# Emergence of Information and the Origins of Life a Tentative Physical Model 

Jacques Ricard*
Institut Jacques-Monod, CNRS, Universités Paris 6 et Paris, Paris cedex 5, France


#### Abstract

Life and living systems requires the existence of information. It is no doubt difficult to understand how this information could have spontaneously emerged in the first living cells. In fact spontaneous random association of molecules cannot give birth to information exactly as the random association of letters has a poor probability to generate a word. The idea developed in this paper is that the information required for the building up of the "primordial cell" does not originate from the association of molecules $x_{i}$, with $p\left(x_{i}\right)$, probability of occurrence, but from the association of $x_{i}$, molecules possessing conditional probabilities $p\left(x_{i} \mid y_{j}, y_{k}, \ldots\right)$. As will be shown in the paper, such a situation implies that molecules of conditional probabilities $p\left(x_{i} \mid y_{j}, y_{k}, \ldots\right)$ do not associate randomly but should follow an order defined by the nature of these conditional probabilities. As a conditional information, , is associated with the corresponding conditional probability of occurrence it follows that a sequence of conditional probabilities possesses a global information. Moreover the system can spontaneously generate an information if $h\left(x_{i} \mid y_{i}, y_{k}, \ldots\right)<h\left(x_{i}\right)$. Hence the global information could have been generated through the "coupling" of many of these individual mathematical uncertainties. The problem discussed in the present paper is precisely the converse of a transfer of information from a source to a destination in a communication channel, which has been extensively studied but the emergence of new information in a system.


Keywords: Origins of life; RNA; DNA; Gibbs-Boltzmann statistics; Molecular biology

Life, as we know it today, requires the existence of information. Many biological macromolecules, ribonucleic acids and proteins, are the expression of this information stored in deoxyribonucleic acids. Such a situation raises difficult problems when considered in the context of the origins of life on earth. As a matter of fact, living organisms require the existence of information and, conversely, in the context of the origin of life, information requires the existence of living systems. In order to avoid this vicious circle, one has to imagine an alternative hypothesis, for instance that the information required to explain the origin of the first living systems originates from purely physical-chemical reactions. The aim of the present paper is to discuss the hypothesis where biological information is generated through purely chemical reactions.

## Information Defined at an Elementary Level

If a material entity, say a molecule $x_{i}$, has a probability of occurrence $p\left(x_{i}\right)$, its information, $h\left(x_{i}\right)$, is defined as
$h\left(x_{i}\right)=-\log p\left(x_{i}\right)$
This expression reduces to

$$
\begin{equation*}
h\left(x_{i}\right) \approx 1-p\left(x_{1}\right) \tag{1b}
\end{equation*}
$$

if $p\left(x_{i}\right)$, is small. The lower this probability and the larger the information brought about by the occurrence of this event. If now two events, $x_{i}$ and $y_{j}$, are mutually independent, the probability of occurrence of this couple $x_{i}, y_{j}$ is then

$$
\begin{equation*}
p\left(x_{i}, y_{j}\right)=p\left(x_{i}\right) p\left(y_{j}\right) \tag{2}
\end{equation*}
$$

If now $x_{i}$ and $y_{j}$ are not independent but if, for instance, $x_{i}$ interacts with $y_{j}$, or conversely, one has then $p\left(x_{i}, y_{j}\right)=p\left(x_{i}\right) p\left(y_{j} \mid x_{i}\right)=p\left(y_{j}\right) p\left(x_{i} \mid y_{j}\right)$

In this expression, $p\left(x_{i} \mid y_{j}\right)$ and $p\left(y_{j} \mid x_{i}\right)$ are conditional probabilities.

