its selection. Nevertheless, one can think of some sort of self-organization process followed by the selection of structures that possess a functional advantage. As in the previous Note [1], the mathematical developments of Section 2 can be omitted on first reading by those unfamiliar with these developments.

2. Prebiotic systems are non-equilibrium dynamic structures able to self-organize and to evolve

Biochemical networks of present day cells can tentatively be considered models of prebiotic systems. These collective structures are, in fact, dynamic open systems that receive molecules from the external milieu, transform these molecules and release some of the corresponding products outside the system. Hence such tentative prebiotic systems could have been networks of connected and encapsulated primordial enzyme reactions. It is no doubt of functional interest, in this perspective, to discuss some of the properties of these systems. An important property of networks of enzyme reactions is that any of these reactions is *per se* a network with its own nodes and links. Hence, exactly as metabolic networks of present day living cells, the tentative prebiotic systems could be considered networks of networks, or *meta-networks*. It is therefore of interest to study some of the dynamic properties of enzyme reactions in this perspective.

2.1. Enzyme reactions as simple biochemical networks

This Section can be omitted on a first reading.

From a phenomenological viewpoint, enzyme reactions as they take place today, are basically of two different types called the sequential and the substitution types. Most enzyme reactions usually involve two substrates, or sometimes three. In the so-called sequential process, the enzyme binds all the substrates before releasing the products. Alternatively, in the substitution mechanisms, the enzyme binds a substrate, releases the corresponding product, then binds the second substrate before releasing the last product. We are going, in the following, to discuss the sequential mechanism, which is, by far, the most important. In the case of an enzyme *E* that binds sequentially the two substrates *A* and *B* before releasing the products *P* and *Q*, the general situation is depicted in Fig. 1 and implies that the enzyme binds either *A* or *B* first and the other substrate, *B* or *A*, afterwards. In this perspective the enzyme can bind *A* first then *B*, or *B* first and then *A*. For that reason the process is called sequential random. A limiting situation is obtained when the free enzyme binds one substrate only, for instance *A* but not *B*. *B* can be bound to the enzyme only after *E* has bound *A*. Then the process is called ordered sequential, (Fig. 1). This model is a limiting case of the general one.

Let us consider the sequential random enzyme process described in Fig. 1A. One can derive the general expression of the probabilities that the enzyme has bound *A*, *B*, or both

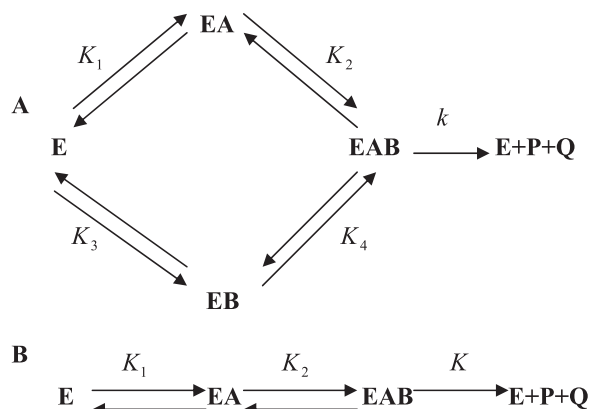


Fig. 1. **A** and **B**. Sequential binding of substrates and release of products to, and from, enzyme *E*. In **A**, it is assumed that the binding of substrates *A* and *B* to the enzyme takes place randomly. In **B** it is postulated that the binding of the substrates *EA* follows a compulsory order.

A and *B*. One finds the general expressions

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

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

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

In these expressions the *u*'s represent the shifts between the standard quasi-equilibrium and the actual non-equilibrium (steady-state) of the system. In Eqs. (1a)–(1c) the constants K_1, K_2, \dots are the ratios $k_1/k_{-1}, k_2/k_{-2}, \dots$, of rate constants. The shifts between the standard quasi-equilibrium and the actual non-equilibrium of the system are shown to be expressed as

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

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

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

One can notice that the *u*'s are of necessity positive and that they introduce non-linear terms in the expression of $p(A), p(B)$ and $p(A, B)$. Moreover, simple inspection of kinetic scheme of Fig. 1A suggests that the binding of *A* and *B* to the enzyme may interact, leading in turn to the increase, or decrease, of the corresponding information (2a)–(2c). One may express in quantitative terms the emergence, or integration, of information through the relation

$$i(A : B) = \log \frac{p(A|B)}{p(A)} = \log \frac{p(B|A)}{p(B)} \quad (3)$$

If $i(A:B)$ adopts negative values which implies that $p(A) > p(A|B)$ and $p(B) > p(B|A)$ the system is *emergent*. If it is the converse that is occurring, the system is *integrated* (3).

