The Scientific Origin of Life

Considerations on the Evolution of Information, Leading to an Alternative Proposal for Explaining the Origin of the Cell, a Semantically Closed System

MARIO VANEECHOUTTE[a]

Department of Clinical Chemistry, Microbiology & Immunology, University Hospital, B 9000 Gent, Belgium

ABSTRACT: We hypothesize that the origin of life, that is, the origin of the first cell, cannot be explained by natural selection among self-replicating molecules, as is done by the RNA-world hypothesis. To circumvent the chicken and egg problem associated with semantic closure of the cell—no replication of information molecules (nucleotide strands) without functional enzymes, no functional enzymes without encoding in information molecules—a prebiotic evolutionary process is proposed that, from the informational *point of view*, must somehow have resembled the current scientific process. The cell was the outcome of interactions of a complex premetabolic community, with information molecules that were devoid of self-replicative properties. In a comparable manner, scientific progress is possible, essentially because of interaction between a complex cultural society and permanent information carriers like printed matter. This may eventually lead to self-replicating technology in which semantic closure occurs anew. Explaining the origin of life as a scientific process might provide a unifying theory for the evolution of information, wherebye at two moments symbolization/encoding of interactions into permanent information occurred: at one moment that of chemical interaction and at another moment that of animal behavior interaction. In one event this encoding led to autonomously duplicating chemistry (the cell), an event that possibly may be one of the outcomes of current scientific progress.

INTRODUCTION

A *cell* is defined here as a semantically closed system (or a system closed to efficient cause[1]). Thus far, it is the only such system that exists on Earth and it is generally agreed that it originated only once. By a *semantically closed system* we mean that the system contains all the information and functionality to duplicate itself. Put differently, semantic closure means here that the information (nucleotide strands) required to produce processors (enzymes) that are able to duplicate this information, depends on the functionality of these processors, that in turn depends on the content of the information that these processors duplicate.

[a]Address for correspondence: Department of Clinical Chemistry, Microbiology & Immunology, Blok A, University Hospital, De Pintelaan 185, B9000 Gent, Belgium. Voice: +32 9 240 36 92; fax: +32 9 240 36 59.

Mario.Vaneechoutte@rug.ac.be

The hypothesis espoused here states that it is virtually impossible that the highly complicated system *cell* developed gradually around simple self-replicating molecules (RNA-hypercycles or autocatalytic peptide networks) by means of natural selection; as is proposed by, for example, the RNA-world hypothesis.[2–7] I propose that the cell was the result of prebiotic events that from an informational *point of view* are comparable to the current scientific process. Science is possible because of the interaction between a behavioral community (culture) and permanent information carriers having unlimited information content (printed texts). Comparably, I argue that it was the interaction between a chemical metabolic community (prebiotic chemistry) with newly developed nucleotide strands as carriers having unlimited information content that led to the cell (the origin of life). Just as printed texts do not self-replicate, but rely on the existence of an underlying metabolism to be replicated, nucleotide strands were replicated by a prebiotic metabolic complex network of chemical interactions. The first cell, some four billion years ago, represented a completely new concept, that of autonomous duplication, which compared to nothing in the society from which it was born.

Present day culture (cultural selection), then, is considered to be in an earlier developmental stage than biology (natural selection), a stage that is best compared with prebiotic chemical interaction at the time when information molecules (nucleotides) were developed. Natural selection (as in biological evolution) is, however, about selection among variations on the theme of autonomous duplication, a theme not yet developed in culture/science/technology. Nevertheless, it is possible that cultural semantic closure can develop, for example, in the form of self replicating machines. As was the case for the first cell, compared to the prebiotic community from which it arose, these machines will compare to nothing in the cultural/scientific society from which they may arise. Just as the cell left its native pond and took off to conquer the Earth, a feat made possible by exponential growth resulting from autonomous duplication, the putative autoreplicative properties of these machines might enable them to leave native Earth and to conquer all parts of the Universe. Just as the prebiotic metabolic chemical interaction led to autonomously duplicating chemistry —once nucleotides were available as permanent carriers of information— cultural behavioral interaction may lead to autonomously replicating technology. Indeed, the availability of printed texts for 500 years has increased complexity (scientific knowledge and technology) exponentially, and may lead to autonomously replicating technology.

