THE ORIGIN OF THE GENETIC CODE

It is important to realize that the DNA molecule is not a code. The sequence of the adenine(A) guanine (G) , thymine(T) and cytosine (C)molecules that is added to the backbone of the DNA molecule , forms the code. A, G ,T and C can be regarded as the letters of the genetic alphabet. The genetic language consist of three letter words and the first two letters of the word are very important. Every word represents either a specific amino acid, a punctuation mark, a tRNA , rRNA and a possible still unknown function..

A DNA molecule that contains a random sequence of A, G, T and C without forming words will be useless. The sequence must follow a certain pattern. DNA is only a code carrier and a random sequence of adenine ,guanine , thymine and cytosine will be meaningless.

A very important message conveyed by the code is the ability it gives cells to select specific l-amino acids. There also other messages some that are still mysteries.
(People not familiar with DNA can "Google " DNA structure and protein synthesis)

"*In science, it is important to distinguish between an observation and an interpretation. Observations are things we measure; while interpretations are the conclusions we derive from those observations. In well-designed experiments the resulting interpretations are the only possible explanations for the observations—but this is a rare occurrence. More often, alternate interpretations are possible*." (Thomas Philips)

Observations made over the two and a half centuries

All the experiments designed to solve the origin of the genetic code assume that life and the genetic code arise spontaneously from chemical reactions that occurred spontaneously on the pre-biotic earth.

The lack of space won't allow a detail discussion of all the biochemistry involved, the reader that is not familiar with the terminology used can search the internet. (Google it)

Important non disputed facts:

(i) Life is associated with blistering fast chemical reactions that occur mainly in cells. Blistering fast mean reaction rates measured in milliseconds.

(ii) The majority (nearly all) of these reactions occur at a very slow rate in the absence of life. Life produces enzymes that catalyse these reactions to occur at a very fast rate.

(iii) The biological catalysts are mainly enzymes . Some are also known as ribozymes (RNA type molecules).

(iv) The enzymes are very large protein molecules . The amino acid sequence of the protein

molecule determines its important three dimensional configuration.

(v) Enzymes are specific for specific reactions. They won't catalyse other reactions,

(vi) Only a small area of the enzyme act as a catalyst. This area is called the active site.

(vii) The three dimensional orientation and shape of the active site is important and are determined by the amino acid sequence as already mentioned.

(viii) The amino acid sequence is determine by the genetic code that resides on the DNA molecule.

(ix) A corrupt code will produce meaningless proteins without any catalytic ability.

The first half of the 19th century was occupied with debates in biological circles about the question if life spontaneously arose from rotten meat, swamp's mud etc. or from pre-existing life forms.

An excellent summary of these events can be obtained at the following link:

http://www.thesilvertablet.net/WINTERWINTER%201859.pdf

JOURNAL ARTICLE Marginalia: Winter 1859 Robert L. Dorit American Scientist
Vol. 98, No. 4 (July-August 2010), pp. 286-288

Robert Dorit 's comment on Pasteur's famous Swan neck experiments clearly expressed the view of most modern biology scholars:

"*Pasteur, of course, was right, but with one major exception. If we think of contemporary organisms in the present, life begets life, and like begets like. But if we look into the past, we quickly realize that there must have been at least one time when Pasteur's dictum did not hold. Some 3.8 to 4 billion years ago, life on Earth emerged from nonlife*"

What Dorit suggests is the spontaneous emerging of life from non-life It is a hypothesis and not an observation. It is a hypothesis that is still accepted today and is called " the veto on "foresight evolution" ".

Is there any experimental observation that supports this hypothesis.?

Lets check.

If we apply the scientific method , the hypothesis is " life emerge spontaneously from non-life." We are now in need of experiments and observations to support the hypothesis.

1953: Was a very important year for the supporters of " the spontaneous emergence of life" hypophosis.

DNA structure discovered.

Francis Crick and James Watson publish their discovery of the structure of the DNA molecule in *Nature* on the 25 April 1953 issue of Nature.

Amino acids and the primordial earth.

The results of the Miller -Urey experiment was published in the 15 May 1953 issue of *Science.*

Stanley Miller synthesised amino acids by mimicking the presumed primordial earth's atmosphere

Crick and Watson 's discovery opened the door that let to the discovery of the genetic code.

; '
The Miller-Urey experiments and similar experiments that followed confirmed that the pre-biotic earth was able to synthesize the building blocks of life. (sugars, amino acids , purines, pyrimidines and phosphate salts)

Expectations were running high that science was on the verge of solving the mystery of life. A RNA world that eventually evolved into a DNA world appeared to be highly probable. .

The next ten years led to the unravelling of the genetic code , the DNA , m RNA, tRNA aixis communication in protein synthesis and other metabolic pathways , all coded for by the genetic code.

