ISSN 2075-1729
www.mdpi.com/journal/life

Peer-Review Record:

The Stereochemical Basis of the Genetic Code and the (Mostly) Autotrophic Origin of Life

Juan C. Fontecilla-Camps

***Life* 2014, *4*, 1013-1025, doi:10.3390/life4041013**

Reviewer 1: Anonymous
Reviewer 2: Anonymous
Reviewer 3: Anonymous
Editor: Niles Lehman (Guest Editor of Special Issue "The Origins and Early Evolution of RNA")

Received: 23 October 2014
First Revision Received: 27 November 2014
Second Revision Received: 9 December 2014
Accepted: 11 December 2014
Published: 16 December 2014

First Round of Evaluation

Round 1: Reviewer 1 Report and Author Response

Summary

In the manuscript "stereochemistry, the genetic code and the (mostly) autotrophic origin of life" the author addresses several fundamental questions concerning the origin of life, notably the early relationships among nucleic acids, amino acids and peptides that led to the genetic code, translation and, ultimately, organismal life as we know it. The author presumes the existence of the "RNA World" (currently the most popular scenario for the origin of informational molecules) prior to the occurrence of DNA, peptides or the genetic code and translation, Arguments are presented that direct, stereospecific interaction between RNA anticodons and amino acids is a plausible route for the beginning of the code. The author differs from most previous advocates of stereospecific code origins by providing arguments that large bulky amino acids with functional "R" groups (not simple ones most readily available protobiologically) were the earliest recognized and incorporated. Furthermore, the author provides suggestions for protometabolic pathways that could have functioned early in Earth's history permitting the synthesis of these complex amino acids (and perhaps other significant building blocks).

Broad Comments (Strengths)

I find that the manuscript offers a good concise overview of the field of life's origins and the early evolution of the genetic code. The following points represent significant and (mostly) novel contributions.

- Strong arguments are provided to support the idea that the earliest amino acids that interacted with nucleic acids and were incorporated into peptides were "bulky" ones with large functional "R" groups. Only such amino acids could be recognized effectively (e.g., by anticodons) and only they would provide useful function to the nucleic acids (allowing natural selection and evolution).
- The recognition of amino acids was stereospecific at the anticodon. This provides both a logical pathway for the genetic code's origin and a rationale for the homochirality of both amino acids and nucleic acids in living system. The arguments the author provides are supported by published data and are theoretically sound.
- I find the discussion of possible protometabolic pathways that could lead to the larger amino acids intriguing. The author suggests that relatively non-specific pathways (mineral or globule catalyzed) could have operated primordially to generate these amino acids (including the basic ones) and "other precursors" (perhaps nucleic acid components). I am pleased that he noted the possibility that globules could serve; most experimental studies have focused on mineral surfaces but I find many aspects of the "composome" scenario (Hunding *et al.*, 2006, Bioessays 28: 399–412) attractive.
- The author provides a plausible scheme for aspects of the expansion of the genetic code beyond the initial phase of direct stereospecific recognition of (bulky) amino acids by anticodons. The aminoacylation site was originally adjacent to the anticodon; subsequently "gene duplication" separated the two sites' allowing the distant anticodon to interact with mRNA codons, thus permitting translation of simpler amino acids.

Broad Comments (Weaknesses)

Although I find the manuscript fully acceptable as it stands, the author might wish to consider the following points in this or subsequent papers.

The author provides a plausible scenario for the protometabolic formation of complex amino acids yet he still adheres to the "RNA World" scenario—avoiding the issue of where the RNA (nucleotide bases, sugar and phosphate linkages and lengthy homochiral "smart" RNA) came from. Of course this avoidance is characteristic of pretty much all of the RNA World literature. From a logical viewpoint is it not much easier to generate fairly complex peptides than a pool of "smart" and pretty long RNA molecules needed for an RNA World?

Response: The reviewer is referring to one of the fundamental problems I have tried to address. I tend to think that by having (i) fewer different potential building blocks, (ii) a potential capacity for replication and (iii) being catalytically active, RNAs could have "functioned" without polypeptides. I find the opposite a lot less likely. How to make and reproduce complex peptides (potentially composed of up to 20 different amino acids) without some kind of synthetic apparatus and coding?

