Submitted to: Journal of Cosmology

Hypothesis section:

Origin of the genetic code: was the original mechanism lost or altered during evolution after the universal genetic code was virtually frozen?

J. T. Trevors

School of Environmental Sciences
University of Guelph
50 Stone Road, E.
Guelph, Ontario
Canada N1G 2W1
E-mail: jtrevors@uoguelph.ca

Tel: 519-824-4120
Fax: 519-837-0442

Key words: bacteria; cells; evolution; frozen genetic code; iterations; mechanism; minimal genome; origin of life

Possible reviewers:

Dr. F. P. R. Nilson, Avd. For Bioteknik (frans_peder@biotek.lu.se)

Dr. J. Demongeot, University of Joseph Fourier of Grenoble (Jacques.demongeot@imag.fr)

Dr. W. S. Chang, University of Texas, (wschang@uta.edu)

Dr. H. Insam, Innsbruck, heribert.insam@uibk.ac.at

Dr. D. Klaus, University of Colorado (david.klaus@colorado.edu)

Abstract: The natural mechanism that organized the corresponding coding between nucleic acids and the corresponding amino acids is still unknown. It is also not known if molecular remnants or relics of this mechanism are present in some living cells as an altered mechanism or the original mechanism was lost during evolution. Prokaryotic organisms may be a plausible location for discovering such a mechanism as they are the ancient species on the Earth. The hypothesis is proposed that the molecular mechanism that generated the universal genetic code was lost, or altered for other functions, once the genetic code was virtually frozen/fixed. By virtually freezing the code, evolution could proceed at a faster pace without generating a new genetic coding system for different species. Different combinations of the code emerged in the evolving species. This is an efficient mechanism of generating new code combinations from an existing genetic code.

Introduction:

The plausibility of arriving at the correct answer as to the location(s) and time(s) (Joseph & Wickramasinghe, 2011) of the origin of life on the Earth and/or elsewhere, is a profound challenge. However, interesting events sometimes occur in research that are unpredictable, and transformative for science and societies (Trevors et al. 2011). It would be an immense achievement if an hypothesis for the origin of life was plausible, testable and had ample supporting evidence. The origin of life from one perspective was a thermodynamic event where overcoming a higher entropy environment to organize a cell (less random than the surrounding environment) capable of regulated growth and division, followed by evolution (Trevors 2010 a,b, 2011a,b, Trevors & Pollack, 2011). This was a molecular organization problem that required the correct genetic instructions for all functions necessary for life sustaining biochemical functions, in the spatial environment of eventually membrane-bound cells.

Although evolution is a well-established theory with supporting evidence, a correct theory on the origin of life with supporting evidence is still a challenge. The origin of life should not be confused with the theory of evolution. Evolution occurs after you have a life form capable of growth and division, as evolution selects at the organism level. The organism either dies, or does not reproduce, or grows and reproduces, with evolution proceeding. The origin of life whether on the Earth or one or more extraterrestrial location(s) (Joseph & Wickramasinghe, 2011) would require a minimal molecular construction kit (including minimal genome) for the organization of the first cell capable of growth and division. One profound aspect of evolution is the capability for some individuals to reproduce under diverse natural selection conditions, and the subsequent diversification of species. The extinction route is also profound as the evolutionary route for many species has been halted.

One possibility is that the first pre-cell(s) assembled using the necessary available components present on the Earth about 4 billion years ago before life capable of growth and division emerged. Another recent hypothesis (Joseph & Wickramasinghe, 2011)

places the origin of the first gene (not life) at an extraterrestrial location about 10 billion years ago, before the formation of the Earth about 4.6 billion years ago. The location (Earth or extraterrestrial) and the time of the origin of genes and life are still debatable and unknown as science requires more supporting evidence on both the estimated time(s) and location(s) of the origin of life. Hopefully, this will be forthcoming as our science knowledge increases. Regardless of the distant time(s) and location(s), mechanism(s) for the organic genetic coding, cellular organization would have been necessary. Secondly, the assembly of the genetic code would require some molecular mechanism for organization and in a location with a non-limiting supply of required molecules and/or the delivery of the required molecules. The possible presence of the correct enzymes and their stability in the microscopic location(s) is also a challenge to solve (Trevors, 2010b). The plausible absence of enzymes for the assembly of the first genetic code means that a physical-chemical mechanism such as thermosynthesis and/or EM radiation, as examples may have been involved in gene assembly (Trevors & Pollack, 2011).

