



Hypothesis

Genome Evolution from Random Ligation of RNAs of Autocatalytic Sets

Felix Broecker

Idorsia Pharmaceuticals Ltd., Heggenheimermattweg 91, CH-4123 Allschwil, Switzerland; felixbroecker@gmx.net

Abstract: The evolutionary origin of the genome remains elusive. Here, I hypothesize that its first iteration, the protogenome, was a multi-ribozyme RNA. It evolved, likely within liposomes (the protocells) forming in dry-wet cycling environments, through the random fusion of ribozymes by a ligase and was amplified by a polymerase. The protogenome thereby linked, in one molecule, the information required to seed the protometabolism (a combination of RNA-based autocatalytic sets) in newly forming protocells. If this combination of autocatalytic sets was evolutionarily advantageous, the protogenome would have amplified in a population of multiplying protocells. It likely was a quasispecies with redundant information, e.g., multiple copies of one ribozyme. As such, new functionalities could evolve, including a genetic code. Once one or more components of the protometabolism were templated by the protogenome (e.g., when a ribozyme was replaced by a protein enzyme), and/or addition modules evolved, the protometabolism became dependent on the protogenome. Along with increasing fidelity of the RNA polymerase, the protogenome could grow, e.g., by incorporating additional ribozyme domains. Finally, the protogenome could have evolved into a DNA genome with increased stability and storage capacity. I will provide suggestions for experiments to test some aspects of this hypothesis, such as evaluating the ability of ribozyme RNA polymerases to generate random ligation products and testing the catalytic activity of linked ribozyme domains.

Keywords: genome evolution; ribozymes; RNA ligase; early Earth; autocatalytic sets; RNA world



Citation: Broecker, F. Genome Evolution from Random Ligation of RNAs of Autocatalytic Sets. *Int. J. Mol. Sci.* **2021**, *22*, 13526. <https://doi.org/10.3390/ijms222413526>

Academic Editor: Ilana Agmon

Received: 15 October 2021

Accepted: 15 December 2021

Published: 16 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The origin of life, i.e., the transition from abiotic molecules to replicating entities subject to Darwinian evolution, remains elusive. There are two main hypotheses that can be summarized as 'RNA-first' (the RNA world hypothesis), or, more broadly, 'replicator-first', and 'metabolism-first'. The former (described in detail below) suggests that the ancestor of the genome arose spontaneously as a self-replicating oligo- or polymer of RNA, and that the metabolism emerged as a consequence of the evolving RNA. The latter hypothesis posits that a metabolism, i.e., a set of chemical reactions and their respective catalysts, within a primordial cell (the protocell) existed and evolved in complexity without any form of genome. Compelling arguments for both scenarios have been made in recent decades theoretically as well as by mathematical modeling and experimentation. Both hypotheses can arrive at some entity with the basic characteristics of life, i.e., the ability to multiply and to undergo Darwinian evolution. However, today's cells contain two components, a replicating genome and a metabolism. A convincing scenario explaining both the evolution of the metabolism and the genome has not been formulated to the best of my knowledge.

In the hypothesis presented here, I suggest that a protometabolism of catalytic RNAs (ribozymes) evolved first and gave rise to the first iteration of the genome, the protogenome. I will start by summarizing three key concepts underlying my hypothesis: the RNA world hypothesis [1], the concept of autocatalytic sets (i.e., a collection of molecules, each of which can be created through catalysis by other molecules within the set; see Figure 1) originally

proposed by Kauffman [2–4] and dry–wet cycles as likely origin and early evolution of RNA polymers and liposome-based protocells (the compartments containing the autocatalytic sets and, subsequently, the protogenome), proposed by Henning and coworkers and Damer and Deamer, respectively [5,6]. I will then lay out my hypothesis on genome evolution, whose key aspect, the emergence of the protogenome from RNA-based autocatalytic sets, is summarized in Figure 1. I will conclude with ideas on how some of the aspects of this hypothesis might be tested experimentally.

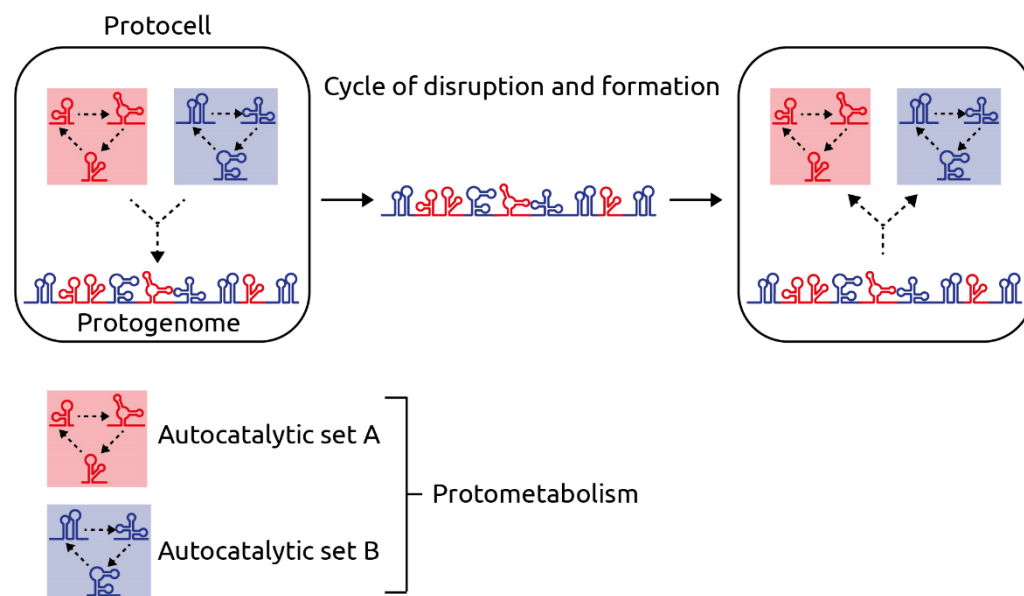


Figure 1. Schematic of the hypothetical protogenome and its function. In the left protocell (e.g., a liposome formed in a dry–wet cycle [6]), autocatalytic sets, composed of ribozymes, comprise a protometabolism. In this example, the protometabolism consists of two autocatalytic sets, A and B (red and blue, respectively). The protogenome, arising from the random ligation mediated by an RNA ligase ribozyme, contains, in one molecule, ribozyme domains of both autocatalytic sets. During a cycle of liposome disruption (‘dry’) and formation (‘wet’) the protogenome enters newly formed liposomes (right). There, it can seed both autocatalytic sets as its ribozyme domains are catalytically active. This process of protometabolism inheritance is more efficient than the stochastic co-encapsulation of individual ribozymes from autocatalytic sets A and B.

1.1. The RNA World

The RNA world hypothesis suggests that there was a stage during Earth’s evolutionary history in which self-replicating RNA molecules proliferated [1]. From these initial RNA replicators, proteins, DNA, and ultimately all life on Earth evolved. The basis of this hypothesis is the observation that RNA is the only biopolymer known to be capable of both storing genetic information (as, for example, in RNA viruses) and of being catalytically active in the form of ribozymes. One can thus imagine a self-replicating RNA molecule that uses itself as a template. The first prerequisite for the emergence of such a replicator is the presence of ribonucleotides as building blocks in the ocean of early Earth.

Life on Earth likely evolved about four billion years ago [7]. The reducing atmosphere of early Earth lacked free oxygen and the primordial ocean presumably contained various organic compounds present in dilute concentrations in the micromolar range [8]. Simple organic molecules like hydrogen cyanide (HCN) or formaldehyde (CH₂O) were likely formed from precursors such as water (H₂O), methane (CH₄), ammonia (NH₃), and hydrogen (H₂). The subsequent chemical reactions that yielded more complex organic molecules such as nucleobases, amino acids and lipid precursors must have required regions that allowed for a concentration of the reactants, creating the ‘warm little pond’ envisioned by Charles Darwin. Such regions may have been hydrothermal fields undergoing wet–dry cycles

as proposed by Damer and Deamer (described below) [6] or hydrothermal vents with strong temperature gradients [9]. The abiotic synthesis of complex organic molecules has been achieved experimentally by simulating the likely conditions on early Earth, starting famously with the Miller–Urey experiment published in 1953 [10]. This experimental setup achieved the abiotic synthesis of various amino acids from a reducing atmosphere composed of CH₄, NH₃, H₂ and water vapor, when energy was provided by electric discharge to simulate lightnings. Even though early Earth's atmosphere, according to more recent data, might have looked different to the one simulated in the Miller–Urey experiment [11], these experiments provided first proof that complex organic molecules can arise abiotically. Following this, other groups have achieved the chemical synthesis of additional amino acids, as well as nucleotides and lipid precursors [12–15] (see [16] for a review). The abiotic chemical synthesis of ribonucleotides from simple precursors under prebiotically plausible conditions in 2009 provided further credence to the existence of the RNA world [14,17]. An additional source of complex organic molecules may have been meteorites; for example, a large array of organic compounds including nucleobases and amino acids have been identified in the Murchison meteorite [18].

