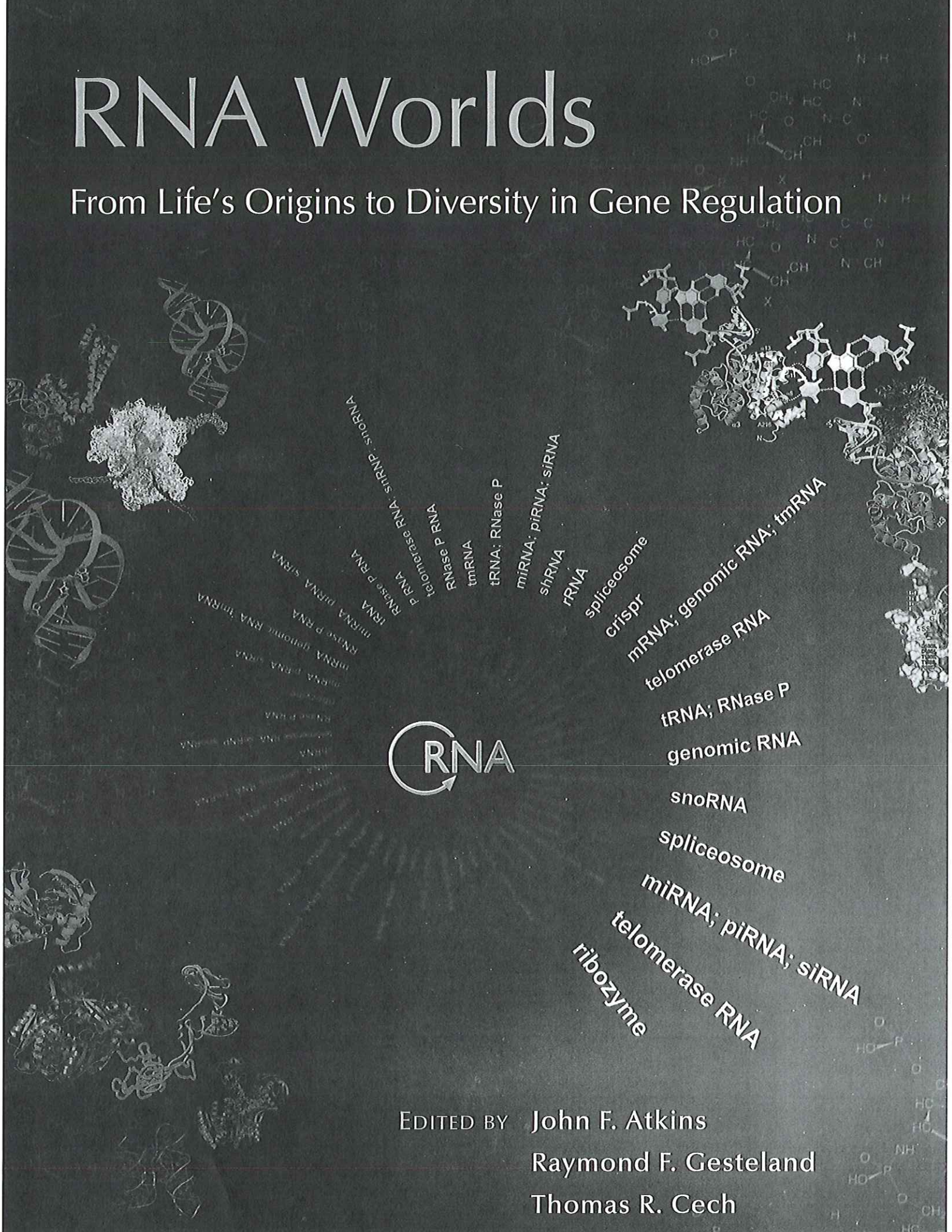
RNA Worlds

From Life's Origins to Diversity in Gene Regulation



EDITED BY John F. Atkins
Raymond F. Gesteland
Thomas R. Cech

Setting the Stage: The History, Chemistry, and Geobiology behind RNA

Steven A. Benner, Hyo-Joong Kim, and Zunyi Yang

Foundation for Applied Molecular Evolution, Gainesville, Florida 32601

Correspondence: sbenner@ffame.org

SUMMARY

No community-accepted scientific methods are available today to guide studies on what role RNA played in the origin and early evolution of life on Earth. Further, a definition-theory for life is needed to develop hypotheses relating to the "RNA First" model for the origin of life. Four approaches are currently at various stages of development of such a definition-theory to guide these studies. These are (a) paleogenetics, in which inferences about the structure of past life are drawn from the structure of present life; (b) prebiotic chemistry, in which hypotheses with experimental support are sought that get RNA from organic and inorganic species possibly present on early Earth; (c) exploration, hoping to encounter life independent of terran life, which might contain RNA; and (d) synthetic biology, in which laboratories attempt to reproduce biological behavior with unnatural chemical systems.

Outline

Cite this article as *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a003541

1 INTRODUCTION

Most scientists have been taught that science develops models for reality by systematically observing, analyzing observations, and designing experiments to test hypotheses that emerge from that analysis. Those making a career in science soon realize that things are not so simple, especially as some of the most interesting questions in science do not lend themselves to such an approach. These include questions about the current reality ("Does alien life exist in the cosmos?") and future reality ("Will human carbon dioxide emissions trigger catastrophic global warming?"). We cannot observe directly most of the cosmos. We cannot observe directly the future. Other approaches are needed to address such questions.

Other approaches are also needed to address equally interesting questions about the historical past. Going backward in time, these include questions like: "How did *Homo sapiens* originate?" "How did multicellular life emerge?" "How was the Earth formed?"

Fortunately, past reality is more accessible than future or distant realities. Ever since the Enlightenment, natural historians have developed strategies to generate models for the past using observations made today, combined with models that can be experimentally tested that (one hopes) apply to all realities at all times. Using these strategies, we today have good models for how the Earth and its Sun were formed, plausible models for how humankind came into being, and rudimentary models for how multicellularity arose.

Unfortunately, one of the most interesting questions about the past has resisted these approaches: "How did life originate?" The best evidence suggests that the antiquity of the key events is greater than the antiquity of any surviving physical record. Hard work in the Olduvai George or elsewhere on Earth does not seem likely to find fossils relevant to this question. Nor can we (yet) constrain our models by observing the formation of life on planets orbiting distant stars that are today undergoing biogenesis.

Still worse, the problem is associated with a resilient semantic question: "What is 'life'?" The question has been long discussed (Koshland 2002), often avoided (Baross and Benner 2007), and infrequently addressed, as when a NASA panel concluded that "life is a self-sustaining chemical system capable of Darwinian evolution" (Joyce et al. 1994).

It is clear that any definition of life must incorporate a "theory of life" (Cleland and Chyba 2000; Benner 2009). The "NASA definition" did so, excluding, for example, nonchemical and Lamarckian systems (Benner et al., 2004). But synthetic biologists attempting to make artificial systems "capable of Darwinian evolution" also find that things are not so simple. Synthetic biologists are today rearranging natural genes in unnatural arrangements to form a cell whose genes come entirely from elsewhere (Lartigue et al. 2009). Others have already generated artificial chemical systems capable of Darwinian evolution by rearranging atoms in terran genetic molecular systems (Yang et al. 2009). Still others have found cellular systems that can grow, divide, and collect material (Yarus 2010). We now know that such advances simply move the bar, because few in the community are prepared to call these "artificial life."

The need to codevelop a definition-theory of life concomitant with our development of a model for its earliest forms on Earth makes understanding life's origins one of the most intellectually interesting challenges in contemporary science. Many approaches might meet this challenge (Fig. 1).

One of the most direct comes from the tradition of natural history, which includes fields like geology, paleontology, and molecular evolution. In this tradition, we start with the life that we know on Earth and whatever physical record that we have about past life (including fossils) and work *backward* in time to create models for simpler forms of life, coupling this with what we know about the chemistry of life (Benner et al. 2002). Conjectural models for past biology can be confirmed in part using the emerging field of paleogenetics, which resurrects genes and proteins from extinct organisms for study in the laboratory (Liberles 2007). Paleogenetics allows experiments to be performed directly on biomolecules that vanished long ago, bringing the power of the experimental method directly to bear on historical questions (Benner 2007).