Pre-cells would have required some of the elements of the Periodic Table and synthesis mechanism(s) to produce the necessary molecules required for life. The conditions on the Earth that would be present during pre-biotic organization would include no oxygen, exposure to the complete electromagnetic (EM) spectrum, thermocycling from heating/cooling and day/night cycles, drying/wetting cycles with accompanying evaporation as a mechanism to concentrate molecules, and diffusion of molecules in the microscopic location (e.g., gel-like cytoplasm) which is one plausible hypothesized micro-location for the origin of life (Trevors & Pollack, 2005, 2011). These conditions need to be better understood, to determine if they had mechanistic roles in the origin of life.

Figure 1 summarizes hypothesized conditions and possible mechanisms during the origin of life including the hypothesis that iterations (or cycles) that may have selected for the best organic genetic coding system that was frozen as our present genetic code, and then carried forward during evolution in different combinations of the 64 RNA codons that specifies one of the 20 amino acids. With a punctuation mark this provides a standard

organic code with 21^{64} or about 4×10^{84} possible genetic codes (Jose et al., 2009). Once virtually frozen the code would not need to be replaced or improved upon as the central molecular machinery was in place to both replicate the code and decode the genetic code during transcription/translation. Codons have been altered during evolution so the code is better described as being virtually frozen/fixed. Certain protozoa and bacteria use genetic codes only minimally different from the near universal code, suggesting that changes to the codon assignments are possible during evolution, and may or may not have an evolutionary advantage in the organisms.

The present code may not have been the best code selected for in the distant past even before evolution at the organism level commenced. But it is the code that was carried forward during evolution. However, the exact mechanism that generated the code is still unknown. Nature may have experimented with numerous different genetic codes and the best code was selected/frozen (Crick, 1968), or only a singular code may have been generated and frozen. The correct answer is not known.

One possibility is to seek out DNA and/or RNA sequences common to many genomes, or RNA relics such as tRNAs (Demongeot et. al., 2009). Jose et al. (2009) suggested that most present day prokaryotes may harbor relics of an ancient RNA world that are two plausible routes between two RNA codes (extended RNA code type I and extended RNA code II) and the current standard genetic code (SGC). The enigma is still how any code originated to generate a life form.

In addition, MicroRNAs (miRs) of 10 to 30 bases have been reported in animals, plants, bacteria and viruses where they have conserved homologies (Demongeot et. al., 2009). This may provide a clue that these sequences were derived from a common distant ancestor(s) and are possible molecular remnants. Some functions provided by these miRs include base-pairing with mRNAs to inhibit their translation of mRNAs (silencing), the degradation of their target sequences and they can also be down-regulated and silenced by other miR duplexes (Demongeot et. al., 2009). The following question can be posed. Were these small RNA sequences involved in a mechanism for the assembly of larger

RNA sequences such as by duplication of the sequence or splicing sequences together to produce codons with a specific function for eventually cells? This begs the question- what functions would be necessary for life? Today we know what genes are required for life in a minimal microbial genome. But how would these genes be decided upon by nature as the necessary genes? The other question is-what was the origin of the initial small RNA sequence? Perhaps another perspective on this enigma is- since there was no mechanism to know which genes were required for the first living cell(s), nature experimented with many iterations of the sequences until one emerged as a functional self-catalytic sequence that also acted as a sequence template for a primitive form of translation. The hypothesis of repeated iterations provides a cycling mechanism where ongoing cycles of organizing a short nucleic acid sequence possibly with the assistance of heating/cooling cycles (thermosynthesis) and EM radiation (especially infrared) providing an energy input for bond stretching and bending (still hypothesized only) allowed a sequence to be organized and then selected with catalytic activity. This could be hypothesized as trial and error iterations until a self-replicating and catalytic sequence emerged. It is also recognized that the iteration mechanism may not be correct as anything can be hypothesized.