In order to check the validity of this idea one can derive, for instance, the expression of $p(A|B)$ and one finds

$$p(A|B) = \frac{K_4[A]}{1 + K_4[A] + u_{EB}^*} \quad (4)$$

where $u_{EB}^* = u_{EB}/K_3[B]$. The expression of u_{EB}^* is then

$$u_{EB}^* = \frac{kk_{-3}(k_{-1} + k_2[B])}{k_{-3}k_{-4}(k_{-1} + k_2[B]) + k_{-1}k_{-2}(k_{-3} + k_4[A])} \quad (5)$$

and the ratio $p(A|B)/p(A)$ assumes the form

$$\frac{p(A|B)}{p(A)} = \frac{K_4[A] + K_1K_4[A]^2 + K_3K_4[A][B] + K_1K_2K_4[A]^2[B] + N(u)}{K_1[A] + K_1K_4[A]^2 + K_3K_4[A][B] + K_1K_2K_4[A]^2[B] + D(u)} \quad (6)$$

In this expression $N(u)$ and $D(u)$ show the role played by the non-equilibrium in the emergence, or integration, of the system. If $N(u)=D(u)=0$ then the system is at thermodynamic equilibrium. Under these conditions of thermodynamic equilibrium, if $K_4 > K_1$, which implies from the first principle of thermodynamics that $K_2 > K_3$, the system is *integrated*. If alternatively, $K_1 > K_4$ and $K_3 > K_2$ the system is *emergent*. Emergence, however, can be enhanced if the system is away from thermodynamic equilibrium, that is if $N(u) \neq 0$ and $D(u) \neq 0$ (4)–(6). These non-equilibrium functions assume the form

$$N(u) = K_4[A]u_{EA} + K_4[A]u_{EB} + K_4[A]u_E \quad (7a)$$

$$D(u) = u_{EA} + K_4[A]u_{EA} + K_1[A]u_{EB}^* + K_1K_2[A][B]u_{EB}^* + u_{EA}u_{EB}^* \quad (7b)$$

The function u_E which appears in the expression of $N(u)$ can be rewritten in terms of either u_{EA} , u_{EB} or u_{EB}^* for one has

$$u_E = \frac{k_{-1} + k_2[B]}{k_1[A]}u_{EA} = \frac{k_{-3} + k_4[A]}{k_3[B]}u_{EB} = \frac{k_{-3} + k_4[A]}{k_{-3}}u_{EB}^* \quad (8)$$

Expression (7a), (7b) can be rewritten as

$$N(u) = K_4[A]u_{EA} + \frac{K_4}{K_1} \frac{k_{-1} + k_2[B]}{k_{-1}}u_{EA} + K_3K_4[A][B]u_{EB}^* \quad (9)$$

or as

$$N(u) = K_4[A]u_{EA} + K_3K_4[A][B]u_{EB}^* + K_4[A] \frac{k_{-3} + k_4[A]}{k_{-3}}u_{EB}^* \quad (10)$$

Non-equilibrium conditions, viz. conditions that maximize the values of the u 's, will increase the degree of emergence if $D(u) - N(u) > 0$. Alternatively, they will increase the degree of integration if $D(u) - N(u) < 0$. Hence the value of the difference $D(u) - N(u)$ is essential to determine how non-equilibrium alters integration, or emergence (8)–(10). The expression of $D(u) - N(u)$ can

be written under either of the following equivalent forms:

$$D(u) - N(u) = \left(1 - \frac{K_4}{K_1} \frac{k_{-1} + k_2[B]}{k_{-1}}\right)u_{EA} + K_1[A]u_{EB}^* + u_{EA}u_{EB}^* \quad (11a)$$

$$D(u) - N(u) = u_{EA} + K_1[A] \left(1 - \frac{K_2}{K_3} \frac{k_{-3} + k_4[A]}{k_{-3}}\right)u_{EB}^* + u_{EA}u_{EB}^* \quad (11b)$$

The difference $D(u) - N(u)$ will be positive and large if

$$K_1 \gg K_4 \text{ and } k_{-1} \gg k_2[B] \quad (12a)$$

$$K_3 \gg K_2 \text{ and } k_{-3} \gg k_4[A] \quad (12b)$$

Under these conditions the expressions of the u 's reduce to

$$u_{EA} = \frac{k}{k_{-2} + k_{-4}}K_1[A] \quad (13a)$$

$$u_{EB} = \frac{k}{k_{-2} + k_{-4}}K_3[B] \quad (13b)$$

$$u_{EB}^* = \frac{k}{k_{-2} + k_{-4}} \quad (13c)$$

and the expression of $D(u) - N(u)$ simplifies to

$$D(u) - N(u) = u_{EA} + K_1[A]u_{EB}^* + u_{EA}u_{EB}^* \quad (14)$$

It appears from this expression that departure of the system from quasi-equilibrium of necessity increases the value of the difference $D(u) - N(u)$ (14). The larger the u values and the greater is the difference $D(u) - N(u)$. It results from this conclusion that departure of the system from quasi-equilibrium of necessity increases its emergent character [2].

The situation is completely different if

$$K_4 \gg K_1 \text{ and } k_2[B] \gg k_{-1} \quad (15a)$$

$$K_2 \gg K_3 \text{ and } k_4[A] \gg k_{-3} \quad (15b)$$

Under these conditions the difference $D(u) - N(u)$ reduces to

$$D(u) - N(u) = \left(1 + u_{EB}^* - \frac{K_4}{K_1} \frac{k_2[B]}{k_{-1}}\right)u_{EA} \quad (16)$$

which is likely to adopt negative values owing to the fact that K_1 is very small relative to K_4 and k_{-1} very small relative to $k_2[B]$. Hence drifting the system away from quasi-equilibrium tends to increase its integrated character (16).

