

A Marvelous Accident: The Birth of Life

Vincenzo Manca*

Dipartimento di Informatica, Università degli Studi di Verona, Ca' Vignal 2 Strada Le Grazie 15, 37134 Verona, Italy

Abstract

A short presentation is given of a recent book on the origin of life with the intent to explain the basic mechanisms and questions to the interested layman. In spite of the elementary nature of the book, a reflexion on the main principles driving the life emergence are discussed from an informational perspective that can be useful in many mathematical and computational trends of biological modeling.

Keywords: Monomers; Biopolymers; Replication; Hypercycle hypothesis; Protogenome; Information theory

Introduction

In this note I will give a brief account of a recent divulgation book (in Italian) written by myself and Marco Santagata [1]. The book is a sort of experiment in scientific communication. Marco is a well-known Italian novelist and an expert of Italian Literature (Dante, Petrarca). I am a mathematician-computer scientist, studying mathematical models of biological phenomena (especially, computational genomics) [2-8]. The topic of the book is the origin of life, with the intent of explaining the basic mechanisms and questions to the interested layman. The experiment is due to the writing method: "I explain a subject to Marco, then he writes and, very often, rewrites any times, until I approve the text, as sufficiently faithful with respect to my ideas". In this manner, the final form is something that Marco understood and I considered to be essentially in agreement with the content I wanted to express. Therefore, I can be confident that there is a good probability that an interested, but non professional, reader can appreciate the text, because he reads what, in principle, he could have written. This method implies some interesting consequences not only of stylistic nature. First, the main interest of the literate (Marco) is in the narration where some actors (molecules, monomers, biopolymers, membranes, enzymes) develop a story that has to be attractive for the reader, but, at the same time, self-contained in the required knowledge, and firmly linked with the scientific nature of the topic. On the other hand, the main interest of the scientist (Vincenzo) is in the principles underlying the main passages, their internal logic, and the search for a unification in the variety and complexity of life phenomena. However, the story seems to follow an implicit trend, and, very often, shows a paradoxical nature essential to the whole evolutive dynamics of life: a continuous antagonism and cooperation between chance and computation. After two prologues (in the sky and in the earth) where, starting from big-bang, atoms appear as produced by stars and basic molecules as produced by planets (earth), seven chapters follow with titles which resemble those of a philosophical treatise, rather than the "seven days" of creation: Aggregation, Replication, Generation, Memorization, Reproduction, Diversification, Evolution. Finally, a note, "the birth of this book", and brief historical remarks conclude the book. The whole content (apart from a short subject index) is under 130 pages. This means that the difficulty of the book was not only writing it, but mostly, selecting what to write. But why a mathematician-computer scientist and a literate-writer decided to discuss about the origin of life? For two reasons: the first one is due to the fact that both of them think that life is too important for being investigated only by biologists; the second one is related to the scientific background of the first author. In fact, I studied for long time replication algorithms for strings (important in computational universality [9] by discovering, in a second moment,

how much replication is fundamental for life. Namely, life started when molecular evolution provided the possibility of representing and processing information at a molecular level. Therefore, life is, in an essential part, a special kind of complex informational process [10]. At the same time, as many pioneering works in cybernetics and computer science prove [11-15], reflexions on life were at the origin of the modern notions of information and computation. An important part of the book are the drawings illustrating the story. Hand-made figures were deliberately chosen, by avoiding "real" illustrations typical of biochemistry or biology manuals. This choice was, in a sense, consequent to the conceptual essentiality that was considered one of the goals of the book. A well-known Italian graphic designer, Guido Scarabottolo, was the author of all the 18 figures (cover included) that accompany the text.

Aggregation, Replication, and Generation

Aggregation is one of the main forces driving the physical reality. The appearance of matter, was realized by means of minimal subatomic particles, successively, of atomic nuclei, and then, of small atoms, such as hydrogen and helium, up to all the atoms of Mendeleev's table. With the decrease of the explosion energy, the same trend led small molecules to form more complex molecules, where the basic principle of repeatability emerged: polymers are constituted by repeating molecular modules. The origin of the repeat mechanism depends on many factors, but surely is due to a tendency to the completeness, the same acting at atomic level and producing the first molecules. Chemical bonds, linking atoms, require stable energetic conditions where stability is due to a reciprocal complementarity. In a wide sense, one component tends to combine with another one that complements it, by reaching a completeness that gives more stability to the constituents. Stability means duration in time, but this means, very often, the conservation of molecular structures that become new elements for further processes of aggregation, a sort of self-referential mechanism, or even more, the appearance of a first form of self-reference, a principle that life shows at any level. The first essential innovation in molecular repeatability are monomers, molecular complexes that are the alphabet of life. These

