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

Peer-Review Record:

The Place of RNA in the Origin and Early Evolution of the Genetic Machinery

Günter Wächtershäuser

***Life* 2014, *4*, 1050-1091, doi:10.3390/4041050**

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

Received: 24 October 2014
First Revision Received: 2 December 2014
Accepted: 9 December 2014
Published: 19 December 2014

First Round of Evaluation

Round 1: Reviewer 1 Report and Author Response

In this massive and dense manuscript, Günter Wächtershäuser furthers his views and opinions on the origin and evolution of life. He reviews some of his previous work and presents an alternative to the dominant 'Ancient RNA world' hypothesis. The alternative views he previously generated are much expanded in this manuscript. In light of recent research developments and argumentation (some of it reviewed), his views should be considered a welcome addition to the many ideas that populate the "origin of life" field of inquiry that counter the dominant paradigm. I have however a number of quibbles that if addressed could increase the accuracy, value and impact of the manuscript. I must note that a careful evaluation of all facets requires expertise in a multitude of disciplines (from prebiotic chemistry and structural biology to evolutionary bioinformatics and biochemistry) and considerable time, none of which I possess. Therefore, my comments will be slanted by my own expertise and will only serve the author as a partial devil's advocate effort

General commentary

Section 1. The place of RNA in LUCA (page 2): In search of features that are more conserved (carrying deep phylogenetic memory) than the sequence of genes, Wächtershäuser focuses on a paper of his in Systematic and Applied Microbiology (1998) that uses gene content and order of microbial genomes to make inferences about the last universal common ancestor (LUCA) of cellular life. He then

mines the significance of some of the conserved chromosomal segments in light of some other evidence. The exercise is at places compelling, but forgets some recent, very global and exhaustive analyses that also use highly conserved biological features to reconstruct the makeup of LUCA (e.g., gene order, 3D molecular structure, molecular functions). See for example PubMed references PMID: 17370266, PMID: 21612591, and PMID: 17908824, which are in line with some conclusions derived from the alignment of Figure 1. The fact that these other analyses make use of hundreds of genomes to infer the ancient biochemistry of LUCA complements and strengthens the preliminary and fragmentary analysis of only 19 of them by the author, which also excludes eukaryotic genomes from the set (understandably, few sequenced genomes were available in 1998, and eukaryotes are in general "master rearrangers").

Response: The three references have been noticed with gratitude and included in the MS.

Given my interest in the bioinformatics of gene content and order, I took the liberty of studying the brief Systematic & Applied Microbiology paper to check the validity of the "reconstruction" methods of Figure 1. The algorithmic implementation that is described is quite raw and does not extract important information that is embedded in the clusters of conserved gene segments. Important algorithms have been devised since the initial work of Sankoff in the sixties and seventies to do exactly that. I refer to the work of Pevzner, Tesler and Bourque as good examples, but also of Warnow. I also recommend visiting GRIMM (http://grimm.ucsd.edu/GRIMM/) and perhaps using the server to confirm or extract additional information from the alignments that are summarized in Figure 1. The Wächtershäuser algorithm makes use of conserved elements of gene content and order but discards information provided by the actual rearrangement operations that erase gene order history. The algorithm does not describe how gene homology was detected, how the limited set of genes was selected, and how the alignments were constructed (I imagine by hand). It is not clear if a guiding tree was used in the alignment (though this is mentioned in Line 65), since this is not made explicit in the 1998 publication. An alignment implies a tree but usually alignment algorithms are greedy and problematic and represent the most important limiting step of a phylogeny (the field is thus moving to the joint alignment and tree reconstruction). Therefore, I do not think the ancestors were properly reconstructed (the tracing of features in ancestor nodes of trees are not described, nor the actual trees). Regardless of all of these limitations, the tight conservation of certain segments is enough to show the existence of a core of ribosomal proteins that is universally present in cellular organisms. This in itself is valuable. Of course, a sample of 19 microbes may not be enough to encompass molecular diversity and the absence of Eukarya may also be problematic for any global evolutionary statement. In other words, the risks of sampling bias are clearly present and should be mentioned in the manuscript.