The RNA world preceding the DNA world hypothesis was postulated during this period.

A simplified summary of the generally proposed hypothesis origin and evolution of the genetic code :

Different tRNA molecules developed via chemical processes from sugars, phosphate salts , purine and pyrimidines on the pre -biotic earth.

Some of the tRNA molecules developed an affinity to form chemical bonds with specific amino acids. Different tRNA molecules has a specificy for one particular amino acid and not for other mono acids.

Some of the RNA molecules evolved into rRNA and gave rise to ribozymes. One of these ribozymes were able catalysed peptide chemical bonding between different amino acids , the origin of proteins (large ,long chain polypeptides).

Identifying the amino acid sequence of successful polypeptides could be identified and transferred to mRNA and from mRNA eventually to DNA through a process of natural selection of the most successful chemical pathways.

The genetic code was deciphered and the pre-biotic earth's ability to produce building blocks

was established. A possible hypothesis for the origin of the genetic code was proposed. The rest of the quest to determine life's mysteries was supposed to be plain sailing but it wasn't.

This post is about the origin and evolution of the code and not about the other encountered problems . These other problems will just be mentioned here, without discussing them:

(i) The pre-biotic earth produced racemic mixtures.

(ii) Ramemic mixtures cause enantiomeric cross inhibition.

(iii) The rate of synthesis and the concentration of building blocks could lead to shortages of building blocks. (if demands exceed supplies)

(iv) Life is homochyralic for l-amino acids and d- sugars and therefore also for d-nucleotides,

(v) The larger organic molecules like sugars and the very large macromolecules are inherent unstable and prone to decay.

The mid 1990s

A new problem appeared.

Tomas Lindahl and other researchers discover that the DNA molecule was not as stable as what was originally thought.

The 2015 Nobel Chemistry Prize honoured three of these researchers that discovered this instability and the proofreading and error correcting machinery that counteract the instability.

The following is a quote from the Press Release of the Royal Swedish Academy that stresses the importance of these researchers work:

"*Each day our DNA is damaged by UV radiation, free radicals and other carcinogenic substances, but even without such external attacks, a DNA molecule is inherently unstable. Thousands of spontaneous changes to a cell's genome occur on a daily basis. Furthermore, defects can also arise when DNA is copied during cell division, a process that occurs several million times every day in the human body.*

The reason our genetic material does not disintegrate into complete chemical chaos is that a host of molecular systems continuously monitor and repair DNA. The Nobel Prize in Chemistry 2015 awards three pioneering scientists who have mapped how several of these repair systems function at a detailed molecular level.

In the early 1970s, scientists believed that DNA was an extremely stable molecule, but Tomas Lindahl demonstrated that DNA decays at a rate that ought to have made the development of life on Earth impossible. This insight led him to discover a molecular machinery, base excision repair, which constantly counteracts the collapse of our DNA."
(The Royal Swedish Academy of Sciences

https://www.nobelprize.org/nobel_prizes/chemistry/laureates/2015/press.html

)

This code carrier instability has an important consequence. It has a negative impact on the understanding of the origin and development of the genetic code. The error correcting software is specific for DNA and without it a DNA molecule is doomed as a code carrier.

Another recent development is the the successful digitising of the genetic code , that led to the "synthetic" cell of the J Craig Venter Institute.

The scientific community was very excited about the synthetic cell that originated in the laboratories of the J Craig Venter Institute but was it really a synthetic life?

No it was not.

They digitalized (computerize) the DNA of an organism known as *M mycoides*. Computers were then used to copy the digitalized code back into synthetic manufactured DNA . These copies were then introduced to *M capricolum* . It is similar to changing a laptop running under Windows to a Macbook running under iOS , by swapping the operation systems.

Quote *:*
" *The complete synthetic **M. mycoides** genome was then isolated from the yeast cell and transplanted into restriction deficient **M. capricolum** recipient cells (which accept the foreign DNA without breaking it down). Following uptake, the synthetic genome begins to encode all the proteins required for life, including restriction enzymes which degrade the native M. capricolum genome. recipient cells (which accept the foreign DNA without breaking it down). Following uptake, the synthetic genome begins to encode all the proteins required for life, including restriction enzymes which degrade the native **M. capricolum** genome." = the beginning of ' **JCVI- M. mycoides Syn1.0** '*

The only difference between the original *M mycoides* and the *synthetic M mycoide*s is that the synthethic DNA molecule containes a watermark

Quote:
"To distinguish the synthetic genome from native DNA, the researchers incorporated four 'watermark' sequences. Replacing one or more cassettes in regions experimentally demonstrated or predicted not to interfere with cell viability, these watermarks contain strings of bases that, in code, spell out a web address to send emails to if you can successfully crack the new code, the names of 46 authors and other key contributors as well as three famous quotations. One of which by James Joyce, perfectly encapsulates the ups and downs of a the 15 year project—"To live, to err, to fall, to triumph, to recreate life out of life."