A similar reasoning can be applied to the kinetic scheme of Fig. 1B. The present situation is thus a limiting one in which the binding of B to the free enzyme is negligible. As previously, the probability of occurrence of the EA complex, $p(A)$, can be expressed in terms of equilibrium concentrations plus u terms that express the deviation of the system from pseudo-equilibrium conditions. One has then

$$p(A) = \frac{K_1[A] + K_1K_2[A][B] + u_{EA}}{1 + K_1[A] + K_1K_2[A][B] + u_{EA} + u_E} \quad (17a)$$

$$p(B) = p(A, B) = \frac{K_1 K_2 [A][B]}{1 + K_1 [A] + K_1 K_2 [A][B] + u_{EA} + u_E} \quad (17b)$$

$$p(B|A) = \frac{K_2 [B]}{1 + K_2 [B] + u_{EA}^*} \quad (17c)$$

with

$$u_E = \frac{k}{k_{-2}} + \frac{kK_2[B]}{k_{-1}} \quad (18a)$$

$$u_{EA} = \frac{k}{k_{-2}} K_1 [A] \quad (18b)$$

$$u_{EA}^* = \frac{k}{k_{-2}} \quad (18c)$$

One can now derive the expression of $p(B|A)/p(B)$. If this ratio is larger than one, the system is integrated (18a)–(18c). If alternatively the ratio is smaller than one the system is emergent. The expression of the ratio $p(B|A)/p(B)$ is

$$\frac{p(B|A)}{p(B)} = \frac{1 + K_1 [A] + K_1 K_2 [A][B] + u_{EA} + u_E}{K_1 [A] + K_1 K_2 [A][B] + K_1 [A] u_{EA}^*} \quad (19)$$

If the system is close to equilibrium, i.e. if the u 's are close to zero then

$$\frac{p(B|A)}{p(B)} = \frac{1 + K_1 [A] + K_1 K_2 [A][B]}{K_1 [A] + K_1 K_2 [A][B]} > 1 \quad (20)$$

Far away from quasi-equilibrium one has

$$D(u) - N(u) = -\frac{k}{k_{-2}} - \frac{kK_2[B]}{k_{-1}} < 0 \quad (21)$$

Hence it follows that a sequential ordered system can only be integrated and that a drift from quasi-equilibrium increases the extent of integration.

2.2. Thermodynamic significance of the concepts of integration and emergence

The thermodynamic and kinetic scheme of Fig. 1A allows one to analyse and understand the thermodynamic significance of the concepts of integration and emergence. Let us assume first that the binding of A to the enzyme does not interfere with that of B and conversely. In such a situation one should have

$$K_1 = K_4 \text{ and } K_2 = K_3 \quad (22)$$

In going from state E to state EAB one can follow two different pathways but the free energy difference between the initial and the final states is independent of the pathway (Fig. 2).

Upon going from a stable state to another stable state one has to overcome an energy barrier corresponding to a transition state (for example, EA^\ddagger). According to the classical transition state theory, the corresponding rate constant, for instance k_1 , depends upon the value of the corresponding free energy, ΔG_1^\ddagger , required to reach the

transition state EA^\ddagger . This classical relationship is [3]

$$k_1 = \frac{k_B T}{h} \exp(-\Delta G_1^\ddagger / RT) \quad (23)$$

where k_B , h , R and T are the Boltzmann constant, the Planck constant, the gas constant and the absolute temperature, respectively. It follows from relation (23) that the higher the free energy ΔG_1^\ddagger and the lower the value of the corresponding rate constant.

Now let us assume that the system is integrated, then one has

$$K_2 > K_3 \text{ and } k_2[B] > k_{-1} \quad (24a)$$

$$K_4 > K_1 \text{ and } k_4[A] > k_{-3} \quad (24b)$$

The free energy profiles for the binding of A to E and B to E will not be affected relative to the situation described in Fig. 2. However, the free energy profiles pertaining to the binding of B to EA and A to EB will be different from those of Fig. 1. Moreover the fact that relations (24a) and (24b) hold implies that the energy level of the ternary EAB state is lower than that of the same ternary state obtained in the absence of any interaction for the binding of A and B to the enzyme. It then follows that integration of the system results in a decrease of the energy level of the ternary EAB complex (Fig. 3).

In the case of emergence, the free energy profiles of Fig. 4 should be altered as to be antagonistic to that shown in Fig. 2. The peaks pertaining to the binding of A and B to the free enzyme E remain unchanged whereas the peaks pertaining to the binding of A and B to the EB and EA complexes should be modified as the relationships

$$K_3 > K_2 \text{ and } k_{-1} > k_2[B] \quad (25a)$$

$$K_1 > K_4 \text{ and } k_{-3} > k_4[A] \text{ apply.} \quad (25b)$$

The corresponding energy profiles are depicted in Figure 4. Relationships (25a) and (25b) imply that the ternary EAB complex possess a higher energy level than those of the binary EA and EB complexes. Hence, contrary to what is obtained for integration, emergence means that the interaction between the two binding processes results in emergence of energy upon going from the state of the free enzyme E up to the state of the ternary EAB complex.

A limiting situation is obtained for ordered sequential mechanisms where one of the pathways, for instance $E \leftrightarrow EB \leftrightarrow EAB$, is lacking. This is likely to occur if the free energy barrier that separates the EB complex from the free enzyme is too high. In that case, one should have $k_{-3} \gg k_3$, $K_3 \approx 0$ and $K_2 \gg K_3$. As free energy is a state function, it follows that $K_4 \gg K_1$. Hence, in agreement with the kinetic results presented above, the system cannot be emergent. It is integrated.

