A̱RJ

Facts Cannot be Ignored When Considering the Origin of Life #1: The Necessity of Bio-monomers Not to Self-Link for the Existence of Living Organisms

Change Laura Tan, Division of Biological Sciences, University of Missouri, Columbia, MO 65211

Abstract

A system for coding and decoding genetic information is essential to every organism. For an organism to survive or to reproduce, its genomic DNA must be replicated, and genes encoded in its genome must be transcribed and translated. The idea of abiogenesis holds that life must have come from non-life spontaneously. First the basic building blocks of life, including amino acids, ribose-based sugars, and nucleotides, or something similar, must have appeared. These monomers must then have self-linked to form polymers, including DNA, RNA, and proteins, or something similar. The resultant polymers then functioned as a genetic information coding and decoding system. In this essay, I argue that the required automatic self-linking of monomers for abiogenesis is incompatible with the genetic information coding and decoding system that is necessary for life.

Introduction

Abiogenesis, the idea that life came from non-life spontaneously from non-living materials, was already prevalent at the time of Aristotle (384–322 BC), but he is largely credited with formalizing the concept (Tan and Stadler 2020). Later the idea was considered by Charles Darwin (1809–1882) (Darwin 1871), Alexander Oparin (1894–1980), and John Haldane (1892–1964), among others, and then explored experimentally by Stanley Miller (1930–2007), Harold Urey (1893–1981), and their followers (Krishnamurthy 2018a , 2018b; Mariscal et al. 2019; Maurel, Malaterre, and Grandcolas 2015; Preiner et al. 2020). The genetic information coding and decoding systems have been deciphered mostly in the last several decades, starting with the discovery of the double strand structure of DNA by Watson and Crick (Watson and Crick 1953b). It is a striking curiosity of history that Miller's seminal work and that of Watson and Crick were published in the same year (Miller 1953; Watson and Crick 1953b), one in *Science* and the other in *Nature*.

In this article, I will show that the only way for the genetic information coding and decoding systems to work is for abiogenesis—as currently proposed—not to work. I will first describe some very basic facts about genetic information coding and decoding, including DNA replication, transcription, and translation. These facts are so familiar and so elementary for us that we become blind to their significance, regarding them the only way of nature. I will then briefly discuss how the basic building blocks of life, i.e., bio-monomers, or similar molecules behave, apart from enzymes found in cells (a more detailed discussion is provided in "Facts Cannot be Ignored When Considering the Origin of Life #2: Challenges in Generating the First Gene-encoding Template DNA or RNA"[Tan 2022b]). After that, I will illustrate how self-linkage of monomers destroy the genetic information coding and decoding systems.

Genetic Information Coding and Decoding Systems

All cellular organisms depend on a functional genetic information, or biological information, coding and decoding system. This includes DNA replication, gene transcription, and gene translation (fig. 1). An astonishing discovery of molecular biology is that each of the three domains of life (Bacteria, Archaea, and Eukaryotes) has its own genome replication system and transcription and translation systems (Cserhati and Shipton 2017; Tan and Tomkins 2015a, 2015b). Organism-specific genetic information coding and decoding systems will be further addressed in "Facts Cannot be Ignored When Considering the Origin of Life #3" (Tan 2022c).