Response: The 1998 paper leaves much to be desired. Its deficiencies reflect the excitement of the first hour. The state of the art at that time may be gleaned from an authoritative paper that came to the opposite conclusion: A.R. Mushegian and E.V. Koonin "Gene order is not conserved in bacterial evolution". TIG 1996, 12, 289–290. The gene cluster table of 1998 was mainly retrieved from the annotations in published genomes and constructed manually with paper and pencil. The state of information technology at that time is reflcted by the fact that the table was folded an individually pasted by hand into each issue by the publisher.

The list of common genes in the alignment of Table 1 is enriched in small and large subunit ribosomal proteins that are the most ancient of the ribosomal set, according to the Caetano-Anolles theorem that is mentioned later in the manuscript. Interestingly, the most ancient of them are clustered toward the 5' end of the genomic sequences that were aligned (S12, S17, S5, S4 and L2, L3 and L24). Could this imply a possible ancient segmental duplication? Also interesting is the placement of the most ancient ribosomal protein S12 between exactly two polymerase (beta and beta prime) and elongation factor (entry and translocation) genes (separated by single and much more derived ribosomal proteins). Could this be an ancient memory of the ribosome mediating translation and replication? A commentary would enhance the value of the section.

Response: The 1998 paper contemplated speculatively a combination of small-scale gene doubling (as evidenced by the immediate neighborhood of EF-Tu/EF-G) and of a large-scale gene cluster doubling (as evidenced by the spacing between secE/secY and rpoH-A/rpoD) with the hope of a future deeper understanding based on folding structures. Now the referee makes an exciting suggestion that may, if executed successfully, go some way to satisfy that hope.

In terms of the genome organization of LUCA (Line 105), there are numerous arguments in favor of an RNA ancestral genome and the late unfolding of DNA as genetic repository (perhaps through viruses). I cannot understand how 2'-hydroxy groups could destabilize the molecule and lead to "intramolecular self-destruction" (Line 107). RNA has the wonderful property of folding in search of energetic and kinetic minima. These processes make use of a frustrated landscape, which is powerful in terms of its biological potential (stability, function, information). It is much more versatile than the rigid DNA alternative, which is enriched in the 'information' capacity. None of them have "self destruction" properties, which would have been weeded out by selection and self-organization very early in "chemical" evolution. In turn, everything about LUCA should be considered quite modern and far away from an initial FeS world. I refer to the work of Daniel Lundin and colleagues in Sweden about the rather late rise of some structures of the ribonuceotide reductase enzymes needed to built the DNA polymers. The Maurel theorem stands if one thinks of single genome molecules. But what if there were many short RNA genomes, as anticipated by Woese, each perhaps linked to different and ancient tRNA-like cofactors? As the author mentions, LUCA is contemporary to a rather complex biochemistry, with numerous structures harboring a multitude of active and allosteric sites in proteins, most of which drive central enzymatic functions.

Response: The self-cleavage of RNA by 2'-OH is a chemical textbook fact. The 2'-OH group has the proper position and orientation for a nucleophilic attack on the phosphate bridge. The kinetics of the reaction is greatly favored by the 5-membered ring structure of the resulting cyclic phospho-bisester. The length of the RNA molecule is not relevant since each nucleophilic attack causes destruction of the chain. The effect should not be confused with the length-dependent "error catastrophe" of accumulating mutations of RNA. Incidentally, some anaerobic ribonucleotide reductases are ancient, while others (aerobic ones) are later inventions.

In Line 114, the role of chromosomal rearrangement in evolution is not well described. What are the effects of genome rearrangements? Several scenarios are possible, including: (1) Genes remain linked

probably because it is more efficient to transcribe genes that produce interacting proteins (part of complexes) than those that are not; (2) Genes remain linked if they are of relatively recent origin (rearrangement has not had the change to split them apart); (3) Genes remain linked if they are part of functional groups historically united by genomic regions (encoding metabolic functions or rRNA?); this includes genes sharing an operon structure for economy purposes in highly reduced organisms such as bacteria; (4) Genes remain linked because they originated when genomic rearrangements were not biochemically motivated and their sequence makeup was later refractory to rearrangement hotspots. There is a rich literature about rearrangement, hotspots and many processes related to these (including domain organization in proteins). Some discussion in this front could be clarifying.