As previously mentioned, occurrence of emergence, or of integration, relies upon the sign and the value of the difference $D(u) - N(u)$. If this difference is positive and large, the system is emergent. Under these conditions, relations (12a) and (12b) apply and the u values are not negligible. This means that the system is away from quasi-equilibrium. Moreover, relationships (12a) and (12b) show that the energy level of the ternary EAB complex is higher

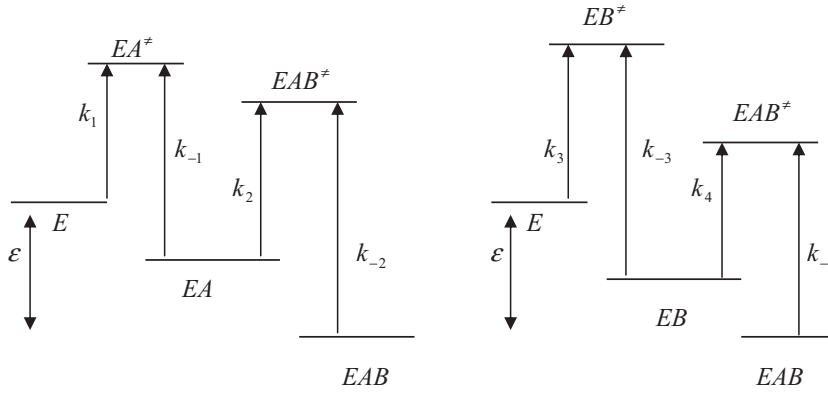


Fig. 2. Free energy profiles for the random and independent binding on a protein of two ligands, A and B. In this process, $k_1 = k_4$, $k_{-1} = k_{-4}$, $k_2 = k_3$ and $k_{-2} = k_{-3}$. The global free energy difference, ϵ , between the initial and the final states is the same for the two processes. For simplicity, the free energy differences between the ground and the transition states are depicted by the corresponding rate constants.

than that of the other stable states. This point has an important implication. It means that the ternary EAB state, owing to non-equilibrium conditions, comes close to the transition state EA^\ddagger . As expected, this means that the catalytic constant k is large. It follows that the enzyme becomes more efficient. *In other words emergence, for an enzyme reaction, means enhancement of catalytic activity.* To a large extent, emergence is due to the lack of thermodynamic equilibrium of the system.

A symmetrical situation takes place for an integrated enzyme network. Away from thermodynamic equilibrium, the difference $D(u) - N(u)$ assumes negative values. Relationships (15a) and (15b) apply and the free energy level of the ternary EAB complex is lower than that of the free enzyme state E . Such a situation has an immediate consequence. It implies that the catalytic

constant is small and therefore that the enzyme is a poor catalyst.

We are now in the position of defining the *identity* of an enzyme reaction, or of a network of enzyme reactions, through the sequence of Aristotelian information of the various steps and through the degree of emergence, or integration, generated by the interaction of ligands bound to proteins. Such sequence is no doubt characteristic of a reaction profile and, perhaps, as specific as a succession of base pairs in DNA. Moreover one could expect that such structures could display random alterations that could increase, or decrease, the performances of the system. “Favourable” changes should lead to networks that, after some time could become dominant relative to less “favoured” systems. In other words, some kind of selection could have been exerted upon these catalytic networks

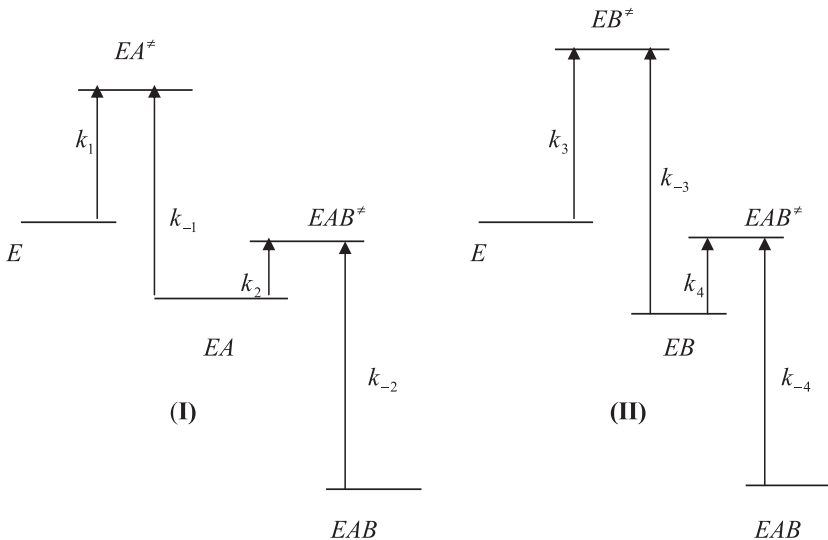


Fig. 3. Free energy profiles of ligand binding to a protein in the case of an integrated system. (I) Relationship $k_2[B] > k_{-1}$ is fulfilled. (II) Relationship $k_4[B] > k_{-3}$ is fulfilled. In (I) and (II) relationships $K_2 > K_3$ and $K_4 > K_1$ are fulfilled. Care must be taken that rate constants are related to the corresponding free energy barrier that separate the ground and the transition states by equation (23) of the main text. This means that the larger the free energy difference between the ground and the transition states and the smaller is the corresponding rate constant. Integration is characterized by a decrease of the final free energy level (EAB) relative to the initial one (E).