The next step in the RNA world scenario is the formation of random RNA oligo- or polymers. Oligomers can form in solution [13,19] and on mineral surfaces [20]; polymers could have emerged in ponds undergoing dry–wet cycles [5], which is also a likely site for the other component required for life to emerge, evolving liposomes (see below). Chemically, initial RNA oligomerization may have been achieved by activated ribonucleotides [21,22] or activation agents [23,24]. Some of the random oligo- or polymers of RNAs could have been catalytically active and become part of autocatalytic sets, possibly together with other organic species such as peptides. It is plausible that the initial prebiotic ribozymes formed in such an environment of repeated dry–wet cycles; however, the exact mechanisms of initial oligonucleotide formation remain elusive. It is assumed here that oligonucleotides have formed abiotically.

The naturally occurring ribozymes in today's world appear to have limited functionalities, mainly catalyzing peptide bond formation in ribosomes, as well as cleavage and ligation of phosphodiester bonds of RNAs [25–28]. In vitro evolution experiments, however, have generated ribozymes with various catalytic activities [29]. The first such identified ribozymes were DNA and RNA ligases [30,31]. One of the RNA ligases has been successfully evolved to an RNA-dependent RNA polymerase (RdRp), which, however, was only able to polymerize a maximum of 14 successive nucleotides [32]. A later iteration could incorporate up to 20 nucleotides [33]. More processive RdRps able to copy up to about 200 nucleotide RNAs have been reported subsequently [34–38]. In addition, a reverse transcriptase ribozyme (RNA-dependent DNA polymerase) has been identified by in vitro evolution, generating DNA products of up to 32 deoxynucleotides [39]. Other in vitro evolved ribozymes catalyze nucleotide synthesis [40], aminoacyl transfer [41], the hydrolysis of carboxyesters [42] and thiophosphates [43], as well as redox reactions [44,45].

Given this rich catalytic repertoire of ribozymes, the RNA world scenario seems plausible and even likely to have existed in the form of ribozyme-catalyzed metabolisms [46]. However, the evolution of the first replicator remains an unsolved problem, as no self-replicating RNA has been identified to date [47]. The closest to an experimentally verified self-replicating RNA system has been a set of two RNA ligase ribozymes that catalyze each other's formation (a simple autocatalytic set), which however requires complex preformed RNA building blocks [48]. Moreover, if the first replicator was a self-replicating RdRp ribozyme, its spontaneous emergence is highly unlikely given that it would consist of about 200 nucleotides [35,49]. In addition, a spontaneously emerging RdRp is likely to be error-prone and therefore unlikely to enable stable self-replication [35,46]. Thus, the largest gap in the RNA world hypothesis remains the emergence of the initial self-replicating RNA.

1.2. Autocatalytic Sets—Evolution of Metabolism without a Genome

The concept of autocatalytic sets (also referred to as collectively autocatalytic sets) provides an explanation on how a metabolism can emerge and evolve without the need of a dedicated information storage device such as the genome. The idea was pioneered by Stuart Kauffman [2,3]. In an autocatalytic set, a set of catalysts (i.e., molecules catalyzing a chemical reaction) are arranged such that the formation of each catalyst is achieved by other catalysts in the set, food sources (building blocks) provided. Thus, the whole set is autocatalytic, as opposed to a single molecule as stated by the RNA world hypothesis. The simplest autocatalytic set is one of two molecules that mutually catalyze the formation of one another. Experimentally, such a system was created by Gerald Joyce and coworkers with two RNA ligase ribozymes catalyzing each other's formation [48,50]. Earlier, von Kiedrowski and coworkers generated cross-catalytic sets of short nucleotide sequences [51,52]. A more complex autocatalytic set composed of up to 16 RNA molecules that assemble into a self-replicating set has been achieved by Lehman and coworkers [53]. New ribozyme functionalities have been successfully generated by recombination [54] and an RNA-based autocatalytic set that couples anabolism and catabolism has been reported [55]. More recently, it was shown experimentally that a ribozyme can diversify and achieve multiple functions within an autocatalytic set [56] and that there is a trade-off between reproductive fitness and variation, providing insights into the evolutionary dynamics of autocatalytic sets [57]. Of note, such an autocatalytic set can be composed of other biomolecules as well. For example, a complex peptide-based autocatalytic set has been reported [58]. Autocatalytic sets also exist in nature and have been, for example, identified in the metabolism of *Escherichia coli* [59], providing evidence that autocatalytic sets evolved naturally.

Mathematical and computational studies on autocatalytic sets, formally described as reflexively autocatalytic and food-generated (RAF) sets (with food being a set of building blocks such as organic molecules on early Earth) have shown that autocatalytic sets have a high probability of emerging even with moderately active catalysts [4]. Moreover, randomly generated autocatalytic sets often contain subsets that are themselves autocatalytic sets, which can become dependent on each other, recombine, and thereby become more complex [60,61]. However, how exactly autocatalytic sets of organic polymers could have evolved and been selected for is still unclear. Vasas et al. (2012) discussed this problem and found that the prerequisite for evolvability are chemical reactions networks with multiple viable cores [62]. Although it remains difficult to imagine theoretical autocatalytic sets capable of evolution by natural selection [63], it is assumed here that RNA-based autocatalytic sets have this ability.

In summary, autocatalytic sets have likely spontaneously emerged from prebiotic, organic precursors present on early Earth, were self-sufficient, replicating, and capable of evolution. RNA oligo- or polymers, whose precursors were likely present on early Earth [14], are a likely catalytic component of the early metabolic networks. However, sustaining such systems in the ocean of early Earth would have been unlikely because of diffusion of the components, especially since the ocean was likely a very dilute 'primordial soup' of organic molecules with concentrations in the micromolar range [6,8]. For the autocatalytic sets to be sustainable and capable of evolution, they needed to be confined in a container [61]. The most likely container for early life to evolve were liposomes, as discussed in the following section.

1.3. Dividing Liposomes

Today's cellular organisms are physically contained by a membrane mainly composed of a bilayer of phospholipids (bacteria and eukaryotes) or ether lipids (archaea). It is thus likely that the precursors of cells, the protocells, were also surrounded by a lipid bilayer, i.e., a liposomal container [64]. Artificially generated liposomes can grow and divide, for example, by incorporating externally provided micelles composed of fatty acids [65,66] or by the ability of internal production of phospholipids [67]. The latter finding suggests that

if a liposomal protocell that has the capability of synthesizing phospholipids (by means of autocatalytic sets) would also be able to divide, i.e., produce offspring. The building blocks for lipid synthesis were likely available on early Earth from abiotic synthesis [15]—but how did liposomes first emerge spontaneously?

An answer to this question has been provided by Damer and Deamer [6]. In areas on early Earth subject to dry–wet cycles, multilamellar lipid layers (during the dry cycle) form liposomes (during the wet cycles). If there are organic molecules in the aqueous phase, they will be distributed into the liposomes, where autocatalytic sets might come into existence. Thus, protocells can form that are able to evolve and, if capable of lipid synthesis, divide during the wet cycles. During the dry cycles, the contents of all the liposomes are combined and re-distributed into the emerging liposomes that form during the wet cycles.

2. Evolution of the Genome from RNA-Based Autocatalytic Sets

Now we have all the components that are required to form a protocell with an RNA-based protometabolism, as laid out above, consisting of RNA oligo- or polymers that form an autocatalytic set.

1. The abiotic synthesis of precursors for RNA and lipid molecules in chemically plausible conditions on early Earth has been experimentally verified (see Section 1.1);
2. Autocatalytic sets have a high probability of forming spontaneously, as per the mathematical RAF theory, and RNA (ribozyme)-based autocatalytic sets have been experimentally generated (see Section 1.2);
3. Liposomes can form spontaneously in dry–wet cycles and can grow, divide, and evolve (see Section 1.3). Such dry–wet cycles also allow for the oligomerization of RNA molecules [5].