The information corresponding to such a situation is then
$h\left(x_{i}, y_{j}\right)=h\left(x_{i}\right)+h\left(y_{j} \mid x_{i}\right)=h\left(y_{j}\right)+h\left(x_{i} \mid y_{j}\right)$
As in the classical communication theory, one can define a new function $i\left(x_{i} y_{j}\right)$, that could be called elementary mutual information. One has
$i\left(x_{i}: y_{j}\right)=h\left(x_{i}\right)+h\left(y_{j}\right)-h\left(x_{i}, y_{j}\right)$
Substituting $h\left(x_{i}: y_{j}\right)$, by its own expression (4) one finds
$i\left(x_{i}: y_{j}\right)=h\left(x_{i}\right)-h\left(x_{i} \mid y_{j}\right)=h\left(y_{j}\right)-h\left(y_{j} \mid x_{i}\right)$
Hence, this simple system will generate information if
$h\left(x_{i} \mid y_{j}\right)>h\left(x_{i}\right)$
and $h\left(y_{j} \mid x_{i}\right)>h\left(y_{j}\right)$
Under these conditions one has $i\left(x_{i}: y_{j}\right)<0$. If, alternatively,
$h\left(x_{i} \mid y_{j}\right)<h\left(x_{i}\right)$
and $h\left(y_{j} \mid x_{i}\right)<h\left(y_{j}\right)$
then one has $i\left(x_{i}: y_{j}\right)>0$. If $i\left(x_{i}: y_{j}\right)<0$, there is emergence of information

[^0]Citation: Ricard J (2014) Emergence of Information and the Origins of Life a Tentative Physical Model. Curr Synthetic Sys Biol 2: 109. doi: 10.4172/2332-0737.1000109
in the system. If conversely, $i\left(x_{i}: y_{j}\right)>0$ there is conduction of information within the same system. This is the last situation that prevails in classical communication theory [1-6].

## Communication Channels, Interaction between Molecules and Emergence of Information

Let us consider the elementary process of information transfer

$$
\begin{equation*}
p(x) \rightarrow p(x, y) \rightarrow p(y) \tag{11}
\end{equation*}
$$

it implies that information is transferred from $x$ to $y$ within the $x y$ "complex". The condition that allows the communication channel to be a perfect one is that
$p(x, y)=p(x) p(y)$
equivalent to $p(x) p(y)-p(x, y)=0$
Under these conditions there cannot exist emergence, but transfer, of information. This is the classical situation that has been studied in details by many authors [1-6]. Now let us assume that x and y interact in such a way that

$$
\begin{equation*}
p(y \mid x)<p(y) \tag{14}
\end{equation*}
$$

then expression (13) becomes

$$
\begin{equation*}
p(x) p(y)-p(x) p(y \mid x)=\delta \tag{15}
\end{equation*}
$$

With $\delta>0$, and does not correspond to the expression of a perfect communication channel. It describes what is occurring in a process of emergence of information. Hence it is implicitly postulated in current communication theory [1-7] that one has always
$h(y)>h(y \mid x)$
Even though this assumption is perfectly correct for the communication of a message in a communication channel, the situation can be completely different in the context of emergence of information through molecular interactions. This is precisely what is going to be discussed in this paper.

Relation (1) of a previous Section implies that
$p(x) \approx 1-h(x)$
$p(y) \approx 1-h(y)$
$p(y \mid x) \approx 1-h(y \mid x)$
Under these conditions one has then
$p(x) p(y)=\{1-h(x)\}\{1-h(y)\}=1-h(x)-h(y)+h(x) h(y)$
and the product of the right-hand side member (equation 15) is
$p(x) p(y \mid x)=\{1-h(x)\}\{1-h(y \mid x)\}=1-h(x)-h(y \mid x)+h(x) h(y \mid x)$
Expression (13) can be rewritten as
$-\{h(y)-h(y \mid x)\}+\{h(y)-h(y \mid x)\} h(x)=\delta$
or as $\{h(x)-1\}\{h(y)-h(y \mid x)\}=\delta$
As $h(x)<1$, it follows that if
$h(y)-h(y \mid x)<0$
the system generates its own information as molecules x and y interact. Under these conditions, as previously pointed out, $\delta>0$. We find here the situation already described in Section 1. Alternatively, if

$$
\begin{equation*}
h(y)-h(y \mid x)>0 \tag{23}
\end{equation*}
$$

molecules $x$ and $y$ do not tend to associate. Relationship (22) implies that the emergence of information originates from an interaction between and $y$.