The role of EM radiation in the origin of life should not be overlooked (Trevors & Pollack, 2011). Infrared radiation (IR) structures water and causes charge separations in water. This would be useful in creating a differential charge in water and eventually across a cell membrane, which eventually was assembled during evolution, as in present day cells. Above absolute zero, atoms in molecules are in continuous motion. If the frequency of the vibration is equal to the frequency of an IR source, the molecule absorbs the radiation. The two major molecular vibrations are bond stretching and bending. As the IR radiation is absorbed, the associated energy is converted into vibrations which need to be better examined for their role in the origin of life.

The role played by IR radiation in water and the organization of pre-biotic cells, and then evolving cells/organisms still remains to be determined. Just because an EM force was present, does not mean it was involved in the organization of pre-biotic cell(s) the first

cell(s) capable of growth and division and subsequent evolution. The counter argument is EM radiation was central to evolution because of the eventual origin of light-capturing pigments in microorganisms, photosynthesis, and oxygen generation that provided an aerobic biosphere on the Earth. It may be correct to investigate and understand if IR radiation had profound effects on bond stretching and bending which may have played a role in structuring water, charge separations, polymer formation and bonding reactions in the initial absence of enzyme catalysts.

Long wave IR radiation does not have the damaging effects that UV and X-rays have on living organisms. The white light wavelength of the EM spectrum is not damaging to living organisms, but necessary for photosynthesis. A rationale question is- did white light and the IR wavelengths contribute to the organization of life, and what was the mechanism? What type of selection did the EM spectrum exert on the pre-cells and then living cells to enable the evolution of light-capturing pigments

Unifying connections in the origin of life:

The unifying connections in science that may lead to a better understanding of the origin of prokaryotic life and all biology, may be quantum mechanics, the laws of thermodynamics, time, energy, mass, light as both a particle and wave, organic genetic instructions and a better understanding of the dual nature of matter when quantum events are manifested as classical events, or conversely classical events are organized quantum events, that are made possible by genetic instructions and the translated proteins in living organisms. The quantum basis of matter is therefore manifested in living cells through the correct, functional, organic, genetic instructions. The genetic instructions may be the connector or bridge between quantum and classical events in living organisms (Trevors & Masson, 2011).

Organic genetic instructions:

The profound and immense confusion around the origin of life is with the origin of the organic, genetic, instructions that are the blueprints for the actual cells. For cells to use genetic instructions, the coded instructions must be decoded to produce the correct proteins, and assemble cells capable of growth and division. It is like Morse code, both the transmitter and the receiver must know the common code is, or no message can be transmitted that contains information. How would a genetic code be organized in a cell without a yet known mechanism to organize it? This is a primary impediment to better understanding the origin of life with a universal genetic instruction code that has existed for billions of years. How does an evolving cell know which protein is required? A possible answer is a type of iteration or trial and error combinations via a cycling mechanism until the best nucleic acid sequence and protein emerges. Any gene and protein is as useless as another if there is no mechanism to know which specific sequence is required and functional? This is an immense knowledge gap.

Gel-like cytoplasm:

Modern day cells have membranes. Therefore, it is often assumed the first cells on the Earth were contained within a microscopic boundary layer or primitive membrane such as lipid vesicles or protein microspheres. Possibly membranes were not originally necessary for the origin of life (Trevors & Pollack, 2011). Pre-biotic vesicles may have been too fragile whereas as a gel structure is robust, cohesive and able to have functions. A hydrogel environment provides a matrix conducive to both non-enzymatic chemical and enzymatic reactions (Trevors & Pollack, 2005). Enzymatic reactions can occur in narrow, nanometer-scale pores, where water is almost certainly structured. Thus, a gel environment may have provided a more stable environment, while at the same time allowing chemical and eventually enzymatic reactions to occur, with enzymes remaining active for longer periods of time. Any enzymes necessary for assembly of the first cell(s) that remained active longer would have been more useful than enzymes with a rapid turnover, especially if the enzymes were synthesized slowly and/or were slow acting, and

with the capability to initially catalyze the transformation of numerous substrates.