Altogether, the dry–wet cycles appear to be a plausible scenario for the first protocells to evolve. Yet, to resemble present-day cells, these protocells require a genome that contains or encodes the information for the protometabolism.

2.1. Protogenome Evolution within the Dry–Wet Cycle Scenario

First, let me point out why a ribozyme-based protometabolism is more likely to evolve increasing complexity than one composed solely of peptides. RNAs are more likely to evolve into more complex molecules (i.e., more specific and more catalytically active ribozymes) because they can be complementary to each other and are less chemically diverse than peptides. This allows for replication of RNA species (e.g., by a polymerase) and sequence-specific cleaving and joining. Thus, it appears more plausible that, initially, ribozymes evolved, while peptide-based catalysts likely remained relatively simple, restricted to short oligomers. Moreover, before the evolution of a genetic code, the sequence-specific assembly of long peptides would have been limited in length due to the chemical diversity of amino acids, as their joining would have required highly specific catalysts (that could, for example, differentiate between aspartic acid and glutamic acid; and the number of amino acids available on early Earth was likely bigger than the 20 proteinogenic amino acids of today's cells). Of note, these additional amino acids were of limited relevance for the evolution of protein biosynthesis whose code likely became 'frozen' in its present form, as originally proposed by Francis Crick [68,69]—I will further elaborate on this topic in the discussion (Section 3). Because of their chemical diversity and non-complementarity, amino acids likely required a genetic code to be assembled into complex enzymes.

Of note, ribozymes can be relatively small. Thus, simple representatives have likely emerged from the random oligomerization of ribonucleotides in dry–wet cycles. For example, the smallest ribozyme identified to date consists of only five nucleotides with a three-nucleotide active center, able to catalyze aminoacylation of another RNA [70], and minimal ligase ribozymes with catalytic cores as small as 18 nucleotides have been identified by in vitro evolution [71]. It is conceivable that such small ribozymes may have formed by the random oligomerization of RNAs in dry–wet cycles.

An early protocell might have contained autocatalytic sets composed of such small ribozymes. As the autocatalytic sets evolved, more complex ribozymes might have emerged. A ribozyme-based protometabolism likely contained the ribozymes required for RNA assembly and copying, i.e., RNA ligases and polymerases. The polymerase might have evolved from a ligase, similar to in vitro evolution experiments that have successfully generated polymerases from ligases [32].

The first iteration of the genome precursor can be explained by means of the RNA ligase. The RNA ligase ribozymes are generally template-dependent and thereby restricted to an RNA substrate with a specific sequence. However, template-independent ligation might have occurred 'accidentally', leading to the random joining of two or more ribozymes. If the random joining of ribozymes occurred in such a way that, for example, it contained the ribozymes required to seed, in one molecule, two or more independent autocatalytic sets that collectively provided an evolutionary advantage over the individual autocatalytic sets, this multi-functional ribozyme protogenome could spread through a population of protocells. It would be reminiscent of a multi-domain protein, in which the individual domains provide a certain combined functionality. Moreover, such an RNA would be able to be replicated by means of an RNA polymerase if it contained two copies of the original ribozyme that was subject to copying by the polymerase (in the example herein designated as 'Rz_{B1}', since Rz_{B1} contains start sites for polymerization (Figure 2)). The resulting protogenome, after polymerization, would have Rz_{B1} sequences at the 5' and 3' ends and the collection of randomly joined ribozymes in between. Importantly, it can combine in one RNA molecule ribozymes of different, potentially independent, autocatalytic sets (red and blue in Figure 2). Such a protogenome would have virus-like properties; the ends are identical in sequence (such as terminal repeats characteristic of, for instance, retroviruses). These ends are required as starting points for the RNA polymerase to generate the complementary RNA ('minus' strand) and the subsequent polymerization of the protogenome copy ('plus' strand) (see Figure 2 for details). Interestingly, the recently discovered retrozymes (a form of retrotransposons) have terminal repeats that contain hammerhead ribozyme domains, encompassing a region of non-coding RNA [72]. Thus, virus-like entities with similarities to the herein proposed protogenome still exist today. The protogenome would have represented an evolutionary disadvantage if it used up too many building blocks for its own replication. As such, versions of it that replicated relatively slowly (e.g., by mutations in the termini that reduced the affinity to the RNA polymerase) might have been selected for in the dry-wet cycle scenario. Various variants of the protogenome (different composition and length) were likely assembled spontaneously initially and were subject to evolutionary competition.

There are three prerequisites for the above scenario to be plausible. First, random ligation products of ribozymes must have a probability to occur. Second, the joined ribozymes must retain their catalytic activity when part of a longer RNA molecule. Third, there must have been an RNA polymerase with sufficient fidelity to copy such a long RNA without running into an error catastrophe.

Regarding the first point, template-independent ligation of single-stranded DNA and RNA molecules has been demonstrated to be catalyzed by T4 DNA ligase and an RNA ligase from the archaeon *Thermococcus kodakarensis*, respectively (both are protein enzymes) [73,74]. RNA ligase ribozymes, especially short ones, have been shown experimentally to be able to accept a broad range of substrates [75], yet, to the best of my knowledge, non-templated ribozyme-catalyzed RNA ligation has not been demonstrated experimentally to date. However, given the fact that ribozymes tend to be generally more promiscuous than protein enzymes [46], it is conceivable that ribozyme RNA ligases have existed that also performed random ligation as a side product.

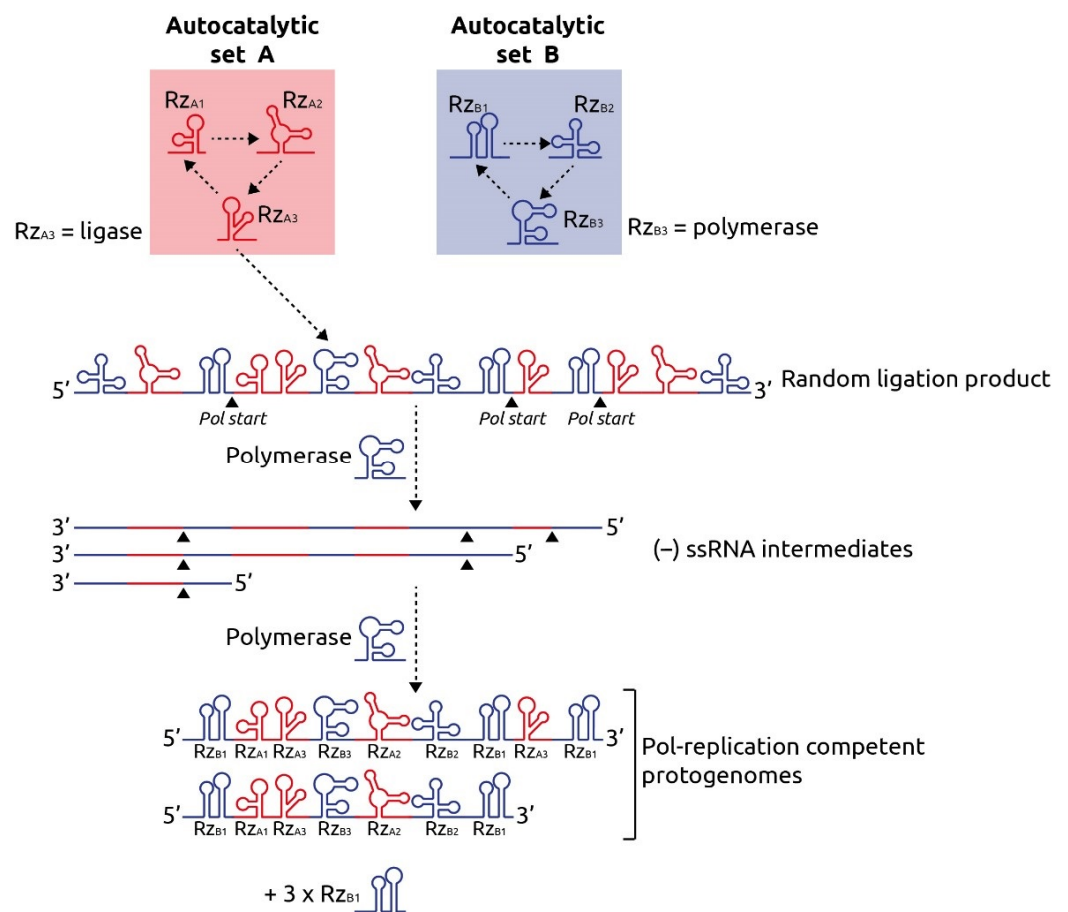


Figure 2. Emergence of the protogenome from RNA-based autocatalytic sets. The protometabolism in this example consists of two independent autocatalytic sets A and B, both comprising three ribozymes (Rz), RZ_{A1} through RZ_{A3} and RZ_{B1} through RZ_{B3}, respectively. In this example, RZ_{A3} is an RNA ligase and RZ_{B3} an RNA polymerase. The RNA ligase produces random ligation products of the ribozymes such as the one depicted. Since the RNA polymerase also catalyzes the replication of RZ_{B1} from an autocatalytic set, there are start signals for the polymerase at any location where RZ_{B1} has been ligated into the RNA (indicated by upward arrowheads). The RNA polymerase generates the minus strand (–) ssRNA replication intermediates that also contain RNA polymerase start signals indicated by arrows. The plus strand copies are shown at the bottom; in this example these are two protogenome versions, both terminated by RZ_{B1} domains at the 5' and 3' ends, and three copies of RZ_{B1}.