This reasoning can be generalized to more complex situations, for instance

$$
\begin{equation*}
\sum_{i} p\left(x_{i}\right) \sum_{j} p\left(y_{j}\right)-\sum_{i} \sum_{j} p\left(x_{i}, y_{j}\right)=\Delta \tag{24}
\end{equation*}
$$

If and interact this equation can rewritten as
$\sum_{i} p(x) \sum_{j} p(y)-\sum_{i} p\left(x_{i}\right) \sum_{j} p\left(y_{j} \mid x_{i}\right)=\Delta$
Equation (21) can be generalized to different values of $x_{i}$ and $y_{j} \mid x_{i}$. One can see that the sign of $\Delta$ depends on the sign of

$$
\begin{equation*}
I(X: Y)=\sum_{j} h\left(y_{j}\right)-\sum_{j} \sum_{i} h\left(y_{j} \mid x_{i}\right) \tag{26}
\end{equation*}
$$

This expression is analogous to the so-called mutual information of the classical information theory [7] and is to be compared to expressions (22) and (23). It follows that if
$I(X: Y)>0$
the system consumes and propagates information, which is the situation expected to occur in a communication channel. Alternatively, if
$I(X: Y)<0$
the system generates information. This situation is due to the possible interaction occurring between $x_{i}$ and $y_{j}$.

## Physical Bases of Emergence

Classical Gibbs-Boltzmann statistics [8,9] allows one to understand the physical nature of the emergence process. Following the GibbsBoltzmann theory, the $y$ molecules are distributed among different energy levels according to the classical relationship
$p\left(y_{j}\right)=\exp \left\{-\left(E_{j}-E_{0}\right) / k_{B} T\right\}$
Where $p\left(y_{j}\right)$ is the probability of occurrence of $y$ molecules on the $j$ th energy level, $E j$ and E 0 the $j$ th and 0 th energy levels, $\mathrm{k}_{\mathrm{B}}$ and T the Boltzmann constant and the absolute temperature, respectively. It follows from this relationship that $h\left(y_{j}\right)$ has an important physical meaning that it does not possess in classical communication theory. One has
$h\left(y_{j}\right)=-\log p\left(y_{j}\right)=\frac{E_{j}-E_{0}}{k_{B} T}$
and it appears that the information of a molecule is directly related to its energy level. It follows from this reasoning that conditional information, $h\left(y_{j} \mid x_{i}\right)$, has a similar meaning i.e.
$h\left(y_{j} \mid x_{i}\right)=\frac{E_{j}-E_{j 0}}{k_{B} T}$
If the interaction between $x_{i}$ and $y_{j}$ is such that
$E_{j i}-E_{j 0}>E_{i}-E_{0}$
then $h\left(y_{j} \mid x_{i}\right)>h\left(y_{j}\right)$
and the interaction between $x_{i}$ and $y_{j}$ generates the "spontaneous" emergence of information (Figure 1)

## Organization of Biochemical Networks

The ideas developed in Section 2 offer a general picture of the organization of chemical and biochemical networks, whether they are primitive or not. In the present paper, we are interested to know how physical interactions between elements of a network generate information. If we consider for instance the joint probability of occurrence of such a network

$$
\begin{equation*}
p\left(x_{1}, x_{2}, x_{3}, x_{4}\right)=p\left(x_{1}\right) p\left(x_{2} \mid x_{1}\right) p\left(x_{3} \mid x_{2}, x_{1}\right) p\left(x_{4} \mid x_{3}, x_{2}, x_{1}\right) \tag{34}
\end{equation*}
$$

it appears to possess both kinds of information for it takes account of occurrence of both node probabilities and node interactions. In fact node interactions define the succession of nodes. It then follows that information generated through mutual interactions of nodes can be expressed from the difference

$$
\begin{equation*}
p\left(x_{1}\right) p\left(x_{2}\right) p\left(x_{3} \backslash x_{4}\right)-p\left(x_{1}\right) p\left(x_{2} \mid x_{1}\right) p\left(x_{3} \mid x_{2}, x_{1}\right) p\left(x_{4} \mid x_{3}, x_{2}, x_{1}\right) \tag{35}
\end{equation*}
$$