DNA replication

DNA replication is the process by which two copies of genomic DNA are made from one copy. DNA replication is not merely synthesizing DNA, i.e., linking deoxyribonucleotides together via phosphodiester bonds to make a polydeoxyribonucleotide chain. DNA replication, in fact, requires the synthesis of a new polydeoxyribonucleotide chain whose nucleotide sequence is the same as that of its parental chain, using a DNA template whose nucleotide sequence is
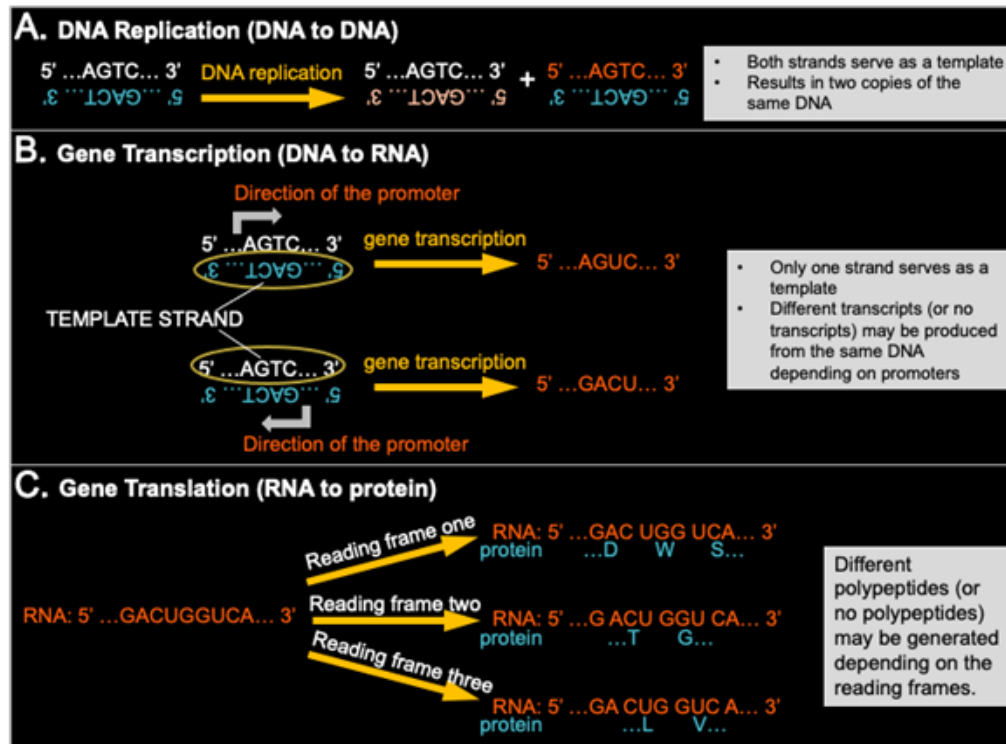
Fig. 1. Genetic information coding and decoding. A, G, T, C represent deoxyribonucleotides in DNA or ribonucleotides in RNA. (A) DNA Replication (DNA to DNA). (B) Gene Transcription (DNA to RNA). Gene promoters are indicated with gray curved arrows. (C) Gene Translation (RNA to protein). Amino acids are represented with single letters. Spaces are added for codon visualization.

reverse and complementary to that of the parental chain (Watson and Crick 1953a) (fig. 1A). This is critical for genetic information maintenance and propagation. Cells use molecular machinery comprised of multiple proteins to replicate their genomic DNA accurately via proofreading of DNA polymerization and mismatch repairing of errors introduced by DNA polymerases. For example, *E. coli* DNA polymerase synthesizes DNA with an error rate (per base pair per round of replication) roughly of 10^{-5} without proofreading, 10^{-7} with proofreading, and 10^{-10} with both proofreading and mismatch repair (Fijalkowska, Schaaper, and Jonczyk 2012).

Accurate DNA replication requires its composition deoxyribonucleotides, 2'-deoxyadenosine (dA), 2'-deoxyguanosine (dG), 2'-deoxycytidine (dC), and thymidine (dT), to be free and to be not free simultaneously. The deoxyribonucleotides are free in the sense that any deoxyribonucleotide, whether it is a dA, a dG, a dC, a dT, can be followed by the same deoxyribonucleotide, any other deoxyribonucleotides, or any combinations of deoxyribonucleotides. The deoxyribonucleotides are not free in the sense that each nucleotide in the daughter DNA must follow the sequence of the parental DNA.

The function of the DNA replication system requires deoxyribonucleotides not to link together by

themselves, either in the form of deoxyribonucleoside monophosphate, diphosphate, triphosphate, or their substitutes. Free, random linking of deoxyribonucleotides by themselves would generate random DNA, mostly unrelated to the parental DNA. The result is the failure of information inheritance. Furthermore, linking the deoxyribonucleotides in a sequence based on their chemical or physical affinities, instead of on the deoxyribonucleotide sequence of the template, also renders information inheritance impossible (in addition to greatly limiting the information content that can be generated).

This is very similar to human languages, of which I will use English as an example. First, for a sentence that is represented with a string of letters to have a specific meaning (i.e., to function), the letters it contains must not link together by themselves. The order of the letters of that sentence should be free to change as required by its meaning, not by the chemical or physical affinities of the letters. Second, the letters in the alphabet do not have intrinsic values to the message being conveyed. The message is determined by the speaker (or writer). Therefore, to use the letters to represent the English language, the letters must be free and not free at the same time. The letters must not link together by themselves, either randomly or by specific physical, chemical, or

other fixed rules other than the rules of the language itself (e.g., grammars and specific spelling of words at a given historical time or in a specific English-speaking country). Random linking of the letters will destroy the message. Linking by their physical, chemical, or other fixed rules other than the rules of the language will also render them useless for communication.