For the second prerequisite there is evidence from naturally existing RNAs as well as from experiments that ribozymes can remain catalytically active when they are part of longer RNA molecules. For example, active hammerhead ribozyme domains are embedded in the larger circular ssRNA genomes of *Avsunviroidae* viroids [76], virus satellite circRNAs [77], and hepatitis delta virus [78], as well as in various retroelements such as short interspersed nuclear elements (SINEs) [79], Penelope-like elements [80], and retrozymes [81]. Moreover, various artificial RNAs with multiple active ribozyme domains have been successfully generated, including two-ribozyme constructs [82–84], RNAs with four [85,86], up to nine [87], and up to ten ribozyme domains [88]. Therefore, a protogenome combining in one molecule the necessary ribozyme activities to start a combination of autocatalytic sets appears to be structurally possible.

With respect to the third point, there are indeed length restrictions to the initial protogenome that are imposed by the fidelity of the RNA polymerase. The error threshold defines the maximum length of an RNA molecule beyond which deleterious mutations will accumulate, as proposed by Manfred Eigen [89]. For ribozymes, and therefore the

protogenome, a relaxed error threshold that increases the error threshold by approximately seven- to eightfold has been suggested, as mutations affect structural properties less than sequence information (as different RNA sequences can adopt a similar fold) [90]. According to this concept, the so far most accurate in vitro-generated RNA polymerase ribozyme with an average fidelity of about 97% [36,37] would allow for the stable replication of an RNA of only up to 250 nucleotides. This would, however, already allow for a number of ribozymes to be assembled in the protogenome, especially if one considers that ribozymes can be as small as five nucleotides. However, if the protogenome were to incorporate complex ribozymes such as the polymerase itself (likely around 200 nucleotides in length), a polymerase with higher fidelity would be required for stable replication. In this regard, the naturally evolved RNA polymerase ribozyme may have had a higher fidelity than the in vitro-generated variants, and fidelity might have been selected for in the protometabolism to allow for copying longer, more complex ribozymes. Of note, Attwater and colleagues successfully increased the average fidelity of an RNA polymerase ribozyme from 91% to above 97% by selecting for ribozymes that incorporate fewer mismatched nucleotides [38]. An RNA polymerase with an evolutionarily plausible fidelity of 99%, for example, would increase the relaxed error threshold to about 750 nucleotides. A protogenome of this length would already be able to contain several complex ribozymes. The error-prone nature of an early RNA polymerase ribozyme may have been the basis for further genetic innovation by introducing random mutations altering the catalytic activity and specificity of ribozymes.

2.2. Protogenome Evolution Outside of the Dry–Wet Cycle Scenario

If a protocell ‘escaped’ the dry–wet cycle scenario and became a free-living entity in the primordial ocean, or in areas such as tidal pools with higher concentrations of the organic molecule food sources, the evolutionary dynamics would have changed. The ability for liposome division must have had already evolved, e.g., via a phospholipid synthesis pathway (see above). The protogenome would have been an advantage for transmitting the metabolic information to the daughter cell during protocell division.

Yet, for the protogenome to become stably inherited during protocell division, the protometabolism had to become dependent on the protogenome. One way might have been that a genetic code evolved in the protogenome that enabled the generation of a peptide catalyst which provided an evolutionary advantage to the protocell. The protogenome, as it would have existed as a quasispecies due to the error-prone replication by an RNA polymerase enzyme, and likely contained redundant information (e.g., several copies of the same ribozyme domain) could have provided a ‘playing field’ for such a genetic novelty to evolve. Of note, small peptides as short as four amino acids have been shown to be efficient catalysts [91–93]. Such a peptide could be encoded by only eight (with a doublet code that might have preceded the triplet code [94]) or twelve nucleotides (with a triplet code). Simple RNA or DNA template-directed peptide formation reactions have been shown experimentally [95–97]. Such encoded peptides do not necessarily need to be catalysts themselves; they might have instead bound to existing ribozymes of the protometabolism and enhanced or modified their catalytic activities. Indeed, it has been proposed that the first template-directed peptides that evolved were cofactors of ribozymes [98]. Such peptides were likely positively charged, with an affinity to the negatively charged RNA. Experimentally, it has been shown, for example, that the activities of an RNA polymerase ribozyme and a hammerhead ribozyme can be strongly enhanced by the presence of positively charged small peptides [99] or a viral RNA-binding protein with net positive charge [100], respectively.

Another dependency mechanism might have been the evolution of addiction modules or toxin–antitoxin systems, in which the antitoxin RNA or templated peptide might have been present in the protogenome and neutralized a toxin in the protometabolism. There is evidence that a certain class of such toxins, Small Alarmone Synthetases, was already

present in the last universal common ancestor (LUCA) as they are found in both bacteria and archaea [101,102].

2.3. Further Evolution of the Protogenome—En Route to LUCA

In order to resemble present-day cellular genomes, the protogenome must have increased in size. As the RNAs of the protometabolism increased in complexity and size, entities resembling present-day transposable or virus-like elements might have emerged, acting cooperatively within the protocell, as proposed by Villarreal and Witzany [103]. They could have had the ability to integrate into the protogenome and thereby provide it with new genetic information. Ancient retroelements with ribozyme activity that are able to integrate into DNA genomes are the group II introns [104], and an artificial RNA composed of two ribozyme domains evolved in vitro had the ability to integrate into ssRNA [83]. Of note, the evolutionary ancestors of present-day viruses and transposable elements, many of which are capable of genomic integration, have likely existed before LUCA [105]. Another mechanism leading to increased protogenome size could have been duplication events of parts of the protogenome. The longer the protogenome became, the higher the fidelity of the RNA polymerase needed to be in order to increase the error threshold; thus, there might have been a co-evolution of increasing protogenome size and RNA polymerase fidelity. Enhanced fidelity might have been achieved, for example, by the evolution of RNA-binding peptides. In addition, it has been recently shown that the processivity of a ribozyme RdRp could be enhanced substantially by introducing a ‘clamping’ domain that prevents dissociation from the RNA template during polymerization [106]. This suggests that other domains, such as a proofreading exonuclease ribozyme, may also be added and could have evolved naturally, enhancing the fidelity of the RdRp. A protein RdRp eventually evolved, likely increasing the possible genome size to that of present-day RNA viruses, which is up to ~30,000 nucleotides. According to Aravind and colleagues, the common ancestor of present-day RdRPs was likely a simple ~40 amino-acid-long peptide with three beta-sheets and a loop with a conserved 5-amino acid motif essential for catalysis, which may have initially been a homodimeric cofactor of a ribozyme [107]. If encoded by a triplet code, this peptide would have required ~120 nucleotides, or ~80 nucleotides with a duplet code that may have preceded the triplet code evolutionarily [94]. The RNA polymerase enzyme likely evolved early, before the emergence of LUCA [108]. Even longer genomes required the evolution of DNA as the carrier of information, which is chemically more stable than RNA. This required the evolution of a reverse transcriptase, perhaps initially as a ribozyme [39] and later as a protein enzyme. The evolution of the DNA genome must have preceded LUCA; the largest RNA viruses, coronaviruses, encode less than twenty genes [109], whereas LUCA’s genome likely contained over 300 genes [110].