Why not have protometabolic pathways generate precursors of both RNA and peptides (including bulky functional amino acids like the basic ones that could interact with nucleic acids). Dehydration reactions could allow some polymerization and true Darwinian evolution could begin when simple peptides and simple nucleic acids began functioning cooperatively (peptide-assisted NA replication; NA assisted translation).

Response: I believe that proto-metabolic pathways did generate RNA precursors and also amino acids. However, as stated above, I think peptide synthesis had to wait until anticodon-bearing nucleic acids emerged, as a condition for the genetic coding era to get started.

An issue with the metabolic pathways (e.g., reverse citric acid cycle) outlined by the author, the process of aminoacylation (and, indeed the whole RNA World concept) is the central role that ATP and other nucleotide phosphates play in these processes, at least nowadays. Where does phosphate bond energy or its equivalent come from and how is it linked to these processes in the model the author has developed?

Response: There is now good evidence that PPi could have been a precursor of ATP (Holm NG, Baltscheffsky H. Links between hydrothermal environments, pyrophosphate, Na(+), and early evolution. Orig Life Evol Biosph. 2011 41: 483–493). Inorganic pyrophosphate was likely to be present in primordial settings where life originated.

The author might want to check out the article by R Griffith (Orig. Life Evol. Biosp. **2009**, *39*, 517–531) which comes to broadly similar conclusions regarding the early evolution of the genetic code but works from an alternative scenario, one that addresses a number of the logistical problems inherent in the RNA world.

Response: I have gone over the article mentioned by the reviewer. Although very interesting, it seems to me that it contains many hypotheses that would be difficult to test. In my article I have heavily relied on the existing literature and I have tried to stay as general as possible to avoid specifics that will not be falsifiable.

Specific Comments

The following minor grammatical changes should be considered:

- on Line 30 replace constrain (v) with constraint (n)
- on Line 54 change catalyze to catalyzing
- on Line 104 change rational (adj) to rationale (n)
- on Line 185 change “which origin” to “the origin of which”

Response: All these errors have been corrected.

Round 1: Reviewer 2 Report and Author Response

This paper presents a synthesis of different ideas surrounding the early evolution of RNA coding and protein biosynthesis. There is a lot to like in this paper, though also a few criticisms. Below are some thoughts.

If I understood the paper correctly, the hypothesis is that there were ancestral RNA molecules that both bound amino acids and catalyzed peptide bond formation, but only for what the author calls “bulky” amino acids. Preceding this combined role, the original RNA-AA binding may have been to increase stability/functionality of RNA. Within the proto-ribosome, then, having the anticodon and acceptor stem near each other would have protected the tRNA-amino acyl bond from hydrolysis. Then, later in evolution and for more sophisticated systems, a key gene duplication event is suggested to have moved the acceptor stem away from the anticodon region, releasing chemical constraints on codon-anticodon recognition and allowing a more “digital” evolution of the code. The system was increasingly able to recognize simpler amino acids, and codon capture expanded the code to include those amino acids, but preserving already existing stereochemical constraints imposed from the original “bulkier” amino acids.

Response: I find the summary of my ideas by the reviewer very good. I would just change “releasing chemical constraints on codon-anticodon recognition” to “releasing chemical constraints on amino acid-anticodon recognition”. Once the anticodon is not required for direct amino acid recognition it seems to me that codon-anticodon recognition will follow (generating a proto-messenger RNA).

This is an interesting scenario that seems like it would be a nice addition to debates on the emergence of life. What is particularly nice is that it addresses the important point of relative abundances of prebiotic inputs to life (*i.e.*, if some environments would have produced a highly diverse mix of compounds, why or how did only some compounds end up in living systems? How did this “mass concentration” work?).

Response: I agree with the reviewer that the degree of “prebiotic inputs to life” is central to any discussion on the origin of life. This is the reason I think significant compartmentalization and selectivity must have been necessary to get only some “primers” (such as pyruvate and citric acid) from the “primordial soup” to get proto-metabolic cycles started. Although it is very difficult to device a scheme to explain this process it must have taken place.

So I do ultimately recommend publication.

