ISSN 1517-6770

SHORT COMUNICATION

The origin of life as a probabilistic event in the Universe

Dimitri Marques Abramov[1*] **& Carlos Alberto Mourao-Junior**[2]

[1]*Mestre e Doutor em Ciências. Laboratório de Neurobiologia e Neurofisiologia Clínica, Instituto Nacional Fernandes Figueira, Fundação Oswaldo Cruz. *E-mail: dimitri.abramov@iff.fiocruz.br.*

[2]*Mestre e Doutor em Ciências. Professor Associado da Universidade Federal de Juiz de Fora, Departamento de Fisiologia. E-mail: camouraojr@gmail.com.*

Abstract. By means of a probabilistic mathematical model, we bring into discussion the origin of life as a stochastic process. We consider only the chance of information emergence in the proteome and genome under the ideal thermodynamic and chemical conditions. For a more realistic model, we used, as a parameter, the information amount in *Nanoarchaeum equitans* genome, the simplest known nowadays, as the equivalent to the first living cell that could have emerged in primitive Earth. We estimated the probability of information emergence by chance as about $10^{-500'000}$. Considering the necessary ideal conditions for information emergence, the probability of the origin of life would be even smaller.

Key words: Evolution, Origin of life, Stochastic processes.

Resumo. **A origem da vida enquanto um evento probabilístico no Universo.** Através de um modelo probabilístico matetmático, nós trazemos à discussão a origem da vida como um processo estocástico. Nós consideramos somente a chance da emergência da informação do proteoma e do genoma sob condições termodinâmicas e químicas ideais. Para um modelo realístico, nós usamos, como parâmetro, o conteúdo informacional no genoma do *Nanoarchaeum equitans*, a mais simples conhecida, como o equivalente a primeira célula viva que poderia ter emergido na terra primitiva. Nós estimamos a probabilidade da emergência como sendo próxima da $10^{-500.000}$. Considerando as condições ideais para emergência da informação, a probabilidade da origem da vida deveria ser ainda menor.

Palavras-Chave: Evolução, Origem da vida, Processos estocásticos.

Was the origin of life a probabilistic event? It is one of the intriguing unresolved questions about our Universe and its nature, which in this work is called "the Original Problem" (OP). A probabilistic viewpoint about biogenesis has been considered since Darwin's evolutionary theory, coined by the paradigm of the Natural Selection (DARWIN, 1859), since the statistical perception about the physical world through Boltzmann's work and, later, through Quantal Mechanics (PRIGOGINE, 1997).

The "Natural Selection" naturally was extrapolated to the OP, when ideas about a

Darwinian prebiotic evolution were developed, basically regarding life as an emergence from an ideal biochemical environment, by chance (FOLLMANN & BROWNSON, 2009, AONO *et al*., 2015). The Stanley Miller's experiment (5) was taken as a crucial empirical evidence in favor of prebiotic evolution.

Observing its complexity, the probability for a cell to emerge by chance must be too small. The skepticism about the Darwin's warm pond is so old as this hypothesis that defies the logical thinking (FOLLMANN & BROWNSON, 2009). In the 40's, the famous astrophysical Fred Hoyle regarded in his novel *Evolution from Space* (HOYLE & WICKRAMASINGHE, 1984) that the probability of emergence of a litlle (and necessary for life) set of enzimes is about $10^{-40.000}$. Hoyle's coauthor, Dr. Wickramasinghe (WICKRAMASINGHE, 2010) had advocated for cosmic panspermia as a possible answer for the Original Problem.

Even if illogical, is the stochastic emergence of the life physically compatible to the all Universe's dynamics? Seth Lloyd (LLOYD, 2002), regarding the estimated number of bits in the Universe (in order of 10^{90}), calculated that the Universe produced a total of 10^{120} transformations (e.g., spin inversions), or computational operations, since the Big Bang. Thus, the probability of any transformation that yet occurred in the Universe is about $1/10^{120}$.

In this essay, we present a probabilistic model to discuss the origin of life as a stochastic event. The model regard the emergence of the necessary and suficient information (encoded in the DNA and also present in the proteome) for a biomolecular machinery express a self-replicant living cell (see methods). For that, we are considering the information in the simplest known living cell as the parameter.

Therefore, the model concerns only the cell information.The problem of the emergence of all chemical, structural and environmental context, called fundamental resources (4), necessary for the putative organization of genomic/proteomic information, is not regarded here. On the contrary of those FR, the probability of the emergence of genomic/proteomic information is theoretically predictable because it is the combination of predefined "bits", that are nucleotides and aminoacids in informative sequences along their respective polimers.

For the subsequent discussion, we consider two assumptions: (1) the Universe is the only existing thermodynamically closed system, once there is nothing encompassing the Universe and there is not a perfectly closed system encompassed within the Universe. (2) The Universe's dynamics is probabilistic, i.e., it is the result of the chance.