3. Discussion

It has been proposed previously that complex ribozymes such as the RNA polymerase might have emerged by the random ligation of smaller hairpin-loop structures, which was supported by in silico studies [111]. A similar scenario was proposed for the evolution of the ribosomal RNAs [112]. Here, I extended these ideas and argued that the ancient genome might have evolved in a similar way, likely in dry–wet scenarios on early Earth, which have been suggested as plausible sites for liposome formation, the evolution of metabolic protocells, and RNA polymers [5,6]. Initial RNA oligo- and polymers may have assembled in such a scenario, assembling into autocatalytic sets contained within liposome containers that form and dissolve repeatedly. Random ligation of ribozymes of RNA-based autocatalytic sets may have occurred by means of an RNA ligase ribozyme, as similar template-independent ligation products have been shown for protein ligases [73,74]. The resulting protogenome described herein linked, in one molecule, the information of several autocatalytic sets, such that it could efficiently seed an evolutionarily advantageous protometabolism to other protocells. It was subsequently replicated by an RNA polymerase, and presumably different versions were in evolutionary competition with each other

initially. Transient compartmentalization, a characteristic of the dry–wet cycle scenario, may have helped in the early evolution of the protogenome by preventing extinction due to parasitic mutants [113]. Complexity of the most fit variants was then increased by random mutations, perhaps leading to the evolution of the genetic code, through integration of early RNA transposable elements, and duplication events. The fidelity of the RNA polymerase ribozyme increased via co-evolution with the protogenome as it became larger. Once the protometabolism became dependent on the protogenome by means of the evolution of a genetic code within the protogenome, or by addition modules, the protogenome would have been stably inherited outside of the dry–wet cycle environment. Of note, a crucial step towards free-living protocells in general, and LUCA in particular, is the evolution of the genetic code and proteins, hallmark features of all known cellular life. Protein biogenesis requires the genetic code and both likely evolved in an RNA world scenario [114]. While the exact evolutionary origin of the genetic code remains elusive, one common theory suggests that the code (the association between codons and amino acids) was arbitrary initially and subsequently evolved by expansion [68,69,115–117]. Once a point was reached when no further expansion was possible, it was ‘frozen’, resulting in the present genetic code. Scenarios for the evolutionary origin of the genetic code have been formulated. For example, the ‘coding coenzyme handle hypothesis’ states that anticodon-like RNA adaptors charged with amino acids served as coenzymes for ribozymes in an RNA world, leading to the evolutionary selection of codon–anticodon pairs [118]. Fitting within this framework, it has been suggested that the initial codons may have been solely composed of two nucleobases (G and C) capable of encoding four amino acids including alanine, followed by an ‘alanine world’ with codons composed of G, C and A nucleobases and the corresponding amino acids all chemically related to alanine [114,119]. Finally, the full set of codons (composed of G, C, A and U) added the remaining proteinogenic amino acids. These scenarios bridge the evolutionary transition from an RNA world with non-coding protogenomes to the present-day cellular world in which the vast majority of catalysts are genome-encoded proteins.

The herein-described hypothesis is supported by computer simulation studies which have confirmed that compartments such as protocells can be favorable for linkage of smaller RNAs to longer strands to occur evolutionarily [120,121] and that in such a system linkage to a protogenome supports more efficient ribozyme evolution [122]. Moreover, simulations indicate that the protogenome needs to be required for protocell replication in order to be stably inherited [122] and that the establishment and maintenance of protogenomes (linked genes) is under positive selection in protocells harboring non-linked genes [123,124].

Some of the aspects of this hypothesis could be tested experimentally. For example, are any of the *in vitro* generated RNA ligase ribozymes [29] able to generate random ligation products similar to protein-based ligases [73,74]? This could be tested by providing the RNA ligase with a pool of random RNA oligomers, or longer ribozymes, followed by deep sequencing of the reaction. It could also be tested if the ribozymes of two different known RNA-based autocatalytic sets, when combined in one RNA molecule, retain their catalytic activities and can be used as a starting point for both autocatalytic sets. These relatively simple experiments could provide further credence for the plausibility of the herein suggested hypothesis on the evolution of the protogenome.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: I would like to thank Karin Moelling (Max Planck Institute for molecular Genetics) for ongoing support and stimulating discussions, Guenther Witzany (Telos-Philosophische Praxis, Buermoos, Austria) and the reviewers and editors for their valuable input.

Conflicts of Interest: I declare no conflict of interest.

References

1. Gilbert, W. Origin of life: The RNA world. *Nature* **1986**, *319*, 618. [[CrossRef](#)]
2. Kauffman, S.A. Cellular Homeostasis, Epigenesis, and Replication in Randomly Aggregated Macromolecular Systems. *J. Cybern.* **1971**, *1*, 71–96. [[CrossRef](#)]
3. Kauffman, S.A. Autocatalytic sets of proteins. *J. Theor. Biol.* **1986**, *119*, 1–24. [[CrossRef](#)]
4. Hordijk, W.; Hein, J.; Steel, M. Autocatalytic Sets and the Origin of Life. *Entropy* **2010**, *12*, 1733–1742. [[CrossRef](#)]
5. Pearce, B.K.D.; Pudritz, R.E.; Semenov, D.A.; Henning, T.K. Origin of the RNA world: The fate of nucleobases in warm little ponds. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 11327–11332. [[CrossRef](#)] [[PubMed](#)]
6. Damer, B.; Deamer, D. Coupled phases and combinatorial selection in fluctuating hydrothermal pools: A scenario to guide experimental approaches to the origin of cellular life. *Life* **2015**, *5*, 872–887. [[CrossRef](#)]
7. Theobald, D.L. A formal test of the theory of universal common ancestry. *Nature* **2010**, *465*, 219–222. [[CrossRef](#)]
8. Stribling, R.; Miller, S.L. Energy yields for hydrogen cyanide and formaldehyde syntheses: The HCN and amino acid concentrations in the primitive ocean. *Orig. Life Evol. Biosph.* **1987**, *17*, 261–273. [[CrossRef](#)]
9. Martin, W.; Baross, J.; Kelley, D.; Russell, M.J. Hydrothermal vents and the origin of life. *Nat. Rev. Microbiol.* **2008**, *6*, 805–814. [[CrossRef](#)]
10. Miller, S.L. A production of amino acids under possible primitive earth conditions. *Science* **1953**, *117*, 528–529. [[CrossRef](#)]
11. Sossi, P.A.; Burnham, A.D.; Badro, J.; Lanzirrotti, A.; Newville, M.; O'Neill, H.S.C. Redox state of Earth's magma ocean and its Venus-like early atmosphere. *Sci. Adv.* **2020**, *6*, eabd1387. [[CrossRef](#)]
12. Oro, J.; Kimball, A.P. Synthesis of purines under possible primitive earth conditions. I. Adenine from hydrogen cyanide. *Arch. Biochem. Biophys.* **1961**, *94*, 217–227. [[CrossRef](#)]
13. Oro, J.; Kamat, S.S. Amino-acid synthesis from hydrogen cyanide under possible primitive earth conditions. *Nature* **1961**, *190*, 442–443. [[CrossRef](#)] [[PubMed](#)]
14. Powner, M.W.; Gerland, B.; Sutherland, J.D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **2009**, *459*, 239–242. [[CrossRef](#)]
15. Patel, B.H.; Percivalle, C.; Ritson, D.J.; Duffy, C.D.; Sutherland, J.D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* **2015**, *7*, 301–307. [[CrossRef](#)]
16. Kitada, N.; Maruyama, S. Origins of building blocks of life: A review. *Geosci. Front.* **2018**, *9*, 1117–1153. [[CrossRef](#)]
17. Becker, S.; Feldmann, J.; Wiedemann, S.; Okamura, H.; Schneider, C.; Iwan, K.; Crisp, A.; Rossa, M.; Amatov, T.; Carell, T. Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides. *Science* **2019**, *366*, 76–82. [[CrossRef](#)]
18. Martins, Z.; Botta, O.; Fogel, M.L.; Sephton, M.A.; Glavin, D.P.; Watson, J.S.; Dworkin, J.P.; Schwartz, A.W.; Ehrenfreund, P. Extraterrestrial nucleobases in the Murchison meteorite. *Earth Planet. Sci. Lett.* **2008**, *270*, 130–136. [[CrossRef](#)]
19. Orgel, L.E. Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99–123. [[CrossRef](#)] [[PubMed](#)]
20. Huang, W.; Ferris, J.P. Synthesis of 35–40 mers of RNA oligomers from unblocked monomers. A simple approach to the RNA world. *Chem. Commun.* **2003**, *15*, 1458–1459. [[CrossRef](#)]
21. Fahrenbach, A.C.; Giurgiu, C.; Tam, C.P.; Li, L.; Hongo, Y.; Aono, M.; Szostak, J.W. Common and Potentially Prebiotic Origin for Precursors of Nucleotide Synthesis and Activation. *J. Am. Chem. Soc.* **2017**, *139*, 8780–8783. [[CrossRef](#)] [[PubMed](#)]
22. Li, L.; Prywes, N.; Tam, C.P.; O'Flaherty, D.K.; Lelyveld, V.S.; Izgu, E.C.; Pal, A.; Szostak, J.W. Enhanced Nonenzymatic RNA Copying with 2-Aminoimidazole Activated Nucleotides. *J. Am. Chem. Soc.* **2017**, *139*, 1810–1813. [[CrossRef](#)] [[PubMed](#)]
23. Sievers, D.; von Kiedrowski, G. Self-Replication of Hexadeoxynucleotide Analogues: Autocatalysis versus Cross-Catalysis. *Chem. Eur. J.* **1998**, *4*, 629–641. [[CrossRef](#)]
24. Edeleva, E.; Salditt, A.; Stamp, J.; Schwintek, P.; Boekhoven, J.; Braun, D. Continuous nonenzymatic cross-replication of DNA strands with in situ activated DNA oligonucleotides. *Chem. Sci.* **2019**, *10*, 5807–5814. [[CrossRef](#)]
25. Cech, T.R. Structural biology. The ribosome is a ribozyme. *Science* **2000**, *289*, 878–879. [[CrossRef](#)]
26. Altman, S. Nobel lecture. Enzymatic cleavage of RNA by RNA. *Biosci. Rep.* **1990**, *10*, 317–337. [[CrossRef](#)]
27. Nielsen, H.; Westhof, E.; Johansen, S. An mRNA is capped by a 2', 5' lariat catalyzed by a group I-like ribozyme. *Science* **2005**, *309*, 1584–1587. [[CrossRef](#)] [[PubMed](#)]
28. Canny, M.D.; Jucker, F.M.; Pardi, A. Efficient ligation of the Schistosoma hammerhead ribozyme. *Biochemistry* **2007**, *46*, 3826–3834. [[CrossRef](#)]
29. Joyce, G.F. Forty years of in vitro evolution. *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 6420–6436. [[CrossRef](#)]
30. Beaudry, A.A.; Joyce, G.F. Directed evolution of an RNA enzyme. *Science* **1992**, *257*, 635–641. [[CrossRef](#)]
31. Bartel, D.P.; Szostak, J.W. Isolation of new ribozymes from a large pool of random sequences. *Science* **1993**, *261*, 1411–1418. [[CrossRef](#)]
32. Johnston, W.K.; Unrau, P.J.; Lawrence, M.S.; Glasner, M.E.; Bartel, D.P. RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension. *Science* **2001**, *292*, 1319–1325. [[CrossRef](#)] [[PubMed](#)]
33. Zaher, H.S.; Unrau, P.J. Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* **2007**, *13*, 1017–1026. [[CrossRef](#)]
34. Wochner, A.; Attwater, J.; Coulson, A.; Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science* **2011**, *332*, 209–212. [[CrossRef](#)]