*Corresponding authors: Vincenzo Manca, Dipartimento di Informatica, Università degli Studi di Verona, Ca' Vignal 2 Strada Le Grazie 15, 37134 Verona, Italy, Tel: +390458027981; Fax: +39 045 802 7068; E-mail: vincenzo.manca@univr.it

Received July 02, 2018; Accepted August 06, 2018; Published August 13, 2018

Citation: Manca V (2018) A Marvelous Accident: The Birth of Life. J Proteomics Bioinform 11: 135-137. doi: [10.4172/0974-276X.1000479](https://doi.org/10.4172/0974-276X.1000479)

Copyright: © 2018 Manca V. 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.

structures are chiral, no plane can split them in two parts each of which is the mirror image of the other one. They consist of four parts: head, tail, body, expression. The first three parts provide the possibility of arranging them in an ordered sequence, based on head-tail bonds. The fourth one is the proper alphabetic component, because the expression can assume values in a finite set of possible values (alphabet). This is the essence of any monomer on which biopolymers of any type are built. Biopolymers of a special kind are, in a natural way, replicating objects. The same principle of complementarity leads monomers to connect with monomers of the same type, but of different nature. In fact, a sort of male-female attraction provides pairing bonds (this is only a metaphor of specific chemical mechanisms). When a biopolymer has all its monomers paired, then the paired monomers, complementary to those of the original sequence, become close enough to get, under suitable conditions, a better stabilization by binding each other contiguously. In this way, a biopolymer is formed that is complementary to the original one. It is a sort of “negative” copy, and when a second process of complementary copy is applied to this negative sequence, a biopolymer equal to the original one is provided. If the first polymer expresses a “biological word” with some interesting biological functionality, then the copy process is a way for escaping from the intrinsic degradation of biopolymers. In fact, a copied word, which can be copied again and again, is something able to survive to its single instances. This is the substantial reason that suggest us to consider replication as the first primordial form of life. However, for technical reasons, when the lengths of replicating polymers are over a threshold (around 100 monomers) the error rate become prohibitive and copies have a percentage of dissimilarity with respect to the original ones that forbids them to be considered true copies. This phenomenon, investigated by the Nobel laureate Manfred Eigen [16,17], is the starting point of a long and complex way toward the life as we know it. Eigen proposed the hypercycle hypothesis, which we revisit in the book, with some specific variants. The main ingredient, to overcome the replication limit, is the molecular cooperation: many biopolymers that are not able to realize a good replication can cooperate for obtaining the right results by means of a collective process. In rough terms, biopolymers, which for fixing ideas we can think as RNA sequences, start to define a “replicative metabolism”, where some biopolymers become enzymes (riboenzymes) of the replication of others, in a reciprocal way. The metaphor is that of an ape who scratches the back of another ape (in a circle arrangement). But such a solution needs an environment where such a process is hosted (selecting, concentrating, and protecting the involved molecules). Membranes, which are based on aggregation of another kind of asymmetric monomers, realize aqueous compartments hosting biopolymers plus basic molecules. Membranes become the new ingredient for realizing this replicative metabolism, where reliable copies of “long” sequences can be produced (even several thousands long). One possible solution allowing long biopolymers to replicate is the concatenated duplication where short RNA fragments are duplicated and then concatenated, that is, the duplication of a concatenation is reduced to a concatenation of duplicates.

But even when good copies are done, a second-level problem arises. How to duplicate the whole duplicating system, before it is destroyed by natural physical degradation? This kind of duplication is more complex because needs to duplicate the reactor and its reagents, in such a way that the copy of the system behaves as the original one, that is, the biological information cumulated in the process does not get lost when a membrane system dies. It seems that, very often, Nature searches for the solution of a problem by searching for the solution of a more complex problem including the original one. This is the so called Polya’s