Response: Genome rearrangements are certainly important during later evolution of the phyla. At the level of LUCA, i.e., prior to the splitting of the domains, it is not clear, if and to what extent rearrangements of the modern style occurred. In this regard we should bear in mind that the LUCA genome may have exhibited sense-antisense coding on both strands as suggested by Rodin and Carter. This position has been adopted and discussed in the present paper. Therefore, speculations concerning possible genome rearrangements may be a bridge too far.

In relation to comments of Line 188 onwards, Di Giulio, Caetano-Anolles and others have suggested that the genetic code started to unfold prior to LUCA but continue to do so once life diversified. The corollary, is that the complete canonical set of amino acids may have not been encoded in LUCA. In terms of metabolic and biochemical competency, there are numerous and interesting studies, which have not been cited (especially in origins and evolution of modern metabolism). This includes the coevolution theory of the genetic code, the coexistence of prebiotic chemistries with modern metabolic reactions (recently studied for nucleotide metabolism), and theories about the origin of translation.

Response: The present analysis comes to the conclusion that the sets of canonical amino acids and bases as well as the genetic triplet code were largely complete at the level of LUCA. The Wong coevolution theory of the genetic code has been discussed in detail. The literature comprises numerous proposals concerning the origin of translation and other aspects of the genetic machinery. A review of all these proposals and many others is beyond the scope of the present paper. The present paper is a research paper and not a review paper. It aims at a comprehensive account of early evolution from the origin of life all the way to LUCA. This puts a systematic constraint on literature selection. An effort (unfortunately fallible) has been made to include all those references that integrate with the main lines of the present account into a coherent account. Contributions by others that have the character of theoretical modules that fit well into this account have been termed "theorems" with names of the main authors attached. A reference to a paper on nucleotide biosynthesis phylogeny is now cited in Section 7.

Section 2: Thermal course of evolution (Page 6): Why a focus on thermal energy? The framework should be on thermodynamics, energy dissipation and information, all of which are linked. What is thermally upward or downward adaptation? Is it conquering niches on Earth or a process involving molecular makeup? Vocabulary and definitions are murky, especially related to the links of environmental thermal fluctuations, energy of folding and stability of polymers. Subsection (1) must be rewritten to help the reader understand the ideas. Should all adaptations comply with maximizing energy

dissipation? How can this be reconciled with upward and downward trends? No references are provided despite the rich literature underscoring the controversial link between physics, information theory and biology. If the focus is conquering planetary niches, then perhaps reference the contrasting views and some of their proponents (deep sea versus surface; thermophilic versus mesophilic, *etc.*). Further elaboration of the Wolfenden theorem appears relevant and its connections to what is known about the origin and evolution of metabolism also of importance.

Response: Based on the valuable criticism Section 2 has been extensively revised. Terminology has been clarified. This Section has a rather restricted purpose. It provides chemical arguments for the proposition that the pioneer organisms could only exist at high temperature and that the subsequent forms of life remained hyperthermophilic for a long time until much later an irreversible evolution generated organisms that required lower and lower temperatures. This conclusion places severe constraints on all aspects of the evolution of the genetic machinery. The fascinating topics of thermodynamics, energy dissipation and information are outside the scope of the paper.

Section 5. Place of RNA in the origin and early evolution of translation/Pre-translational coding of peptides. The view that is presented in this crucial segment posits a pre-translational mechanism (side-by-side tRNA mediated coding) that very much resembles the ribosomal entropic system, which could be very advanced. What if the pre-translational coding was assembly line-like and mediated by the aminoacyl-tRNA synthetases themselves? See PMID: 23991065 for one such alternative and a related previous model (PMID: 22210458). This would also match commentary of a possible early origin of non-ribosomal peptide synthesis machinery compared to that of the ribosome.