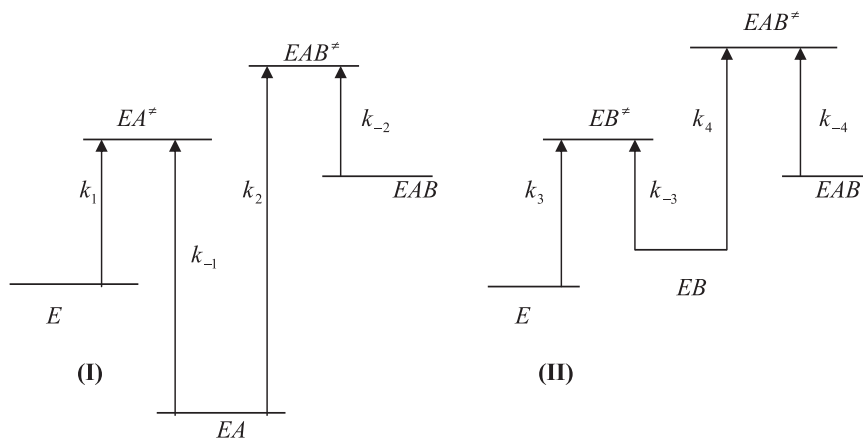


Fig. 4. Free energy profiles of ligand binding to a protein in the case of an emergent system. **(I)** Relationship $k_{-1} > k_2[B]$ is fulfilled. **(II)** Relationship $k_{-3} > k_4[A]$ is fulfilled. In **(I)** and **(II)** relationships $K_3 > K_2$ and $K_1 > K_4$ are fulfilled. Care must be taken that the larger the free energy difference between the ground and transition states and the smaller is the corresponding rate constants. Emergence is characterized by an increase of the final state (EAB) relative to the initial one (E).

leading, from sequential ordered systems, to sequential random ones.

3. Networks of catalysed chemical reactions

It has already been outlined that it is probably not unsound to consider that the first prebiotic systems could have been encapsulated networks of catalysed chemical reactions. In this perspective, it is logical to consider that metabolic networks of present day living systems should display some similarities with these tentative prebiotic networks and could be considered models of these networks. Such prebiotic networks should possess the following features that can also be found in present day metabolic networks:

- a network of catalysed chemical reactions is in fact a network of networks as any catalysed chemical reaction is *per se* an elementary network. For this reason such global network can be considered a meta-network. Hence this meta-network is a global structure of which the nodes are the catalysed chemical reactions and the links the transport processes of metabolites from a catalyst to another one;
- the links that associate nodes should be considered directed processes that are slow relative to substrate binding, catalysis and product release of individual chemical reactions. Put in other words, there should exist a time-hierarchy between the “fast” chemical processes and the “slow” diffusion of chemicals within the primitive cell from catalyst to catalyst. Such representation should be modelled with directed graphs;
- in a global process a given catalysed reaction could possibly play a major, or a minor, role. The importance of this role is expressed by the probability of occurrence of this node, viz. the probability of occurrence of this reaction. This probability of occurrence is, as we shall see, related to the degree of connections of this node. As a matter of fact, it is evident that a catalysed

reaction that is connected to many others should play a more important role than another one with few connections;

- it follows from the previous remarks that the probability of occurrence of a node, viz. a catalysed chemical reaction, should be related to its degree of connections. This means that the probability of occurrence of a node is an expression of the overall network topology;
- meta-networks should comply with the laws of thermodynamics. More precisely they should fulfil two requirements: first they should be considered open systems with an input and an output of matter; second they should not violate the first principle of thermodynamics;
- these two requirements have important implications. Both of them show there cannot exist a “general science of networks” [4] for networks of social relationships, for instance, do not have to comply with thermodynamics.

As outlined above, a biochemical network can be considered a network of networks, or a meta-network for each node is itself a network. This situation is likely to be associated with a time hierarchy in the overall system, the steps of substrate binding and release to and from the catalyst being much faster than the transfer of a metabolite from catalyst to catalyst. This condition is likely to be fulfilled in most cases in such a way it is sensible to consider each catalysed reaction a macro-state, or a node, of the system. Within each macro-state one can distinguish different micro-states which correspond to the free catalyst and the various complexes it may form with the substrates. The probability that catalyst E_i has bound, for instance, substrate A_i is

$$p(A_i)_{E_i} = \frac{[E_i A_i] + [E_i A_i B_i]}{Y_i} \quad (26)$$

where

$$Y_i = [E_i] + [E_i A_i] + [E_i B_i] + [E_i A_i B_i] \quad (27)$$

Similarly the probability that the network, considered as a whole, has bound substrate A_i is

$$p(A_i)_N = \frac{[E_i A_i] + [E_i A_i B_i]}{Y_T} \quad (28)$$

where

$$Y_T = Y_1 + Y_2 + Y_3 + \dots \quad (29)$$

It follows from relations (26) and (28) that the relationship between $p(A_i)_{Ei}$ and $p(A_i)_N$ is