However, I also found the paper a little difficult to read, and think its clarity could be significantly improved. The summary above required me to through the paper several times to figure out how all the parts fit together.

My biggest recommendation would be to include a central summary figure with a good legend early in the paper that explains the overall scenario and highlights which parts of the hypothesis are new. The scenario has a bunch of “moving parts” that are now mostly scattered throughout the paper, and having such a centralized figure would provide a great aid to the reader in understanding how all the various underlying arguments fit together. Similarly (see below), better explaining various individual arguments and improving overall flow would also improve clarity.

Response: I have now included a figure as suggested by the reviewer. I have also added an extensive legend.

Smaller comments:

Abstract, Line 24: relating to the above point, the transition “besides conferring new functionalities…” is somewhat abrupt, and more generally I found the abstract a little hard to follow, only really getting some of the points once I had read through the rest of the paper. If this could generally be smoothed out, it would really strengthen the paper.

Response: I have made an effort to render the abstract easier to read.

Page 2, Lines 72–74: “In this scenario...the amino acid binds to this sequence and gets acylated by the adapter, which then dissociates carrying it along.” This sentence is unclear. Does the author mean that an RNA molecule with dual functionality recognizes and binds one amino acid and then catalyzes the reaction with a second amino acid before dissociating the (di)peptide? Or is this meant to prime the reader to the point about risk of hydrolysis? Or something else? Please make clearer what is meant.

Response: This sentence has been clarified.

Page 3, Lines 91–92 (and later on throughout the manuscript): “As indicated by both spark-tube … possible early settings”.

This depends on the prebiotic scenario we consider to be most likely. For spark-tube (atmospheric) and meteorites (interstellar + parent body processing) this appears a fair statement, but much less is known about chemistry at hydrothermal vents. Especially as the author does not appear to be necessarily arguing in favor of atmospheric or interstellar chemistry, but more generally discussing the difficulties that need to be considered in any scenario, it may be better phrased as an IF/THEN statement (*i.e.*, IF we consider spark-tubes *etc*. THEN the following problem holds). And perhaps mention a little bit about hydrothermal vents. But at least place a caveat.

Response: I have included IF/THEN statements. I agree that less is known about hydrothermal vents. I also think that the repertoire of organic molecules emerging from these vents must have been much more restricted than those of atmospheric and sidereal sources. I find hydrothermal vents interesting in the origin of life debate because they provide non-solar energy sources like hydrogen and carbon monoxide, which along with proton gradients could have been essential for life at its beginnings.

Page 3, Lines 93–94: “Binding of these … was unlikely to provide … derived peptides”. Why not? Please elaborate.

Response: The idea is that small amino acids such as Gly and Ala could not be functionally relevant either as handles (in the sense of Szathmary, ref. 18) or as components of short peptides. For instance, all the basic residues are bulky. Maybe the expression “derived peptides” is not clear enough. What I meant is that even if amino acid polymerization started with nucleotides bound with Gly, Ala and other small amino acids, the resulting peptides would not be functional (i.e., catalytic) because of both lack of relevant chemical functions and folding problems. I have tried to clarify this point.

Page 4, Line 125: “Neither option seems especially plausible” Why not? Please elaborate.

Response: It has to do with implausible sequences of events: option 1: anticodons are arbitrarily assigned (in a stereochemical sense) to small amino acids followed by stereochemically assigned anticodons to larger amino acids. The question here is how could the two sets of events be connected? If small amino acids could not be recognized by anticodons, how did these base triplets emerge? And, what unlikely coincidence made them compatible with a stereochemically based recognition process of large amino acids before it happened? Option 2: small amino acids are incorporated into polypeptides using a different primordial code that eventually evolves to a triplet-based code. This new code allows for the stereochemical recognition of large amino acids. Besides posing the same problem for the connection between triplets first coding for small

amino acids and then for larger amino acids this option has the added difficulty of implying a change in the nature of the code, which is not easy to envisage. I have included a sentence to better explain this.

Page 4, Lines 129–134 are again a little confusing. It is not immediately clear what "This apparent contradiction" is referring to as the directly preceding sentence provides a hypothesis that the author is proposing. This whole sequence seems central to the argument of the paper, but is confusing. Please make clear and precise what exactly the "contradiction" is, and improve overall flow, so that readers can understand better what the hypothesis in this paper solves and how.