BACKGROUND

The author is well aware that the ideas developed here are highly unusual. They were developed (1) by considering evolutionary events (prebiotic, biological, and cultural) as the evolution of information and interaction—from the detached *point of view* of information, (2) by trying to find formal analogies between biology and culture, and (3) by critically studying the available hypotheses concerning the origin of life.

DEFINITIONS

The words we use most often lead to confusion, because the same word may have different connotations to different people. Thus, it is first necessary to explain how I define the most important concepts used in this paper.

Information. An elegant and highly applicable description of the concept *information* is due to Gregory Bateson: "A difference which makes a difference."[8] Of all the molecules that surround molecule *A*, only those that can interact with *A* can be said to contain information. There are many different molecules surrounding *A*, but only a few are able to make a difference; that is, to influence the chemical *behavior* of *A*. In comparison, of all the sounds and visual signals reaching the perception system of an animal, only some influence its behavior and can be said to provide information. This definition implicitly represents the contextual dependence of information. In the above example, those signals not influencing the behavior of animal *A*, may provide information to animal *B*.

Biology and Culture. Biology is basically the study of autonomously duplicating chemistry. Chemistry requires direct physical/material contact to make a difference, that is, to let information interaction occur. The interactions (information exchange) between organic molecules and cells are material/chemical. Culture became possible when animals started to influence each others behavior by means of sonic and visual signals, which do not require direct contact and can be exchanged over long distances. This is a profoundly new manner of making a difference—of transmitting information.

Transient Information. During behavioral interaction (chemical and cultural) information is transient, since the difference that can make a difference no longer exists after the interaction. During chemical behavioral interaction molecules are transformed (reappearing later in the chain of interactions as a consequence of ongoing interactions). During cultural behavioral interaction, sonic and visual signals persist over a limited time interval, and mental processes are required to cause signals with comparable information content to reappear later in the interconnected chain of behavioral interactions.

Permanent Information. Just as nucleotides enabled chemical behavioral interaction to be encoded, symbolic language enabled the encoding of cultural behavioral interaction. From an informational point of view, written (to a lesser extent), printed, and electronic texts compare most closely to nucleotide strands. They represent permanent differences that continue to exist whether or not interaction is taking place. They have the potential to make differences at many instances (e.g., nucleotides when translated and text when read or interpreted). Also, because they are symbolic, they have unlimited information content,[7] which means that a limited set of conventions (symbols or syntax) enables many differences to be represented. Another important characteristic is that permanent information can be recombined (randomly), exponentially increasing the information space that can be searched. Although it is impossible to mix processes, interactions, and processors, their encoded representations can be recombined endlessly. It is impossible to mix processes: One cannot mix the activity of enzyme *x* with that of enzyme *y*; one cannot take the bacterium *Escherichia coli*, mix it with a piece of human tissue, and retain a functional process; one cannot mix the ideas of two people by mixing their brain processes.

However, we can mix a functional *E. coli* cell with human enzymes, by inserting into its genome the code for such a human enzyme. Evolution itself created many new enzymes by the same process. In culture, having printed texts, many different lineages of information can continually come together, with true backup copies of the original ideas remaining available for proofreading.

Processors. Processors are entities that can repeat the same activity several times without being changed by the interaction. Processors (enzymes, transistors) are digital catalysts. Analog catalysts like some coenzymes, ribozymes, and thioesters are transformed during the process of interaction. Polypeptidic enzymes and transistors still exist after the interaction.

Life. Life cannot be understood by studying a single living organism. Since all currently living cells are the descendants of the ancestor cell—the first and still the only autonomously duplicating system that was ever developed—life is to be considered as a single four-billion-year old billion-billion-cellular organism, consisting of all the descendants of the first cell. Cooperation and competition between these cells, and between temporary colonies of cells (multicellular organisms), happen continuously as is also the case between cells within multicellular organisms. Because of the possibility of exponential growth, as a consequence of autonomous duplication, the organism *life* continually changes its environment and has to adapt to these self-induced changes. Evidently, the environment (biosphere) can be understood as a creation of this organism and as a part of it (consistent with the modern version of the Gaia hypothesis[9]). The question *What is Life* cannot be answered by studying a single living organism, because all extant creatures can be understood only in the context of their relations with other extant organisms and by considering the past evolutionary interactions and events.