$$p(A_i)_N = p(Y_i) p(A_i)_{Ei} \quad (30)$$

where $p(Y_i)$ is the probability of occurrence of node Y_i in the meta-network. If substrate A_i is specific for enzyme E_i , viz. if it binds to enzyme E_i only, then the conditional probability that E_i has bound A_i , given it has already bound B_i , is the same whether E_i is isolated or included in the network sequence. Then one has

$$p(A_i|B_i)_{Ei} = p(A_i|B_i)_N \quad (31)$$

and it becomes possible to define functions $h(A_i)_N$ and $h(A_i|B_i)_N$ as

$$h(A_i)_N = -\log[p(Y_i) p(A_i)_{Ei}] \quad (32)$$

$$h(A_i|B_i)_N = -\log p(A_i|B_i)_N \quad (33)$$

Hence the amount of Aristotelian information released, or consumed, at the level of node Y_i is

$$\begin{aligned} I(A_i : B_i)_N &= h(A_i)_N - h(A_i|B_i)_N \\ &= \log \frac{p(A_i|B_i)_{Ei}}{p(A_i)_{Ei}} - \log p(Y_i) \end{aligned} \quad (34)$$

This expression is equivalent to

$$I(A_i : B_i)_N = I(A_i : B_i)_{Ei} - \log p(Y_i) \quad (35)$$

It then appears that the Aristotelian information released, or consumed, at the level of a node is equal to the intrinsic Aristotelian information of that node, considered in isolation, plus a term that expresses how this node is integrated in the network. The second contribution is in fact a measure of the effect of context of this node on its information. Hence the mean Aristotelian information per node of the meta-network can be written as

$$\langle I(A_i : B_i) \rangle = \sum_i p(Y_i) I(A_i : B_i)_{Ei} - \sum_i p(Y_i) \log p(Y_i) \quad (36)$$

The first term of the right-hand side of this expression represents the mean contribution of the Aristotelian information of the micro-states in a set of unconnected macro-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 is dependent upon the topology of the network made up of these macro-states each of them having a probability $p(Y_i)$ dependent upon the network topology. Expression (36) provides a good criterion for the

identity of the meta-network exactly as a DNA sequence allows one to characterize a living organism (Fig. 5).

In order to illustrate how a meta-network is organized, it is of interest to consider, as an example, the simple ideal situation of a regular open system made up of four enzyme reactions (Fig. 5). As already outlined previously, the transition constant, τ_i , between two macro-states is defined from the product of the probability of occurrence, $p(A_i, B_i)$, of the ternary complex, $E_i A_i B_i$, the catalytic constant, k_i , and the apparent diffusion constant, k_i^{D*} . One has then [5]

$$\tau_i = k_i k_i^{D*} p(A_i, B_i) \quad (37)$$

In this expression, k_i^{D*} and $p(A_i, B_i)$ are mathematically defined as

$$k_i^{D*} = k_i^D \frac{\partial[S_i]}{\partial x} \quad (38)$$

and

$$\begin{aligned} p(A_i, B_i) &= \frac{K_{1i} K_{2i} [A_i] [B_i]}{1 + K_{1i} [A_i] + K_{3i} [B_i] + K_{1i} K_{2i} [A_i] [B_i] + u_{EA} + u_{EB} + u_E} \end{aligned} \quad (39)$$

In expression (38) k_i^D is the real diffusion constant and $\partial[S_i]/\partial x$ the concentration gradient of substance S_i . In the case of the network shown in Fig. 6, the differential equations that describe the dynamics of the macro-states are

$$\begin{aligned} \frac{dp(Y_1)}{dt} &= \frac{v_i}{Y_T} + \tau_4 p(Y_4) - \tau_1 p(Y_1) \\ \frac{dp(Y_2)}{dt} &= \tau_1 p(Y_1) - \tau_2 p(Y_2) \end{aligned} \quad (40)$$

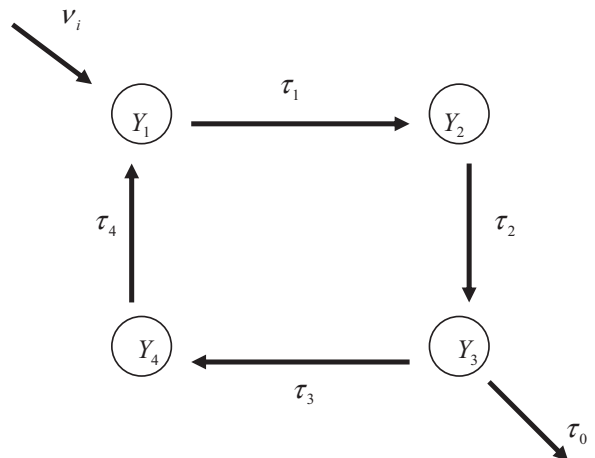


Fig. 5. An ideal regular open metabolic network. The nodes Y_1, \dots, Y_4 are the individual enzyme reactions. The τ 's are the apparent transition constants (see main text). The system is open and displays an input and output of matter.