Information content expressed by this difference may be defined by the differences between information and conditional information $h\left(x_{2}\right)-h\left(x_{2} \mid x_{1}\right) h\left(x_{3}\right)-h\left(x_{3} \mid x_{2}, x_{1}\right)$ and $h\left(x_{4}\right)-h\left(x_{4} \mid x_{3}, x_{2}, x_{1}\right)$. One has then
$i\left(x_{1}: x_{2}: x_{3}: x_{4}\right)=h\left(x_{2}\right)-h\left(x_{2} \mid x_{1}\right)+h\left(x_{3}\right)-h\left(x_{3} \mid x_{2}, x_{1}\right)+h\left(x_{4}\right)-h\left(x_{4} \mid x_{3}, x_{2}, x_{1}\right)$ or
$i\left(x_{1}: x_{2}: x_{3}: x_{4}\right)=h\left(x_{2}\right)+h\left(x_{3}\right)+h\left(x_{4}\right)-h\left(x_{2} \mid x_{1}\right)-h\left(x_{3} \mid x_{2}, x_{1}\right)-h\left(x_{4} \mid x_{3}, x_{2}, x_{1}\right)$
One can also notice that if $\mathrm{x} 1, \mathrm{x} 2, \ldots$. interact one can possibly have
$h\left(x_{2}\right)<h\left(x_{2} \mid x_{1}\right)$
$h\left(x_{3}\right)<h\left(x_{3} \mid x_{2}, x_{1}\right)$
$h\left(x_{4}\right)<h\left(x_{4} \mid x_{3}, x_{2}, x_{1}\right)$


Figure 1: An interaction between $x_{i}$ and $y_{j}$ may result in an increase of $h\left(y_{j} \mid x_{i}\right)$ relative to $h\left(y_{j}\right)$. In the perspective of molecular interactions, $h\left(y_{i}\right)$ and $h\left(y_{j} \mid x_{i}\right)$ are equivalent to energies.

It follows from these relationships and from the differences that appear in equations (36) that the production of information by the system is associated with positive values of $i\left(x_{1}: x_{2}: x_{3}: x_{4}\right)$. This situation will be obtained if relationships (38) are met. One can conclude from these results that the organization of a network corresponding to the joint probability $\mathrm{p}\left(x_{1}, x_{2}, x_{3}, \ldots\right)$ relies upon two kinds of factors i.e. the number of nodes and the way these nodes are connected. This leads to define the information stored by networks as

$$
\begin{equation*}
i\left(x_{1}: x_{2}: x_{3}: x_{4}: \ldots\right)=\sum_{i} h\left(x_{i}\right)-\sum_{i} \sum_{j} h\left(x_{i} \mid x_{j}\right)=\sum_{i} h\left(x_{i}\right)-h\left(x_{1}, x_{2}, x_{3}, \ldots\right) \tag{39}
\end{equation*}
$$

The terms $h\left(x_{i}\right)$ are the contributions of the nodes to the global information of the system without taking into account the interactions that may exist between them. The term $\mathrm{h}\left(x_{1}, x_{2}, x_{3}, \ldots\right)$, or $\sum \sum h\left(x_{i} \mid y_{j}\right)$ , expresses how the nodes, and their interactions, contribute to this information. It is then evident that $\mathrm{i}\left(x_{1}: x_{2}: \ldots\right)$ possesses negative values if $x_{i}$ and $y_{j}$ interact and if relationships (38) apply.