35. Tjhung, K.F.; Shokhirev, M.N.; Horning, D.P.; Joyce, G.F. An RNA polymerase ribozyme that synthesizes its own ancestor. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 2906–2913. [[CrossRef](#)] [[PubMed](#)]
36. Attwater, J.; Wochner, A.; Holliger, P. In-ice evolution of RNA polymerase ribozyme activity. *Nat. Chem.* **2013**, *5*, 1011–1018. [[CrossRef](#)] [[PubMed](#)]
37. Horning, D.P.; Joyce, G.F. Amplification of RNA by an RNA polymerase ribozyme. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 9786–9791. [[CrossRef](#)]
38. Attwater, J.; Raguram, A.; Morgunov, A.S.; Gianni, E.; Holliger, P. Ribozyme-catalysed RNA synthesis using triplet building blocks. *Elife* **2018**, *7*, e35255. [[CrossRef](#)] [[PubMed](#)]
39. Samanta, B.; Joyce, G.F. A reverse transcriptase ribozyme. *Elife* **2017**, *6*, e31153. [[CrossRef](#)]
40. Unrau, P.J.; Bartel, D.P. RNA-catalysed nucleotide synthesis. *Nature* **1998**, *395*, 260–263. [[CrossRef](#)]
41. Lohse, P.A.; Szostak, J.W. Ribozyme-catalysed amino-acid transfer reactions. *Nature* **1996**, *381*, 442–444. [[CrossRef](#)]
42. Chun, S.M.; Jeong, S.; Kim, J.M.; Chong, B.O.; Park, Y.K.; Park, H.; Yu, J. Cholesterol Esterase Activity by in Vitro Selection of RNA against a Phosphate Transition-State Analogue. *J. Am. Chem. Soc.* **1999**, *121*, 10844–10845. [[CrossRef](#)]
43. Saran, D.; Held, D.M.; Burke, D.H. Multiple-turnover thio-ATP hydrolase and phospho-enzyme intermediate formation activities catalyzed by an RNA enzyme. *Nucleic Acids Res.* **2006**, *34*, 3201–3208. [[CrossRef](#)] [[PubMed](#)]
44. Tsukiji, S.; Pattnaik, S.B.; Suga, H. An alcohol dehydrogenase ribozyme. *Nat. Struct. Biol.* **2003**, *10*, 713–717. [[CrossRef](#)] [[PubMed](#)]
45. Tsukiji, S.; Pattnaik, S.B.; Suga, H. Reduction of an aldehyde by a NADH/Zn²⁺-dependent redox active ribozyme. *J. Am. Chem. Soc.* **2004**, *126*, 5044–5055. [[CrossRef](#)] [[PubMed](#)]
46. Janzen, E.; Blanco, C.; Peng, H.; Kenchel, J.; Chen, I.A. Promiscuous Ribozymes and Their Proposed Role in Prebiotic Evolution. *Chem. Rev.* **2020**, *120*, 4879–4897. [[CrossRef](#)] [[PubMed](#)]
47. Piette, B.M.A.G.; Heddle, J.G. A Peptide-Nucleic Acid Replicator Origin for Life. *Trends Ecol. Evol.* **2020**, *35*, 397–406. [[CrossRef](#)]
48. Lincoln, T.A.; Joyce, G.F. Self-sustained replication of an RNA enzyme. *Science* **2009**, *323*, 1229–1232. [[CrossRef](#)]
49. Kun, Á.; Szilágyi, A.; Könnnyű, B.; Boza, G.; Zachar, I.; Szathmáry, E. The dynamics of the RNA world: Insights and challenges. *Ann. N. Y. Acad. Sci.* **2015**, *1341*, 75–95. [[CrossRef](#)] [[PubMed](#)]
50. Kim, D.E.; Joyce, G.F. Cross-catalytic replication of an RNA ligase ribozyme. *Chem. Biol.* **2004**, *11*, 1505–1512. [[CrossRef](#)] [[PubMed](#)]
51. Sievers, D.; von Kiedrowski, G. Self-replication of complementary nucleotide-based oligomers. *Nature* **1994**, *369*, 221–224. [[CrossRef](#)] [[PubMed](#)]
52. Patzke, V.; von Kiedrowski, G. Self replicating systems. *Arkivoc* **2007**, *5*, 293–310. [[CrossRef](#)]
53. Vaidya, N.; Manapat, M.L.; Chen, I.A.; Xulvi-Brunet, R.; Hayden, E.J.; Lehman, N. Spontaneous network formation among cooperative RNA replicators. *Nature* **2012**, *491*, 72–77. [[CrossRef](#)] [[PubMed](#)]
54. Hayden, E.J.; Riley, C.A.; Burton, A.S.; Lehman, N. RNA-directed construction of structurally complex and active ligase ribozymes through recombination. *RNA* **2005**, *11*, 1678–1687. [[CrossRef](#)] [[PubMed](#)]
55. Arsène, S.; Ameta, S.; Lehman, N.; Griffiths, A.D.; Nghe, P. Coupled catabolism and anabolism in autocatalytic RNA sets. *Nucleic Acids Res.* **2018**, *46*, 9660–9666. [[CrossRef](#)] [[PubMed](#)]
56. Jeancolas, C.; Matsubara, Y.J.; Vybornyi, M.; Lambert, C.N.; Blokhuis, A.; Alline, T.; Griffiths, A.D.; Ameta, S.; Krishna, S.; Nghe, P. RNA diversification by a self-reproducing ribozyme revealed by deep sequencing and kinetic modelling. *Chem. Commun.* **2021**, *57*, 7517–7520. [[CrossRef](#)] [[PubMed](#)]
57. Ameta, S.; Arsène, S.; Foulon, S.; Saudemont, B.; Clifton, B.E.; Griffiths, A.D.; Nghe, P. Darwinian properties and their trade-offs in autocatalytic RNA reaction networks. *Nat. Commun.* **2021**, *12*, 842. [[CrossRef](#)]
58. Ashkenasy, G.; Jagasia, R.; Yadav, M.; Ghadiri, M.R. Design of a directed molecular network. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 10872–10877. [[CrossRef](#)]
59. Sousa, F.L.; Hordijk, W.; Steel, M.; Martin, W.F. Autocatalytic sets in *E. coli* metabolism. *J. Syst. Chem.* **2015**, *6*, 4. [[CrossRef](#)]
60. Hordijk, W.; Steel, M.; Kauffman, S. The structure of autocatalytic sets: Evolvability, enablement, and emergence. *Acta Biotheor.* **2012**, *60*, 379–392. [[CrossRef](#)] [[PubMed](#)]
61. Hordijk, W.; Naylor, J.; Krasnogor, N.; Fellermann, H. Population Dynamics of Autocatalytic Sets in a Compartmentalized Spatial World. *Life* **2018**, *8*, 33. [[CrossRef](#)]
62. Vasas, V.; Fernando, C.; Santos, M.; Kauffman, S.; Szathmáry, E. Evolution before genes. *Biol. Direct.* **2012**, *7*, 1–14. [[CrossRef](#)]
63. Charlat, S.; Ariew, A.; Bourrat, P.; Ferreira Ruiz, M.; Heams, T.; Huneman, P.; Krishna, S.; Lachmann, M.; Lartillot, N.; Le Sergeant d’Hendecourt, L.; et al. Natural Selection beyond Life? A Workshop Report. *Life* **2021**, *11*, 1051. [[CrossRef](#)]
64. Hanczyc, M.M.; Szostak, J.W. Replicating vesicles as models of primitive cell growth and division. *Curr. Opin. Chem. Biol.* **2004**, *8*, 660–664. [[CrossRef](#)]
65. Berclaz, N.; Müller, M.; Walde, P.; Luisi, P.L. Growth and Transformation of Vesicles Studied by Ferritin Labeling and Cryotransmission Electron Microscopy. *J. Phys. Chem.* **2001**, *105*, 1056–1064. [[CrossRef](#)]
66. Hanczyc, M.M.; Fujikawa, S.M.; Szostak, J.W. Experimental models of primitive cellular compartments: Encapsulation, growth, and division. *Science* **2003**, *302*, 618–622. [[CrossRef](#)] [[PubMed](#)]
67. Castro, J.M.; Sugiyama, H.; Toyota, T. Budding and Division of Giant Vesicles Linked to Phospholipid Production. *Sci. Rep.* **2019**, *9*, 165. [[CrossRef](#)]
68. Crick, F.H. On protein synthesis. *Symp. Soc. Exp. Biol.* **1958**, *12*, 138–163. [[PubMed](#)]
69. Crick, F.H. The origin of the genetic code. *J. Mol. Biol.* **1968**, *38*, 367–379. [[CrossRef](#)]