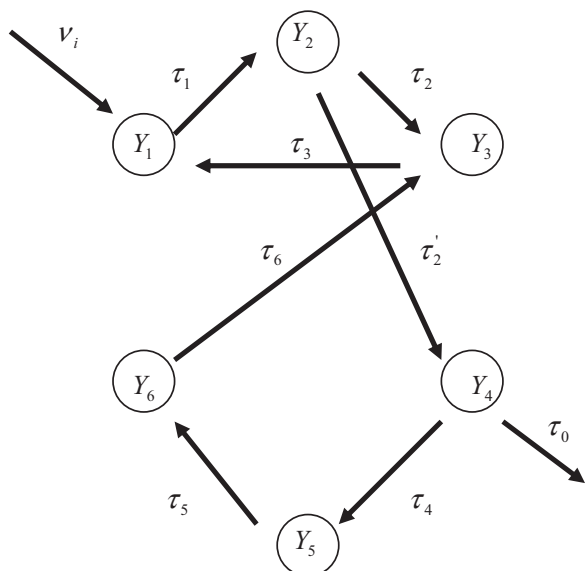


Fig. 6. A fuzzy-organized network of enzyme reactions. The probability of occurrence of the nodes (the enzyme reactions) depends upon their degree of connections.

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

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

Moreover a conservation equation should hold for the probabilities of occurrence of the macro-states, namely

$$p(Y_1) + p(Y_2) + p(Y_3) + p(Y_4) = 1 \tag{41}$$

Under steady state conditions, differential equations degenerate into algebraic expressions. Moreover, under these conditions, the input and output rates should be equal viz.

$$\frac{v_i}{Y_T} = \tau_0 p(Y_3) \tag{42}$$

Three steady state equations plus the conservation equation are sufficient to allow the derivation of the expression of the probabilities of occurrence of the macro-states. Substituting the third steady state equation by the conservation equation allows one to write

$$\begin{bmatrix} \tau_1 & -\tau_2 & 0 & 0 \\ 0 & \tau_2 & -(\tau_0 + \tau_3) & 0 \\ 0 & 0 & \tau_3 & -\tau_4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} p(Y_1) \\ p(Y_2) \\ p(Y_3) \\ p(Y_4) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{43}$$

Solving this Cramer system yields

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

$$p(Y_2) = \frac{\tau_1 \tau_3 \tau_4 + \tau_0 \tau_1 \tau_4}{\tau_1 \tau_2 \tau_3 + \tau_1 \tau_2 \tau_4 + \tau_1 \tau_3 \tau_4 + \tau_2 \tau_3 \tau_4 + \tau_0 \tau_2 \tau_4 + \tau_0 \tau_1 \tau_4} \tag{44}$$

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

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

It appears from these results that the transition constant of the output, τ_0 , plays a part in the expression of the probability of occurrence of the nodes and that they tend to increase the probability of the nodes located up-stream the output of the network. It would then be incorrect to consider the network a closed, isolated, system.

Let us consider now the fuzzy-organized system of Fig. 6 where the nodes are enzyme reactions. The reasoning followed previously allows one to derive the expressions of the probabilities of occurrence of the nodes. One finds

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

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

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

$$p(Y_4) = \frac{\tau_1 \tau_2' \tau_3 \tau_5 \tau_6}{\Delta} \tag{45}$$

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

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

with

$$\Delta = \tau_5 \tau_6 (\tau_4 + \tau_0) (\tau_1 \tau_3 + \tau_1 \tau_2 + \tau_2 \tau_3 + \tau_2' \tau_3) + \tau_1 \tau_2' (\tau_3 \tau_4 \tau_5 + \tau_3 \tau_4 \tau_6 + \tau_3 \tau_5 \tau_6 + \tau_4 \tau_5 \tau_6) \tag{46}$$

In these open systems the probability of occurrence of any node cannot be arbitrarily defined. It is in fact defined by the topology of the network considered as a collective entity. Any probability of occurrence of a node depends upon the number of pathways leading to that node. As we shall see in the next Section these conclusions imply a dramatic change of our views about the mode of description of metabolic networks subjected to the thermodynamics of open systems.

4. Discussion

Even simple enzyme reactions display some kind of behaviour that can be considered an evolutionary process.

Obviously, this type of evolution is not of neo-Darwinian type for there is, in a catalysed chemical reaction, no random mutation of the structure of a genetic material that is to be selected by the external milieu. The type of evolution we are referring to is based on some kind of self-organization of the system. Of particular interest in this perspective are the physical processes that lead to the *emergence*, or enhancement, of catalytic activity. This situation may occur if the enzyme displays random binding of its two substrates (see Fig. 1), viz. the enzyme should bind its substrates in either order. Emergence can occur even if the system is close to equilibrium provided that

$$K_1 > K_4 \text{ and } K_3 > K \quad (47)$$

The degree of emergence, however, cannot be very large under these sole conditions. As a matter of fact the extent of emergence can be considerably amplified if the system is far away from thermodynamic equilibrium. As shown in Section 2.1, this implies that the u values are not negligible and then simple relationships (47) should be replaced by

$$K_1 > K_4 \text{ and } k_{-3} > k_4[A] \quad (48a)$$

$$K_3 > K_2 \text{ and } k_{-1} > k_{-2}[B] \quad (48b)$$

To a large extent it is a drift from equilibrium that is responsible for the enhancement, or the emergence of the catalytic function.