Evolution and Complexity. The most general and straightforward definition of evolution is *change over time*. We observe that more interactions and different pathways become possible with time. This increased flexibility can be considered to be complexity (a concept closely related to or synonymous with *intelligence*). The increase of complexity is not a necessity, but is almost inevitably a consequence of competition and cooperation between the descendants of the autonomous duplicator. This follows from the fact that evolution is open ended toward higher complexity: more complex systems are able to exploit new niches that were beyond reach before this level of complexity was reached. For example, multicellular animals can feed on complete unicellulars, but not vice versa. This open niche explains, for instance, the tremendous radiation observed once the concept of a multicellular animal had evolved.

Selection, Cultural Selection, and Natural Selection. Selection is a general principle. It occurs when variations on a theme exist. None, one, more or all variations will be able to exist in a given environment. Radioactive decay is an example of selection among variations on the theme of physically stable atomic configurations in a universe (environment) with certain parameters for fundamental laws. The difference with cultural and natural selection is that there is no amplification/replication of the fit variations. In cultural selection, different answers to a problem (variations on a theme) may be valued differently by the environment (*in casu* the scientific community). The most valued hypothesis are amplified, other hypotheses may disappear. Cultural evolution differs from natural selection differs in that the amplification efficiency of hypotheses is not encoded in the hypotheses: unlike cells, hypotheses do not self-replicate, but are replicated by scientists, presses, and computers.

In natural selection, the only theme is amplification efficiency itself, and whereby selection among the different variations on the theme of autonomous replication automatically leads to amplification. At first sight this is a powerful principle, but with respect to developing higher complexity, biology is hampered because of limitations in searching information space: only those variations that do not diminish autonomous duplication efficiency can exist. On the other hand, the evolution of cultural information does not depend on its ability to replicate autonomously. Therefore, any idea or recombination of ideas imaginable is possible. Hence, the difference between cultural selection and natural selection: if science were to proceed by natural selection, this would mean that the texts produced by science should also contain all the necessary information on how to make a new text. Any changes to these texts that would undermine the ability to self reproduce would disable the text from spreading any further.

As an aside, in defining these generally used, but difficult concepts, not a single neologism has been used. All that has been done is to rethink the concepts we use, broadening their content where needed (e.g., the concept of behavior) and narrowing it elsewhere.

THE ORIGIN OF LIFE: PROBLEMS WITH THE RNA-WORLD MODEL

The discovery of catalytic activity of RNA-molecules (ribozymes)[2] has led to a revival of the idea suggested in 1968 by Francis Crick: that a single biopolymer, like RNA, might have served both information and catalytic roles and, thus, have propelled evolution toward the first cell by means of natural selection.[10] In contrast, the hypothesis suggested here states that no such autonomous duplication existed before the first cell and, thus, natural selection started only with the first cell.

Despite searching quadrillions of molecules, it is clear that a spontaneous RNA-replicator is unlikely to occur.[11] Reports of nucleotide[2,4,6] and peptide[12] self-replication still depend upon human intervention (for instance, by changing the environmental conditions between two rounds of replicaton or by denaturing the double strands). The problem of denaturing the double-nucleotide strand in a non-enzymatic manner has been overlooked and has contributed to a failure to establish molecular self-replication.

Even if these practical problems could be overcome, the RNA-world places the burden of both replication and variable information content on the same molecule, thus, the COSMIC-LOPER (*c*apability *o*f *s*earching *m*utation *s*pace *i*ndependent of *c*oncern *o*ver *l*oss *o*f *p*roperties *e*ssential to *r*eplicaton)[11] is very limited. Indeed, as explained above (see the difference between natural and cultural selection), introducing natural selection too early is a limitation rather than a gain. I propose that the original role of nucleotides was not self-replicative, so that there was high recombinatorial freedom for the information they carried, as is the case in current human culture using printed texts. It should be stressed here that many of the important findings of RNA-world research (see e.g., Ref. 13) need not be dismissed, as long as the catalytic role of ribozymes is restricted to metabolic and translational, nonreplicational functions.

One consequence of this model is that evolution could try out an exponentially larger number of possibilities (high COSMIC-LOPER) and could proceed much faster than by means of natural selection. In comparison, cultural selection, since the introduction of printed texts some 500 years ago, has increased complexity (scientific knowledge and technology) exponentially, whereas it took natural selection roughly two billion years to go from prokaryote to eukaryote, one billion years to proceed to the first multicellular animals, and one billion years to the first symbol using animal (humans), only some half a million years ago.