For example, we should be allowed to say that "Tom had a fish for breakfast" if Tom did have a fish for breakfast. In the unfortunate situation when the opposite occurred, we should be allowed to rearrange the letters and say that "A fish had Tom for breakfast." There are many ways to randomly arrange the same letters, but most of the arrangements would be gibberish. To arrange these letters in a fixed manner will also make only gibberish. For instance, if the letter "a" has to be followed by the letter "b", which has to be followed by the letter "c", so on and so forth down the English alphabet, the same letters in "Tom had a fish for breakfast" can make only three English words: "a", "I", "hi", provided that spaces can be inserted freely in between letters. In such restricted condition, the alphabet becomes useless for the English language.

Gene transcription

Gene transcription generates RNA molecules, but it is not merely linking ribonucleotides together via phosphodiester bonds to make a poly-ribonucleotide chain. Gene transcription synthesizes RNA molecules whose sequences are determined by the sequences of their DNA templates (fig. 1B). For each gene, only one strand of the double-stranded DNA serves as a template. From the same segment of DNA, none or multiple different transcripts can be generated, depending on the presence or absence of promoters, number of promoters, directions of promoters, and location of promoters.

Successful gene transcription necessitates that ribonucleotides do not link together by themselves, either in the form of nucleoside monophosphate, diphosphate, or triphosphate.

Using the analogy of human languages again, gene transcription is to copy segments of DNA, except replacing "T (thymine)" with "U (uracil)." The usefulness of a scribe depends on how accurately he/she can copy the text in front of him/her, with the correct starting and ending points, not how able he/she is to add, omit, or rearrange the words or letters.

Gene translation

Gene translation generates polypeptides (or proteins), but it is not merely linking amino acids together via peptide bonds to make a polypeptide chain. Functional gene translation makes polypeptides whose sequences are determined by the sequences of RNAs being translated (fig. 1C). Different polypeptides (or no polypeptide) may be generated depending on the presence or absence of appropriate translation starting sites and their reading frames.

Successful gene translation demands that the amino acids do not link together by themselves, either in their free form or in their active form (e.g., in the aminoacyl-tRNA form).

To stay with our analogy of human languages, one can surely write any essays. However, the usefulness of a translator depends on how well he/she can accurately translate the meaning of the original language, say Chinese, to the target language, say English. To write an essay in English is not the same as to translate a Chinese essay into English. Faithful to the meaning of the original language is crucial for a good translation.

Making DNA whose sequence is determined by the sequence of the parental DNA, making RNA whose sequence is determined by the sequence of its DNA template, and making proteins whose sequence is determined by the sequence of the RNA being translated is essential for genetic information coding and decoding and vital for each and every organism. No exceptions have ever been observed.

In short, for genetic information coding and decoding, including DNA replication, transcription, and translation systems to function, the compositional monomers of DNA, RNA, and proteins must be free and not free at the same time to ensure that the encoded information can be decoded and be transmitted adequately.

The paradox of genetic information coding and decoding

Bio-monomers being free and not free simultaneously, this is the paradox of genetic information coding and decoding. It has three components: (1) the necessity of particular sequences of nucleotides in DNAs and RNAs and amino acids in proteins for their functionality, (2) the lack of any chemical or physical laws, apart from living organisms, to account for the sequences of DNAs, RNAs, and proteins, and (3) the essentiality of accurately transferring the sequential information from DNA to DNA during DNA replication, from DNA to RNA during transcription, and from RNA to proteins during translation.

Self-linkage-depending Abiogenesis Destroys Genetic Information Coding and Decoding Systems

Abiogenesis refers to the idea that life spontaneously came from non-life. Several abiogenesis concepts have

been proposed, including the RNA first, protein first, DNA first, and RNA-protein-DNA together models (Krishnamurthy 2018a, 2018b; Mariscal et al. 2019; Maurel, Malaterre, and Grandcolas 2015; Preiner et al. 2020). In all models, abiogenesis begins with making basic building blocks of life (bio-monomers), including amino acids, ribose-based sugars, and nucleotides, or something similar (e.g., the assumed proto-nucleic acid-generating monomers (Fialho et al. 2021)), from simple compounds (e.g., CH_4, NH_3, H_2, H_2O [Miller 1953]) that presumably existed on the pre-biotic earth. These monomers then joined together to form biopolymers, including DNA, RNA, and proteins, or something similar. These biopolymers eventually self-organized, generating a genetic information coding and decoding system and, ultimately, life as we know it. This is why the Miller-Urey Experiment (Miller 1953) has been so highly regarded.