The concept of an "RNA world" (Gilbert 1986) is itself an outcome of this backwards-in-time process (Rich 1962; Crick 1968; White 1976; Visser and Kellogg 1978). The most direct evidence for an RNA World, crystallographic evidence that the RNA in the ribosome catalyzes peptide bond formation (Harms et al. 2001; Moore and Steitz 2003), is applied within a historical argument, one that infers that all ribosomes in terran biology use RNA to synthesize peptide bonds, and therefore the last common ancestor of all ribosomes used RNA to synthesize peptide bonds. Extrapolating back in time using genomic data has offered us a glimpse of the metabolic complexity of the RNA world (Benner et al. 1989; Benner et al. 1993; Koonin 2003).

Alternatively, we might work forwards in time. We can today make an inventory of molecules observed in nearby regions of our galaxy that are right now forming stars, planets, and (perhaps) life, adding molecules found in comets and meteorites that may have delivered to a nascent Earth (Pizzarello 2004; Chyba and Sagan 1992). We might complete our inventory by guessing what other molecules might have been generated on primitive Earth, exploiting

Cite this article as *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a003541

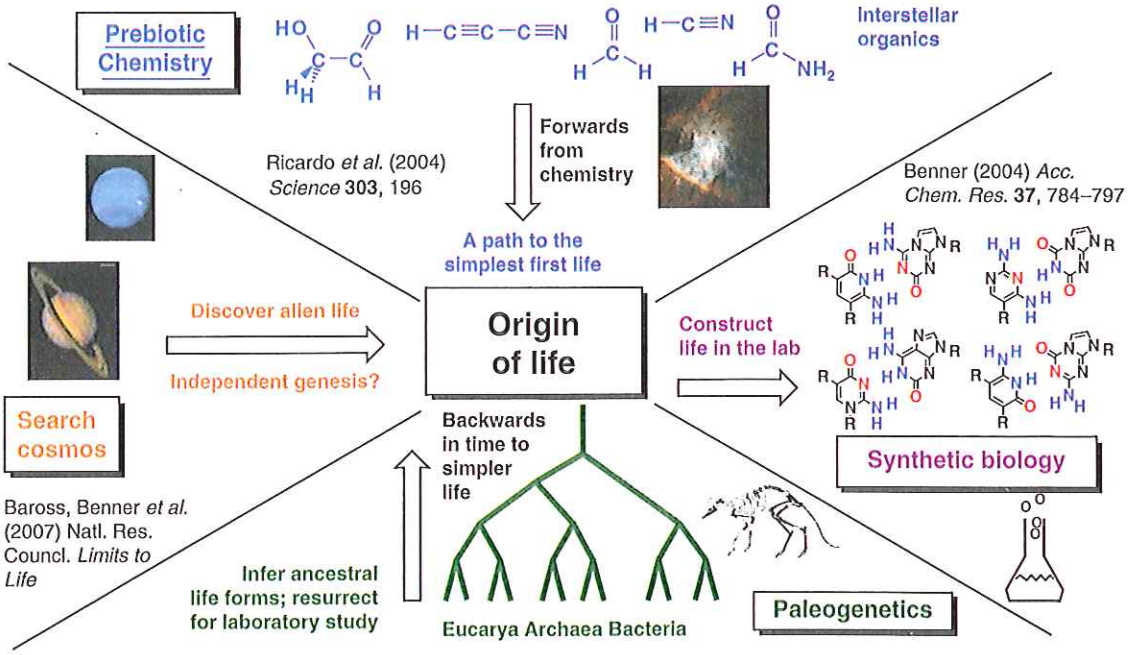


Figure 1. The origin of life as a historical question cannot be studied directly. There are, however, many indirect ways to approach the question. Four of these are illustrated here. The bottom wedge represents approaches that work backwards in time from contemporary biology to more ancient forms of life, The top wedge represents approaches that work forward in time from organic species presumably available on early Earth to the first Darwinian chemical systems. The left wedge represents approaches that hope to discover an alien or weird form of life by exploration, a form of life whose structure might constrain models for how terran life emerged. The right wedge represents efforts in the laboratory to create artificial Darwinian systems, systems that might further constrain models of how terran life emerged, even if they are not constrained by current models for the environment on early Earth. Benner et al. (2007) *Adv Enzymol Mol Biol Protein Evol* **75:** 1–132.

our best models for its minerals, atmosphere, and ocean at that time. Given this inventory, experiments can be imagined whereby we create a chemical system capable of Darwinian evolution (perhaps RNA) from this inventory (Benner et al. 2006; Rich 1962).

A third approach, exploration, has driven discovery and paradigm change throughout human existence. We might go to other planets to see what we can find. If, for example, Mars or Titan hold life forms having histories independent of life on Earth, observation there might jolt our concept of what life is, how it might emerge, and whether RNA is universally a key component.

Alternatively, an alternative form of life might exist here in Earth in a "shadow biosphere," evading our detection so far because we have not looked in the right place or in the right way (Benner 1999; Cleland and Copley 2005; Davies et al. 2009). A shadow biosphere might even hold RNA-only organisms (Benner 1999), which would not necessarily be detectable using probes that look for ribosomal RNA. Unfortunately, exploration on other planets remains expensive, slow, and dangerous, as terran exploration has always been. Nor can exploration easily find what we do not know how to seek or how to recognize should we encounter it.

Therefore, a fourth approach might return to synthetic biology (Benner and Sismour 2005). Failing to find a second example of life, we might create one in a laboratory on Earth. Again, we might start with nucleic acids as catalysts, focusing on a cartoon of how RNA life might have emerged on Earth. In this cartoon, a nucleic acid emerged on early Earth able to catalyze the template-directed synthesis of its copy (Rich 1962). We might attempt in the laboratory to create such a nucleic acid.

Unfortunately, confident as we are of RNA catalysis and the *possibility* that RNA could catalyze its own template-directed replication (which, if accompanied by strand displacement, might meet a definition for life), no one has actually *showed* that RNA can do so. The closest appears to be work by Bartel, Zaher, Unrau and others (Zaher and Unrau 2007), much discussed elsewhere in this collection, which has yielded an RNA molecule able to add about two dozen monomers to a primer before it falls apart in

the solutions containing the high concentrations of Mg^{2+} for its activity.

This sets the stage for two questions whose answers would forward the "RNA first" model for the origin of life: (a) How is functional behavior distributed within RNA sequence space? and (b) How might RNA molecules within that space emerge spontaneously?

The second question requires that we address the challenge of creating oligomeric nucleic acids prebiotically, solving various associated problems (chirality, destruction by water, entropy of polymer assembly from dilute building blocks; Shapiro 2007). The first question tells us how grim the challenge is. If only one in 10^{30} RNA molecules 100 nucleotides in length can spark Darwinian evolution, the challenge is bigger than if 10% of all RNA molecules 20 nucleotides in length can do so.

Both of these questions are experimentally accessible today, and are being pursued (Rajamani et al. 2008), perhaps without the full intensity that one might desire. The two ingredients for success in science (funding and enthusiasm) are presently missing in much of the scientific community to pursue such questions with intensity. Even the Howard Hughes Medical Institute and the Templeton Foundation, long sources of support for fundamental questions of these types, have turned their attention elsewhere in recent years.

These four approaches set the stage for a collection that places RNA at the center of our model for life as a universal. As the Table of Contents shows, Earth today is still very much an "RNA world." More than half of the pages of this collection describe what RNA does today in the terran biosphere, especially with the proteins that it encodes and synthesizes, the DNA that instructs it and the metabolism that supports it. These can be approached by scientific methods that we learned in school. Accordingly, this article will emphasize the historical, alien, and synthetic parts of the RNA world, approachable only by more exotic methods.

2 MOVING BACKWARD IN TIME

The RNA World model may be examined using just about every strategy available to draw inferences about the historical past through observations made on modern biosphere. Perhaps the most remarkable consequence of those observations is that all known life on Earth is built from the same encoded biopolymers (RNA, DNA, and proteins) built from the same building blocks. This historical reality allows us to align the sequences of these biopolymers, first to ask whether they are "homologous" (related by common ancestry) and, if they are, to construct evolutionary trees that show their family relationships.

These alignments led to the discovery that ribosomal RNA molecules from all known life on Earth were homologous, and that their alignment could guide the construction of a "universal" tree of life (Woese 1998). Conversely, the ribosomal RNA tree can be used to defines species, even in microorganisms in which classical definitions of species fail (Woese 2004). By extension, all ribosomes on Earth descended from a common ancestor. The antiquity of the ribosome is strong evidence for the use of RNA in early terran life (see Lambowitz and Zimmerly 2010; Moore and Steitz 2010).