Central to the assembly of the first cell(s) would have been the mechanisms of energy production, storage and utilization. What advantages would a gel-like cytoplasm have offered for these processes, compared to an aqueous cytoplasm? First, in the pre-cell(s), all components would be held in a flexible but not too viscous gel that physically maintained components in a microscopic physical location over long periods of time. Without physical contact, reactions cannot proceed. Further, the gel state brought some order to an otherwise disordered environment, was simpler to encapsulate with an evolving cytoplasmic membrane, diffusion was possible, and solutes could be concentrated that were necessary for the evolution of cellular metabolism and cell structures (Trevors & Pollack, 2005, 2011).

A cohesive gel cytoplasm with an initial distinct boundary can also allow hydrogen and some other gases and molecules to enter *via* diffusion. This entry may be significant as hydrogen was ubiquitous on the early Earth, and may have been a universal energy source. Present-day bacterial cells can utilize hydrogen in some biochemical reactions and some coliform bacteria still utilize hydrogen as an energy source, but not for growth. Growth requirements are now satisfied by other substrates. A simple organizing gel, permeable to hydrogen by diffusion, may have been the precursor to the origin of the first cell(s), which plausibly used hydrogen as a universal energy source (Trevors & Pollack, 2005). The gel environment also facilitates processes such as growth and division. A cohesive gel-like cytoplasm is an easier environment to partition into two entities without cytoplasm streaming away. The mechanism that divides a single bacterial cell into two offspring cells, for example, accomplishes this activity with no loss of genetic material or cytoplasmic contents.

IR radiation and structured water (exclusion zone (EZ) water).

It is hypothesized that EM radiation generates an exclusion zone (EZ) with a negative charge inside of gels and also outside, immediately adjacent to the gel boundary (Trevors & Pollack, 2011). The gel may have been attached to a mineral surface but this is not necessary. The gel with the EZ is the required structure. The charge differential in the -

100 to -200 mv range at the EZ boundary may have been a plausible location for the latter assembly of a primitive cytoplasmic membrane. (Trevors and Pollack, 2011). Present day bacterial cytoplasmic membranes (e.g., *E. coli*) also have a charge differential in the – 85 to - 154 mv range depending on the pH; with an average change of about -22 mV per pH unit. This charge differential is in the same magnitude as that generated in the exclusion zone. This may be coincidence or it may have significance in establishing a charge boundary for eventual membrane organization. It is hypothesized that the initial processes in the organization of pre-biotic life were physical-chemical events with no organic, genetic instructions present. As organization proceeded, the organizing microscopic gel structure made the transition to a gel with a charged boundary layer to which phospholipids, proteins and a more rigid cell wall were attached once organic, genetic instructions were available and enclosed in the gel-like cytoplasm (Trevors & Pollack, 2005, 2011).

The minimal genome:

The enigma of what genes are essential for minimal prokaryotic has been a challenge (Delaye and Moya, 2010; Gil et al., 2004; Hutchinson & Montague, 2002; Islas et al., 2004; Itaya, 1995; Koonin 2000; Mushegian 1999; Mushegian & Koonin, 1996; Smalley et al., 2003; Stano et al., 2011). The minimization of a bacterial genome is a useful exercise to determine the minimal or core genome (genome that contains minimal genes for cell growth and division; complete cell cycle) essential cell growth and division. However, the exact definition of a universal minimal genome is still being debated. Pan-genome (set of all genes present in a group or genus of organisms) analysis of 573 bacterial genomes revealed about 250 genes belonged to the bacterial core genome (genes encoding for translation, replication and energy homeostasis) (Lapierre & Gogarten, 2009). This indicated that during several billion years of evolution, the core genome remained in the genomes studied. This core genome can be understood as the set of central genes upon which the remaining genomes are organized (Lapierre & Gogarten, 2009). The core genome provides the genes that can be duplicated and mutated during

evolution leading to new genes, as well as some other genes being acquired by gene transfer events in bacteria (i.e., transformation, conjugation and transduction).

The origin of the mechanism to store the correct, genetic information in a code-decode mechanism where the correct code is translated at the ribosome into the correct protein, is not known. This is a profound enigma as a testable, plausible, hypothesis with experimental data is not easily available. When the genetic information that is needed for the correct functional protein is not known, then it is irrelevant which protein is translated. Maybe this provides a clue to the origin of the genetic instructions and the corresponding functional proteins that must be correct for the corresponding substrate(s). The microscopic size of the pre-biotic gel cytoplasm would provide a stable environment for the countless interactions between simple prebiotic nucleic acids sequences, peptides and simple substrates which then became the eventual substrates for the first protein catalyzed reactions. However, the problem of the genetic code origin is still not solved.