Response: The contradiction I was referring to is contained in ref. 14 and is discussed as Options 1 and 2 in the preceding answer. Then I used a reference to Yarus' work where he argues than proto-metabolic pathways may have generated large amino acids before the emergence of the genetic code (a concept I fully agree with). I have changed this paragraph to improve readability.

Page 5, Line 150 (and more generally all of the beginning of Section 3.1.): "…alanine …most abundant in the soup". This depends on the scenario we assume to be most likely. See comment before.

Response: It is expected that the "primordial soup" scenario will favor Ala as one of the major amino acids present in proto-metabolisms. Other scenarios will give different answers and, as stated in the manuscript, I believe all the amino acids used for polypeptide synthesis were always of proto-biotic origin. My point here is that the "primordial soup" is a very popular concept that does not explain the transition between abiotic and proto-biotic amino acids syntheses.

Page 5, Line 154 and on: Moreover, I don't think the follow-up argument about modern alanine biosynthesis is very strong. All of the reactions mentioned by the author are transamination reactions, in which the only thing that alanine gets from the 'more complex' amino acids is the amino group. As a group the transaminases are generally known to have broad specificity, so a scenario in which one of the mentioned transaminases replaced an ancestral reaction of pyruvate + NH_4 to alanine is entirely plausible. Indeed, some organisms are know to use alanine dehydrogenase (ec 1.4.1.1) to drive pyruvate + NH_4, even using it as the main input for NH_4 assimilation into the cell when glutamate systems are disabled.

Response: The remarks of the reviewer are well taken. I have removed this section.

So instead of supporting the author's "however, no clear link …" point from Line 148, the above scenario could just as well be hypothesized to argue against it.

Response: My point here is that the way amino acids are synthesized in spark-tube experiments (and possibly in outer space) are very unlikely to be related to the way biology does it. So, how may the transition between the two, required if the primordial soup provides the first amino acids, be explained if completely unrelated?

However, I'm not sure this particular point about alanine is key as the overall narrative in the paper has much more context. Indeed the broad specificity of the transaminases could support the author's argument about enzyme promiscuity. Either way I would suggest significantly rewriting this first part of Section 3.1, as the current argument has weaknesses that could be seen to undermine the whole.

Response: I had also thought about transaminases in this context, especially when Glu is used as an amine source because that's a very common way of providing amino groups to metabolites.

Pages 5–6 (the rest of Section 3.1): Many others have written about promiscuous enzymes and their potential roles in early metabolic evolution (e.g., Copley, Jensen, Tawfik, Fani, Pereto and others), it might be good to at least cite and possibly briefly discuss some of that work in this section.

Response: Promiscuous enzymes are at the end of my list. I think that, by definition, primordial catalysts (minerals followed by nucleotides?) could not have helped having low specificity. In fact, I think this was a great advantage during evolution because they were, albeit less efficiently, able to catalyze in parallel a large number of similar reactions. More specialized catalysts would not have done the job. Being already complex proteins promiscuous enzymes probably represented the next-to-the-last step towards the first "modern" living organism.

Page 8, Lines 211–214: "However, direct amino acid… an "analogic" device". This sentence is hard to read with some circular aspects (amino acid…recognition could only be effectual is the amino acid…was recognized). Please simplify and improve readability.

Response: I have tried to do so.

Also, "analogic" should be "analog".

Response: Corrected.

Relatedly, regarding the general hypothesis of a "analog" to "digital" transition, a good paper to cite and be aware of is [Walker and Davies 2013, "algorithmic origins of life"], which talks in detail about transitions from analog to digital logic and the connection to the emergence of life.

Response: I have gone over that paper and although I agree there are points in common with my proposition the authors have a much more encompassing view about "analog" and "digital" aspects in biology than the transition of RNA adaptors I discuss in my paper.

Round 1: Reviewer 3 Report and Author Response

This manuscript makes the unusual argument that simple amino acids were probably not used for early coded peptides. Instead, relatively non-specific catalysis produced a variety of more complex amino acids, which could interact with RNA, and become specifically acylated to it. After hairpin tRNA gene duplication, these early RNAs could have both anticodons and (distant) acylations. Only at this point were the simple amino acids incorporated into an emergent process of translation.