Evidence for an earlier biosphere on Earth that relied on RNA before the emergence of encoded proteins comes from the structure of the ribosome. An RNA component of the ribosome appears to be in direct contact with reacting atoms as the peptide bond is formed (Moore and Steitz 2003). Given the homology among all ribosomal RNA, one can infer from the ribosomal structures from even a few organisms that the synthesis of *all* peptide bonds on Earth is catalyzed by RNA (Harms et al. 2001), and that their common ancestral ribosome also used RNA to catalyze peptide bond synthesis. This completes the use of structural biology to define the concept of the RNA world, a use that began with the structure of transfer RNA (Kim et al. 1974), which prompted Francis Crick to comment that tRNA looked like a molecule trying to be a catalyst (Crick 1968).

Paleogenetic resurrections have enriched this concept. For example, the structures of many proteins that interact with the ribosome and ribosomal RNA are also homologous in many forms of life. Elongation factors, which present charged tRNA to the ribosome, are an example. Evolutionary trees based on their alignment are models for the familial history of elongation factors. Taking the next step, once the those trees are available, it is possible to infer the sequence of ancestral elongation factors and, through the magic of recombinant DNA biotechnology, bring them back for study in the laboratory.

Gaucher et al. did exactly this for bacterial elongation factors dating back perhaps three billion years (Gaucher et al. 2003; Gaucher et al. 2008). Inferring the sequences of ancestral elongation factors, these authors inferred the sequences of various candidate ancestral elongation factors, synthesized genes encoding these ancient proteins, and resurrected the now-extinct elongation factors in the laboratory. These resurrected proteins are, of course, physical manifestations of a hypothesis that these sequences actually existed three billion years ago and worked with a functioning ribosome that also existed at that time.

Paleogenetics experiments allowed these historical hypotheses to be tested in the laboratory, first by seeing

Cite this article as *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a003541

if the hypothetical ancestral elongation factors actually work. The result was positive: The putative ancestral elongation factors work.

This, in turn, allowed a discovery: The temperature optima for the ancestral elongation factors deep in the eubacterial tree was rather high, about 65°C (Gaucher et al. 2003). The temperature optimum of *modern* elongation factors, measured in vitro, is the same as the ambient temperature at which their hosts live. Exploiting a favorite axiom from natural history ("The present is the key to the past"), these observations with resurrected elongation factors implies that ancestral bacteria deep in the bacterial tree living billions of years ago lived at 65°C.

Further resurrections of ancestral elongation factors throughout the eubacterial tree inferred a temperature history of life, all assuming a functioning, homologous ribosome (Gaucher et al. 2008). Perhaps not so far in the future, entire ancestral ribosomes will be resurrected, taking paleogenetics back in time about as far as the contemporary record might allow. If we are more fortunate, we might be able to infer the sequence of membrane proteins that diverged before the last common ancestor (Linkkila and Gogarten 1991) and use paleogenetics strategies to go still farther back in time.

Experiments and observations with the ribosome were not the only ones to drive biological chemists to conclude that an RNA world was a historical reality. Already in 1976, White noted that RNA fragments attached to various cofactors (Fig. 2, magenta moieties) were widely distributed in modern terran life, and were most likely present in the last common ancestor of life on Earth (White 1976). White suggested that these RNA fragments were remnants of an RNA world. Just two years later, Visser and Kellogg (1978) used these observations to account for the reactivity and distribution of biotin in modern metabolism.

By the end of the 1980s, sufficient sequence data were available to take the next step. Here, comparative genomics was used to infer the entire genetic complement of the organism that is represented in the universal tree as the organism connecting the archeal, bacterial, and eucaryal kingdoms (Benner et al. 1989; Benner et al. 1993; Anantharaman et al. 2002; Koonin 2003). Correlations between these with biosignatures in the geological record, including the emergence of atmospheric dioxygen, were used to infer a metabolism encoded by the "protogenome" found in that organism (Benner et al. 1989). From there, it was argued that the noninvolvement of the RNA portions of the RNA cofactors in the core of their functional chemical reactivity implied that they arose at a time when RNA was the only biopolymer available. This, in turn, was taken to imply that the RNA world was metabolically complex, containing many RNA enzymes able to catalyze, for example, the transfer phosphate groups, reduction and oxidation reactions, and the formation of carbon–carbon bonds.

3 MOVING FORWARD IN TIME ON A ROCKY EARTH

Such narratives nicely illustrate why many are confident that an RNA World existed as a historical reality, and why

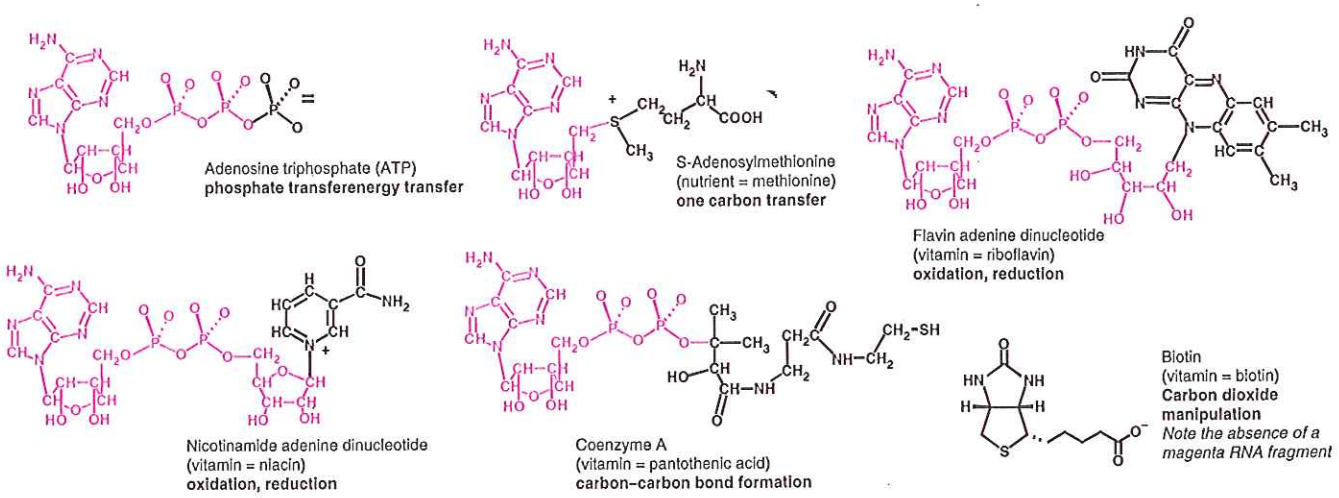


Figure 2. Shown in magenta are RNA fragments attached to many cofactors that are widely distributed in modern terran metabolism, and therefore placed in the last common ancestor of all known life on Earth. Because those RNA fragments do not participate in the chemistry of the metabolic reaction, they are not likely to have arisen convergently, but rather reflect an episode of life on Earth when RNA was the only encoded component of biocatalysis, and used these fragments as "handles." Under this hypothesis, the *absence* of a magenta RNA cofactor on biotin implies that it arose *after* the RNA World.

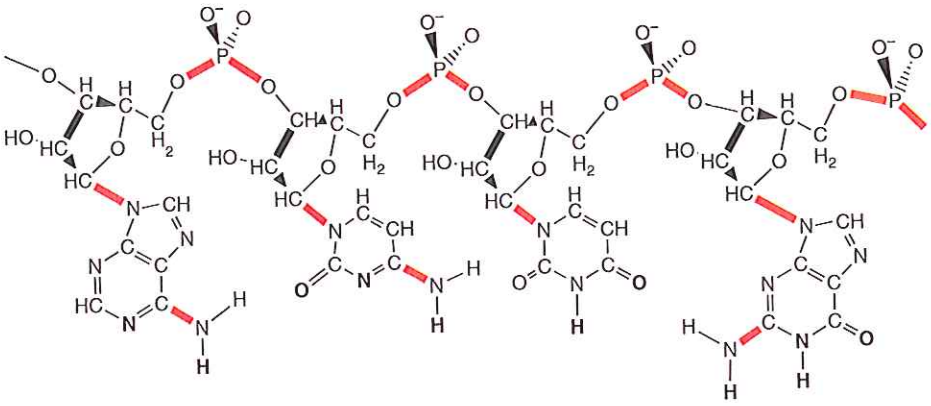


Figure 3. The red bonds in this general structure for RNA are all thermodynamically unstable with respect to hydrolysis in water, and suffer from a standard organic reaction mechanism by which they can hydrolyze.

some agree that metabolism was complex in the RNA biosphere. However, no matter how convincing these narratives, they do not force the conclusion that RNA was the *first* encoded biopolymer on Earth. To support this conclusion, one must find a model that accounts for the emergence of polymeric RNA without the intervention of prior biology.