The code-decode genetic system was then fixed so the correct genetic instructions at the correct time were transcribed and translated (gene expression). Also, replication of the instructions was possible. No plausible hypothesis to date has been put forth that addresses the exact origin (and time, location) of genetic instructions. However, genes are expressed in dead bacterial cells for short periods of time (minutes). Therefore, a form of primitive gene expression may have existed in prebiotic cells if the code was translated and the matching enzymes and substrates were present in a gel cytoplasm, and the microscopic environmental conditions were not too harsh.

If the genetic code proceeded through countless iterations until the best code was obtained for cell growth and division which was then frozen, what was the mechanism? The iteration process or mechanism was then lost from the cell(s) once the genetic code was fixed and evolution proceeded. One can ask- are any remnants of this iteration mechanism present in any living organisms, especially bacteria? To date none have been discovered. From an evolutionary perspective it would be best to jettison this mechanism

once the genetic code was fixed. A hint however may be present in the complex machinery of central intermediary metabolism. Numerous biochemical reactions in cells are components of biochemical cycles that are a form of iteration (repetition of a process). Hence, iteration mechanisms are part of the evolutionary history of biochemical reactions in cells. A genetic instruction iteration process, yet to be discovered, may have been central to the origin of the best, now virtually fixed, genetic code.

It is also evolutionary interesting that prokaryotes do not exhibit a trend to increase their genome sizes.. In microbial cells the energetic cost of protein synthesis is about 37 times the cost of DNA replication. Any significant increase in the genome size results in an immense increase in the bioenergetics of protein synthesis, which is thermodynamically not possible for living microbial cells. Eukaryotic cells are not at any disadvantage because abundant mitochrondria per cell (including core mitochrondrial genome that codes for electron transport metabolism) are present. There is no significant increased energy penalty for an expanded genome as the number of proteins eukaryotic organisms can assemble during evolution, can increase without an energy penalty.

This knowledge can be applied to the origin of the first bacterial cells. It is logical to hypothesize the first bacterial cells would contain small core genomes, and the corresponding translated proteins were possible from a bioenergetic perspective. This provides a possible clue that organization of the first genetic instructions in the first cell(s) used a mechanism that selected for a bioenergetically favourable genetic code and the translated proteins. As discussed in this article a possible mechanism may have been an immense number of cycles where different combinations of a very small nucleic acid code (possibly produced by self-catalytic cutting and splicing mechanism) and simple peptides finally were selected as bioenergetically possible combinations for active and stable enzyme conformations. The selection mechanism and how the functions were selected for, are still research challenges. The mechanism that provided the interface between the linear sequence of nucleic acids (first genome) and the translated proteins that produced the first living phenotype may have required many different iterations

before the code was virtually frozen and the organic genetic code solution to life emerged.

This can also be viewed as anabolic metabolism to organize the first living cell. The situation is one where scientists are challenged with questions from events that occurred billions of years ago at still unknown locations(s), and very plausibly where replication of genetic instructions was not perfect, and enzymes may have been slower and less specific or absent for some period of time. It is a profound challenge to obtain the correct answer for the origin of life when one considers the no genome without the presence of the correct enzymes, and no functional enzymes in the absence of the corresponding correct genomes. Possibly, an iteration selection mechanism placed the first correct gene sequence and protein sequence at the same microscopic location, and the code was expanded by more iterations, and then virtually frozen using a minimum number of amino acids (i.e. 20). The non-life to life transition then followed in a stable and simple microscopic environment. This transition was also dependent on the prebiotic availability of both the code and amino acids and if functional proteins were produced and maintained in a microscopic location such as a gel or primitive cytoplasm.

Acknowledgements: This research was supported by a NSERC (Canada) Discovery Award to J.T.T.

References:

Crick, F.H.C. (1968). The origin of the genetic code. J. Mol. Biol. 38, 367-379.

Delaye, L, Moya A. (2010). Evolution of reduced prokaryotic genomes and the minimal cell concept: variations on a theme, BioEssays 32, 281-287.