Let us consider, for instance, the ideal processes shown in Figure 2. In Figure 2a it is assumed that the intermediate $C$ interacts with A and that D interacts with B . Both C and D play the part of inhibitors. In Figure $2 \mathrm{~b}, \mathrm{C}$ and E are inhibitors of the process. The probability of occurrence of the first network is
$p(A, B, C, D)=p(C) p(D) p(A \mid C) p(B \mid D)$
Similarly, the process shown in Figure 2 b has a probability of occurrence defined by
$p(A, B, C, D, E)=p(B) p(D) p(E) p(A \mid C) p(C \mid E)$
One can write from equation (40)
$\sum h\left(x_{i}\right)=h\left(x_{A}\right)+h\left(x_{B}\right)+h\left(x_{C}\right)+h\left(x_{D}\right)$
$\sum h\left(x_{i} \mid x_{j}, x_{k}, \ldots\right)=h\left(x_{A} \mid x_{C}\right)+h\left(x_{C}\right)+h\left(x_{B} \mid x_{D}\right)+h\left(x_{D}\right)$
and
$i\left(x_{1}: x_{2}:, \ldots\right)=h\left(x_{A}\right)-\left(x_{A} \mid x_{C}\right)+h\left(x_{B}\right)-h\left(x_{B} \mid x_{D}\right)$
It then appears that the system generates information if
$h\left(x_{A} \mid x_{C}\right)>h\left(x_{A}\right)$
$h\left(x_{B} \mid x_{D}\right)>h\left(x_{B}\right)$
A somewhat similar situation is expected to occur for equation (44) and Figure 3. One has

$$
\begin{align*}
& \sum h\left(x_{i}\right)=h\left(x_{A}\right)+h\left(x_{B}\right)+h\left(x_{C}\right)+h\left(x_{D}\right)+h\left(x_{E}\right)  \tag{45a}\\
& \sum h\left(x_{i} \mid x_{j}, x_{k}, \ldots\right)=h\left(x_{A} \mid x_{C}\right)+h\left(x_{B}\right)+h\left(x_{C} \mid x_{E}\right)+h\left(x_{D}\right)+h\left(x_{E}\right)  \tag{45b}\\
& \text { and } \\
& i\left(x_{1}: x_{2}:, \ldots\right)=h\left(x_{A}\right)-h\left(x_{A} \mid x_{C}\right)+h\left(x_{C}\right)-h\left(x_{C} \mid x_{E}\right) \tag{46}
\end{align*}
$$

As previously, emergence of information will be obtained if
$h\left(x_{A} \mid x_{C}\right)>h\left(x_{A}\right)$
$h\left(x_{C} \mid x_{E}\right)>h\left(x_{C}\right)$

## Information, Objective Reality of Equations and Closure of Biochemical Systems

Information that possibly exists in a macromolecule, made up of a succession of smaller molecules, relies upon the sequence of the

Citation: Ricard J (2014) Emergence of Information and the Origins of Life a Tentative Physical Model. Curr Synthetic Sys Biol 2: 109. doi: 10.4172/2332-0737.1000109
elements present in this macromolecule. This is precisely what is occurring with "classical" informational biomolecules such as DNA, RNA and proteins. It follows that one cannot expect the random association of different monomers to generate an informational polymer. Hence, one could wonder about the mechanism responsible for the emergence of an "order" in the succession of the elements of a sequence such as $\sum h\left(x_{i} \mid x_{i}, x_{k}, \ldots.\right)$. In fact the "position" of an element in the sequęnce relies upon some of the elements of this sequence. For instance the position of $x_{i}$ in the sequence depends upon that of $x j$ and $x k$. As will be shown below, it is the nature of the conditional information, or of the conditional probability, that contributes to define the sequence of the multimolecular system that is being synthesized.

Expressions such as $h\left(x_{1}, x_{2}, x_{3}, \ldots\right)$, above can possibly be considered a specific information that corresponds to a real, multimolecular, structure. For instance, the two sequences of events


Figure 2: Two models of feedback processes. a-C interacts with $A$ and $D$ interacts with B. b-C interacts with $A$ and $E$ interacts with $C$. In either situation (a) or (b), $p(A), p(B), \ldots$ represent probabilities of occurrence of reaction intermediates.