The community is today twice divided. The first divide separates some who attempt to find such a model from others who consider that such a model cannot be found (Shapiro 2007). At the center of those who doubt that RNA emerged prebiotically are many experiments defining the intrinsic instability of RNA in water. For example, the red bonds in Figure 3 are all thermodynamically unstable with respect to hydrolysis in water, a substance often believed essential for life (Baross and Benner 2007). Even if prebiotic processes managed to assemble (up a free energy gradient) several dozen ribonucleotides of the same chirality, the product RNA might have promptly fallen apart. Still worse, reactivity intrinsic in the 2′-OH group of RNA will lead to cleavage even in the absence of water. Indeed, the most successful ribozymal RNA polymerases to date (Lawrence and Bartel 2003; Zaher and Unrau 2007) do not indefinitely replicate RNA because they themselves fall apart in water in the presence of divalent magnesium cations, which are required at substantial concentration for their reactivity. It is not surprising that Joyce and Orgel called RNA a "prebiotic chemist's nightmare" (Joyce and Orgel 1999; Joyce and Orgel 2006).

These observations alone are sufficient for some thoughtful authors to abandon RNA as a candidate for the first genetic molecule (Larralde et al. 1995; Shapiro 2007). Some have abandoned *all* genetic biopolymers, suggesting instead that something like Darwinian evolution must have been supported by a set of small organic molecules dissipating free energy in a cycle that, through its operation, can adapt to changing conditions and evolve in a Darwinian sense (Kauffman 1986; Smith and Morowitz 2004). Others, notably the late Leslie Orgel, disputed this view (Orgel 2008). This topic is being debated at a special session in the April 2010 Astrobiology Science Conference in Houston, a debate that has an unpredictable outcome.

Unfortunately, RNA encounters prebiotic synthesis problems long before it becomes an oligomer. As discussed in Robertson and Joyce 2010, approaches exist to generate various nucleobases from precursor molecules that the community has come to view as "plausibly prebiotic." Unfortunately, three of these nucleobases (not uracil) suffer hydrolytic deamination in water (Levy and Miller 1998).

Still worse, ribose is unstable. Just a decade ago, Stanley Miller and his group quantitated the instability of ribose by measuring the rate of decomposition of ribose under a variety of conditions. At pH 7 and 100°C, ribose decomposes with a half-life of only 73 minutes; its half-life is 44 year at 0°C (Larralde et al. 1995). These observations led Miller to "preclude the use of ribose and other sugars as prebiotic reagents except under very special conditions. It follows that ribose and other sugars were not components of the first genetic material, and that other possibilities, such as the peptide nucleic acids, and other non-sugar based backbones, should be examined."

Unfortunately, uncharged nonsugar backbones do not appear to be able to support Darwinian evolution (Richert et al. 1996; Benner and Hutter 2002). Other carbohydrates, such as threose, can scaffold a biopolymer with pairing properties (Horhota et al. 2005), but suffer from the same instability problems as ribose. The hunt is on for other biopolymers that might both support RNA-like replication and are stable under the conditions in which they are formed.

Cite this article as *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a003541

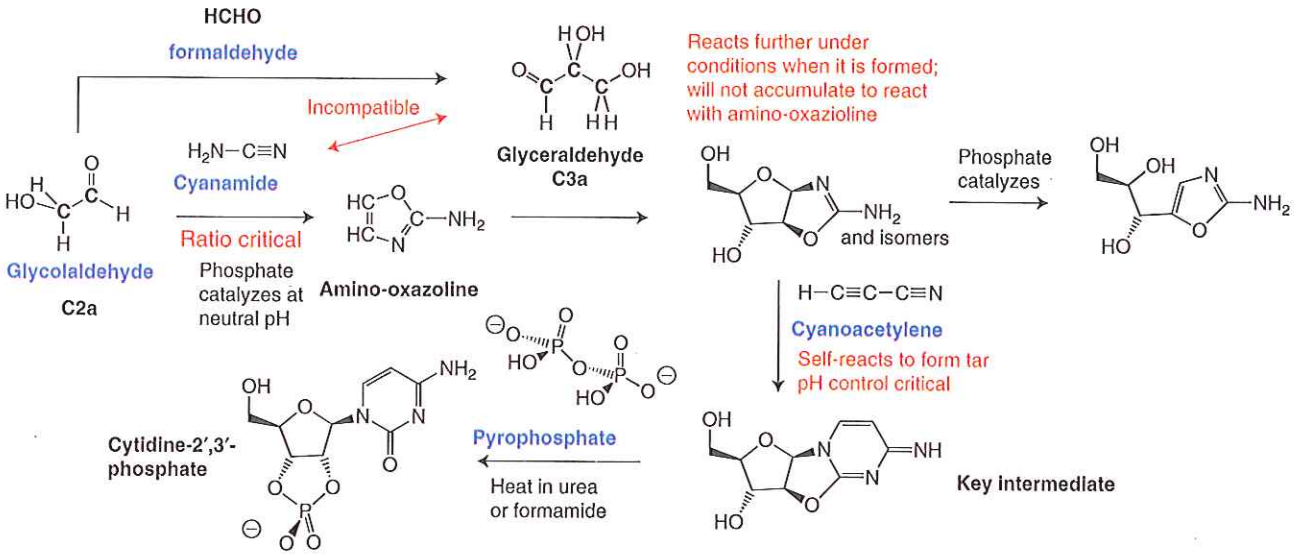


Figure 4. A scheme to obtain one of the red bonds in Figure 3 (Powner et al. 2009). For those accepting Shapiro's critique of an "RNA first" model for life's origin (Shapiro 2007), this scheme offers much to criticize. Compounds such as glyceraldehyde are not formed under conditions in which they accumulate. Glyceraldehyde and cyanamide are incompatible. Hands-on intervention by intelligent chemists must control their ratios and availabilities. To get the 2′,3′-cyclic phosphate of cytidine. pyrophosphate and the key intermediate must be heated in urea or formamide (certainly available on early Earth), but as a different reaction medium. Some regard this as an example of "synthetic organic chemistry," not "prebiotic chemistry," others disagree (Benner 2009).

Even the community that has *not* abandoned RNA as the first Darwinian molecule is divided, however. The division is illustrated by a line of recent work of Sutherland and his coworkers, illustrated in Figure 4 (Powner et al. 2009). Their approach focused on generating in water the bond that joins ribose to a cytosine heterocycle, one of the red bonds in Figure 3. They developed a laboratory route that forms this bond early, with later completion of the synthesis of both the heterocyclic and ribose rings. In this route, 2-aminooxazole and glyceraldehyde react in water, freshly prepared cyanoacetylene is added, and the product mixture is recovered and suspended in a urea solution with pyrophosphate, which is then placed on a filter, evaporated, and heated. Phosphate serves as a catalyst and buffer critical for several of these steps.

Even among those searching for a prebiotic origin of RNA, Sutherland's approach is not universally regarded as satisfactory. It requires unstable molecules such as glyceraldehyde as precursors; glyceraldehyde might be made from the condensation of glycolaldehyde and formaldehyde, which is indisputably prebiotic, but reacts further under conditions in which it is formed. The solution of cyanoacetylene must be freshly prepared, as cyanoacetylene polymerizes to form tar. Even if the precursors are stipulated, several of them are incompatible; for example, cyanamide destroys glyceraldehyde. Further, the ratio of glycolaldehyde and cyanamide is critical to the success of the first step.

Attempting to manage these issues, Sutherland suggested that 2-aminooxazole might have been formed in a location different from glyceraldehyde, sublimed from that location into a planetary atmosphere, to be later rained into a separate pond holding the glyceraldehyde in the correct ratio (Anastasi et al. 2007). Critics of this type of prebiotic chemistry find this an excessive *Deus ex machina*, but also note that if amino-oxazole is to be formed in a different locale than glyceraldehyde, the effort to find pH neutral conditions to form amino-oxazole (recognizing the instability of glyceraldehyde at high pH) was unnecessary. If this were not sufficient, the final step requires yet another change of environments, to urea or formamide as a reaction mixture, not water.

Shapiro criticized such syntheses, noting that they are analogous to a golfer "who having played a golf ball through an 18-hole course, then assumed that the ball could also play itself around the course in his absence" (Shapiro 2007). Sutherland himself recognized the potential power of such criticisms, writing: "One can imagine a number of scenarios that would result in heating, and progressive dehydration followed by cooling, rehydration, and irradiation" to get the desired cytidine. "Comparative assessment of these scenarios is beyond the scope of this work" (Powner et al. 2009). To which Shapiro might reply: "Indeed."