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3026460/

[Restriction deficient means *: M capricolum* contains ensymes that destoy foreign DNA. These enzymes were restricted and the *JCVI- M. mycoides Syn1.0* genome produce its own restrictors that destroyed the host's DNA.]

Important to notice that *M mycoides Syn 1.0* contains exactly the same code as *M mycoides* , . Exactly the same sequence of adenine, guanine, thtymine and cytosine as the native *M mycoides,* except for the added "watermark".

A great achievement but is was an intellectual guided and not a spontaneous event.

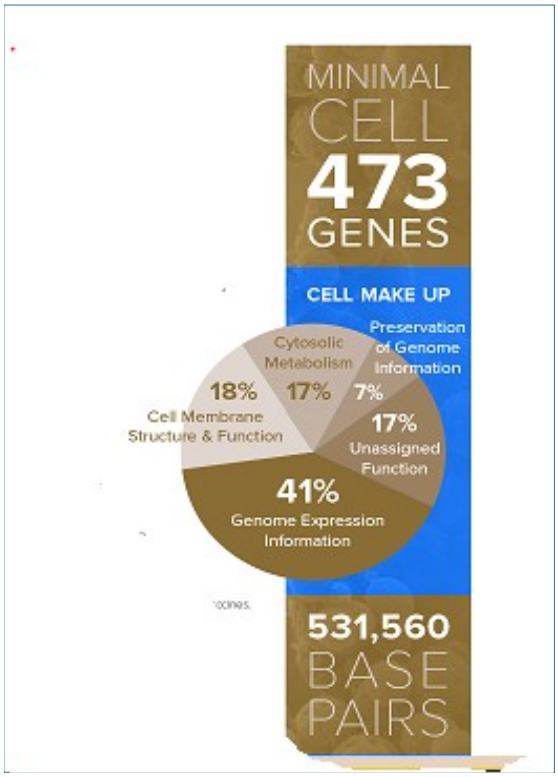
The *JCVI- M. mycoides Syn1.0 's* genome has since been manipulated further.. Its genome size was continuously trimmed until *JCVI-Syn3.0* emerged as a minimal cell.

Quote:
"*A minimal cell is usually defined as a cell in which all genes are essential. This definition is incomplete, because the genetic requirements for survival, and therefore the minimal genome size, depend on the environment in which the cell is grown. The work described here has been conducted in medium that supplies virtually all the small molecules required for life. A minimal genome determined under such permissive conditions should reveal a core set of environment-independent functions that are necessary and sufficient for life. Under less permissive conditions, we expect that additional genes will be required."*

(JCVI
http://www.jcvi.org/cms/research/projects/minimal-cell/overview/

click full text or http://science.sciencemag.org/content/351/6280/aad6253.full?ijkey=77AGRUAdvXIP2&keytype=ref&siteid=sci

Minimal cell summary/
17 % appeared to be junk code but wasn't .



What is life. My definition is : Life is continuous rapid occurring chemical reactions . These reactions occur at blistering speeds measured in milliseconds. The majority of these reactions will occur extremely slowly, or not at all, without enzymes and ribozymes. Life needs a code to control and guide these reactions. Without the code life is not possible.

The JCVI concedes that a cell resembles a computer.

(JCVI

"*A biological cell is very much like a computer—the genome is the software that encodes the instructions of the cell and the cellular machinery is the hardware that interprets and runs the genome software*"
http://www.jcvi.org/cms/research/projects/minimal-cell/overview)

The synthetic cell however has major consequences on our understanding of the origin and further development of the genetic code.

The inherent instability of DNA and fact that it seems that life needs a minimum amount of genetic coding to be able to survive, mystify the origin of life even further.

A major observation is that on top of this minimum amount of necessary code, code that protects the DNA molecule against decay , is also a priority .. . DNA therefore needs to carry code that protects it against itself..

Three important questions arise from observations about DNA's stability and the minimal cell:

(i) Can a complicated code developed on an unstable, prone to fail, code carrier (like a continuously crashing hard disk drive)? How do an unstable code carrier remember the lengthy sequences of a protein molecule's amino acids and the necessary words , grammar and sentences necessary for it to function properly. if it frequently crashes?

(ii) Is it possible for hardware to function without software? Could life (hardware) evolved without previous existing software?

(iii) DNA requires code to be able to produce enzymes that is able to protect its integrity. These enzymes are specific and can only protect the DNA molecule. This observation implies foreknowledge and foresight that goes against a spontaneous unguided origin of the code. (vialation of " the veto on "foresight evolution" ").