Figure 3: Emergence and storage of information in the two models of Figure
2. a-Two feedback processes exerted through $h\left(x_{A} \mid x_{C}\right)$ and $h\left(x_{B} \mid x_{C}\right)$. If $h\left(x_{A} \mid x_{C}\right)>h\left(x_{A}\right)$ and $h\left(x_{B} \mid x_{D}\right)>h\left(x_{B}\right)$ there is emergence of information. b-The two feedback processes are exerted through $h\left(x_{A} \mid x_{C}\right)$ and $h\left(x_{C} \mid x_{E}\right)$. If $h\left(x_{A} \mid x_{C}\right)>h\left(x_{A}\right)$ and $h\left(x_{C} \mid x_{E}\right)$ there is emergence of information.
presented in Figures 2 and 3 can be expressed by joint probabilities $p(A, B, C, D), p(A, B, C, D, E)$ and their corresponding information
$h\left(x_{A}, x_{B}, x_{C}, x_{D}\right)=h\left(x_{A} \mid x_{C}\right)+h\left(x_{B} \mid x_{D}\right)+h\left(x_{C}\right)+h\left(x_{D}\right)$
$h\left(x_{A}, x_{B}, x_{C}, x_{D}, x_{E}\right)=h\left(x_{A} \mid x_{C}\right)+h\left(x_{B}\right)+h\left(x_{C} \mid x_{E}\right)+h\left(x_{D}\right)+h\left(x_{E}\right)$
Expressions (48a) and (48b) describe specific sequences of events that define the global information of the corresponding system. Moreover, in this perspective, an expression such as $h\left(x_{A}, x_{B}, \ldots\right)$, can possibly define the information of a multimolecular system.

According to molecular biology RNA and DNA play the part of "templates" that control mutual interactions and respective binding of amino acids according to a certain order that define the information on the corresponding polypeptide chain. In fact one cannot explain the spontaneous random association of monomers to generate information. In the case of protein synthesis, information is transferred from DNA to proteins thanks to specific interactions between tRNAs and amino acids. The basic idea which is developed in the present paper is that information in primordial systems is generated thanks to the properties of conditional probabilities. Thus, for instance, if we consider a population of molecular entities $A, B, C, D, E, F$ defined by their conditional probabilities of occurrence $p(A \mid B, F) p(B \mid A, C) p(C \mid B, D) p(D \mid C, E) p(E \mid D, F) p(F \mid E, A)$, these entities will form a complex possessing a definite information for its elements possess a definite sequence, namely $A, B, C, D, E, F$. Moreover one can conclude that the system thus formed has a "circular" structure (Figures 4 and 5). This situation can be explained if the joint probability that defines the system $p(A, B, C, D, E, F)$ is equal to the product

$$
\begin{equation*}
p(A, B, C, D, E, F)=p(A \mid B, F) p(B \mid A, C) p(C \mid B, D) p(D \mid C, E) p(E \mid D, F) p(F \mid E, A) \tag{49}
\end{equation*}
$$

It is the very nature of the molecular interactions involved in these conditional probabilities that defines the sequence $A, B, C, D, E, F$.

The corresponding information of the system can then be expressed

$$
\begin{align*}
& \text { as } \\
& \begin{array}{r}
h\left(x_{A}, x_{B}, x_{C}, x_{D}, x_{E}, x_{F}\right)=h\left(x_{A} \mid x_{B}, x_{F}\right)+h\left(x_{B} \mid x_{A}, x_{C}\right)+h\left(x_{C} \mid x_{B}, x_{D}\right)+h\left(x_{D} \mid x_{C}, x_{E}\right) \\
+ \\
+h\left(x_{E} \mid x_{D}, x_{F}\right)+h\left(x_{F} \mid x_{E}, x_{A}\right)
\end{array} \tag{50}
\end{align*}
$$

and it is clear that both expressions (49) and (50) represent the global information of a sequence of elements, namely conditional probabilities and conditional informations, respectively.