Thus, the remaining part of the community still considering an "RNA first" model for life's origins is looking

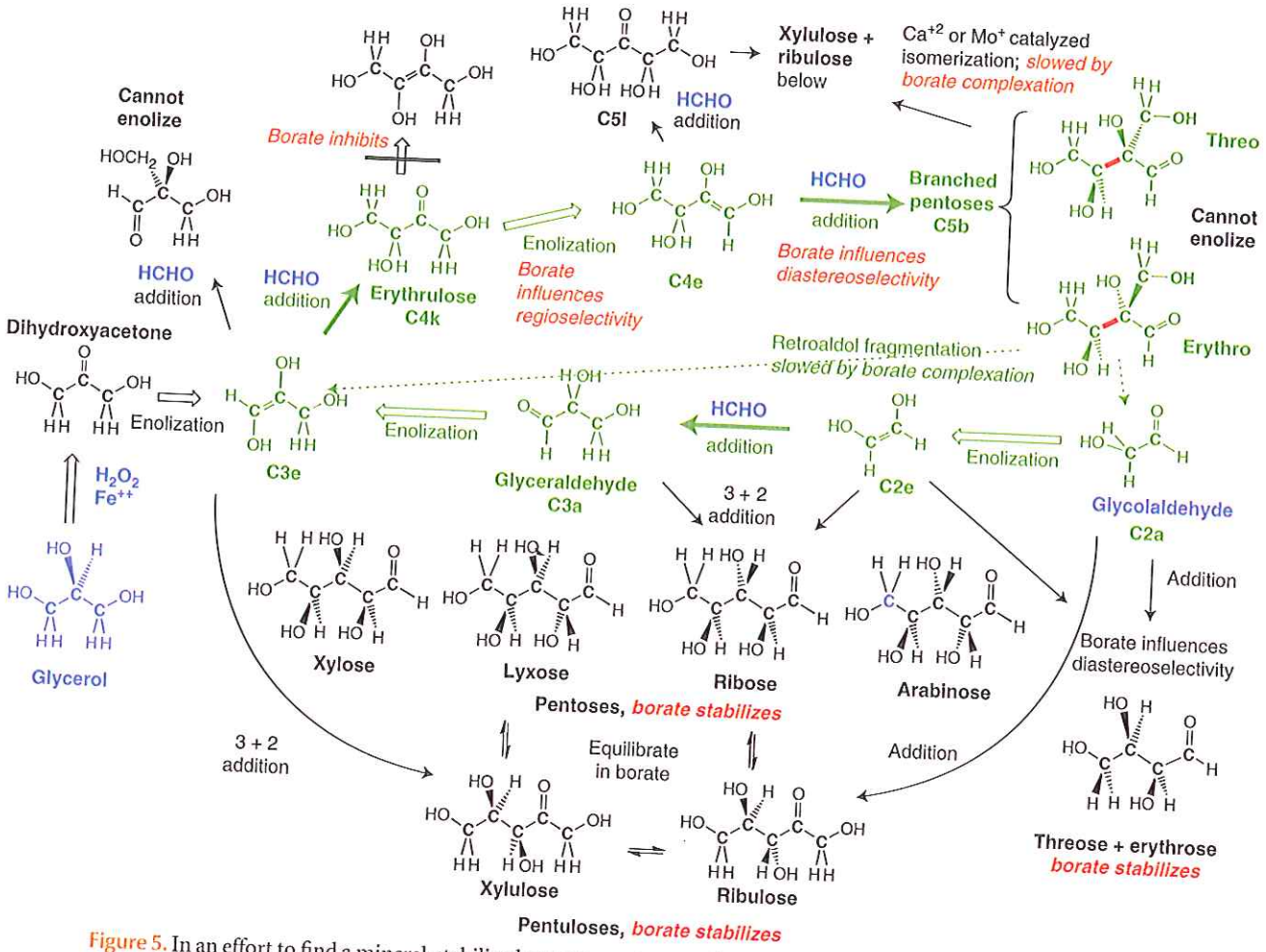


Figure 5. In an effort to find a mineral-stabilized route to pentoses and pentuloses, HJ Kim et al. (in prep.) propose a "premetabolic cycle" under the control of mineral borate. Compounds and reactions involved in the cycle are shown in green; the cycle fixes formaldehyde (HCHO) operating clockwise. Prebiotic compounds in blue feed the cycle. Leakage from the cycle indicated by black arrows emerging from green compounds; in each case but one, these lead directly or indirectly to pentoses and pentuloses, which are stabilized by borate. Whether this represents long sought experimental support for a "metabolism first" model for the origin of life, or the way in which pentoses and pentuloses emerged prebiotically, is disputable.

for ways to stabilize unstable intermediates, especially if such stabilization also guides productively their further reaction and avoids reactions that lead to tar. To these ends, some have added moieties, such as phosphate, to ribose (Xiang et al. 1994; Muller et al. 1990). Others propose organic scaffoldings (Persil and Hud 2007). Still others propose that minerals stabilize reactive intermediates, including minerals that contain borate (Ricardo et al. 2004).

Although these efforts have focused on making RNA under a "gene first" model for the origin of life, they curiously have come as close as any to providing a working example of a cycle reminiscent of models that put "metabolism first." For example, Figure 5 shows a cycle that starts with glycolaldehyde, abundant in the cosmos (Hollis et al. 2001), and fixes formaldehyde (formed by electrical discharge through an atmosphere of moist carbon dioxide) in the presence of borate minerals to give stereoisomeric branched pentoses. These branched pentoses are themselves unable to enolize, and form stable complexes with borate. They can, however, undergo calcium-catalyzed rearrangement to give pentuloses directly, or suffer retroaldolcleavage to give glycolaldehyde and glyceraldehyde, which can either assemble in the presence of borate to form pentoses, which are themselves stabilized by borate, or add more formaldehyde to form more branched pentose.

Experimental work with this cycle has uncovered several of the challenges that must be met for such cycles to

Cite this article as *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a003541

support prebiotic synthesis, let alone to serve as a "metabolism first" evolving system. Key to these is "leakage," the loss of material from the cycle to give undesired products. If carbon leaks from the cycle faster than it is fixed, the cycle disappears.

In the cycle in Figure 5, a principal source of leakage arises from the addition of formaldehyde to the enediol of erythrulose at the less hindered center. In this particular case, experiments show that the leakage is productive; the resulting 3-pentulose isomerizes to give pentuloses (ribulose, xylulose) that are stabilized by borate as well. Leakage through the reaction of two glycolaldehyde molecules gives threose, an alternative genetic carbohydrate, which is also stabilized. These are both fortunate outcomes, as all of these carbohydrates have potential as building blocks for genetic materials. Other cycles need not be so fortunate.

4 EXPLORATION

Since the third edition of *The RNA World* appeared (Gesteland et al. 2006), exploration of our solar system has discovered substantial amounts of perchlorate on Mars and oceans of methane on Titan, the largest moon of Saturn. The first is interesting as it offers one explanation for the failure of previous landers on Mars, including the 1976 Viking lander, to detect organic species (Benner et al. 2000). As a first step in its analysis of Martian soil, Viking heated a sample to 500°C. Organics heated with perchlorate at this temperature ignite, a property useful in fireworks. At Martian temperatures, however, perchlorate is entirely compatible with organic species. Thus, optimists still note that organics may exist in the surface of Mars, and RNA is not excluded from these.

Another dramatic climax in our exploration of the Solar System was the Cassini-Huygens landing on Titan. There, at 94 K (−179°C), oceans of methane were observed. Unfortunately, RNA is unlikely to dissolve in liquid methane (although the experiment seems not to have been performed). The alkalinity of the alternative fluid proposed for Titan, subsurface water-ammonia eutectics, would make an RNA world unlikely there as well.

The search for remnants of the RNA world on Earth (Benner 1999; Cleland and Copley 2005) has also not advanced much. Several authors have now given serious consideration to pursuing it further (Davies et al. 2009), and the first ideas suggesting how to explore for a "shadow biosphere" have been presented. For example, as 70% of the volume of a typical eubacterial cell is filled with the machinery to make proteins, a surviving riboorganism might be considerably smaller than the smallest modern bacteria. It may, therefore, be found in environments where size is a constraint.