The two presumptions, normally unstated, of abiogenesis are that life was not created by God and that uniformitarianism is true. Uniformitarianism is the basis of evolution. According to creation, initial circumstances could have been different, during which time God created the world. Ironically, the uniformitarianism is practically denied by its practitioners whenever it is inconvenient, including when claiming that there was a time when no life existed on earth and that there was no oxygen (thus the early earth had a reducing atmosphere) when the first organism and its components were slowly formed via chemical evolution (Krishnamurthy 2018a, 2018b; Malaterre 2015; Mariscal et al. 2019).

For abiogenesis studies carried out today in our laboratories to be of any use for understanding how life came about, uniformitarianism has to work. Thus, chemical laws applied yesterday, apply today, and will apply tomorrow. There is no reason to suspect that they will ever change. The basic building blocks of life, including amino acids, ribose sugars, and nucleobases, or something similar, either self-link or not. Were they able to self-link before, e.g., in the pre-biotic timeframe, then we should see them self-linking today, inside or outside cells. Had they not been able to self-link before, then they should not be self-linking today. Their chemical properties should have remained the same.

But this presents a catch-22 problem. Abiogenesis requires the bio-monomers (or similar chemicals) to self-link to form biopolymers, generating a genetic information coding and decoding system. But self-linking bio-monomers (or similar chemicals) are incompetent to gain any sort of coding and decoding system. In fact, self-linking bio-monomers (or similar chemicals) destroy the genetic information coding and decoding systems.

For example, instead of duplicating and passing on a copy of genomic DNA via DNA replication, a cell with self-linking nucleotides would be filled with DNA randomly assembled by self-linking nucleotides and be more likely than not to pass on a random DNA molecule to its daughter cells. For example, the isomer number of a di-deoxynucleotide dAdC is 16,200 (a detailed calculation is described in "Facts Cannot be Ignored When Considering the Origin of Life #2" [Tan 2022b]). If you want to have a genome of two base pairs, with one strand 5'-dAdC-3' and the other 5'-dGdT-3' (i.e., 3'-dTdG-5'), the number of isomers of that genome will be 145,800,000, ignoring

- all the isomers of D-2-deoxyribose that are not a D-2-deoxyribose,
- all the isomers of A/T/G/C that are not A/T/G/C, and
- all the isomers of linking the two sugar molecules through non-phosphodiester bond.

One correct sequence in a sea of 145,800,000 isomers for a genomic DNA that is two basepairs long! In comparison, the human genome contains about three billion basepairs, and the genome of *E. coli* (*Escherichia coli*), a bacterium inside the human gut, is about 4.6 million basepairs.

The problem is even worse with RNA than with DNA. For a single-stranded RNA with two nucleotides AC, the number of isomers is 129,600. If you wish to have a double-stranded RNA with two base pairs, of which one is AC, you will have 18,662,400,000 ways of linking its constituting components together. Only one of these would be allowed if cellular enzymes were performing the synthesis.

Likewise, instead of making proteins encoded in its genome, a cell with self-linking amino acids would be filled with polypeptides randomly assembled by self-linking amino acids.

We do have genetic information coding and decoding systems, at least one for bacteria, another for archaea, and yet another for eukaryotes (Tan and Tomkins 2015a, 2015b). This is only possible because amino acids, ribose sugars, and nucleotides do not self-link.

Conclusion

The function and usefulness of genetic information coding and decoding systems depend on the fact that amino acids, ribose sugars, and nucleotides do not self-link to form biopolymers, including proteins and nucleic acids. However, abiogenesis needs amino acids, ribose sugars, and nucleotides (or similar chemicals) to self-link to form biopolymers. If such self-linkage had occurred, then it would have destroyed the systems of genetic information coding and decoding and the life that depends on the systems.

References

Cserhati, Matyas , and Warren Shipton. 2017. "The Irreducibly Complex Ribosome is a Unique Creation in the Three Domains of Life." *Journal of Creation* 31, no. 2 (August): 94–102.