This reasoning can be applied to the information of the whole system
$i\left(x_{A}: x_{B}: x_{C}: x_{D}: x_{E}: x_{F}\right)=\sum_{i} h\left(x_{i}\right)-h\left(x_{A}, x_{B}, x_{C}, x_{D}, x_{E}, x_{F}\right)$
where
where
$\sum h\left(x_{i}\right)=h\left(x_{A}\right)+h\left(x_{B}\right)+h\left(x_{C}\right)+h\left(x_{D}\right)+h\left(x_{E}\right)+h\left(x_{F}\right)$
It follows that
$i\left(x_{A}: x_{B}: x_{C}: x_{D}: x_{E}: x_{F}\right)=h\left(x_{A}\right)-h\left(x_{A} \mid x_{B}, x_{F}\right)+h\left(x_{B}\right)-h\left(x_{B} \mid x_{A}, x_{C}\right)$
$+h\left(x_{C}\right)-h\left(x_{C} \mid x_{B}, x_{D}\right)+h\left(x_{D}\right)-h\left(x_{D} \mid x_{C}, x_{E}\right)+h\left(x_{E}\right)-h\left(x_{E} \mid x_{D}, x_{F}\right)+h\left(x_{F}\right)-h\left(x_{F} \mid x_{E}, x_{A}\right)$
One can realize that the expressions of $h\left(x_{A}, x_{B}, x_{C}, x_{D}, x_{E}, x_{F}\right)$ and $i\left(x_{A}: x_{B}: x_{C}: x_{D}: x_{E}: x_{F}\right)$ have a definite geometric representation (Figures 4 and 5). This representation displays a "closure" as requested in living systems that display an"inside" and an "outside". It is remarkable that $h\left(x_{A}, x_{B}, x_{C}, x_{D}, x_{E}, x_{F}\right)$ and $i\left(x_{A}: x_{B}: x_{C}: x_{D}: x_{E}: x_{F}\right)$, in Figures 4 and 5, are a geometric representation of the distribution of information within a multimolecular complex (Figure 4) and how the
interactions within the complex affect the informations of its elements (Figure 5). Moreover, this system could attract and bind other "potential biomolecules" in a definite order thus transferring its information to this new macromolecule. Hence it is tempting to assume that nonbiological systems can spontaneously synthesize molecular edifices similar to those shown in Figures 4 and 5. One could speculate that both of them could possess the information required for the building up of the first "living" systems.

## Discussion

The classical solution of the problem of the origins of life raises a number of difficult questions [10-14]. These difficulties are present in many theories, in particular in the theory of the existence of a primitive RNA world in which the first living organisms on earth were solely made up of RNAs. One of the difficulties of this classical vision is that, according to the literature, no one so far has been able to observe the replication of RNAs in the absence of enzymes.

Any process of emergence of a living system implies that


Figure 4: The system $h\left(x_{A}, x_{B}, x_{C}, x_{D}, x_{E}\right)$ defines the spatial localization of conditional information as a closed geometric structure. This situation is obtained if any element of the system is dependent upon the other elements as shown (52). Such a situation is expected to occur in the case of the probability of occurrence of a system defined by equation (51). Information is generated through the interaction between the elements of the system.


Figure 5: The system $i\left(x_{A}: x_{B}: x_{C}: x_{D}: x_{E}\right.$ ) defines the spatial localization of both information and conditional information. As previously, this situation is obtained if any element of the system is dependent upon two other elements in the system.
the information carried up by a macromolecule, or a pluri-molecular system, is already present before the first living organism appears. Today information, under the form of DNA molecules, has an obvious biological origin in such a way that we need life to define information of biological systems and we need information to define life. There is some kind of circular statement, in the context of the problem of the origin of life, in defining living systems through their information, and in defining information through living systems.

There have been several attempts at explaining emergence of life from purely chemical events. It is highly probable however that the building up of a "living" system requires the existence of information. We are then faced with the idea that the first living systems were the products of the expression of some information that was not biological in its essence.

There exists today an information theory which is in fact the theory of the communication of a message. The communication of a message is based on the equation
$I(X: Y)=H(X)-H(X \mid Y)$
which is in fact a generalization of equation (26) of the present paper. In expression (54) above, $I(X: Y)$ is the co-called mutual information transferred in a communication channel from a source to a destination. It is then clear that, in this case, $I(X: Y)$ has to be positive, otherwise no information could have ever been transferred from a source to a destination. This implies in turn that $H(X)$ is of necessity larger than $H(X \mid Y)$. If $\mathrm{I}(X: Y)$ were negative this would imply that the system would generate its own information. This is indeed impossible in the case of a communication channel but this is quite possible in the case of a biochemical system involving interacting molecules.