A number of enzymes do not display random binding of their substrates, viz. the substrates bind to the protein following a compulsory order. In that case, and whatever the system is close to a steady state, or not, it can only display *integration*. The concept of emergence is important for it implies that owing in part to non-equilibrium conditions it has thermodynamic implications. It has been shown in particular in Section 2.2 that conditions (48a) and (48b) imply that the free energy level of the ternary enzyme-substrates *EAB* state should be higher than that of the free enzyme itself. The higher the energy level of the *EAB* state and the closer it should be to that of the *EAB*[‡] transition state. From a thermodynamic viewpoint this implies that the free energy barrier, which has to be overcome during catalysis, is reduced. It then appears that in the case of an enzyme process the term emergence means emergence, or enhancement, of a catalytic process. Hence it is tempting to consider that non-equilibrium self-organization of simple catalytic processes could have been the basis of their evolution.

In fact, the networks we have been referring to are not only isolated enzyme reactions but even more networks of enzyme reactions, which we have called meta-networks. In such a system, each node (also called macro-state) is in fact an enzyme reaction connected to other reactions according to a certain topology. In this model description, the links connecting the nodes correspond to the transfer of metabolites from catalyst to catalyst. The distinction we have made between macro- and micro-states is fully justified if there is a time-hierarchy in the system, which is likely to be the case, as the time-scale of the events taking place during the enzyme reaction are much faster than the

migration of metabolites from catalyst to catalyst. In this model description, any macro-state is characterized by several micro-states that represent the successive states of a catalyst interacting with its substrates. The information of a node of such meta-network is the sum of the individual informations of the corresponding micro-states plus a topological information that expresses how the global network topology affects the information of one of its nodes.

The model description of networks of biochemical reactions raises the general question of the best description of such networks. Recently a number of authors [6–14] have discussed the organization and functional properties of networks in a completely different perspective. This description, implicitly or explicitly, postulates that there exists a general mode of formal representation of networks whatever their nature. In this perspective, for instance, one could describe and analyse in the same manner networks of social relationships in a human society, for instance, and metabolic networks as well [6–14]. This is quite an optimistic claim for metabolic networks are subjected to thermodynamics whereas networks of social relationships are not. Sequencing genomes of living systems has allowed one to identify their genes, proteins and metabolites. Today databases offer this information. On this basis, it is possible to construct a graph of which the nodes are the metabolites and the links the chemical reactions that connect them. Using this mode of network description, Barabasi and his colleagues [11,13,14] have concluded that real networks, and in particular biochemical networks, are scale-free. This means that a large number of nodes are poorly connected whereas a small number of them, called “hubs”, appear highly connected.

This mode of network description is purely static and does not tell anything about the real probability of occurrence of the nodes. Moreover it does not take account of the fact that a network of catalysed chemical reactions should comply with the laws of thermodynamics. In this perspective, such system should possess the following properties:

- it should be considered an open system with an input and an output of matter;
- the free energy change upon going from a node to another one should be the same regardless the pathway thus leading to 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. It is then clear that these thermodynamic requirements are not taken into account in Barabasi’s description of metabolic networks [11]. The very fact there could exist highly connected nodes with a poor probability of occurrence is hardly compatible with the fact, already mentioned, that the probability of occurrence of a node is directly related to its degree of connexion. Probability of occurrence and degree of connexion cannot be considered independent variables. In this perspective, it might seem surprising that a metabolic network, which should comply with the laws of thermodynamics of open systems, could be described, in the same

manner as, for instance, the film actor collaboration network [8] which has little to do with thermodynamics...

Last, the author of the present articles would like to stress again that the scenario proposed for the origins of life is conjectural. However, these conjectures are physically sound and therefore worth being explored together with other possibilities [15].

References

- [1] J. Ricard, Systems biology and the origins of life – part I, *C. R. Biologies* 333 (2010) 761–768.
- [2] G. Nicolis, I. Prigogine, *Self-organization in Nonequilibrium Systems*, John Wiley and Sons, New York London, 1977.
- [3] G.W. Castellan, *Physical Chemistry*, Second Edition, Addison Westley, Amsterdam, London, 1973.
- [4] A.L. Barabasi, *Linked: the New Science of Networks*, Perseus Publishing, New York, 2002.
- [5] J. Ricard, *Emergent Collective Properties, Networks and Information in Biolog.* Elsevier, Amsterdam, New York, 2006.
- [6] R. Albert, A.L. Barabasi, Statistical mechanics of complex networks, *Rev. Mod. Phys.* 74 (2002) 47–97.
- [7] S. Milgram, The small world problem, *Psychol. Today* 2 (1967) 60–67.
- [8] M.E.J. Newman, The structure of scientific networks, *Proc. Natl. Acad. Sci. U S A* (2001) 404–409.
- [9] S. Redner, How popular is your paper? An empirical study of the citation distribution, *Eur. J. Phys* 134 (1988) 131–134.
- [10] D.J. Watts, S.H. Strogatz, Collective dynamics of “small world” networks, *Nature* 333 (1998) 440–442.
- [11] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, A.L. Barabasi, The large scale organization of metabolic networks, *Nature* 407 (2000) 651–654.
- [12] D.E. Fell, A. Wagner, The small world of metabolism, *Nature Biotech.* 18 (2000) 1121–1122.
- [13] A.L. Barabasi, R. Albert, Emergence of scaling in random networks, *Science* 286 (1999) 509–512.
- [14] A.L. Barabasi, R. Albert, Mean field theory for scale-free networks, *Physica A* 272 (1999) 173–197.
- [15] J.A. Doudna, J.W. Szostak, RNA catalysed synthesis of complementary strand RNA, *Nature* 339 (1979) 519–522.