Another problem with the RNA-world hypothesis is known as Eigen's paradox[7]: the simplest cell known today contains a chromosome with 2,000 genes, most of these encoding for very different functionalities and with none of these genes themselves containing sufficient information to cover the complex process of autonomous duplication. Eigen realized that a society of self-replicating, competing RNA-hypercycles will outcompete each other when brought together in a cell, instead of merging into a chromosome. (Eigen's paradox is resolved (?) only by the rather artificial stochastic corrector model[7]).

HYPOTHESIS

The possible congruence between culture after the introduction of printed texts with prebiotic chemistry after the development of nucleotide strands.

Christian de Duve[3] has argued convincingly that the enzymatically driven metabolism of biology is functionally congruent with the prebiotic catalysis driven by, for example, thioesters. Although enzymes are very different from thioesters, they fulfill the same functional role. I argue that this congruence can be found back in current society, where metabolic and information functions of living beings are being conveyed to technological counterparts: computation, pattern recognition, speech, vision (by computers and robots), locomotion (cars, airplanes, missiles, and robots), energy provision (steam engines, nuclear plants, and photovoltaic cells).

Another congruence, is the observation that the rate of current transformations increased exponentially after the introduction of permanent information carriers (written, printed, and electronic texts) that are copyable in high numbers with high fidelity, that have unlimited information content, and that can be recombined endlessly. Congruently, it is generally agreed that complex enzymes (digital catalysts) could have been developed only after nucleotide strands existed, whereafter enzymes gradually replaced the original (analog) catalysts.

Congruently, science is possible because of the interaction between scientists and permanent information carriers, and this interaction between scientists and printed matter depends on the ongoing activities of many other people—farmers in the first place. Farmers in turn exploit plants and animals to produce food; animals and plants can thrive only because of bacterial metabolism. This is just to say that scientific activity and knowledge is only possible because it thrives on a complex underlying network of chemical, biological, and cultural interactions. Congruently, I propose that certain elements of a complex premetabolic network (the existence of which was proposed by de Duve[3]) started to develop symbolic language in the form of nucleotide strands (think of humans as their counterpart) and that this may have started a process comparable to the scientific process.

BRIEF PROPOSAL OF A MODEL FOR THE DEVELOPMENT OF THE FIRST CELL BY MEANS OF A SCIENTIFIC PROCESS

The following is intended only to draw a possible picture of the course of events. Imagine a large membrane irregularly making contact with a solid substrate, creating a microcosmos between membrane and substrate in which the presence of gaps allows for interaction with the environment outside of the membrane. Prebiotic metabolism[3] develops in this microcosmos. At some moment, the constituents of this network enable the production of information molecules, possibly connected to the outside of the membrane. Initially RNA nucleotides fulfilled this function (in combination with catalytic functions) and enabled the development of enzymes such as RNA-DNA polymerases (reverse transcriptases) that lead to DNA strands. (Polymerases are indeed supposed to be among the earliest enzymes[14]). These strands played a role comparable to that of printed texts that represent present day scientific hypotheses; for example, hypotheses about how to construct more efficient technology (enzymes). Billions of these DNA-protogenes were produced, most without any functionality. Comparably, many of our hypotheses do not lead to increased technological functionality. Those strands that encoded for enzymes with higher efficiency increased the efficiency of the local society as a whole, which resulted in a higher probability that these *genes* were reproduced more successfully. In comparison, the most technologically advanced societies gain greater economic advantages, without necessarily destroying other societies, upon which they rely for more basic, metabolic needs; and eventually both societies may profit from this knowledge. In the end, many informationally different, large protochromosomes—formed by ligation/combination of protogenes and containing an assembly of genes that encoded for very different functions, were attached to the membrane and were surrounded by enzymes (free and membrane-attached) that were encoded by neighboring and/or other chromosomes. Occasionally, blebbing (somehow comparable to obcell formation[15,16]) occurred; that is, splitting and circularization a piece of the membrane—a phenomenon still observed in present day bacteria such as meningococcus). It can be imagined that, on several of these occasions, the closing membrane internalized a protochromosome and enzymes. At one of billions of such occasions the protochromosome that was enclosed, could have been composed of the essential genes carrying the information required to duplicate the entire system. At this occasion it was of primordial importance that some of the active enzymes and ribozymes enclosed (DNA-RNA polymerases and ribosomes) were capable of translating the chromosomal information into enzymes with metabolic (e.g. pyrophosphate synthase) and replicative (DNA-DNA polymerase) functions.