70. Turk, R.M.; Chumachenko, N.V.; Yarus, M. Multiple translational products from a five-nucleotide ribozyme. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4585–4589. [[CrossRef](#)]
71. Nomura, Y.; Yokobayashi, Y. Systematic minimization of RNA ligase ribozyme through large-scale design-synthesis-sequence cycles. *Nucleic Acids Res.* **2019**, *47*, 8950–8960. [[CrossRef](#)] [[PubMed](#)]
72. Cervera, A.; Urbina, D.; de la Peña, M. Retrozymes are a unique family of non-autonomous retrotransposons with hammerhead ribozymes that propagate in plants through circular RNAs. *Genome Biol.* **2016**, *17*, 135. [[CrossRef](#)] [[PubMed](#)]
73. Kuhn, H.; Frank-Kamenetskii, M.D. Template-independent ligation of single-stranded DNA by T4 DNA ligase. *FEBS J.* **2005**, *272*, 5991–6000. [[CrossRef](#)] [[PubMed](#)]
74. Zhang, L.; Tripathi, A. Archaeal RNA ligase from *Thermococcus kodakarensis* for template dependent ligation. *RNA Biol.* **2017**, *14*, 36–44. [[CrossRef](#)]
75. Dhar, N.; Weinberg, M.S.; Michod, R.E.; Durand, P.M. Molecular trade-offs in RNA ligases affected the modular emergence of complex ribozymes at the origin of life. *R. Soc. Open Sci.* **2017**, *4*, 170376. [[CrossRef](#)]
76. Di Serio, F.; Li, S.F.; Matoušek, J.; Owens, R.A.; Pallás, V.; Randles, J.W.; Sano, T.; Verhoeven, J.T.J.; Vidalakis, G.; Flores, R.; et al. ICTV Virus Taxonomy Profile: *Avsunviroidae*. *J. Gen. Virol.* **2018**, *99*, 611–612. [[CrossRef](#)]
77. de la Peña, M.; Ceprián, R.; Cervera, A. A Singular and Widespread Group of Mobile Genetic Elements: RNA Circles with Autocatalytic Ribozymes. *Cells* **2020**, *9*, 2555. [[CrossRef](#)] [[PubMed](#)]
78. Wadkins, T.S.; Been, M.D. Ribozyme activity in the genomic and antigenomic RNA strands of hepatitis delta virus. *Cell. Mol. Life Sci.* **2002**, *59*, 112–125. [[CrossRef](#)] [[PubMed](#)]
79. Ferbeyre, G.; Smith, J.M.; Cedergren, R. Schistosome satellite DNA encodes active hammerhead ribozymes. *Mol. Cell. Biol.* **1998**, *18*, 3880–3888. [[CrossRef](#)] [[PubMed](#)]
80. Cervera, A.; De la Peña, M. Eukaryotic penelope-like retroelements encode hammerhead ribozyme motifs. *Mol. Biol. Evol.* **2014**, *31*, 2941–2947. [[CrossRef](#)]
81. de la Peña, M.; Cervera, A. Circular RNAs with hammerhead ribozymes encoded in eukaryotic genomes: The enemy at home. *RNA Biol.* **2017**, *14*, 985–991. [[CrossRef](#)] [[PubMed](#)]
82. Shahi, S.; Banerjee, A.C. Multitarget ribozyme against the S1 genome segment of reovirus possesses novel cleavage activities and is more efficacious than its constituent mono-ribozymes. *Antivir. Res.* **2002**, *55*, 129–140. [[CrossRef](#)]
83. Kumar, R.M.; Joyce, G.F. A modular, bifunctional RNA that integrates itself into a target RNA. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 9738–9743. [[CrossRef](#)] [[PubMed](#)]
84. Ivanov, S.A.; Vauléon, S.; Müller, S. Efficient RNA ligation by reverse-joined hairpin ribozymes and engineering of twin ribozymes consisting of conventional and reverse-joined hairpin ribozyme units. *FEBS J.* **2005**, *272*, 4464–4474. [[CrossRef](#)]
85. Perriman, R.; Graf, L.; Gerlach, W.L. A ribozyme that enhances gene suppression in tobacco protoplasts. *Antisense Res. Dev.* **1993**, *3*, 253–263. [[CrossRef](#)] [[PubMed](#)]
86. Cameron, F.H.; Jennings, P.A. Multiple domains in a ribozyme construct confer increased suppressive activity in monkey cells. *Antisense Res. Dev.* **1994**, *4*, 87–94. [[CrossRef](#)] [[PubMed](#)]
87. Chen, C.J.; Banerjee, A.C.; Harmison, G.G.; Haglund, K.; Schubert, M. Multitarget-ribozyme directed to cleave at up to nine highly conserved HIV-1 env RNA regions inhibits HIV-1 replication—potential effectiveness against most presently sequenced HIV-1 isolates. *Nucleic Acids Res.* **1992**, *20*, 4581–4589. [[CrossRef](#)]
88. Ohkawa, J.; Yuyama, N.; Takebe, Y.; Nishikawa, S.; Taira, K. Importance of independence in ribozyme reactions: Kinetic behavior of trimmed and of simply connected multiple ribozymes with potential activity against human immunodeficiency virus. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 11302–11306. [[CrossRef](#)] [[PubMed](#)]
89. Eigen, M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **1971**, *58*, 465–523. [[CrossRef](#)] [[PubMed](#)]
90. Kun, A.; Santos, M.; Szathmáry, E. Real ribozymes suggest a relaxed error threshold. *Nat. Genet.* **2005**, *37*, 1008–1011. [[CrossRef](#)]
91. Griswold, K.S.; Miller, S.J. A peptide-based catalyst approach to regioselective functionalization of carbohydrates. *Tetrahedron* **2003**, *59*, 8869–8875. [[CrossRef](#)]
92. Kolundzic, F.; Noshi, M.N.; Tjandra, M.; Movassaghi, M.; Miller, S.J. Chemoselective and enantioselective oxidation of indoles employing aspartyl peptide catalysts. *J. Am. Chem. Soc.* **2011**, *133*, 9104–9111. [[CrossRef](#)] [[PubMed](#)]
93. Peris, G.; Jakobsche, C.E.; Miller, S.J. Aspartate-catalyzed asymmetric epoxidation reactions. *J. Am. Chem. Soc.* **2007**, *129*, 8710–8711. [[CrossRef](#)] [[PubMed](#)]
94. Koonin, E.V.; Novozhilov, A.S. Origin and evolution of the genetic code: The universal enigma. *IUBMB Life* **2009**, *61*, 99–111. [[CrossRef](#)]
95. Bruick, R.K.; Dawson, P.E.; Kent, S.B.; Usman, N.; Joyce, G.F. Template-directed ligation of peptides to oligonucleotides. *Chem. Biol.* **1996**, *3*, 49–56. [[CrossRef](#)]
96. Tamura, K.; Schimmel, P. Peptide synthesis with a template-like RNA guide and aminoacyl phosphate adaptors. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8666–8669. [[CrossRef](#)] [[PubMed](#)]
97. Di Pisa, M.; Hauser, A.; Seitz, O. Maximizing Output in RNA-Programmed Peptidyl-Transfer Reactions. *ChemBiochem* **2017**, *18*, 872–879. [[CrossRef](#)]
98. Wolf, Y.I.; Koonin, E.V. On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. *Biol. Direct.* **2007**, *2*, 14. [[CrossRef](#)]