These assumptions yield two logical deployments: (1) if the Universe is the only closed system, everything has its origin inside the Universe and everything depends uniquely on the Universe's resources; (2) if the Universe is a probabilistic system, every deterministic process is naturally probabilistic too, because the first process in the Universe, that was the precursor of the all other processes, purely happened by chance. Thus, even being the process X set by a chain of causes, this process is already probabilistic per se, because it is statistically dependent on the first process.

In light of these assumptions and their logical deployments, non-living (non-self-replicant) entities in the universe had originated from chains of probabilistically dependent events P(A|B): in a chain of dependent probabilistic events, if A is determined by B, and B is determined by C. The probability that event A happens is dependent on the probability that event B happens, and the probability that event B happens is dependent on the probability that event C happens, and so on. Therefore, the product of the probabilities of each event taken independently is equal to the probability that the final event A happens after the occurrence of a chain of consecutively dependent events.

Self-replicant entities, producing many copies of themselves, enlarge the probabilities of transformation and the conditional probabilities don't obey the above linear chain of statistical dependency.

Thus, before the life origin, the probability of emergence of genome information is equal to the probability for all nucleotides that form such genome to be in the correct sequence, having this genome appeared at once or in N steps, each one probabilistically dependent on the preceding ones.

The present model explores this probability of encoded information in a genome/proteome to emerge among all the possible combinations of monomers (among all possible codes) to discuss the admissibility of the origin of life as a stochastic event.

THE MODEL

The model predicts that the genome and proteome are formed by combinations of any nucleotides and amino acids with equal binding probability between them. The genome/proteome chosen was the thermophilic prokaryotic cell *Nanoarchaeum equitans*, an archeabacteria, the simplest known living being (HUBER *et al*., 2002; WATERS *et al.*, 2003; DAS *et al.*, 2006), which is regarded as being a living fossil (DI GIULIO, 2006). *N. equitans* lives in extreme conditions, similar to the primitive Earth environment, one billion years ago. This cell has a simple metabolome, which does not synthetize lipids, amino acids or nucleotides; however, it is obviously able to perform self-replication. *N. equitans* holds 466340 pairs of informative bases (95% of all the cell DNA), which form its 540 genes. This microorganism uses all the twenty known amino acids in its proteome.

The model accesses the chance for the genome and proteome to be expressed in parallel (in order of magnitude), as follows:

$P(g) = b^{-g}$, where g is the amount of informative DNA nucleotides in the genome, and b is the chance for any DNA base to be raffled ($b = 0.25$, for four nucleotide types).

$P(p) = a^{-p}$, where *p* is the amount of amino acids in the proteome, corresponding to the number of encoded pieces of information in the genome ($p = g/3$) and *a* is the chance for any amino acid to be raffled ($a = 0.05$, for twenty amino acids).

Therefore, we calculated the total probability of parallel emergence of *N. equitans* proteome and genome:

$E = P(g) * P(p)$;

where *E* is the probability for the piece of information in the genome/proteome in the *N. equitans* to emerge.

Considering that the 466340 nucleotide pairs, g = 155446.67 » 155446, the probabilities for the genome and the proteome to be raffled are $P(g) = 0.25^{466340}$ » $10^{-280737}$ and $P(p) = 0.05^{155446}$ » $10^{-202236}$. Thus, E » $10^{-482973}$.

Regarding the known constitution of the structure and metabolism of the genome and proteome of cells at general, based on DNA and polipeptides, the probability for the origin of life to be a stochastic event by the means of the chance of information emergence in the simplest known cell, inside an ideal dissipative system with all fundamental resources, is of about $1 / 10^{500K}$, a number astronomically near to zero.

As we have previously pointed out, this model is informational, i.e., it considers only the probability of information emergence, previously regarding all the structural dynamic requiring conditions (the fundamental resources) as ideal, since these elements cannot be estimated. We also disregard the necessary amount of specimens for the emerging species to be able to persevere. It is unlikely that one single emerging cell can originate an ecologically stable population for beginning biotic evolution. Therefore, the probability of origin of life is a product of E: the emergence of ten cells means a probability ten times lower.

The main discussion about the validity of this simple probabilistic model appears if we consider that the putative prebiotic processes create a proto-metabolism and a proto-genome that would be steps for life emergence (Wachtershauser, 1990; Lee *et al*., 1997; Aono *et al.*, 2015). Before the proto-system became self-replicant, the final/total probability is equal to the product of the probabilities of those steps, because they are dependent on the preceding ones. After the first self-organized and self-replicant system emerged, the probabilities of subsequent system transformations increase, in function of the number of self-replicant units. Nevertheless, that emergent self-organized and self-replicant system was the first life form: the OP would have been solved.

Therefore, considering that the first life form was equivalent to the *N. equitans*, the probability of the origin of life would be remarkably smaller than E.