5 SYNTHETIC BIOLOGY

Since the third edition of *The RNA World*, synthetic biological approaches, represented by the right wedge in Figure 1, have made progress (Benner 2009). As discussed in other articles in this collection, experimentalists are changing the structure of RNA to help manage parts of its unfortunate reactivities. Amino groups replacing the ribose hydroxyl groups are being used to facilitate template-directed polymerization (Stutz et al. 2007) (Fig. 6). Threose is being contemplated as an alternative to ribose (Schoning et al. 2000). Many in the community have recognized that the RNA World model might be supported if nucleic acids of *any* kind can support Darwinian evolution in the laboratory.

One limitation of the RNAWorld has been hypothesized to be the paucity of building blocks in the RNA biopolymer (Benner et al. 1999). Certainly, proteins, with 20 amino acids, have a richer diversity of functionality than RNA, and this functionality is useful for binding and catalysis. With these thoughts in mind, several groups have sought to increase the number of nucleotides in nucleic acids (Hirao 2006; Benner 2004; Henry and Romesberg 2003).

The challenge of synthesizing components of an artificially expanded genetic information systems and showing that they bind to each other with expanded Watson-Crick specificity has been met for several systems (Henry and Romesberg 2003; Benner 2004). Accordingly, the bar has moved. Today's challenge has been to develop enzymatic processes that allow artificial genetic systems to be copied, and their copies to be copied indefinitely in a PCR format, without unidirectional loss of the unnatural components of the molecule (Johnson et al. 2004). This process of repeated copying, with imperfections, with the imperfections themselves being copy-able, is the essence of Darwinism at the molecular level.

Today, three examples are now known in which six nucleotide letters can be incorporated generally into PCR without substantial loss of the components with essentially independent selection of sequence (Sismour et al. 2004; Sismour et al. 2005; Yang et al. 2007). In principle, any of these systems can support in vitro evolution.

A particularly successful example of these is the six letter PCR incorporating 6-amino-5-nitropyridin-2-one and its complement imidazo(1,2-c)pyrimidin-5(1H)-one, trivially designated Z and P (Fig. 6). Protein polymerases support "six letter PCR" with this system. Further, mechanisms have been found by which Z:P pairs are mutated to give C:G pairs and C:G pairs are mutated back to Z:P pairs. As with natural mutation, this process involves

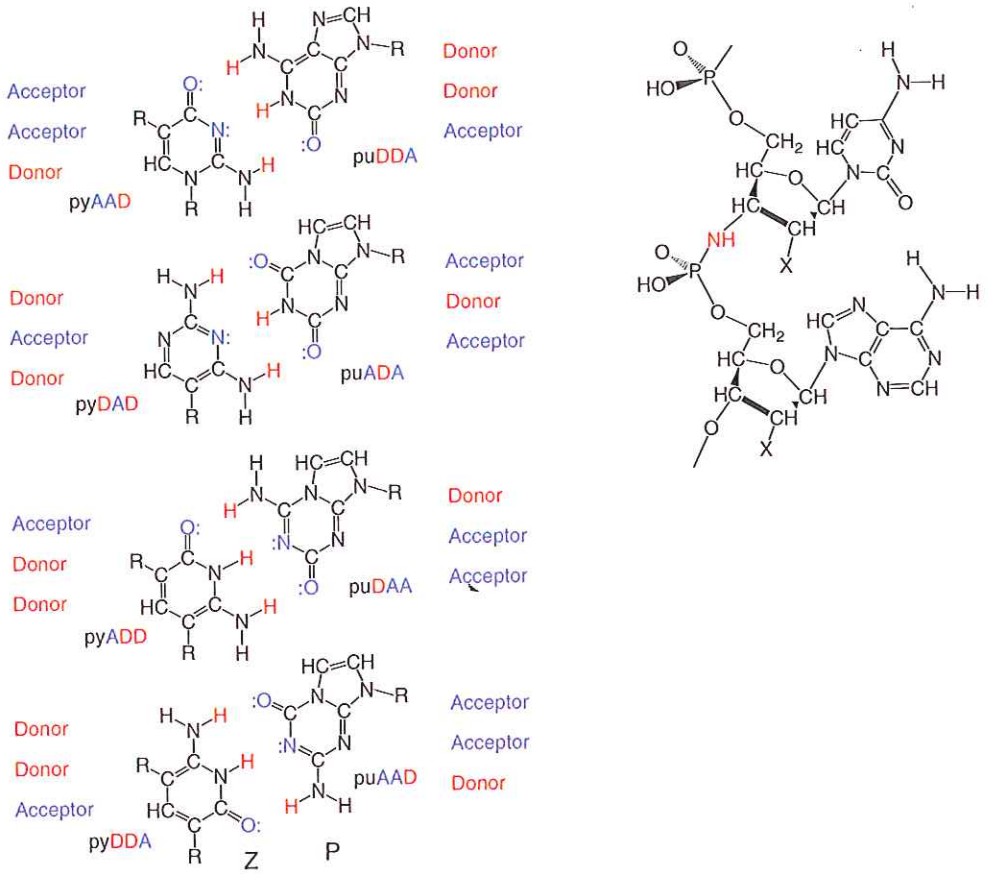


Figure 6. Some unnatural nucleic acid structures that have been developed by synthetic biologists as RNA-like (but not RNA) molecules possibly capable of supporting Darwinian evolution, including the expanded genetic alphabet (left) that supports six-letter PCR and the phosphoramidate linkage that supports enzyme-free primer extension (Stutz et al. 2007).

the protonation and deprotonation of the nucleobase heterocycles. Thus, this artificial genetic system is capable of supporting Darwinian evolution, and has been shown to do so in the laboratory (Fig. 7) (Yang and Benner 2009).

Although this system certainly can support the test of the hypothesis that nucleic acids having more letters can deliver more functional diversity than nucleic acids having fewer (Reader and Joyce 2002), it does not constitute a "nucleic acids only" system; it depends on polymerases that have been delivered by four billion years of natural biological evolution. Joseph Piccirilli suggested some time ago that perhaps the RNA World had more nucleotides of this type, and therefore had access to greater catalytic diversity than it presently seems to have access to with just adenine, guanine, cytosine, and uracil. In this model, the genetic biopolymer evolved to lose building blocks, becoming more streamlined, perhaps reflecting the fact that four nucleotides are optimal in a genetic system (Szathmary 1999).

6 THE STATE OF THE RNA WORLD: 2009

As subsequent articles in this collection show, we today live in an RNA World. Further, it seems very likely that catalytic RNA played a pervasive role in earlier life. These inferences make even more intriguing the failure of prebiotic experiments to generate RNA easily, as well as our failure to have tripped across RNA-only organisms in a shadow biosphere, and the difficulty in getting synthetic systems based on RNA or RNA-like to catalyze their own reproduction via monomer addition. We are missing something in our models for reality.

What might be wrong? For example, we may have misread the historical import of the modern record. The synthesis of oligonucleotides by monomer addition, for example, is difficult in the laboratory, even though it is today "universal" in terran biology. But perhaps the first RNA replication did not use monomer addition. Ligation, recombination (Hayden et al. 2005), and other processes are certainly available to obtain longer RNA molecules

Cite this article as *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a003541