Inventor’s paradox [18] present also in mathematics: solving a problem by including it in a bigger problem that can be afforded with more powerful means. Therefore, the membrane approach introduces a new problem, the “generation” problem: providing a copy of a membrane with all the replicative metabolism hosted inside it. The most probable solution of such a kind of problem is based on protogenomes, that is, long biopolymers consisting of a concatenation of the copies of all RNA fragments and enzymes acting in the duplicative metabolism. In this sense a protogenome is a sort of primordial biological memory containing, at a potential level, all the (RNA) biopolymers hosted by the membrane. In this way, when a membrane splits in two equal membranes (by a superficial tension mechanism) each of the two copies of the protogenome (only two copies) becomes the biological memory of each daughter cell, directing a process similar to that of the mother cell. The passage from the protogenome potentiality to the effective metabolism is obtained by coping fragments of the protogenome (where parts are tied within a compact linear structure) that realize the functions of single “free” RNA strings. But, again, the initial problem of biopolymers replication is transformed in another problem, that of protogenome duplication (only two copies of it need). A long story of more than one billion of years concerns with the solution of this problem: passing from protogenomes to genomes, that is, reliable biological memories able to guarantee a structural and dynamical similarity between the daughter protocell and the mother protocell. When this situation is reached, then generative lines become reproductive lines in the full sense, where the biological knowledge reached by individuals can be passed (with acceptable variations) along the generations. But, why membranes split? In the book a tentative answer is outlined, but even this phenomenon is a consequence of the replicative metabolism inside membranes.

Memorization and Reproduction

The two chapters about memorization and reproduction enter in the internal dynamics of the cell: the passage from RNA to DNA and the protein synthesis. In a sense this story is along the usual track of traditional manuals. However, here the emphasis is given to the role of DNA as biological memory with a double linear structure. Two aspects are relevant in this regard. The first one is the informational perspective in considering DNA. The second one is related to the incredible capability of spatial compression that DNA molecule can realize. Let us consider all the possible DNA strings of length 100 (very short sequences). How many are they? Obviously they are 4100 and it is easy to realize that $4^{100} > 10^{60}$ (being $2^{10} > 10^3$). This evaluation tells us, by easy calculations, that for realizing all these sequences a matter is required that is contained in a mass of billions and billions of planets equal to the earth. This implies that the 100-long DNA strings present in nature are an infinitesimal part of all the possible 100-long DNA strings. Therefore, DNA of living organisms was selected by means of a specialized process of informational nature (information selects some among many possibilities [19]). At the same time, the strings occurring in living systems, which are the product of an informational selection, are also instruments for selecting specific possibilities in the huge of evolution paths. Other simple calculations tell us that if we put together all DNA inside our cells of a human body (in the average 1014 cells) we get a sequence having a length equal to 600 times the distance between the earth and the sun (human cell genome is about 2 meters long). Genomes of living organisms are long millions or billions of nucleotides, because cells need a big biological memory, taking trace not only of the cell functions, but of the previous evolution story since LUCA up to the present time. Our DNA contains the information of our biological functions, but also our past, and the possible next

steps of our evolution. All this information has to enter in the space of cell nucleus. The structure of DNA is a direct consequence of the efficient DNA replication algorithm, and of the efficient packing of DNA structure at microscopic level. In the book the geometry of DNA helix is explained as a consequence of the packing problem, by using the notion of monomeric triangle. In fact having two paired sequences of objects, each object of this structure is obliged to be concatenated with the next one and to be paired with its complementary object. This means that the minimal module of a bilinear structure is a triangle: monomer-concatenated-paired. By repeating paired monomeric triangles, we obtain the whole bilinear structure. But, if we analyze the spatial possibilities to wrap a stripe of these triangles, using very reasonable conditions of uniformity and stability, we discover that among all possible spiral arrangements only one is possible, where concatenated triangles rotate along their pairing axis. Moreover, this arrangement implies several other consequences that are crucial in the whole DNA packing process.