99. Tagami, S.; Attwater, J.; Holliger, P. Simple peptides derived from the ribosomal core potentiate RNA polymerase ribozyme function. *Nat. Chem.* **2017**, *9*, 325–332. [[CrossRef](#)]
100. Moelling, K.; Mueller, G.; Dannull, J.; Reuss, C.; Beimling, P.; Bartz, C.; Wiedenmann, B.; Yoon, K.; Surovoy, A.; Jung, G. Stimulation of Ki-ras ribozyme activity by RNA binding protein, NCp7, in vitro and in pancreatic tumor cell line, capan 1. *Ann. N. Y. Acad. Sci.* **1994**, *733*, 113–121. [[CrossRef](#)] [[PubMed](#)]
101. Harms, A.; Brodersen, D.E.; Mitarai, N.; Gerdes, K. Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology. *Mol. Cell* **2018**, *70*, 768–784. [[CrossRef](#)] [[PubMed](#)]
102. Fraikin, N.; Goormaghtigh, F.; Van Melderen, L. Type II Toxin-Antitoxin Systems: Evolution and Revolutions. *J. Bacteriol.* **2020**, *202*, e00763-19. [[CrossRef](#)] [[PubMed](#)]
103. Villarreal, L.P.; Witzany, G. Social Networking of Quasi-Species Consortia drive Virolution via Persistence. *AIMS Microbiol.* **2021**, *7*, 138–162. [[CrossRef](#)] [[PubMed](#)]
104. Lambowitz, A.M.; Zimmerly, S. Group II introns: Mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.* **2011**, *3*, a003616. [[CrossRef](#)]
105. Krupovic, M.; Dolja, V.V.; Koonin, E.V. The LUCA and its complex virome. *Nat. Rev. Microbiol.* **2020**, *18*, 661–670. [[CrossRef](#)]
106. Cojocaru, R.; Unrau, P.J. Processive RNA polymerization and promoter recognition in an RNA World. *Science* **2021**, *371*, 1225–1232. [[CrossRef](#)]
107. Iyer, L.M.; Koonin, E.V.; Aravind, L. Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Struct. Biol.* **2003**, *3*, 1. [[CrossRef](#)] [[PubMed](#)]
108. Lazcano, A.; Fastag, J.; Gariglio, P.; Ramírez, C.; Oró, J. On the early evolution of RNA polymerase. *J. Mol. Evol.* **1988**, *27*, 365–376. [[CrossRef](#)] [[PubMed](#)]
109. Nelson, C.W.; Ardern, Z.; Goldberg, T.L.; Meng, C.; Kuo, C.H.; Ludwig, C.; Kolokotronis, S.O.; Wei, X. Dynamically evolving novel overlapping gene as a factor in the SARS-CoV-2 pandemic. *Elife* **2020**, *9*, e59633. [[CrossRef](#)]
110. Weiss, M.C.; Sousa, F.L.; Mrnjavac, N.; Neukirchen, S.; Roettger, M.; Nelson-Sathi, S.; Martin, W.F. The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* **2016**, *1*, 16116. [[CrossRef](#)]
111. Briones, C.; Stich, M.; Manrubia, S.C. The dawn of the RNA World: Toward functional complexity through ligation of random RNA oligomers. *RNA* **2009**, *15*, 743–749. [[CrossRef](#)] [[PubMed](#)]
112. Root-Bernstein, M.; Root-Bernstein, R. The ribosome as a missing link in the evolution of life. *J. Theor. Biol.* **2015**, *367*, 130–158. [[CrossRef](#)] [[PubMed](#)]
113. Matsumura, S.; Kun, Á.; Ryckelynck, M.; Coldren, F.; Szilágyi, A.; Jossinet, F.; Rick, C.; Nghe, P.; Szathmáry, E.; Griffiths, A.D. Transient compartmentalization of RNA replicators prevents extinction due to parasites. *Science* **2016**, *354*, 1293–1296. [[CrossRef](#)] [[PubMed](#)]
114. Kubyshkin, V.; Budisa, N. The Alanine World Model for the Development of the Amino Acid Repertoire in Protein Biosynthesis. *Int. J. Mol. Sci.* **2019**, *20*, 5507. [[CrossRef](#)]
115. Barbieri, M. Evolution of the genetic code: The ambiguity-reduction theory. *Biosystems* **2019**, *185*, 104024. [[CrossRef](#)]
116. Seligmann, H.; Warthi, G. Genetic Code Optimization for Cotranslational Protein Folding: Codon Directional Asymmetry Correlates with Antiparallel Betasheets, tRNA Synthetase Classes. *Comput. Struct. Biotechnol. J.* **2017**, *15*, 412–424. [[CrossRef](#)]
117. Ribas de Pouplana, L.; Torres, A.G.; Rafels-Ybern, À. What Froze the Genetic Code? *Life* **2017**, *7*, 14. [[CrossRef](#)] [[PubMed](#)]
118. Szathmáry, E. Coding coenzyme handles: A hypothesis for the origin of the genetic code. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 9916–9920. [[CrossRef](#)] [[PubMed](#)]
119. Kubyshkin, V.; Budisa, N. Anticipating alien cells with alternative genetic codes: Away from the alanine world! *Curr. Opin. Biotechnol.* **2019**, *60*, 242–249. [[CrossRef](#)]
120. Shah, V.; de Bouter, J.; Pauli, Q.; Tupper, A.S.; Higgs, P.G. Survival of RNA Replicators Is Much Easier in Protocells Than in Surface-Based, Spatial Systems. *Life* **2019**, *9*, 65. [[CrossRef](#)] [[PubMed](#)]
121. Shah, V. Comparing Protocell and Surface-Based Models of RNA Replicator Systems and Determining Favourable Conditions for Linkage of Functional Strands. Master's Thesis, McMaster University, Hamilton, ON, Canada, 2019.
122. Szilágyi, A.; Kun, A.; Szathmáry, E. Early evolution of efficient enzymes and genome organization. *Biol. Direct* **2012**, *7*, 38. [[CrossRef](#)] [[PubMed](#)]
123. Smith, J.M.; Szathmáry, E. The origin of chromosomes. I. Selection for linkage. *J. Theor. Biol.* **1993**, *164*, 437–446. [[CrossRef](#)] [[PubMed](#)]
124. Szilágyi, A.; Kovács, V.P.; Szathmáry, E.; Santos, M. Evolution of linkage and genome expansion in protocells: The origin of chromosomes. *PLoS Genet.* **2020**, *16*, e1009155. [[CrossRef](#)] [[PubMed](#)]