Would be a self-organized and self-replicant system substantially simpler than *N. equitans* possible? First, we must define the threshold to "substantially simpler": an emergent system with emergence probability equivalent to the Universe's computational load, Lloyd's number (Lloyd, 2002), near Dembsky's universal bound of probability. A genome with probability of 10^{-120} would have only 250 nucleotides, which is equivalent to about 28% of one gene from the *N. equitans*. The smallest known viral genome, from the *circoviridae* family (Belyi *et al.,* 2010), has about 2000 nucleotides, and viruses are not self-replicant. Therefore, it is conceivable that the information relative to 250 nucleotides is not sufficient to generate any life form. A system one hundred times simpler would still be too big, whose probability (10^{-5000}) is still incompatible to the estimated magnitude of the Universe.

Observing Lloyd's number (Lloyd, 2002), we can notice that the mathematical probability of the origin of life as a stochastic phenomenon could be regarded as a physical impossibility.

As an alternative discussion, we visited England's theory of emergent adaptation of complex systems by energy dissipation (England, J.L. 2013; Perunov *et al*., 2016): if, on the one hand, the Universe tends to the thermodynamic equilibrium (maximum entropy), on the other hand, the open systems inside the Universe would tend to spontaneous processes of dissipation-driven self-organization and adaptation. These processes result in a progressive enlargement of diversity and complexity. The replication of complex dissipative systems, in the nature, seems to be a rule: reproduction is a way to dissipate energy, for the species stability improvement and for the increase on the probability of new state transitions (i.e., evolution). Another alternative for this problem is the RNA World Theory, what regards a possible prebiotic evolution of self-replicators made by RNA molecules with catalytic properties, exponential growth and pottentially able to darwinian evolution (Robertson & Joyce, 2012; Robertson & Joyce, 2014). But it still speculative lacking evidences to explain the lifeforms with DNA and typical enzymes.

Thus, we consider that the Origin of Life as a stochastic phenomenon is incompatible with the presented assumptions and their logical deployments. Therefore, unknown properties of the Universe organization could be necessary for the origin of life in conditions for survival and evolution.

REFERENCES

Aono, M.; Kitadai, N. & Oono, Y. 2015. A Principled Approach to the Origin Problem. **Origins of life and evolution of the biosphere** 45 (3):327-38.

Belyi, V.A.; Levine, A.J. & Skalka, A.M. 2010. Sequences from ancestral single-stranded DNA viruses in vertebrate genomes: the parvoviridae and circoviridae are more than 40 to 50 million years old**. Journal of Virology** 84 (23): 12458–12462.

Darwin, C. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. **London, John Murray**, 312 p.

Das, S.; Paul, S.; Bag, S.K. & Dutta, C. 2006. Analysis of *Nanoarchaeum equitans* genome and proteome composition: indications for hyperthermophilic and parasitic adaptation. **BMC genomics** 7:186.

Di Giulio, M. 2006. *Nanoarchaeum equitans* is a living fossil. **Journal of theoretical biology** 242 (1):257-260.

England, J.L. 2013. Statistical physics of self-replication. **The Journal of chemical physics** 139(12):121923.

Follmann, H. & Brownson, C. 2009. Darwin's warm little pond revisited: from molecules to the origin of life. **Die Naturwissenschaften** 96 (11):1265-1292.

Hoyle, F. & Wickramasinghe, C. 1984. Evolution from space. **New York: Touchstone**; 129 p.

Huber, H.; Hohn, M.J.; Rachel, R.; Fuchs, T.; Wimmer, V.C. & Stetter, K.O. 2002. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. **Nature** 417 (6884):63-67.

Lee, D.H.; Severin, K. & Ghadiri, M.R. 1997. Autocatalytic networks: the transition from molecular self-replication to molecular ecosystems. **Current Opinion Chemical Biology** 1 (4):491-6.

Lloyd, S. 2002. Computational capacity of the universe. **Physical review Letters** 88(23):1-4.

Miller, S.L. 1953. A production of amino acids under possible primitive earth conditions. **Science** 117 (3046): 528-529.

Perunov, N.; Marsland, R.A. & England, J.L. 2016. Statistical Physics of Adaptation. **Physical Review X** 6(2):1-12.

Prigogine, I. 1997. The end of certainty: time, chaos and the new laws of nature**. New York: The Free Press**; 208 p.

Robertson, M.P. & Joyce, G.F. 2012. The origins of the RNA world. Cold Spring Harb **Perspectives in Biology** 4(5):1-22.

Robertson, M.P. & Joyce, G.F. 2014. Highly efficient self-replicating RNA enzymes. **Chemical Biology** 21(2):238-245.

Wachtershauser, G. 1990. Evolution of the first metabolic cycles. **Proceedings of the National Academy of Sciences of the United States of America**. 87 (1):200-204.

Waters, E.; Hohn, M.J.; Ahel, I.; Graham, D.E.; Adams, M.D. & Barnstead M. 2003. The genome of Nanoarchaeum equitans: insights into early archaeal evolution and derived parasitism. **Proceedings of the National Academy of Sciences of the United States of America** 100 (22):12984-12988.

Wickramasinghe C. 2010. The astrobiological case for our cosmic ancestry. **International Journal of Astrobiology** 9 (2):119-129.

Recebido: 03/03/2017

Revisado: 26/05/2017

Aceito: 19/06/2017