Diversification and Evolution

The last two chapters are centered on two main problems. How species arise and how multicellular organisms realize the incredible paradox of germinal cells? A germinal genome contains the project of the whole organism to which the germinal cell belongs. We know that a part cannot contain the totality. A case where this principle is falsified are the infinite sets, as Galileo Galilei discovered (Galileo's paradox) by considering the even numbers that can be put in 1-to-1 correspondence with the set of all natural numbers. Of course, living organisms are not infinite, but they realize something very similar to Galileo's paradox. The genome of the a germinal cell is multivalent, a sort of multi-genome able to be restricted and modulated during the embryogenesis of the organism. Genomes give the possibility of motivating, in a precise and rigorous manner, the main principles of Darwin's Theory of Evolution [20]. It is probable (it is my belief) that the criticisms to the evolution theory, as intrinsically qualitative, will be confuted in the near future, when quantitative (mathematical and computational) analyses of genomes will be developed. A very important example of this perspective is Fisher's Theorem of Natural Selection [21], claiming that the evolution rate of a population is proportional to the genetic variability within the population. Analyses of genome regularities could disclose specific mechanisms of genome evolution, and rigorous measures of genome distances or dissimilarities, could determine quantitative thresholds over which speciations occur. The book suggests only this new perspective, but of course, it does not give details. However, an important point is discussed, concerning with many important passages of the story outlined in the book. What is the relationship between computation and chance in life? Of course this is an old debate, related to the chance/necessity dichotomy of the famous Monod's book [22]. Here, necessity is replaced by information and computation, new scientific categories developed by Information Theory and Computer Science in the last century, now crucial in science. The main thesis of our book is that chance is a sort of "mathematical necessity" complementing computation, for the intrinsic property that random and chaotic processes have of generating what mathematicians call 'ergodic dynamics'. In such a kind of dynamics the "space of possibilities" is visited in a homogeneous way, avoiding preferential

regions, by means of uniform and complete explorations. This is the randomness of brownian motions, observed by Robert Brown in the zig-zag of pollen grains suspended in water. In problems with large sets of possible solutions the algorithms that develop a computation based on random explorations gain a big advantage with respect to algorithms with fixed, even if sophisticated, strategies. Probably, this is the main trick of evolution, therefore randomness and chaos are a proper part of the internal logic of life. Genomes are the structures where I hope that, in the near future, these intuitions could become precise scientific laws giving new knowledge about the marvelous accident which we are part of.

References

1. Manca V, Santagata M (2018) Un meraviglioso accidente. La nascita della vita. Mon- dadori, Milano.
2. Manca V, Franco G (2008) Computing by polymerase chain reaction. *Math Biosci* 211: 282-298.
3. Manca V (2013) Infobiotics: information in biotic systems. Springer (2013).
4. Manca V (2015) Information theory in genome analysis. In *Membrane Computing, LNCS 9504*, 3-18, Springer.
5. Manca V (2016) Infogenomics: genomes as information sources. chap. 21, 317-324 Elsevier, Morgan Kauffman.
6. Bonnici V, Manca V (2015) Infogenomics tools: A computational suite for informational analysis of genomes. *J Bioinfo Proteomics Rev* 1: 8-14.
7. Bonnici V, Manca V (2016) Informational laws of genome structures. *Sci Rep* 6: 28840.
8. Manca V (2017) The Principles of Informational Genomics. *Theoretical Computer Science* 701: 190-202.
9. Turing AM (1936) On Computable Numbers, with an Application to the Entscheidungsproblem. *P Lond Math Soc* 42: 230-265.
10. Schrödinger E (1944) What is Life? The physical Aspect of the Living Cell. Cambridge University Press.
11. Rosenblueth A, Wiener N, Bigelow J (1943) Behavior, purpose, and teleology. *J Philos Sci* 10: 18-24.
12. Wiener N (1948) Cybernetics or control and communication in the animal and the machine, Hermann, Paris.
13. McCulloch W, Pitts W (1943) A Logical Calculus of Ideas Immanent in Nervous Activity. *Bull Math Biol* 5: 115-133.
14. Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc Nat Acad Sci* 79: 2554-2558.
15. Von Neumann J (2000) The Computer and the Brain. Yale University Press, 2nd Edition.
16. Eigen M (1971) Selforganization of matter and evolution of biological Macromolecules. *Naturwissenschaften* 58: 465-523.
17. Eigen M, Schuster P (1979) The Hypercycle: A Principle of Natural Self-Organization. Springer.
18. Pólya G (1945) How to solve it. Princeton University Press.
19. Shannon CE (1948) A mathematical theory of communication. *Bell Sys Tech J* 27: 623-656.
20. Darwin C (1859) The Origin of Species. John Murray, London.
21. Fisher RA (1922) On the Mathematical Foundation of Theoretical Statistics. *Philos Trans R Soc Lond B Biol Sci* 222: 309-368.
22. Monod J (1972) Chance and Necessity: Essay on the Natural Philosophy of Modern Biology, Vintage Books, New York.