The first cell, life, was born and natural selection (selection among variations on the theme of autonomous duplication) commenced. Because the kind of information that can be contained is limited due to constraints imposed by the functionality needed for self-replication, the evolution towards higher complexity—which had been extremely fast since the introduction of nucleotides as permanent carriers of information—slowed down drastically. On the other hand, competition between autonomously duplicating systems was now possible and (bio)diversity increased.

DISCONGRUENCES?

One might argue that the scientific process cannot be compared to simple chemistry because science is not random and science is done by goal directed beings (humans), unlike molecules. First, it can be easily argued that science can be considered to be a largely random process. Second, it is clear that humans are goal directed beings, unlike molecules. However, if for the sake of the argument, we imagine that self-replicating technology will result from our activities, it is clear that this was never—and still is not—our goal. This is true for most of our inventions. The invention of writing some 5,000 years ago was not to make scientific activity possible, but was goal directed towards facilitating the inventory of life stock and property. The introduction of the press in Western Europe was intended, in part, to spread the Bible; its impact on scientific activity was not foreseen. Computers were invented to increase computation speed, not to result in robots or the internet, and so forth. Nevertheless, each of these goals enabled new knowledge and insights that caused other applications and goals emerge. Although none of these goals intended to develop self-replicating technology, it is conceivable that in the near future, we may find applications for this technology (see e.g., Ref. 17) and we may try to develop this. The fact that, at present, we are conveying rapidly many biological functionalities to machines, indicates that this difficult enterprise becomes less and less unthinkable.

CONCLUSION: THE POSSIBILITY FOR A GRAND UNIFYING THEORY OF THE EVOLUTION OF INFORMATION

Explaining the origin of life as a scientific process might provide a unifying theory for the evolution of information, in which at two events symbolization/encoding of interactions into permanent information occurred—at one moment that of chemical behavioral interaction, at another moment that of animal behavioral interaction—and in which on one occasion this encoding led to autonomously duplicating chemistry (the cell), an event that possibly may be one of the outcomes of current scientific progress.

REFERENCES

1. ROSEN, R. 1991. Life itself. A Comprehensive Enquiry Into the Nature, Origin and Fabrication of Life. Columbia University Press. New York.
2. CECH, T.R. 1986. A model for the RNA-catalyzed replication of RNA. Proc. Natl. Acad. Sci. USA **83:** 4360–4363.
3. DE DUVE, C. 1995. Vital Dust. Life as a Cosmic Imperative. Basic Books. New York.
4. FERRIS, J. 1994. Chemical replication. Nature **369:** 184–185.
5. JOYCE, G.F. 1989. RNA evolution and the origins of life. Nature **338:** 217–224.
6. ORGEL, L.E. 1992. Molecular replication. Nature **358:** 203–209.

7. SMITH, J.M. & E. SZATHMARY. 1995. The Major Transitions in Evolution. W.H. Freeman. Oxford, UK.
8. BATESON, G. 1979. Mind and Nature.
9. WESTBROEK, P. 1998. Homage to Earth. Nature **391:** 550–551.
10. WATSON, J.D. 1993. Prologue: early speculations and facts about RNA templates. *In* The RNA World. R.F. Gesteland & J.F. Atkins, Eds. :*xv–xxiii*. Cold Spring Harbour Laboratory Press.
11. BENNER, S. 1999. Old views of ancient events. Nature **283:** 2026.
12. YAO, S., I. GHOSH, R. ZUTSHI & J. CHMIELEWSKI. 1998. Selective amplification by auto- and cross-catalysis in a replicating peptide system. Nature **396:** 447–450.
13. ZHANG, B. & T.R. CECH. 1997. Peptide bond formation by *in vitro* selected ribozymes. Nature **390:** 96–99.
14. STEITZ, T.A. 1998. A mechanism for all polymerases. Nature **391:** 231–232.
15. BLOBEL, G. 1980. Intracellular membrane topogenesis. Proc. Natl. Acad. Sci. USA **77:** 1496.
16. CAVALIER-SMITH, T. 1987. The origin of cells: a symbiosis between genes, catalysts, and membranes. Cold Spring Harbour Symp. Quant. Biol. **52:** 805–824.
17. TIPLER, F. 1995. The Physics of Immortality: Modern Cosmology, God and the Resurrection of the Dead. Anchor.