Darwin, Charles. 1871. "From Darwin's letter to the botanist and explorer Joseph D. Hooker, February 1, 1871." *Darwin Correspondence Project*. https://www.darwinproject.ac.uk/letter/DCP-LETT-7471.xml

Fialho, David M., Suneesh C. Karunakaran, Katherine W. Greeson, Isaac Martinez, Gary B. Schuster, Ramanarayanan Krishnamurthy, and Nicholas V. Hud. 2021. "Depsipeptide Nucleic Acids: Prebiotic Formation, Oligomerization, and Self-Assembly of a New Proto-Nucleic Acid Candidate." *Journal of the American Chemical Society* 143, no. 34 (1 September): 13525–13537.

Fijalkowska, Iwona J., Roel M. Schaaper, and Piotr Jonczyk. 2012. "DNA Replication Fidelity in *Escherichia coli*: A Multi-DNA Polymerase Affair." *FEMS Microbiology Reviews* 36, no. 6 (5 April): 1105–1021.

Krishnamurthy, Ramanarayanan. 2018a. "Experimentally Investigating the Origin of DNA/RNA on Early Earth." *Nature Communications* 9, no. 1 (12 December): 5175.

Krishnamurthy, Ramanarayanan. 2018b. "Life's Biological Chemistry: A Destiny or Destination Starting from Prebiotic Chemistry?" *Chemistry* 24, no. 63 (13 November): 16708–16715.

Mariscal, Carlos, Ana Barahona, Nathanael Aubert-Kato, Arsev Umur Aydinoglu, Stuart Bartlett, María Luz Cárdenas, Kuhan Chandru et al. 2019. "Hidden Concepts in the History and Philosophy of Origins-of-Life Studies: A Workshop Report." *Origins of Life and Evolution of Biospheres* 49, no. 3 (September): 111–145.

Malaterre, Christophe. 2015. "Chemical Evolution and Life." *BIO Web of Conferences* 4. https://doi.org/10.1051/bioconf/20150400002.

Miller, Stanley L. 1953. "A Production of Amino Acids Under Possible Primitive Earth Conditions." *Science* 117, no. 3046 (15 May): 528–529.

Preiner, Martina, Silke Asche, Sidney Becker, Holly C. Betts, Adrien Boniface, Eloi Camprubi, Kuhan Chandru et al. 2020. "The Future of Origin of Life Research: Bridging Decades-Old Divisions." *Life* 10, no. 3, 20. https://www.mdpi.com/2075-1729/10/3/20/htm.

Tan, Change Laura, and Rob Stadler. 2020. *The Stairway To Life: An Origin-Of-Life Reality Check*. Bucharest, Romania: Evorevo Books.

Tan, Change, and Jeffrey P. Tomkins. 2015a. "Information Processing Differences Between Archaea and Eukarya—Implications for Homologs and the Myth of Eukaryogenesis." *Answers Research Journal* 8 (25 March): 121–141. https://answersingenesis.org/biology/microbiology/information-processing-differences-between-archaea-and-eukarya/.

Tan, Change, and Jeffrey P. Tomkins. 2015b. "Information Processing Differences Between Bacteria and Eukarya—Implications for the Myth of Eukaryogenesis." Answers Research Journal 8 (25 March): 143–162. https://answersingenesis.org/biology/microbiology/information-processing-differences-between-bacteria-and-eukarya/.

Tan, Change Laura. 2022a. "Facts Cannot be Ignored When Considering the Origin of Life. #2: Challenges in Generating the First Gene-encoding Template DNA or RNA." *Answers Research Journal* 15, in press.

Tan, Change Laura. 2022b. "Facts Cannot be Ignored When Considering the Origin of Life. #3: Necessity of Matching the Coding and the Decoding Systems. *Answers Research Journal* 15, in press.

Watson, J. D., and F. H. Crick. 1953a. "Genetical Implications of the Structure of Deoxyribonucleic Acid." *Nature* 171, no. 4361 (20 May): 964–967. http://www.ncbi.nlm.nih.gov/pubmed/13063483.

Watson, J. D., and F. H. Crick. 1953b. "Molecular Structure of Nucleic Acids; A Structure for Deoxyribose Nucleic Acid." *Nature* 171, no. 4356 (25 April): 737–738. http://www.ncbi.nlm.nih.gov/pubmed/13054692.