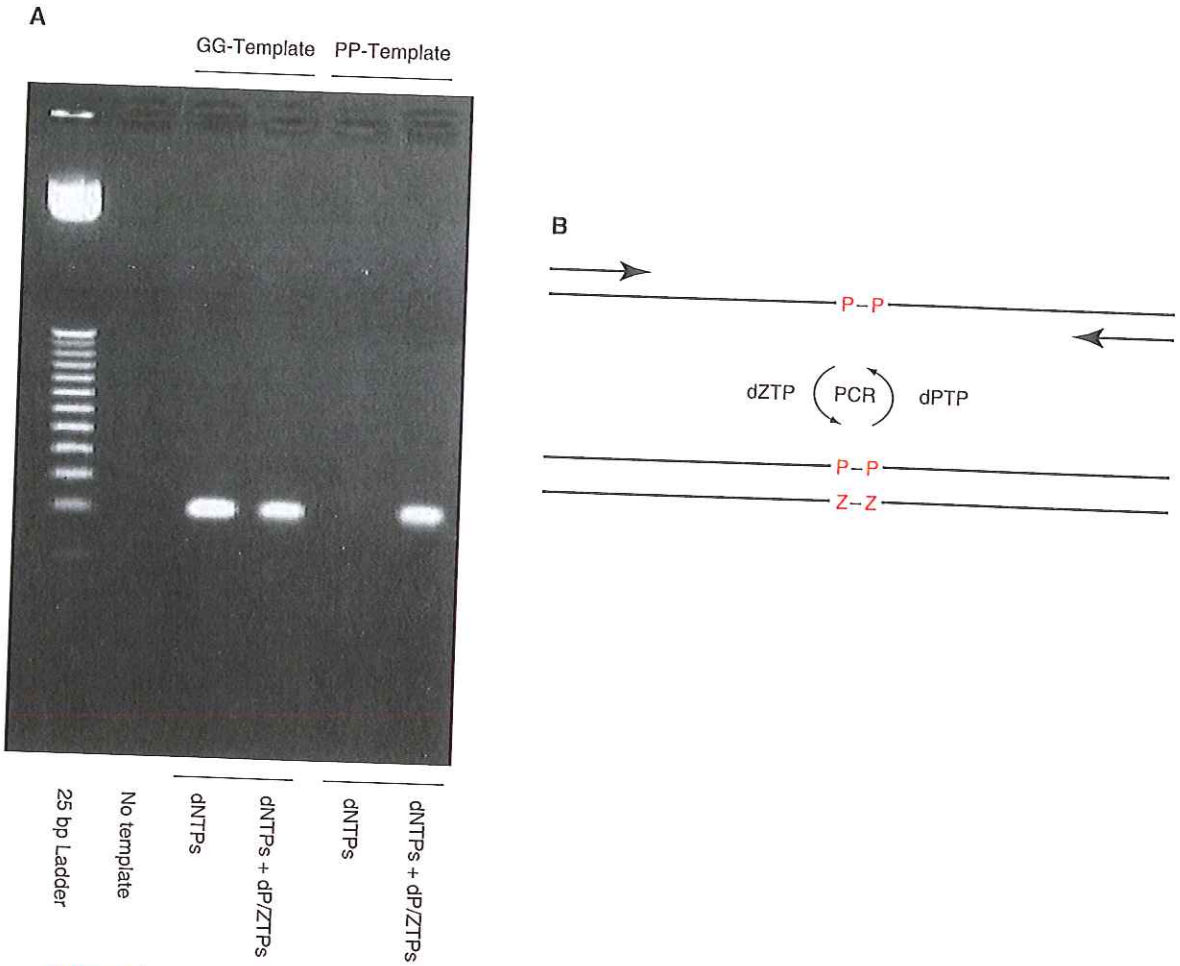


Figure 7. "Six letter PCR" showing products of amplification of an amplicon containing two adjacent nonstandard nucleotides.

om shorter ones, which seem at present to be much easier obtain prebiotically. These alternative processes may have :en lost in all lineages of terran life that we know about.

It is also, of course, possible that we have not modeled rrectly the early environment on Earth. Hot versus cold, ıter or not, the nature of available minerals, and places on e planet (surface or deep oceans) that could possibly have ıdled life's formation are all uncertain. For example, vents like formamide, which lack certain of the RNA-stroying properties of water, may have dominated a gely water-free environment where the first RNA mole-es were assembled.

For these reasons, synthetic biology might offer the most ıstructive paradigm for future effort. Synthetic biology ; a grand challenge, here the construction of an RNA em that catalyzes its own replication in the context of sen metabolism, isolation systems, and environ-ıtal conditions. Pursuit of that challenge drags scientists)ss uncharted terrain where they must ask and answer unscripted questions. If theory driving that pursuit is inadequate, the synthesis fails, and fails in a way that cannot be ignored. In contrast, if observations contradict an accepted theory, they are (as often as not) ignored (Benner 2009). For this reason, synthesis drives discovery and paradigm change in ways that analysis cannot. And the incremental advances in the historical, alien, origins parts of the RNA world since the third edition of this collection, suggest that discovery and paradigm change both will be needed before the RNA World becomes established as a statement about the historical past.

REFERENCES

Anantharaman V, Koonin EV, Aravind L. 2002. Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucl Acids Res* **30:** 1427–1464.

Anastasi C, Crowe MA, Sutherland JD. 2007. Two-step potentially prebiotic synthesis of r-D-cytidine-5'-phosphate from D-glyceraldehyde-3-phosphate. *J Am Chem Soc* **129:** 24–25.

Baross J, Benner SA, Cody GD, Copley SD, Pace NR, Scott JH, Shapiro R, Sogin ML, Stein JL, Summons R, et al. 2007. *The limits of organic life in planetary systems.* The National Academies Press, Washington DC.

Benner SA. 1999. How small can a microorganism be? *Size limits of very small microorganisms: proceedings of a workshop, steering group on astrobiology of the space studies board.* National Research Council, pp. 126–135.

Benner SA. 2004. Understanding nucleic acids using synthetic chemistry. *Acc Chem Res* **37:** 784–797.

Benner SA. 2007. The early days of paleogenetics. Connecting molecules to the planet. *Experimental Paleogenetics,* (ed. D.A., Liberles), pp. 3–19, Academic Press.

Benner SA. 2009. *Life, the universe and the scientific method,* Gainesville FL, FfAME Press.

Benner SA, Hutter D. 2002. Phosphates, DNA, and the search for nonterrean life: A second generation model for genetic molecules. *Bioorg Chem* **30:** 62–80.

Benner SA, Sismour AM. 2005. Synthetic biology. *Nature Rev Genetics* **6:** 533–543.

Benner SA, Ellington AD, Tauer A. 1989. Modern metabolism as a palimpsest of the RNA world. *Proc Nat Acad Sci* **86:** 7054–7058.

Benner SA, Ricardo A, Carrigan MA. 2004. Is there a common chemical model for life in the universe? *Curr Opinion Chem Biol* **8:** 672–689.

Benner SA, Sassi SO, Gaucher EA. 2007. Molecular paleosciences. Systems biology from the past. *Adv Enzymol Related Areas Mol Biol Protein Evol* **75:** 1–132.

Benner SA, Burgstaller P, Battersby TR, Jurczyk S. 1999. Did the RNA world exploit an expanded genetic alphabet? In *The RNA world,* 2nd ed. (ed. RF Gesteland et al.), pp. 163–181. Cold Spring Harbor Laboratory Press, Cold Springs Harbor, NY.

Benner SA, Caraco MD, Thomson JM, Gaucher EA. 2002. Planetary biology. Paleontological, geological, and molecular histories of life. *Science* **293:** 864–868.

Benner SA, Carrigan MA, Ricardo A, Frye F. 2006. Setting the stage: The history, chemistry and geobiology behind RNA. In The *RNA world* 3rd ed. (ed. R.F. Gesteland et al.), pp. 1–21. Cold Spring Harbor Laboratory Press, Cold Springs Harbor, NY.

Benner SA, Cohen MA, Gonnet GH, Berkowitz DB, Johnsson K. 1993. Reading the palimpsest. Contemporary biochemical data and the RNA world. In *The RNA world* (ed. R. Gesteland, J. Atkins), pp. 27–70. Cold Spring Harbor Laboratory Press, Cold Springs Harbor, NY.

Benner SA, Devine KG, Matveeva LN, Powell DH. 2000. The missing organic molecules on Mars. *Proc Nat Acad Sci* **97:** 2425–2430.

Chyba C, Sagan C. 1992. Endogenous production, exogenous delivery and impact-shock synthesis of organic-molecules: An inventory for the origins of life. *Nature* **355:** 125–132.

Cleland CE, Chyba CF. 2000. Defining 'life'. *Orig Life Evol Biosphere* **32:** 387–393.

Cleland CE, Copley SD. 2005. The possibility of alternative microbial life on Earth. *Int J Astrobiol* **4:** 165–173.

Crick FHC. 1968. The origin of the genetic code. *J Mol Biol* **38:** 367–379.

Davies PCW, Benner SA, Cleland CE, Lineweaver CH, McKay CP, Wolfe-Simon F. 2009. Signatures of a shadow biosphere. *Astrobiol* **9:** 241–249.

Gaucher EA, Govindarajan S, Ganesh OK. 2008. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* **451:** 704–U2.

Gaucher EA, Thomson JM, Burgan MF, Benner SA. 2003. Inferring the paleoenvironment during the origins of bacteria based on resurrected ancestral proteins. *Nature* **425:** 285–288.

Gesteland R.F., Cech T.R., Atkins J.F. editors. 2006. *The RNA World* 3rd Edition. Cold Spring Harbor Laboratory Press, Cold Springs Harbor, NY.

Gilbert W. 1986. The RNA World. *Nature* **319:** 818.

Harms J, Schluenzen F, Zarivach R, Bashan A, Gat S, Agmon I, Bartels H, Franceschi F, Yonath A. 2001. High resolution structure of the large ribosomal subunit from a Mesophilic Eubacterium. *Cell* **107:** 679–688.

Hayden EJ, Riley CA, Burton AS, et al. 2005. RNA-directed construction of structurally complex and active ligase ribozymes through recombination. *RNA* **11:** 1678–1687.

Henry AA, Romesberg FE. 2003. Beyond A, C, G and T: Augmenting Nature's alphabet. *Curr Opin Chem Biol* **7:** 727–733.

Hirao I. 2006. Unnatural base pair systems for DNA/RNA-based biotechnology. *Curr Opin Chem Biol* **10:** 622–627.

Hollis JM, Vogel SN, Snyder LE, Jewell PR, Lovas FJ. 2001. The spatial scale of glycolaldehyde in the galactic center. *Astrophys J* **554:** L81–L85 Part 2.

Horhota A, Zou K, Ichida JK, Yu B, McLaughlin LW, Szostak JW, Chaput JC. 2005. Kinetic analysis of an efficient DNA-dependent TNA polymerase. *J Am Chem Soc* **127:** 7427.

Jermann TM, Opitz JG, Stackhouse J, Benner SA. 1995. Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. *Nature* **374:** 57–59.

Johnson SC, Sherrill CB, Marshall DJ, Moser MJ, Prudent JR. 2004. A third base pair for the polymerase chain reaction. Inserting isoC and isoG. *Nucl Acids Res* **32:** 1937–1941.

Joyce GF. 1994. Foreward. In *Origins of life: The central concepts* (ed. D.W. Deamer et al.), pp. xi–xii. Jones & Bartlett, Boston.

Joyce GF, Orgel LE. 2006. in *The RNA world* (ed. R.F. Gesteland et al.), pp. 23–56, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Joyce GF, Orgel LE. 1999. Prospects for understanding the origin of the RNA world. in *The RNA World,* 2nd edition (ed. R.F. Gesteland et al.), pp. 49–77, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Kauffman SA. 1986. Autocatalytic sets of proteins. *J Theor Biol* **119:** 1–24.

Kim SH, Suddath FL, Quigley GJ, Mcpherson A, Sussman JL, Wang AHJ, Seeman NC, Rich A. 1974. 3-Dimensional tertiary structure of yeast phenylalanine transfer-RNA. *Science* **185:** 435–440.

Koonin EV. 2003. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nature Reviews Microbiol* **1:** 127–136.

Koshland DE. 2002. The seven pillars of life. *Science* **295:** 2215.

Lambowitz AM, Zimmerly S. 2010. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol* **2:** a003616.

Larralde R, Robertson MP, Miller SL. 1995. Rates of decomposition of ribose and other sugars. Implications for chemical evolution. *Proc Natl Acad Sci* **92:** 8158–8160.

Lartigue C, Vashee S, Algire MA, Chuang RY, Benders GA, Ma L, Noskov VN, Denisova EA, Gibson DG, Assad-Garcia N, et al. 2009. Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science,* published online.

Lawrence MS, Bartel DP. 2003. Processivity of ribozyme-catalyzed RNA polymerization. *Biochemistry* **42:** 8748–8755.

Levy M, Miller S. 1998. The stability of the RNA bases: Implications for the origin of life. *Proc Natl Acad Sci* **95:** 7933–7938.

Liberles D.A. editor 2007. *Experimental PALEOGENETICS,* Academic Press.

Linkkila TP, Gogarten JP. 1991. Tracing origins with molecular sequences – rooting the universal tree of life. *Trends in Biochemical Sciences* **16:** 287–288.

Moore PB, Steitz TA. 2003. The structural basis of large ribosomal subunit function. *Annual Review Biochem* **72:** 813–850.

Moore PB, Steitz TA. 2010. The roles of RNA in the synthesis of protein. *Cold Spring Harb Perspect Biol* **2:** a003780.

Muller D, Pitsch S, Kittaka A, Wagner E, Wintner CE, Eschenmoser A. 1990. Chemistry of α-aminonitriles—aldomerisation of glycolaldehyde phosphate to rac-hexose 2,4,6-triphosphates and (in presence

Cite this article as *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a003541

of formaldehyde) rac-pentose 2,4-diphosphates—rac-allose 2,4,6-triphosphate and rac-ribose 2,4-diphosphate are the main reaction-products. *Helv Chim Acta* **73:** 1410–1468.

Orgel LE. 2008. On the implausibility of metabolic cycles on prebiotic Earth. *PLOS Biol* **6:** e18.

Persil O, Hud NV. 2007. Harnessing DNA intercalation. *Trends in Biotech* **25:** 433–436.

Pizzarello S. 2004. Chemical evolution and meteorites: An update. *Origins Life Evol Biosphere* **34:** 25–34.

Powner MW, Gerland B, Sutherland JD. 2009. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459:** 239–242.

Rajamani S, Vlassov A, Benner S, Coombs A, Olasagasti F, Deamer D. 2008. Lipid-assisted synthesis of RNA-like polymers from mononucleotides. *Orig Life Evol Biosphere* **38:** 57–74.

Reader JS, Joyce GF. 2002. A ribozyme composed of only two different nucleotides. *Nature* **470:** 841–844.

Ricardo A, Carrigan MA, Olcott AN, Benner SA. 2004. Borate minerals stabilize ribose. *Science* **303:** 196.

Rich A. 1962. On the problems of evolution and biochemical information transfer. in *Horizons In Biochemistry* (eds. M. Kasha, B. Pullmann), pp. 103–126. Academic Press, New York.

Richert C, Roughton AL, Benner SA. 1996. Nonionic analogs of RNA with dimethylene sulfone bridges. *J Am Chem Soc* **118:** 4518–4531.

Robertson MP, Joyce GF. 2010. The origins of the RNA world. *Cold Spring Harb Perspect Biol* **2:** a003608.

Schoning KU, Scholz P, Guntha S, Wu X, Krishnamurthy R, Eschenmoser A. 2000. Chemical etiology of nucleic acid structure: The α-threofuranosyl-(3′→2′) oligonucleotide system. *Science* **290:** 1347–1351.

Shapiro R. 1988. Prebiotic ribose synthesis: A critical analysis, *Origins of Life* **18:** 71–85.

Shapiro R. 2000. A replicator was not involved in the origin of life. *IUBMB Life* **49:** 173–176.

Shapiro R. 2007. A simpler origin for life. *Scientific American* **296:** 46–53.

Sismour AM, Benner SA. 2005. The use of thymidine analogs to improve the replication of an extra DNA base pair: A synthetic biological system. *Nucl Acids Res* **33:** 5640–5646.

Sismour AM, Lutz S, Park J-H, Lutz MJ, Boyer PL, Hughes SH, Benner SA. 2004. PCR amplification of DNA containing non-standard base pairs by variants of reverse transcriptase from human immunodeficiency virus-1. *Nucl Acids Res* **32:** 728–735.

Smith E, Morowitz HJ. 2004. Universality in intermediary metabolism. *Proc Natl Acad Sci* **101:** 13168–13173.

Stutz JAR, Kervio E, Deck C, Richert C. 2007. Chemical primer extension: Individual steps of spontaneous replication. *Chemisty & Biodiversity* **4:** 784–802.

Szathmary E. 1999. The origin of the genetic code: Amino acids as cofactors in an RNA world. *Trends Genet* **15:** 223–229.

Visser CM, Kellog RM. 1978. Biotin. Its place in evolution. *J Mol Evol* **11:** 171–178.

White HBIII. 1976. Coenzymes as fossils of an earlier metabolic state. *J Mol Evol* **7:** 101.

Woese C. 1967. The evolution of the genetic code. In *The genetic code*, pp. 179–195. Harper & Row, New York.

Woese CR. 1998. The universal ancestor. *Proc Natl Acad Sci* **95:** 6854–6859.

Woese CR. 2004. A new biology for a new century. *Microb Mol Biol Rev* **68:** 173–186.

Xiang Y-B, Drenkard S, Baumann K, Hickey D, Eschenmoser A. 1994. Chemistry of α-amino nitriles. 12 Exploratory experiments on thermal-reactions of α-amino nitriles. *Helv Chim Acta*, **77:** 2209.

Yang Z, Benner SA. 2009. Darwinian systems based on an expanded genetic alphabet. *Nature Chem Biol* submitted.

Yang Z, Sismour AM, Sheng P, Puskar NL, Benner SA. 2007. Enzymatic incorporation of a third nucleobase pair. *Nucl Acids Res* **35:** 4238–4249.

Yarus M. 2010. Getting past the RNA world: the initial Darwinian ancestor. *Cold Spring Harb Perspect Biol* **2:** a003590.

Zaher HS, Unrau PJ. 2007. Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* **13:** 1017–1026.