Demongeot J, Glade, N., Moreira, A., Vial, L. (2009). RNA relics and origin of life. International Journal Molecular Sciences, 10, 3420-3441.

Gil, R., Silva, F.J., Pereto, J., Moya, A. (2004). Determination of the core of a minimal bacterial gene set. Microbiology Molecular Biology Reviews, 68, 518-537.

Hutchison III, C.A., Montague, M.G. (2002). Mycoplasmas and the minimal genome concept. In Molecular Biology and pathogenicity of Mycoplasmas. (Eds: Razin and Herrmann, Kluwer, New York. pp. 221- 253.

Islas, S., Becerra, A., Luissi, P.L., Lazcano, A. (2004). Comparative genomics and the gene complement of a minimal cell. Origins Life Evolution Biosphere, 34, 243-256.

Itaya, I. (1995). An estimation of minimal genome size required for life. FEBS Letters. 362, 257-260.

Jose, M.V., Govezensky, T., Garcia,J.A., Bobadilla, J.R. (2009). On the evolution of the standard genetic code: vestiages of critical scale invariance from the RNA world in current prokaryote genomes. PLOS One. 4(2) e340.

Joseph, R., Wickramasinghe, C. (2011). Genetics indicates extra-terrestrial origins for genetic life and the first gene. Journal Cosmology, in press.

Koonin, E.V. (2000). How many genes can make a cell: the minimal gene concept. Annual Review Genomics Genetics, 1, 99-116.

Lapierre, P., Gogarten, J.P. (2009). Estimating the size of the bacterial pan-genome. Trends Genetics, 25, 107-110.

Mushegian A. (1999). The minimal genome concept. Current Opinions Genetics Development, 9, 709-714.

Mushegian, A.R., Koonin, E.V. (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proceedings National Academy of Science, 93, 10268-10273.

Poetsch, A., Wolters, D. (2008). Bacterial membrane proteins. Proteomics, 8,4100-4122.

Smalley, D.J. Whitely, M., Conway, T. (2003). In search of the minimal *Escherichia coli* genome. Trends Microbiology, 11, 6-8.

Stano, P., Ferri, F., Luisi, P.L. (2011). Semi-synthetic minimal living cells. In Chemical Synthetic Biology (Eds. P. Luigi and C. Chiarabelli) John Wiley and Sons.

Trevors, J.T. **(**2010a)**.** Researching the transition from non-living to the first microorganism: methods and experiments are major challenges. Journal of Microbiological Methods. 81, 259-263.

Trevors, J.T. (2010b). Suitable microscopic entropy for the origin of microbial life: microbiological methods are challenges. Journal of Microbiological Methods, 83, 341-344.

Trevors, J.T. (2011a). The composition and organization of cytoplasm in prebiotic cells. International Journal Molecular Sciences 12, 1650-1659.

Trevors, J.T. (2011b). Hypothesized origin of microbial life in a prebiotic gel and the transition to a living biofilm and microbial mat. Comptes Rendus Biologies Academie des Sciences, Paris 334,269-272.

Trevors, J.T, (2004). Evolution of cell division in bacteria. Theory in Biosciences. 123/1, 3-15.

Trevors, J.T. (2003). The possible origin of a membrane in the subsurface of the Earth. Cell Biology International, 27, 451-457.

Trevors, J.T., Masson, L. (2011). How much bacterial cytoplasm and cell size can a bacterial genome control? Journal of Microbiological Methods. 84, 147-150.

Trevors, J.T., Masson, L. **(**2011). Quantum microbiology. Current Issues Molecular Biology 13, 43-50.

Trevors, J. T., Pollack, G.H. (2011). Origin of life hypothesis: a gel cytoplasm lacking a bilayer membrane, with infrared radiation producing exclusion zone (EZ) water, hydrogen as an energy source and thermosynthesis for bioenergetics. Biochemie (in press).

Trevors, J.T.**,** Pollack, G. H. (2005). Hypothesis: the origin of life in a hydrogel environment. Progress Biophysics. Molecular Biology, 89,1-8.

Trevors, J. T., Pollack, G. H.,Saier, M. H., Masson, L. (2011). Transformative research: definitions, approaches and consequences. C. R. Chimie (submitted).

List of Figures: