The antiquity of RNA-based evolution

Gerald F. Joyce

Departments of Chemistry and Molecular Biology and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA (e-mail: gjoyce@scripps.edu)

All life that is known to exist on Earth today and all life for which there is evidence in the geological record seems to be of the same form — one based on DNA genomes and protein enzymes. Yet there are strong reasons to conclude that DNA- and protein-based life was preceded by a simpler life form based primarily on RNA. This earlier era is referred to as the 'RNA world', during which the genetic information resided in the sequence of RNA molecules and the phenotype derived from the catalytic properties of RNA.

The RNA molecule has a pervasive role in contemporary biology, especially with regard to the most fundamental and highly conserved cellular processes. It is involved as a primer in DNA replication, a messenger that carries genetic information to the translation machinery, and a catalyst that lies at the heart of the ribosome. RNA instructs the processing of precursor messenger RNAs during splicing and editing, and mediates numerous other transactions of RNA and proteins in the cell. Catalytic RNAs (ribozymes) assist in RNA processing events and the replication of viral genomes. Individual nucleotides serve as important signalling molecules and their coenzyme derivatives participate in most of the reactions of central metabolism. It is as if a primitive civilization had existed prior to the start of recorded history, leaving its mark in the foundation of a modern civilization that followed. Although there may never be direct physical evidence of an RNA-based organism, because the RNA world is likely to have been extinct for almost four billion years, molecular archaeologists have uncovered artefacts of this ancestral era, none more pronounced than the recently reported crystal structure of the ribosome[1–3]. This structure reveals the face of the RNA world in the active role that RNA has in protein synthesis.

In the laboratory, biochemists have come to appreciate the remarkable structural and functional versatility of RNA. Despite containing only four different chemical subunits, RNA folds into a variety of complex tertiary structures, analogous to structured proteins, and catalyses a broad range of chemical transformations (see review in this issue by Doudna and Cech, pages 222–228). RNA evolution in the laboratory, which can be viewed as a model of RNA evolution in the RNA world, has been used to obtain many new RNA enzymes. These include RNAs that catalyse nucleotide synthesis[4] RNA polymerization[5], aminoacylation of transfer RNA[6] and peptide bond formation[7]. It seems likely that RNA has the capability to support life based on RNA genomes that are copied and maintained through the catalytic function of RNA. There are substantial gaps, however, in scientific understanding concerning how the RNA world arose, the degree of metabolic complexity that it attained, and the way that it led to DNA genomes and protein enzymes.

The dawn of darwinian evolution

The worlds of prebiotic chemistry and primitive biology lie on opposite sides of the defining moment for life, when darwinian evolution first began to operate (Fig. 1). Before that time, chemical processes may have led to a substantial level of complexity. Depending on the nature of the prebiotic environment, available building blocks may have included amino acids, hydroxy acids, sugars, purines, pyrimidines and fatty acids. These could have combined to form polymers of largely random sequence and mixed stereochemistry (handedness). Some of the polymers may have had special properties, such as adherence to a particular mineral surface, unusual resistance to degradation, or the propensity to form supramolecular aggregates. Eventually every polymer, no matter how stable, would have succumbed to degradation.

A special class of polymers are those that are capable of self-replication. Although polymer self-replication is often interpreted as involving residue-by-residue copying of the polymer — a view biased by familiarity with nucleic acid replication in biology — all that is actually required is that the polymer gives rise to additional polymer molecules of the same sequence. If the rate of production of new copies exceeds the rate of degradation of existing copies, then a particular polymer sequence will persist over time.

Natural environments are subject to fluctuating conditions, ranging from diurnal and seasonal variation to

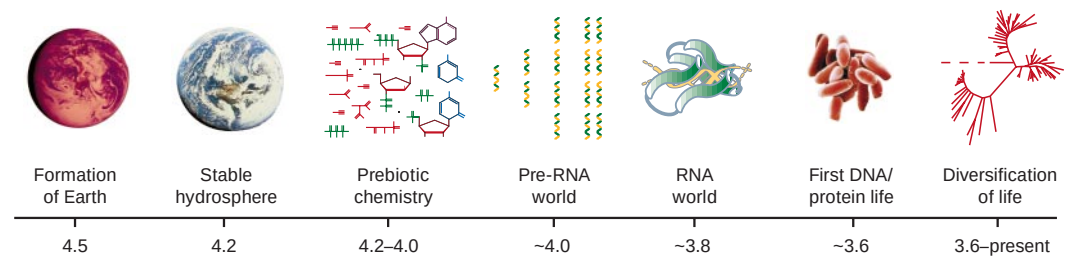

Figure 1 Timeline of events pertaining to the early history of life on Earth, with approximate dates in billions of years before the present.

unpredictable and potentially cataclysmic events. When the environment is altered, the special properties associated with a particular polymer may no longer apply and the capacity for self-replication may be lost. Persistence in a changing environment requires a more general mechanism for self-replication that allows the polymer sequence to change somewhat over time, but retain its heritage in most of the sequence that is unchanged. The polymer must be replicated in essentially the same manner regardless of its sequence. Variation will arise owing to inevitable copying errors, and those variants too must be amenable to replication.

Once a general mechanism existed for self-replication, allowing the introduction of variation and the ability to replicate those variants, darwinian evolution began to operate. This marked the beginning of life. The special properties of a particular polymer sequence then were defined by its net rate of accumulation (rate of production minus rate of degradation), and sequences that were associated with the most favourable survival rates would have come to dominate their locale. From that point onward, the natural history of life on Earth played out as a succession of dominant polymer sequences and their associated functional properties.

A cluttered path to RNA

RNA is a polymer of variable sequence that is amenable to self-replication by a templating mechanism[8]. Different sequences have different chemical properties, but almost all sequences are able to form Watson–Crick duplex structures that facilitate the production of new copies. If the building blocks of RNA were available in the prebiotic environment, if these combined to form polynucleotides, and if some of the polynucleotides began to self-replicate, then the RNA world may have emerged as the first form of life on Earth[9,10]. But based on current knowledge of prebiotic chemistry, this is unlikely to have been the case. Ribose, phosphate, purines and pyrimidines all may have been available, although the case for pyrimidines is less compelling[11,12]. These may have combined to form nucleotides in very low yield[13,14], complicated by the presence of a much larger amount of various nucleotide analogues. The nucleotides (and their analogues) may even have joined to form polymers, with a combinatorial mixture of 2′,5′-, 3′,5′- and 5′,5′-phosphodiester linkages, a variable number of phosphates between the sugars, D- and L- stereoisomers of the sugars, α- and β-anomers at the glycosidic bond, and assorted modifications of the sugars, phosphates and bases (Fig. 2). It is difficult to visualize a mechanism for self-replication that either would be impartial to these compositional differences or would treat them as sequence information in a broader sense and maintain them as heritable features.

The chief obstacle to understanding the origin of RNA-based life is identifying a plausible mechanism for overcoming the clutter wrought by prebiotic chemistry. Several avenues of investigation are being pursued. Perhaps there were special conditions that led to the preferential synthesis of activated β-D-nucleotides or the preferential incorporation of these monomers into polymers. For example, the prebiotic synthesis of sugars from formaldehyde can be biased by starting from glycoaldehyde phosphate, leading to ribose 2,4-diphosphate as the predominant pentose sugar[15]. This reaction can occur starting from dilute aqueous solutions of reactants at near-neutral pH when carried out in the presence of certain metal-hydroxide minerals[16]. The polymerization of adenylate, activated as the 5′-phosphorimidazolide, yields 2′,5′-linked products in solution, but mostly 3′,5′-linked products in the presence of a montmorillonite clay[17]. Thus, through a series of biased syntheses, fractionations and other enrichment processes, there may have been a special route to a warm little pond of RNA.

Another approach is to hypothesize that life did not begin with RNA; some other genetic system preceded RNA, just as it preceded DNA and protein (Fig. 1). This approach has met with substantial progress in recent years, despite the lack of guidance from known metabolic pathways in biology regarding the chemical nature of a precursor to RNA. A systematic investigation of potentially natural nucleic acid analogues containing various sugars and linkage isomers has led to the recognition of some intriguing pairing systems[18]. Most notable is the threose nucleic acid (TNA) analogue based on α-L-threofuranosyl units joined by 3′,2′-phosphodiester linkages[19] (Fig. 3a). This analogue forms stable Watson–Crick pairs with itself and with RNA. From the point of view of overcoming the clutter of prebiotic chemistry, TNA is more advantageous than RNA because of its relative chemical simplicity. Threose is one of only two aldotetroses (four-carbon sugars) and can only be joined at the 2′ and 3′ positions. Additionally, it is not difficult to imagine how a 'TNA world' might have made the transition to an RNA world while preserving the continuity of genetic information (see below).

There are other interesting candidates for a potential predecessor to RNA. Peptide nucleic acid (PNA) consists of a peptide-like backbone of *N*-(2-aminoethyl)glycine units with the bases attached through a methylenecarbonyl group[20] (Fig. 3b). Aminoethylglycine has been synthesized in spark discharge reactions from nitrogen, ammonia, methane and water[21], although the prebiotic synthesis of an entire PNA monomer has not been achieved. PNA forms Watson–Crick-like duplex structures with itself and with RNA. Even though it is non-chiral, PNA is susceptible to cross-inhibition of the opposing enantiomers when directing the polymerization of activated D,L-ribonucleotides[22,23]. Furthermore, PNA monomers can undergo an intramolecular *N*-acyl transfer reaction that would prevent any conventional mechanism for their polymerization[24]. Two other proposals for what might have come before RNA are glycerol-derived nucleic acid analogues[25–29] (Fig. 3c) and pyranosyl-RNA (containing 4′,2′-linked β-D-ribopyranosyl units)[30,31] (Fig.

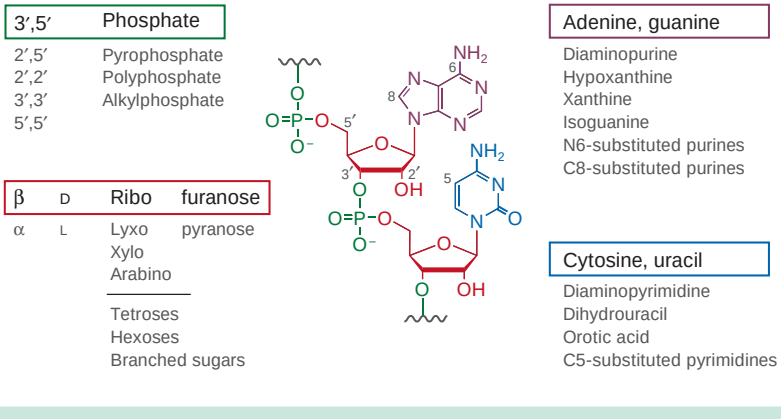

Figure 2 Prebiotic clutter surrounding RNA. Each of the four components of RNA (coloured green, red, purple and blue) would have been accompanied by several closely related analogues (listed in black type), which could have assembled in almost any combination. All possible building blocks for each of the components should be regarded as sorting independently; for example, the phosphodiester linkage may have comprised either a 3′,5′ linkage involving a phosphate or a 2′,5′ linkage involving a pyrophosphate.

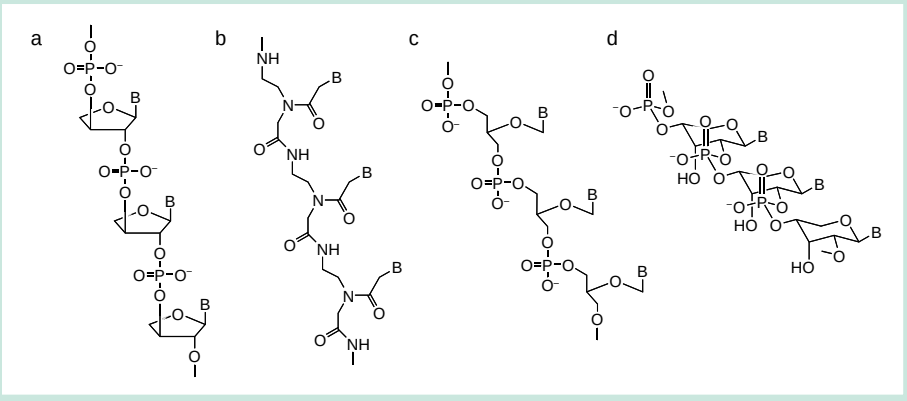

Figure 3 Candidate precursors to RNA during the early history of life on Earth. **a**, Threose nucleic acid; **b**, peptide nucleic acid; **c**, glycerol-derived nucleic-acid analogue; **d**, pyranosyl-RNA. B, nucleotide base.

3d), although neither has garnered sufficient experimental support to be considered a strong candidate.

It is also possible that RNA-based life was preceded by a replicating, evolving polymer that bore no resemblance to nucleic acids. Self-replication without darwinian evolution has been demonstrated for certain peptides[32] and even small organic compounds[33]. Why not cast the net broadly and consider any polymer that is capable of self-replication? A critical issue then becomes whether there is a sufficient diversity of polymer sequences that can be replicated faithfully to provide the basis for darwinian evolution. Nucleic acids have the great advantage that their potential to act as a template is sequence independent, but the templating properties of a particular nucleic acid molecule are highly sequence specific. Peptide replication based on templating within a complex of α-helices offers more restricted choices of distinct self-replicating entities, but perhaps enough to sustain a lineage of compounds in the face of a changing environment. A more radical suggestion is that the first form of life was not based on organic polymers at all, but rather on inorganic clays[34]. Information would be represented by the distribution of charges or shapes along the surface of the clay, and replication would involve copying that information to newly formed clay layers. Suggestions of this kind challenge chemists to think more broadly about the nature of heritable chemical information and to devise experiments to test these ideas.

The transition to RNA from whatever might have preceded it would have had a very different character depending on whether the predecessor was a nucleic acid-like molecule. If the predecessor was able to cross-pair with RNA then the transition may have been a gradual one. Genetic information could have been preserved by 'transcription' of the pre-RNA to RNA, conferring selective advantage based on the function of the transcribed molecules. Once the RNA became self-replicating, it could have usurped the role of genetic material and the pre-RNA would have become expendable. If the predecessor was not a nucleic acid-like molecule, the appearance of RNA might have involved either a 'translation' process, adapting pre-RNA-based information to RNA-based information, or a 'genetic takeover'[35] in which none of the genetic information in pre-RNA was passed on to RNA.

Catalytic activity that resided in a pre-RNA molecule, even if that molecule was very similar to RNA, would not be expected to carry over to RNA without further evolutionary refinement. However, a pre-RNA catalyst that resembled RNA might be pre-adapted to evolve a specific function when prepared as the corresponding RNA because of preserved features of its secondary and tertiary structure. The catalytic potential of TNA, PNA and other proposed precursors to RNA has not yet been explored, but any cogent hypothesis regarding pre-RNA life must consider whether that prior genetic system could have facilitated the appearance of RNA. Once RNA appeared and became beneficial to a system undergoing darwinian evolution, further evolutionary innovation pertaining to the synthesis and utilization of RNA would be expected to follow. In this way, pre-RNA life may have helped to overcome the problems of clutter in prebiotic synthesis by providing solutions discovered through natural selection. Eventually RNA molecules would have become responsible for ensuring the availability and replicability of RNA, ushering in the era of RNA-based darwinian evolution.

RNA-catalysed RNA replication

The general features of RNA-based life can be inferred by considering the requirements for darwinian evolution and the biochemical properties of RNA. The central process of the RNA world was the replication of RNA, presumably catalysed by RNA. The most widely studied, but by no means exclusive model for RNA replication involves template-directed polymerization of activated mononucleotides. Alternatively, replication may have involved the joining of oligonucleotides or even larger subunits[36], perhaps by modular assembly rather than organization along a linear template. The standard model, however, is most congruent with known biological systems and illustrates the requirements for RNA-catalysed RNA replication.

Nucleotides can be activated in several ways. The biological strategy of using nucleoside 5′-triphosphates (NTPs) is especially appealing because the linkage between the α- and β-phosphates provides a strong thermodynamic driving force (standard free energy of hydrolysis at pH 7 of about –10 kcal mol^{-1}), yet is kinetically stable in typical aqueous environments ($k_{hydrolysis}$ ~10^{-10} min^{-1} at pH 7 and 37 °C)[37,38]. Furthermore, polymerization of NTPs is accompanied by release of inorganic pyrophosphate, a small molecule that can readily diffuse away from the reaction centre, avoiding product inhibition. Because of its kinetic stability, however, the α,β-phosphoester reacts slowly with the 3′-hydroxyl of RNA unless the reaction is catalysed. The uncatalysed rate of joining two adjacent template-bound oligonucleotides, one bearing a 2′,3′-hydroxyl and the other a 5′-triphosphate, is only ~10^{-7} min^{-1} at pH 7 and 37 °C (ref. 39). This is comparable to the rate of hydrolysis of a single RNA phosphodiester under the same reaction conditions[40].

There is no known ribozyme in biology that catalyses the template-directed polymerization of NTPs, but such molecules have been obtained using test-tube evolution. Like the evolution of organisms in nature, evolution of RNA in the laboratory involves repeated rounds of selective amplification, linking the survival of an RNA species to its fitness[41–43]. In the laboratory, fitness is defined by the experimenter, for example, based on the ability of RNA to catalyse a particular chemical reaction. Molecules that have been selected as a consequence of their function are then amplified using standard molecular biology techniques, typically reverse transcription followed by amplification using the polymerase chain reaction and then forward transcription. Random mutations may be introduced during the amplification process in order to maintain variation in the population. Through repeated rounds of selective amplification and

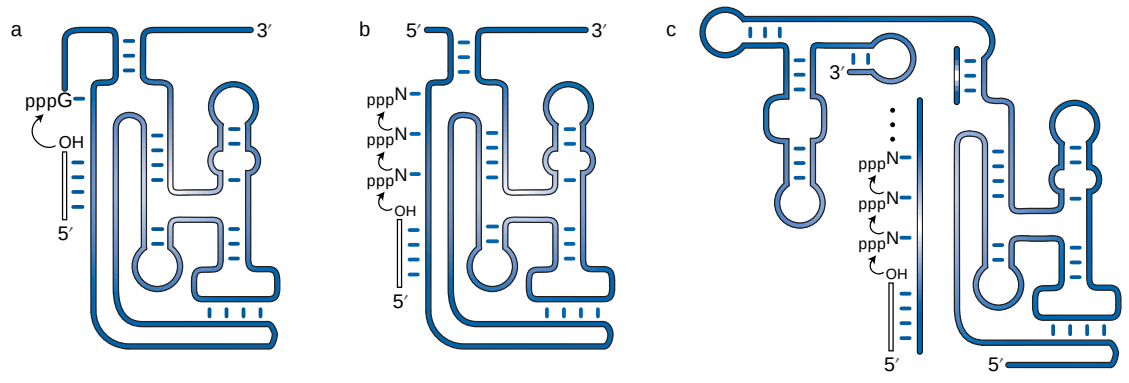

Figure 4 Successive phases in the *in vitro* evolution of an RNA polymerase ribozyme. **a**, Class I ligase ribozyme catalyses template-directed joining of the 3′ end of an RNA primer (open line) to the 5′ end of the ribozyme[44]. **b**, Class I ligase also catalyses addition of three nucleoside 5′-triphosphates (NTPs) to the 3′ end of the primer, directed by an internal template[46]. **c**, Class I-derived polymerase catalyses addition of up to 14 NTPs on an external RNA template[5].

mutation, a population of RNA molecules can be evolved to perform a defined task, provided that they have the capacity to do so.

There are several examples of *in vitro*-evolved ribozymes that catalyse the template-directed joining of an oligonucleotide 3′-hydroxyl and oligonucleotide 5′-triphosphate. The best studied of these is the class I ligase, first isolated almost ten years ago[44] (Fig. 4a). It contains an internal template region that binds a complementary RNA substrate, and directs attack of the 3′-hydroxyl of the substrate on the 5′-triphosphate of the ribozyme, forming a 3′,5′-phosphodiester[45]. The ribozyme contains ~120 nucleotides and operates with a catalytic rate of ~100 min^{-1}, corresponding to a rate enhancement of ~10^9-fold compared to the uncatalysed reaction. The class I ligase also catalyses the polymerization of NTPs, adding up to three residues to the 3′ end of an RNA primer in a template-directed manner[46] (Fig. 4b).

The class I ligase was used as a starting point for further evolution experiments, resulting in a ribozyme with much more robust NTP polymerization activity[5] (Fig. 4c). The final evolved ribozyme contains ~200 nucleotides and catalyses extension of an RNA primer on an external RNA template, adding up to 14 successive nucleotides in 24 hours. It is general with respect to the template sequence, yet operates with an average fidelity of ~97% per nucleotide in copying the template sequence to that of a complementary product. This activity is not sufficient to support the RNA-catalysed replication of RNAs that are as large as the catalyst itself. However, there does not seem to be any fundamental obstacle to achieving the required level of activity, provided there exists a sufficiently powerful evolutionary search procedure.

There are likely to be many different RNA molecules that are capable of catalysing the template-directed polymerization of NTPs. The hc ligase ribozyme, also obtained by *in vitro* evolution[47], has no significant structural or sequence similarity to the class I ligase, but also catalyses the extension of an RNA primer on an external RNA template[48]. It adds only one or two nucleotides, but so far has been selected for ligation rather than polymerization activity. Two other *in vitro*-evolved ribozymes, the L1 and R3 ligases, catalyse formation of a 3′,5′-phosphodiester on an internal, but not external, template[49,50]. The latter ribozyme is notable because it contains only three of the four nucleotides, completely lacking cytidine. Other ligases have been obtained that catalyse formation of a 2′,5′- rather than 3′,5′-phosphodiester, the ligases themselves being composed of 3′,5′-linked RNA[45,51]. A thorough exploration of the vast number of possible RNA sequences would be expected to produce numerous 3′,5′ ligases, many of which could be evolved into NTP polymerases. Each of these would be tolerant of substantial sequence variation by replacing base pairs in stem regions with other Watson–Crick pairs and substituting non-critical residues within loop regions by different nucleotides.

Although a very large number of RNA polymerase ribozymes might be possible, collectively they would comprise only a tiny fraction of the huge number of possible RNA sequences. For RNA molecules that contain 100 nucleotides, there are 4^{100} (~10^{60}) possible sequences. A pool of one copy each of these molecules would have a mass greater than 10^{13} times that of the Earth. A pool of one copy each of all possible 40mers, with a mass of 26 kg, just might be achievable, but it is not clear if 40 nucleotides are sufficient to provide robust RNA polymerase activity. The ribozyme would be required not only to perform the chemistry of polymerization, but also to do so with sufficient fidelity to maintain the selected sequence information over successive generations. Occasional mutations are needed to maintain variability in an evolving population, but too many mutations make it impossible to retain an advantageous genotype. There is a well-established theoretical framework for assessing the effect of genome size, replication rate and replication fidelity on the ability to maintain heritable genetic information[52]. As a rule of thumb, the error rate of replication per nucleotide must be no more than about the inverse of genome length, corresponding to 99% fidelity for replication of a 100mer and 97.5% fidelity for replication of a 40mer. There may be polymerase ribozymes that meet these requirements, although such molecules have not yet been demonstrated.

The above discussion ignores other obstacles to RNA-catalysed RNA replication, such as maintaining a supply of activated mononucleotides, ensuring that the ribozyme will recognize its corresponding genomic RNA while ignoring other RNAs in the environment, overcoming stable self-structure within the template strand, separating the template and product strands, and operating in a similar manner on the product strand to generate new copies of the template. Additional genetic information might be required to overcome these obstacles, but a longer genome would necessitate an even higher fidelity of replication. Mitigating against these demands is the likelihood that RNA polymerase activity first arose in a pre-RNA world. The earliest RNA polymerases need not have been responsible for replicating entire RNA genomes, but merely for generating RNAs that enhanced the fitness of pre-RNA-based life. Further evolutionary innovation could have occurred by exploring sequences related to these functional polymerases, rather than a much broader search of all possible sequences.

Metabolic function in the RNA world

Although the central process of the RNA world was the replication of RNA genomes, some form of metabolism must have supported the process. In keeping with the second law of thermodynamics, the increase in order that occurs in a genetic system is achieved through the expenditure of high-energy starting materials that are converted

to lower-energy products. The incorporation of NTPs into an RNA polymer would qualify as a simple metabolism, although one would need to account for the source of the high-energy NTPs. Some of the starting materials may have been provided by the environment, for example, ribose and other sugars, inorganic polyphosphate, and the building blocks of purines and pyrimidines. Chemical processes in the environment may have led to more complex compounds, drawing on natural energy sources such as sunlight, electric discharges and geothermal activity. These reactions would not be considered part of metabolism, because they would not be carried out by genetically encoded catalysts. The final touches, however, leading to specific chemical organization, probably would have required the assistance of evolved catalysts. In the RNA world, those catalysts are assumed to have been ribozymes.

Ribozymes that catalyse some of the steps of nucleotide synthesis have been obtained by *in vitro* evolution (Fig. 5). One such ribozyme catalyses the formation of a nucleotide from a pyrimidine and activated ribose[4]. This is a notoriously difficult reaction in prebiotic chemistry[53], but was achieved starting with a pool of random-sequence RNAs that were tethered to 5-phosphoribosyl-1-pyrophosphate and allowed to react with 4-thiouracil. The evolved ribozyme performs the reaction with a catalytic rate of ~0.1 min^{-1} and a rate enhancement of $>10^7$-fold compared to the uncatalysed reaction. Another *in vitro*-evolved ribozyme catalyses 5′-phosphorylation of polynucleotides, using ATP-γ-S (or ATP) as the phosphate donor[54]. It operates with a catalytic rate of ~0.2 min^{-1} (or 0.003 min^{-1} with ATP) and a catalytic rate enhancement of ~10^9-fold. Yet another ribozyme catalyses activation of the 5′-phosphate by attachment of a 5′,5′-pyrophosphate-linked nucleotide[55]. This linkage is less energetic than the α,β-phosphoanhydride of an NTP. However, ribozymes have been obtained that catalyse template-directed ligation of RNA driven by release of adenylate from a terminal adenosine-5′,5′-pyrophosphate[56].

There are several important reactions in nucleotide synthesis that have not yet been carried out with a ribozyme (Fig. 5). The formation of ribose from simple aldehydes would be a significant achievement, requiring a ribozyme that catalyses a substrate-specific aldol condensation. The formation of alkylphosphates, such as glycerol phosphate or phosphorylethanolamine, would be notable, especially if the source of phosphate was a mineral or some other compound that was abundant in the environment. RNA is adept at catalysing phosphoryl transfer reactions, so it is not difficult to imagine how the phosphate, once mobilized, then would be transferred to ribose or other compounds. The synthesis of purines (for example, from cyanates) and pyrimidines (for example, from carbamoyl phosphate and aspartate) would be important as well.

The possibility of a more complex RNA-based metabolism is purely conjectural. That said, one could imagine that all of the reactions of central metabolism, now catalysed by protein enzymes, were once catalysed by ribozymes. It has been suggested that the nucleotide-derived coenzymes, which have a prominent role in most of these reactions today, are remnants of an earlier RNA-based metabolism[57]. Another extreme but opposite point of view is that the only catalytic function of RNA in the RNA world was to direct the synthesis of encoded polypeptides. The synthesis of nucleotides and the template-directed polymerization of RNA may have been the responsibility of pre-RNA catalysts, with RNA merely serving as a messenger, aminoacyl adaptor and peptidyl transferase catalyst, just as it does today.

How does one assess the likelihood that a putative RNA-based function did in fact exist in the RNA world? First, it must fall within the capabilities of RNA, preferably bolstered by an experimental demonstration. Ribozymes have been obtained through *in vitro* evolution that catalyse a broad range of chemical reactions, including acyl transfer[58,59], *N*- and *S*-alkylation[60,61], carbon–carbon bond formation[62,63], amide bond formation[64] and Michael addition[65]. But controlling a free radical within a hydrophobic pocket is likely to be

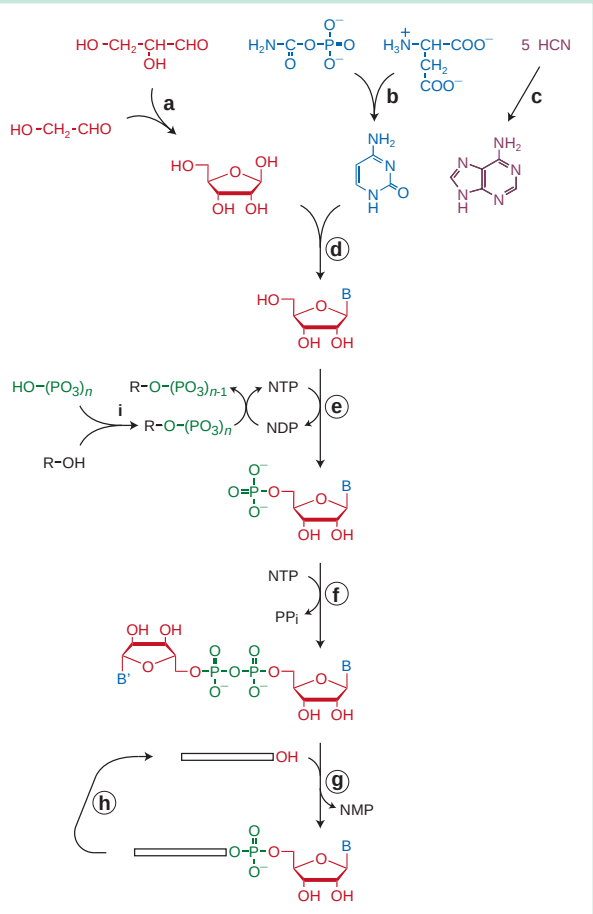

Figure 5 Hypothetical pathway for RNA-catalysed synthesis of RNA. A circled letter indicates reactions that have been demonstrated experimentally. **a**, Aldol condensation of glycoaldehyde and glyceraldehyde to form ribose. **b**, Transfer of the carbamoyl group of carbamoyl phosphate to aspartate and subsequent cyclization to form a pyrimidine. **c**, Pentamerization of HCN to form a purine. **d**, Addition of a purine or pyrimidine (B) to ribose to form a nucleoside. **e**, Phosphorylation of a nucleoside to form a nucleotide. **f**, Activation of a nucleotide by transfer of the nucleotide portion of NTP. **g**, Addition of a nucleotide to the 3′ end of an RNA primer (open line). The two RNA substrates are bound at adjacent positions on a complementary template (not shown). **h**, Successive nucleotide additions resulting in further primer extension. **i**, Phosphoryl transfer from an alkyl polyphosphate to NDP, regenerating NTP. NMP is converted to NDP in a similar manner. The ultimate source of phosphate is a polyphosphate mineral.

beyond the capabilities of RNA, leading some to suggest that ribozymes never were responsible for the conversion of ribose to deoxyribose[66,67]. Second, the RNA-catalysed reaction must have provided some selective advantage that was not contingent on future evolutionary developments. For example, a ribozyme that catalysed formation of ribosyl-1-amine from ribosyl-1-pyrophosphate and glutamine may have been selected based on the utility of the amino-sugar, but not with regard to its eventual role as an intermediate in purine biosynthesis.

If both of these requirements are met, then one should consider whether there is any evidence for the RNA-based function in contemporary biology or the geological record. It is possible, of course, that the function existed in the RNA world, but left no trace that can be detected today. Conversely, an RNA-based function that exists in contemporary biology need not have arisen in the RNA world. If, however, that function is widely distributed across all three kingdoms of life and uses RNA in a way that is not uniquely beholding to the

chemical properties of RNA, then the argument that it is a remnant of the RNA world becomes more persuasive[68].

The instructed synthesis of proteins is a strong candidate for a function that existed in the RNA world. Similar rationale has been applied to suggest that tetrapyrrole biosynthesis may have arisen in the RNA world[68]. The C5 pathway for the synthesis of 5-amino-levulinic acid leading to the tetrapyrroles is well represented in all three kingdoms. It uses glutamyl-transfer RNA for the synthesis of glutamate-1-semialdehyde[69], although other glutamyl esters would do just as well. In contrast, self-splicing introns are broadly distributed in contemporary biology but take special advantage of the base-pairing properties of RNA, and thus should be viewed more agnostically. Other RNA-based functions for which there is no evidence in biology, such as nucleotide synthesis and RNA polymerization, are assumed to have existed in the RNA world based on first principles, but it is important to recognize that this assumption is not supported by available historical evidence.

It is often said, again based on first principles rather than historical evidence, that RNA-based life must have entailed some form of cellular compartmentalization[70,71]. This would be advantageous for keeping together an RNA replicase ribozyme and its corresponding genomic RNA, and more generally for retaining the fruits of an RNA-based metabolism for the benefit of the system that produced them. The notion of cellular compartmentalization should not be taken too literally — although all cells in contemporary biology are surrounded by a membrane composed of amphipathic lipids, there are other ways to achieve the preferential association of an ensemble of compounds. It has been proposed, for example, that there were organizing centres, analogous to modern ribosomes, where various RNAs and small molecules came together through non-covalent or transient covalent interactions[72]. Small organic molecules could have been esterified to RNA and these carrier-linked metabolites could have been held in close proximity through RNA–RNA interactions or organization along a surface. The same outcome could be achieved by passive compartmentalization on the surface of fine particulate matter, within aerosol particles in the atmosphere[73], or within the pores of a rock. Even if the RNA world (or pre-RNA world) synthesized compartments, those compartments might have been assembled from something other than complex phospholipids, such as nucleic acids, alternating polypeptides that form β-sheets[74], or simple terpenoids[75].

Transition to the DNA–protein world

Although RNA is well suited as a genetic molecule and can evolve to perform a broad range of catalytic tasks, it has limited chemical functionality and thus may not be equipped to meet certain challenges and opportunities that arise in the environment. An important innovation of life on Earth was the development of a separate macromolecule that would be responsible for most catalytic functions, even though that molecule contained subunits that were poorly suited for replication. The invention of protein synthesis, instructed and catalysed by RNA, was the crowning achievement of the RNA world, but also began its demise.

RNA is capable of performing all of the reactions of protein synthesis (Fig. 6). The messenger, transfer and ribosomal RNA molecules that exist in all known organisms direct the assembly of specific polypeptide sequences, instructed by corresponding RNA sequences. The activation of amino acids in the form of aminoacyl adenylates, and subsequent transfer of the amino acids to the 2′(3′) terminus of tRNAs, are catalysed in modern biology by the set of 20 aminoacyl-tRNA synthetase proteins. These reactions also have been achieved with *in vitro*-evolved ribozymes. A ribozyme that contains ~110 nucleotides catalyses addition of either leucine or phenylalanine to its own 5′ terminus, forming an aminoacyl-nucleotide anhydride[76]. Another ribozyme catalyses aminoacylation of its own 2′(3′) terminus, using various aminoacyl-adenylate substrates[77,78]. Yet another ribozyme catalyses aminoacylation of tRNAs that are either covalently attached to the ribozyme or provided as a separate substrate[79].

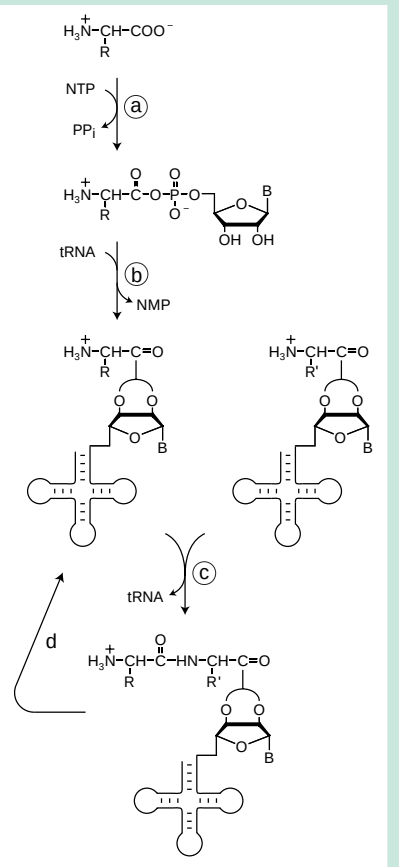

Figure 6 Hypothetical pathway for RNA-catalysed protein synthesis. A circled letter indicates reactions that have been demonstrated experimentally. **a**, Activation of an amino acid by formation of an aminoacyl-nucleotide anhydride. **b**, Transfer of an activated amino acid to the 2′(3′) terminus of tRNA. The semicircle between the 2′- and 3′-oxygens indicates that the amino acid migrates rapidly between these two positions. **c**, Peptidyl transfer resulting in formation of a dipeptide. The two aminoacyl-tRNA substrates are bound at adjacent positions on a complementary template (not shown). **d**, Successive peptidyl transfer reactions resulting in formation of a polypeptide.

The final step of protein synthesis involves binding aminoacyl and peptidyl oligonucleotides at adjacent positions along an RNA template and catalysing peptide bond formation through attack of the α-amine of the amino acid on the carbonyl of the peptidyl ester. From a chemical perspective this is the easiest step, proceeding spontaneously once the reactants are brought into close proximity. It has been shown, for example, that when 2′(3′)-glycyl adenosine is bound to a complementary template, peptide bond formation ensues, giving rise to diglycine, which then cyclizes to form a diketopiperazine[80]. The modern ribosome achieves high template occupancy and precise orientation of the aminoacyl- and peptidyl-tRNAs, and may use additional catalytic strategies in promoting peptide bond formation[81]. It is not difficult to visualize how RNA alone could carry out this reaction; in fact, the crystal structure of the ribosome reveals a peptidyl transferase site that is composed entirely of RNA (ref. 82, and see review in this issue by Moore and Steitz, pages 229–235).

In vitro evolution has been used to develop ribozymes that catalyse peptide bond formation. Beginning with a pool of random-sequence RNAs with phenylalanine tethered to their 5′ end, molecules were selected based on their ability to react with a methionyl-adenylate substrate to form a tethered dipeptide[7]. One of the evolved ribozymes contains ~190 nucleotides and has a catalytic rate of ~0.1 min^{-1}. It accepts several different aminoacyl adenylates, preferring methionine, and can be made to operate with multiple turnover by providing the phenylalanine substrate tethered to a short oligonucleotide rather than to the ribozyme itself[83]. Just as the class I ligase ribozyme has been evolved to function as an RNA polymerase, this peptidyl transferase ribozyme might be evolved to form multiple peptide bonds in succession.

It is not known whether the invention of protein synthesis preceded or followed the invention of DNA genomes. The primary advantage of DNA over RNA as a genetic material is the greater chemical stability of DNA, allowing much larger genomes based on DNA. Protein synthesis may require more genetic information than can be maintained by RNA. However, the original aminoacyl adaptor molecules may have been smaller in size and fewer in number than contemporary tRNAs[84], and if the aminoacylation and peptidyl transfer activities were far less sophisticated than in the modern ribosome, an RNA genome of only a few thousand nucleotides might have been sufficient for protein synthesis. The chief argument in favour of proteins before DNA is that ribozymes seem to be incapable of catalysing the reduction of deoxyribose to ribose through the same mechanism used by all known ribonucleoside reductase proteins[66,67]. It is conceivable, however, that RNA used a different mechanism, perhaps involving reduction of an attached purine followed by acid-catalysed elimination of the ribose 2′-hydroxyl (A. Eschenmoser, personal communication).

The template-directed polymerization of DNA is more difficult than for RNA because the 3′-hydroxyl of DNA has substantially lower acidity compared to that of RNA[85]. Nonetheless, ribozymes are capable of deprotonating a DNA 3′-hydroxyl, allowing nucleophilic attack on phosphate to form a 3′,5′-phosphodiester linkage[86]. It is not difficult to imagine that a ribozyme could function as a DNA polymerase. Such a molecule might arise in nature or in the laboratory as an evolutionary descendant of an RNA polymerase ribozyme. RNA-based information could be reverse transcribed to DNA for safe keeping, then read back to RNA by a DNA-dependent RNA polymerase. Eventually the DNA molecules became the objects of replication, completing the transition to the DNA–protein world.

A largely open question concerns the origin of the genetic code. The aminoacylation of RNA initially must have provided some selective advantage unrelated to the eventual development of a translation machinery. It has been proposed, for example, that aminoacylation protected RNA from degradation[87], anchored RNA in an advantageous environment[87], marked genomic RNAs for replication[87,88], or enhanced the catalytic properties of RNA[89]. Some amino acids would have been especially useful in these roles, leading to selective aminoacylation with one or perhaps a few related amino acids. Different RNAs may have been aminoacylated with different amino acids, providing the basis for a family of precursors to the modern aminoacyl-tRNAs. The RNA component of these precursor molecules may have been as simple as a stem-loop structure[84,90] or as elaborate as a ribozyme that catalyses its own aminoacylation[77,79].

The next step towards the origin of the genetic code was the formation of peptide bonds between amino acids that were attached to RNA. The products of this reaction must have conferred some selective advantage, even though the peptides probably would have been too small and too heterogeneous in sequence to function as catalysts. Instead, they might have served as cofactors for ribozymes[91,92] or been more effective than amino acids for any of the roles suggested above. RNA-catalysed peptide bond formation would have resulted in a large number of possible peptide sequences, and even this mixture may have been useful[87]. However, the development of a crude mechanism for controlling the diversity of possible peptides would have been advantageous, and progressive refinement of that mechanism would have provided further selective advantage. It is reasonable to postulate that, like the modern translation apparatus, the ancestral translation system made use of messenger-like RNA molecules to gather aminoacyl-RNAs in a specific order through Watson–Crick pairing interactions. It is not clear, however, how the detailed assignments of the genetic code were made.

RNA has several features that make it suitable as the basis for a simple darwinian system: it contains only four different subunits with very similar chemical properties, its subunits polymerize readily when activated and bound to a complementary template, it is a polyanion that is readily soluble in water almost irrespective of sequence, it forms simple secondary structures that are highly tolerant of sequence variation, and it can adopt entirely different structures following the acquisition of a few critical mutations[93,94]. These same features make it less sophisticated compared to its DNA and protein successors. The lower reactivity but greater stability of DNA makes it a better choice for the genetic material, whereas the greater chemical diversity of the subunits of proteins, including anionic, cationic and hydrophobic groups, makes protein a better choice as the basis for catalytic function. However, those more sophisticated molecules could not have arisen without the foundation that had been laid by RNA.

Outlook

The reign of the RNA world on Earth probably began no more than about 4.2 billion years ago and ended no less than about 3.6 billion years ago[95]. It may have occupied only a small portion of that interval, with the pre-RNA world having come before. Insight into the origin and operation of the RNA world is largely inferential, based on the known chemical and biochemical properties of RNA. In the best of circumstances those inferences are supported by examining the role of RNA in contemporary biology. Without that support one must be careful not to draw detailed conclusions regarding these historical events. Future studies will sharpen the picture of ancestral RNA-based life through combined efforts in prebiotic chemistry, *in vitro* evolution, biochemical analysis and molecular phylogenetics. It should be possible to formulate more precise boundary conditions regarding the environmental conditions of the early Earth and the types of chemical reactions that would have occurred under those conditions. Additional catalytic RNAs are likely to be found in biology and undoubtedly many more will be discovered through test-tube evolution. The construction of artificial RNA-based life from synthetic oligonucleotides is a distinct possibility[71], and there even is a chance that a remnant of the RNA world will be found lurking in some special contemporary microenvironment. □

1. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289,** 905–920 (2000).
2. Wimberly, B. T. *et al.* Structure of the 30S ribosomal subunit. *Nature* **407,** 327–338 (2000).
3. Yusupov, M. *et al.* Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292,** 883–896 (2001).
4. Unrau, P. J. & Bartel, D. P. RNA-catalysed nucleotide synthesis. *Nature* **395,** 260–263 (1998).
5. 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* **292,** 1319–1325 (2001).
6. Lee, N., Bessho, Y., Wei, K., Szostak, J. W. & Suga, H. Ribozyme-catalyzed tRNA aminoacylation. *Nature Struct. Biol.* **7,** 28–33 (2000).
7. Zhang, B. & Cech, T. R. Peptide bond formation by *in vitro* selected ribozymes. *Nature* **390,** 96–100 (1997).
8. von Kiedrowski, G. A self-replicating hexadeoxynucleotide. *Angew. Chem.* **25,** 932–935 (1986).
9. Gilbert, W. The RNA world. *Nature* **319,** 618 (1986).
10. Joyce, G. F. RNA evolution and the origins of life. *Nature* **338,** 217–224 (1989).
11. Ferris, J. P., Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis III. Synthesis of pyrimidines from cyanoacetylene and cyanate. *J. Mol. Biol.* **33,** 693–704 (1968).
12. Robertson, M. P. & Miller, S. L. An efficient prebiotic synthesis of cytosine and uracil. *Nature* **375,** 772–774 (1995).
13. Lohrmann, R. & Orgel, L. E. Prebiotic synthesis: phosphorylation in aqueous solution. *Science* **161,** 64–66 (1968).
14. Fuller, W. D., Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis. VI. Synthesis of purine nucleosides. *J. Mol. Biol.* **67,** 25–33 (1972).
15. Müller, D. *et al.* Chemie von α-Aminonitrilen. Aldomerisierung von Glykolaldehydphosphat zu *racemischen* Hexose-2,4,6-triphosphaten und (in Gegenwart von Formaldehyd) *racemischen* Pentose-2,4-diphosphaten: *rac.*-Allose-2,4,6-triphosphat und *rac.*-Ribose-2,4-diphosphat sind die Reaktionshauptprodukte. *Helv. Chim. Acta* **73,** 1410–1468 (1990).
16. Krishnamurthy, R., Pitsch, S. & Arrhenius, G. Mineral induced formation of pentose-2,4-bisphosphates. *Origins Life Evol. Biosph.* **29,** 139–152 (1999).
17. Ferris, J. P. & Ertem, G. Oligomerization of ribonucleotides on montmorillonite: reaction of the 5′-phosphorimidazolide of adenosine. *Science* **257,** 1387–1389 (1992).
18. Eschenmoser, A. Chemical etiology of nucleic acid structure. *Science* **284,** 2118–2124 (1999).
19. Schöning, K.-U. *et al.* Chemical etiology of nucleic acid structure: the α-threofuranosyl-(3′→2′) oligonucleotide system. *Science* **290,** 1347–1351 (2000).
20. Nielsen, P. E., Egholm, M., Berg, R. H. & Buchardt, O. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* **254,** 1497–1500 (1991).
21. Nelson, K. E., Levy, M. & Miller, S. L. Peptide nucleic acids rather than RNA may have been the first genetic molecule. *Proc. Natl Acad. Sci. USA* **97,** 3868–3871 (2000).
22. Joyce, G. F. *et al.* Chiral selection in poly(C)-directed synthesis of oligo(G). *Nature* **310,** 602–604 (1984).
23. Schmidt, J. G., Nielsen, P. E. & Orgel, L. E. Enantiomeric cross-inhibition in the synthesis of oligonucleotides on a nonchiral template. *J. Am. Chem. Soc.* **119,** 1494–1495 (1997).
24. Eriksson, M. *et al.* Sequence dependent *N*-terminal rearrangement and degradation of peptide nucleic acid (PNA) in aqueous solution. *New J. Chem.* **22,** 1055–1059 (1998).

25. Spach, G. Chiral versus chemical evolutions and the appearance of life. *Origins Life Evol. Biosph.* **14,** 433–437 (1984).
26. Joyce, G. F., Schwartz, A. W., Miller, S. L. & Orgel, L. E. The case for an ancestral genetic system involving simple analogues of the nucleotides. *Proc. Natl Acad. Sci. USA* **84,** 4398–4402 (1987).
27. Schneider, K. C. & Benner, S. A. Oligonucleotides containing flexible nucleoside analogues. *J. Am. Chem. Soc.* **112,** 453–455 (1990).
28. Merle, Y., Bonneil, E., Merle, L., Sagi, J. & Szemzo, A. Acyclic oligonucleotide analogues. *Int. J. Biol. Macromol.* **17,** 239–246 (1995).
29. Chaput, J. C. & Switzer, C. Nonenzymatic oligomerization on templates containing phosphoester-linked acyclic glycerol nucleic acid analogues. *J. Mol. Evol.* **51,** 464–470 (2000).
30. Pitsch, S., Wendeborn, S., Jaun, B. & Eschenmoser, A. Why pentose- and not hexose nucleic acids? Pyranosyl-RNA ('p-RNA'). *Helv. Chim. Acta* **76,** 2161–2183 (1993).
31. Pitsch, S. *et al.* Pyranosyl-RNA ('p-RNA'): base-pairing selectivity and potential to replicate. *Helv. Chim. Acta* **78,** 1621–1635 (1995).
32. Lee, D. H., Granja, J. R., Martinez, J. A., Severin, K. & Ghadiri, M. R. A self-replicating peptide. *Nature* **382,** 525–528 (1996).
33. Tjivikua, T., Ballester, P. & Rebek, J. Jr A self-replicating system. *J. Am. Chem. Soc.* **112,** 1249–1250 (1990).
34. Cairns-Smith, A. G. The origin of life and the nature of the primitive gene. *J. Theor. Biol.* **10,** 53–88 (1966).
35. Cairns-Smith, A. G. & Davies, C. J. in *Encyclopaedia of Ignorance* (eds Duncan, R. & Weston-Smith, M.) 391–403 (Pergamon, Oxford, 1977).
36. James, K. D. & Ellington, A. D. The fidelity of template-directed oligonucleotide ligation and the inevitability of polymerase function. *Origins Life Evol. Biosph.* **29,** 375–390 (1999).
37. Kirby, A. J. & Younas, M. The reactivity of phosphate esters. Diester hydrolysis. *J. Chem. Soc. B* 510–513 (1970).
38. Admiraal, S. J. & Herschlag, D. Catalysis of phosphoryl transfer from ATP by amine nucleophiles. *J. Am. Chem. Soc.* **121,** 5837–5845 (1999).
39. Rohatgi, R., Bartel, D. P. & Szostak, J. W. Kinetic and mechanistic analysis of nonenzymatic, template-directed oligoribonucleotide ligation. *J. Am. Chem. Soc.* **118,** 3332–3339 (1996).
40. Li, Y. & Breaker, R. R. Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2′-hydroxyl group. *J. Am. Chem. Soc.* **121,** 5364–5372 (1999).
41. Robertson, D. L. & Joyce, G. F. Selection *in vitro* of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* **344,** 467–468 (1990).
42. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249,** 505–510 (1990).
43. Ellington, A. D. & Szostak, J. W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346,** 818–822 (1990).
44. Bartel, D. P. & Szostak, J. W. Isolation of new ribozymes from a large pool of random sequences. *Science* **261,** 1411–1418 (1993).
45. Ekland, E. H., Szostak, J. W. & Bartel, D. P. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **269,** 364–370 (1995).
46. Ekland, E. H. & Bartel, D. P. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382,** 373–376 (1996).
47. Jaeger, L., Wright, M. C. & Joyce, G. F. A complex ligase ribozyme evolved *in vitro* from a group I ribozyme domain. *Proc. Natl Acad. Sci. USA* **96,** 14712–14717 (1999).
48. McGinness, K. E. & Joyce, G. F. RNA-catalyzed RNA ligation on an external RNA template. *Chem. Biol.* **9,** 297–307 (2002).
49. Robertson, M. P. & Ellington, A. D. *In vitro* selection of an allosteric ribozyme that transduces analytes into amplicons. *Nature Biotechnol.* **17,** 62–66 (1999).
50. Rogers, J. & Joyce, G. F. The effect of cytidine on the structure and function of an RNA ligase ribozyme. *RNA* **7,** 395–404 (2001).
51. Landweber, L. F. & Pokrovskaya, I. D. Emergence of a dual-catalytic RNA with metal-specific cleavage and ligase activities: the spandrels of RNA evolution. *Proc. Natl. Acad. Sci. USA* **96,** 173–178 (1999).
52. Eigen, M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58,** 465–523 (1971).
53. Orgel, L. E. & Lohrmann, R. Prebiotic chemistry and nucleic acid replication. *Acc. Chem. Res.* **7,** 368–377 (1974).
54. Lorsch, J. & Szostak, J. W. *In vitro* evolution of new ribozymes with polynucleotide kinase activity. *Nature* **371,** 31–36 (1994).
55. Huang, F. & Yarus, M. Versatile 5′ phosphoryl coupling of small and large molecules to an RNA. *Proc. Natl Acad. Sci. USA* **94,** 8965–8969 (1997).
56. Hager, A. J. & Szostak, J. W. Isolation of novel ribozymes that ligate AMP-activated RNA substrates. *Chem. Biol.* **4,** 607–617 (1997).
57. White, H. B. III Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **7,** 101–104 (1976).
58. Lohse, P. A. & Szostak, J. W. Ribozyme-catalysed amino-acid transfer reactions. *Nature* **381,** 442–444 (1996).
59. Jenne, A. & Famulok, M. A novel ribozyme with ester transferase activity. *Chem. Biol.* **5,** 23–34 (1998).
60. Wilson, C. & Szostak, J. W. *In vitro* evolution of a self-alkylating ribozyme. *Nature* **374,** 777–782 (1995).
61. Wecker, M., Smith, D. & Gold, L. *In vitro* selection of a novel catalytic RNA: characterization of a sulfur alkylation reaction and interaction with a small peptide. *RNA* **2,** 982–994 (1996).
62. Tarasow, T. M., Tarasow, S. L. & Eaton, B. E. RNA-catalysed carbon–carbon bond formation. *Nature* **389,** 54–57 (1997).
63. Seelig, B. & Jäschke, A. A small catalytic RNA motif with Diels-Alderase activity. *Chem. Biol.* **6,** 167–176 (1999).
64. Wiegand, T. W., Janssen, R. C. & Eaton, B. E. Selection of RNA amide synthases. *Chem. Biol.* **4,** 675–683 (1997).
65. Sengle, G., Eisenführ, A., Arora, P. S., Nowick, J. S. & Famulok, M. Novel RNA catalysts for the Michael reaction. *Chem. Biol.* **8,** 459–473 (2001).
66. Freeland, S. J., Knight, R. D. & Landweber, L. F. Do proteins predate DNA? *Science* **286,** 690–692 (1999).
67. Stubbe, J., Ge, J. & Yee, C. S. The evolution of ribonucleotide reduction revisited. *Trends Biochem. Sci.* **26,** 93–99 (2001).
68. Benner, S. A., Ellington, A. D. & Tauer, A. Modern metabolism as a palimpsest of the RNA world. *Proc. Natl Acad. Sci. USA* **86,** 7054–7058 (1989).
69. Schön, A. *et al.* The RNA required in the first step of chlorophyll biosynthesis is a chloroplast glutamate tRNA. *Nature* **322,** 281–284 (1986).
70. Luisi, P. L. About various definitions of life. *Origins Life Evol. Biosph.* **28,** 613–622 (1998).
71. Szostak, J. W., Bartel, D. P. & Luisi, P. L. Synthesizing life. *Nature* **409,** 387–390 (2001).
72. Gibson, T. J. & Lamond, A. I. Metabolic complexity in the RNA world and implications for the origin of protein synthesis. *J. Mol. Evol.* **30,** 7–15 (1990).
73. Dobson, C. M., Ellison, G. B., Tuck, A. F. & Vaida, V. V. Atmospheric aerosols as prebiotic chemical reactors. *Proc. Natl Acad. Sci. USA* **97,** 11864–11868 (2000).
74. Brack, A. & Orgel, L. E. β structures of alternating polypeptides and their possible prebiotic significance. *Nature* **256,** 383–387 (1975).
75. Ourisson, G. & Nakatani, Y. The terpenoid theory of the origin of cellular life: the evolution of terpenoids to cholesterol. *Chem. Biol.* **1,** 11–23 (1994).
76. Kumar, R. K. & Yarus, M. RNA-catalyzed amino acid activation. *Biochemistry* **40,** 6998–7004 (2001).
77. Illangasekare, M., Sanchez, G., Nickles, T. & Yarus, M. Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* **267,** 643–647 (1995).
78. Illangasekare, M. & Yarus, M. Small-molecule-substrate interactions with a self-aminoacylating ribozyme. *J. Mol. Biol.* **268,** 631–639 (1997).
79. Saito, H., Kourouklis, D. & Suga, H. An *in vitro* evolved precursor tRNA with aminoacylation activity. *EMBO J.* **20,** 1797–1806 (2001).
80. Weber, A. L. & Orgel, L. E. Poly(U)-directed peptide bond formation from the 2′(3′)-glycyl esters of adenosine derivatives. *J. Mol. Evol.* **16,** 1–10 (1980).
81. Barta, A. *et al.* Mechanism of ribosomal peptide bond formation. *Science* **291,** 203a (2001) (published online at www.sciencemag.org/cgi/content/full/291/5502/203a).
82. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289,** 920–930 (2000).
83. Zhang, B. & Cech, T. R. Peptidyl-transferase ribozymes: *trans* reactions, structural characterization and ribosomal RNA-like features. *Chem. Biol.* **5,** 539–553 (1998).
84. Schimmel, P. & Henderson, B. Possible role of aminoacyl-RNA complexes in noncoded peptide synthesis and origin of coded synthesis. *Proc. Natl Acad. Sci. USA* **91,** 11283–11286 (1994).
85. Izatt, R. M., Hansen, L. D., Rytting, J. H. & Christensen, J. J. Proton ionization from adenosine. *J. Am. Chem. Soc.* **87,** 2760–2761 (1965).
86. Sugimoto, N., Tomka, M., Kierzek, R., Bevilacqua, P. C. & Turner, D. H. Effects of substrate structure on the kinetics of circle opening reactions of the self-splicing intervening sequence from *Tetrahymena thermophila*: evidence for substrate and Mg^{2+} binding interactions. *Nucleic Acids Res.* **17,** 355–371 (1989).
87. Orgel, L. E. The origin of polynucleotide-directed protein synthesis. *J. Mol. Evol.* **29,** 465–474 (1989).
88. Weiner, A. M. & Maizels, N. 3′ terminal tRNA-like structures tag genomic RNA molecules for replication: implications for the origin of protein synthesis. *Proc. Natl Acad. Sci. USA* **84,** 7383–7387 (1987).
89. Wong, J.-T. Origin of genetically encoded protein synthesis: a model based on selection for RNA peptidation. *Origins Life Evol. Biosph.* **21,** 165–176 (1991).
90. Schimmel, P. & Ribas de Pouplana, L. Transfer RNA: from minihelix to genetic code. *Cell* **81,** 983–986 (1995).
91. Roth, A. & Breaker, R. R. An amino acid as a cofactor for a catalytic polynucleotide. *Proc. Natl Acad. Sci. USA* **95,** 6027–6031 (1998).
92. Joyce, G. F. Nucleic acid enzymes: playing with a fuller deck. *Proc. Natl Acad. Sci. USA* **95,** 5845–5847 (1998).
93. Fontana, W. & Schuster, P. Continuity in evolution: on the nature of transitions. *Science* **280,** 1451–1455 (1998).
94. Schultes, E. A. & Bartel, D. P. One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science* **289,** 448–452 (2000).
95. Joyce, G. F. The rise and fall of the RNA world. *New Biol.* **3,** 399–407 (1991).

Acknowledgements

I thank T. Cech, A. Eschenmoser, R. Krishnamurthy, L. Orgel, N. Paul, P. Schimmel and W. Shih for helpful comments on the manuscript. I acknowledge funding from the National Aeronautics and Space Administration and the Skaggs Institute for Chemical Biology.