Anybody who disagree, should develop a code and use this code to create messages on an unstable code carrier. (A hard disk with a problematic writing device that frequently, randomly scratches the disc surface, continuously making writing errors with the added continuous damage to areas necessary to store memory).

Viruses, both RNA , DNA and viroids (plant pathogens) are software copies , that can only operate in living cells. (they are infective and harmless outside living cells).

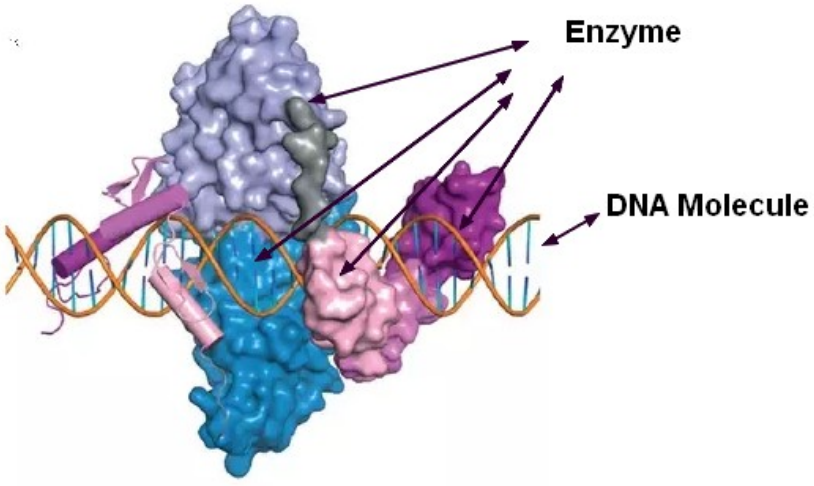
Viral DNA and RNA are coated with a protein or protein lipid layer to protect the nucleic acid of the virus. Against destruction outside living cells.
The code contained on naked nucleic acids can also only be executed in living cells (varoids causing plant diseases). (comparable to external memory devices, which code can only be accessed when they are connected to computing devises ("desktops' ."laptops", tablets or smart phones)

Returning to the scientific method.

In spite of the pre – 1990 experiments pointing to the spontaneous emergence of life, data from experiments during the 1990s and later were and are currently still not so supportive of a spontaneous event or events that could cause the appearance of life from non life.

The genetic code fits and functions exceedingly well on the DNA molecule, but it won't function without the protection code.

The following drawing gives an indication off the complexity enzymes. It shows how a nucleases enzyme that separates the strings of the double helix of a DNA molecule.



The drawing indicates how the nuclease enzyme moves along the DNA molecule. , separating the two strings.

The nuclease enzymes forms an important part in all DNA functions. Is it possible to develop code (copy accuracy and memory required) that contain the necessary instructions to sustain life, on unreliable code carriers. Unreliable code carriers will lead to code corruption and memory loss.

SUMMARY

The post 1990s research seems to indicate that DNA needs to carry a minimum about of code

to function . This code include error correction , without which the DNA molecule would be a useless code carrier.

The following are two quotes from a review article by kroonin and Novozhill

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3293468/

IUBMB Life. 2009 Feb; 61(2): 99–111. doi: 10.1002/iub.146
Origin and evolution of the genetic code: the universal enigma
Eugene V. Koonin* and Artem S. Novozhil

Quote

"The genetic code is nearly universal, and the arrangement of the codons in the standard codon table is highly non-random "

"a fundamental but often overlooked question is why the code is (almost) universal. Of course, the stereochemical theory, in principle, could offer a simple solution, namely, that the codon assignments in the standard code are unequivocally dictated by the specific affinity between amino acids and their cognate codons. As noticed {previously}, however, the affinities are equivocal and weak, and do not account for the error-minimization property of the code "

Quote

"Summarizing the state of the art in the study of the code evolution, we cannot escape considerable skepticism. It seems that the two-pronged fundamental question: "why is the genetic code the way it is and how did it come to be?", that was asked over 50 years ago, at the dawn of molecular biology, might remain pertinent even in another 50 years. Our consolation is that we cannot think of a more fundamental problem in biology. "

Kroonin and Novozhill did not mention why is the code seems specifically designed for DNA, especially the error correcting ability.

No known cell functions on RNA alone.

Cells need a minimum amount of code to function. (The *JGVI syn-3.0* actually contains the minimum amount of code necessary to keep the native (wild) *M mycoides* going in a laboratory. (with the optimal availability of nutrients (building blocks) . More genes are probably necessary in the wild) . It *(JCVI SYN 3,0*)was derived from and have exactly the same code that resides in *M mycoides.*

Life without error correction of the code seems impossible, therefore it is highly probable that the very first DNA molecule had to have error correcting enzymes. ….But where did or how did these error correcting enzymes specific for DNA, originate from. From evolutionary foresight, from an intelligent entity or from possible, not experimental proven, random events?