In order for such a system to generate information one has to assume there exists some interaction between $x_{i}$ and $y_{i}$ in such a way that
$\sum_{i} h\left(x_{i} \mid y_{j}\right)>\sum_{i} h\left(x_{i}\right)$
Under these ${ }^{i}$ conditions, the system, which does not play the part of a communication channel, generates its own information under the form of a macromolecular complex that can associate in a definite order other molecules, thus giving birth to a primitive "living" organism. This order is, in fact, defined and imposed by the very nature of the elements of conditional information that constitute the system. It is because the position of $x_{i}$ is defined by the values of $y_{i}$, which are precisely associated with this $x_{i}$ unit, that the sequence of all the $x_{i}$ possesses an information. As shown above, if the conditional events associate, this association could not be random but of necessity
$(A|B, F \gamma B| A, C \chi C \mid B, D)(D|C, E\rangle E|D, F \gamma F| E, A)$ It is, in fact, the conditional character of these events that imposes their sequence as well as the possible "closed character" of the sequence. In the frame of the present theory, $\mathrm{i}\left(x_{i}: y_{j}\right)$ should adopt negative values. It is then not surprising that physical interactions between $x_{i}$ could generate an information.

According to these theoretical considerations, novel interactions between $x_{i}$ and $y_{i}$ take place and the system becomes more and more complex. This increase of complexity, which has a physical origin, follows a number of requirements defined by a mathematical model. It is clear, in such a system, that both the emergence of information and the spatial organization of the system are predicted by equations of purely physical phenomena.

Citation: Ricard J (2014) Emergence of Information and the Origins of Life a Tentative Physical Model. Curr Synthetic Sys Biol 2: 109. doi: 10.4172/2332-0737.1000109

Page 6 of 6

The morphological shape of the system is also predicted by the nature of the equation that describes the process of emergence. In such systems, both the emergence of shape, of a closure of the system for instance, and information are controlled by purely physical events.

## Reference

1. Shannon CE (1948) A mathematical theory of communication. Bell System Technical Journal 27: 379-423.
2. Shannon CE (1949) The Mathematical Theory of Communication. University of Illinois Press, Urbana.
3. Kullback S (1959) Information Theory and Statistics. Wiley and Sons, New York.
4. Cover JM, Thomas JA (1991) Elements of Information Theory. Wiley and Sons, New York.
5. Callager, RG (1964) Information Theory. In: Margenau H and Murphy H (Eds) The Mathematics of Physics and Chemistry. Van Nostrand, New York.
6. Adami C (1998) Introduction to artificial life. Springer-Verlag, New York.
7. Ricard J (2006) Emergent collective properties, Networks and Information in Biology. Elsevier.
8. Setlow RB, Pollard EC (1962) Molecular Biophysics. Addison-Westley.
9. Hill TL (1987) Statistical Mechanics. Principles and applications. Dover Publications Inc. New York.
10. Cech TR (1986) A model for the RNA-catalyzed replication of RNA. Proc Natl Acad Sci U S A 83: 4360-4363.
11. Cech TR, Herschlag D, Piccirilli JA, Pyle AM (1992) RNA catalysis by a group I ribozyme. Developing a model for transition state stabilization. J Biol Chem 267: 17479-17482.
12. Doudna JA, Szostak JW (1989) RNA-catalysed synthesis of complementarystrand RNA. Nature 339: 519-522.
13. Fox SW, Dose K (1977) Molecular evolution and the Origin of Life. Marcel Dekker, New York.
14. Raeymaekers $L$ (2002) Dynamics of Boolean networks controlled by biologically meaningful functions. J Theor Biol 218: 331-341.

[^0]:    *Corresponding author: Jacques Ricard, Institut Jacques-Monod, CNRS, Universités Paris 6 et Paris 7, 2, Place Jussieu, 75251 Paris cedex 5, France, Tel : 0143909650 ; E-mail: jkricard@aol.com

    Received September 12, 2013; Accepted February 11, 2014; Published February 15, 2014
    Citation: Ricard J (2014) Emergence of Information and the Origins of Life a Tentative Physical Model. Curr Synthetic Sys Biol 2: 109. doi: 10.4172/23320737.1000109

    Copyright: © 2013 Ricard J. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