Response: *The author expresses his gratitude for the two additional references, which have been included in the text.*

Section 9. General overview. While the grand finale “Nothing in evolution makes sense except in the light of chemical predetermination” is impressive and summarizes the championing work of the author, chemical predetermination may apply to the very early stages of evolution and not to many of the stages described in the manuscript. How much chemical predetermination can there be with polymers as these explore a minute fraction of the space they make possible? In the enthalpic-entropic gradient that is proposed, there is also a gradient from “predetermination” to historical contingency. The boundaries of such a gradient is murky and the final statement may not apply to much of modern biochemistry, once proteins start to achieve stable complex structure and much earlier that the time of LUCA. In fact, the genome rearrangements that underlie Figure 1 are testament to the historical contingencies that were already at play in LUCA and not to chemical predetermination.

Response: The statement has been clarified.

Specific commentary

- Title. The title does not encompass appropriately the wide subject matter covered by the manuscript. I suggest “The place of RNA in the origin and early evolution of life”.
- Line 39. And what is the place of proteins in all that?

- Line 39 and Section 1 “The place of RNA in LUCA”. The word retrodiction implies reconstruction of ancestors from extant information.
- Line 40. Interpolation of LUCA and competing theories of the origin of life. In my view, LUCA is not a theory of origin of life. It is a theory of origin of diversified cellular life.
- Line 52. Not only genes are gained and lost with time (Woese theorem), but also mutations accumulate to saturation in genes making homology statements difficult. This has been made explicit abundantly in the molecular evolutionary literature. A focus on gene content and order makes only sense if such homologies at gene level are preserved (if not there will be artifactual losses in the alignments).This should be highlighted in the text because it is relevant to the validity of the alignment of Figure 1, making its significance more important.
- Line 62: Change “multiply” to “multiple”.
- Line 406. “LUCA protein cycle”. Why bring LUCA to a link with the proposed “protein cycle”. LUCA is just the last of a chain of ancestor of diversified life. If the protein cycle requires a coupling of the enthalpic activation of amino acids and the entropic biosynthetic action of a primordial ribosome, then perhaps it is better to call it “ribocellular protein cycle”, since this coupling marks the start of modern cells, not necessarily “modern diversified cells” (*i.e.*, LUCA).

Response

The author prefers the present title, because the paper is concerned with the origin of the genetic machinery. Theories on the origin of life or on other aspects of early evolution, such as cellularization, serve merely as starting points.

- *In the introduction the term “or proteins” has been added after “RNA” ,*
- *and the term “retrodict” has been defined.*
- *The term “interpolate” has been clarified.*
- *The problem of mutational saturation is now included in the discussion of Figure 1.*
- *The term “multiply impaired” has been replaced by a clearer wording.*
- *LUCA is discussed only in Section 1.*
- *The protein cycle is discussed in a separate Section 4, which is concerned with the course of evolution before LUCA.*

Round 1: Reviewer 2 Report and Author Response

The manuscript comprises an exciting account on the origin of life with emphasis on the emergence and function of RNA. The article demonstrates that the iron-sulfur surface metabolism theory at high temperatures has a much better explanatory power than the RNA World theory. I have only a few comments, which might improve the manuscript.

Please explain what do you mean with Kandler, Wolfenden *etc.* Theorem. Is a theorem Kandler’s theory or proposal?

Response: There are many relevant literature references. Among these there are specific proposals concerning a circumscribed problem that are included to fill a logical gap, as a theoretical module so to speak. These have been designated by the term “theorem” with the added name(s) of the main

author(s). This makes it clear that the account given is comprehensive in the sense that major independent contributions by other scientists integrate readily with the overall account given.

Line 716: The biosyntheses of His and Trp are not so closely related. They only have in common that they use phosphoribosediphosphate as substrate and therefore share two related enzymes.

Response: The relationship of His and Trp biosyntheses has been toned down.

In the iron sulfur world the surface metabolism prevents the diffusion of the fixed products into the ocean. But without compartmentation soluble intermediates would escape. Furthermore, the establishment of early bioenergetics based on electrochemical Na^+ gradients would have been impossible.

Response: Section 8 has been revised in order to address the issues involved in the last query.

Second Round of Evaluation

Round 2: Reviewer 1 Report

Most commentary has been addressed and new changes have improved the exposition. The manuscript represents a nice piece of work with many exciting ideas for further exploration.