Response: The interpretation of the reviewer is exact. Although I agree that my contention is unusual I also think that the results of Miller's experiments have been over-interpreted. Taken face value, all they show is that molecules such as amino acids can be synthesized abiotically. Conversely, they do not provide plausible pathways for their biological synthesis. Not surprisingly, in spark-tube experiments the simpler molecules are more abundant. However, I think that relative complexity must have been the landmark of biological processes right from the start.

Because the inferred evolutionary process is unusual, this notion is worth description, and could be published. However, I have a few suggestions.

(1) A graphic depicting the suggested process would make the proposed path easier to first comprehend, then remember. Furthermore, having to specifically present the pathway will likely clarify the implied course of successive selections leading to a translation apparatus.

Response: I have now included a graph as suggested by the reviewer. I have also added an extensive legend to the figure.

(2) Doesn't the idea require that primordial catalysts exist to carry our substantial fractions of the pathways in the Figures? That seems a rather rigorous requirement; are there any examples that make this more plausible?

Response: Several reactions concerning one of the examples I chose, the tricarboxylic acid cycle (Figure 2) have been carried out using colloidal semiconductor particles (see refs. 34–36). I think this is a good example of possible primordial catalysts.

(3) I do not understand the argument in Lines 156–157, and would like to see the numbers presented better explained.

Response: It was my intention to argue that the relative molar ratios of amino acids found in spark-tube experiments are not consistent with the sequence of their metabolic synthetic pathways as they are observed in biology. One would have expected that if simpler amino acids were the first to be incorporated into biological processes they would have been precursors of more complicated ones. I took Ala as an example because it is normally synthesized from the more complex amino acids glutamate, cysteine and valine. The numbers the reviewer refers to are the molar ratios of these amino acids (except Cys) relative to Ala in spark-tube experiments. I agree that this should have been better explained. As another reviewer has commented that this paragraph was not central and could be taken out I have done just that.

(4) The legend to Figure 2 contains a risible misspelling, "alpha-ketoisocrapoate". In the Figure 1 legend "dephosphorylation = dephosphorilation"?

Response: I have removed the "crap" from "alpha-ketoisocaproate". I have also corrected "dephosphorylation"

(5) The title is unusually uninformative; a revised one which gave some hint of the nature of the hypothesis would be more helpful to a prospective reader.

Response: I understand the reviewer's comment but it can also be interpreted as: "what is this author trying to tell us?" And attract the reader's attention. However, I have changed it a bit.

Second Round of Evaluation

Round 2: Reviewer 2 Report and Author Response

The author has significantly improved the manuscript, and I recommend publication.

The answers to my comments provided by the author have helped clear up gaps in my own understanding of the background and context, and modifications to the text have significantly improved clarity and flow. In particular the addition of Figure 3 with its extensive legend provide the reader with

a centralized summary and graphical representation of the hypothesis, greatly aiding understanding of the key points.

As previously stated, I think the manuscript provides a nice synthesis and interesting hypothesis that deserves to be part of the ongoing debates of how life emerged.

A few remaining minor comments:

(1) Lines 194–195, Page 5 (Section 3.1) and Figure 1: The author names the route for AA synthesis the “tricarboxylic acid cycle (TCA)-semialdehyde” route, but none of the substrates in Figure 1 are intermediates of the TCA cycle or tricarboxylic acids ... A brief clarification in text, or addition to the figure (or its legend) that the starting substrates in Fig. 1 are aminated versions of intermediates of the pathways in Figure 2 might make the interconnections between the chemistries more clear. I leave this up the author.
(2) Line 217, Page 6: typo “polyepetides”.
(3) Lines 310–311, Page 10/11: grammatical error “crystal structure of the ribosome [which] proteins interact...” should it be [which shows that]?

Response: I have made the changes suggested by the reviewer (highlighted in yellow). I thank him/her for very constructive comments.

Round 2: Reviewer 3 Report and Author Response

The ideas of the MS are notably clarified in the revision, particularly by the added Figure and its text.

Response: This reviewer has now accepted my changes. I thank him/her for constructive criticism.