Constructing an RNA world

David P. Bartel and Peter J. Unrau

A popular theory of life’s origins states that the first biocatalysts were not made of protein but were made of RNA or a very similar polymer. Experiments are beginning to confirm that the catalytic abilities of RNA are compatible with this “RNA world” hypothesis. For example, RNA can synthesize short fragments of RNA in a template-directed fashion and promote formation of peptide, ester and glycosidic linkages. However, no known activity fully represents one presumed by the “RNA world” theory, and reactions such as oxidation and reduction have yet to be demonstrated. Filling these gaps would place the hypothesis on much firmer ground and provide components for building minimal forms of RNA-based cellular life.

Since the time of Darwin and his ‘warm little pond’, scientists have been wrestling with the challenge of formulating a plausible scenario for the origin and early evolution of life. How did a self-replicating assembly of molecules emerge on the early Earth and give rise to cellular life? From this perspective, even the simplest free-living bacteria in contemporary biology are of staggering complexity. *Aquifex aeolicus*, a simple autotrophic bacterium, has 152 open reading frames and at least 55 RNA genes. Because hundreds of these genes are common to all branches of life, phylogenetic comparison cannot whittle away the genomes of cellular life to a tractable set of genes representing the first cell. The fossil record for life appears to go back 3.5–3.85 billion years – nearly as far back as the oldest terrestrial rocks. Yet some of the oldest fossils resemble modern cyanobacteria, providing few clues as to how the earliest life differed from that of today.

Because approaching the origin of life by looking back through phylogenetic or palaeontologic evidence has proved difficult, one might hope to work forward instead, starting with the molecules and conditions thought to prevail on the early Earth. However, only a minute fraction of the possible combinations of molecules, conditions, catalytic surfaces and circumstances can be explored. The study of prebiotic chemistry provides key insights into the plausibility of a given scenario, but it cannot hope to recreate a scenario *de novo*. Without constraints from phylogeny and palaeontology, and without a guide through the maze of prebiotic possibilities, theorists formulating scenarios for the early evolution of life have started with the basic principles of replication and evolution.

The RNA world

One hypothesis is that early life was based on RNA. This is, biocatalysis was performed by catalytic RNAs rather than by protein enzymes. The appeal of RNA-based life is that catalytic RNAs, which could have served as their own genes, would have been much simpler to duplicate than proteins. According to this theory, RNA first promoted the reactions required for life with the help of metals, pyridines, amino acids and other small-molecule cofactors. Then, as metabolism became more complex, RNA developed the ability to synthesize coded polypeptides that served as more sophisticated cofactors. DNA eventually replaced RNA as the genetic polymer, and protein replaced RNA as the prominent biocatalyst. The conversion to protein catalysis is not considered complete; RNA retains a central role in protein synthesis, perhaps including catalysis of peptidyl transfer. Remnants of ancestral ribozymes are also thought to persist as nucleotides within many cofactors, such as NAD^+\textsuperscript{1}, NAPD\textsuperscript{+}, FAD, coenzyme A, coenzyme B\textsubscript{12}, ATP and S-adenosylmethionine\textsuperscript{2}.

Now that the ‘RNA world’ hypothesis has been canonized within current biology textbooks, its status as a hypothesis is easily forgotten. Problems remain, particularly the implausibility of prebiotic RNA synthesis and stability\textsuperscript{3–5}. Indeed, most professional advocates of an RNA world are doubtful that life began with RNA per se. Instead, they propose that life began with an RNA-like polymer, yet to be identified, that possessed the catalytic and templating features of RNA but miraculously lacked RNA’s undesirable traits, most notably, its intractable prebiotic synthesis\textsuperscript{6–9}. The era of this RNA-like polymer is the ‘pre-RNA world’, which presumably gave rise to the RNA world in a manner analogous to that in which the RNA world gave rise to the protein-nucleic-acid world of today. Relegating the RNA world to a crucial intermediate in the early evolution of life rescues the hypothesis from the ridiculed status of prebiotic chemistry but still does not place it on firm footing. The ‘RNA world’ scenario hinges on some rather far-fetched assumptions about the catalytic capability of RNA. For example, RNA polymerase ribozymes must have been responsible for replicating the ribozymes of the RNA world, including themselves (via their complementary sequences). RNA replication is a very challenging set of reactions – far more challenging than those yet known to be catalyzed by RNA. And replication is just the beginning. To successfully make the transition to the protein-nucleic-acid world, RNA must have been able to promote coded polypeptide synthesis and a host of metabolic reactions.

The ‘RNA world’ theme park

Can RNA catalyse the reactions needed for self-replication on the early Earth? Can RNA-based life achieve the metabolic sophistication needed to give birth to the protein-nucleic-acid world? In beginning to answer these questions, it has been useful to look beyond the ribozymes found in contemporary biology – primarily because only seven different types have been found. Two types perform self-splicing reactions, four perform self-cleavage and one trims off the 5’ end of pre-RNA\textsuperscript{10}. All perform phosphodiester transfer or phosphodiester hydrolysis at RNA and, sometimes, DNA linkages, although valiant efforts have extended reactions catalysed by ribozymes to include ionic reactions at carbon centres\textsuperscript{11}. Although the reactions of natural ribozymes are fascinating and impressive, they do not approach the sophistication of the key reactions assumed by the ‘RNA world’ hypothesis.

The ability to explore the repertoire of RNA catalysis dramatically improved with the development of *in vitro* randomization, selection and amplification methods\textsuperscript{12–14}. Ribozymes with new or enhanced activities can be isolated from large libraries of ribozyme variants\textsuperscript{15–20}, and entirely new ribozymes can be isolated from large pools of random-sequence molecules (Table 1). During *in vitro* selection of ribozymes, the desired sequences are enriched on the basis of their ability to modify themselves in such a way...
that they can be separated from the inactive molecules. When examining reactions that do not involve nucleic acid substrates, the desired substrate can be linked to each of the random-sequence-pool molecules, and molecules that contain the desired substrate are selected\(^{22–26}\). Next, the selected molecules are amplified, and the selection–amplification procedure is repeated until sequences with the desired activity dominate the pool. Before the technology of ex vivo selection existed, it was easy to proclaim boldly that RNA could catalyse the reactions required in the RNA world — no one expected experimental verification. However, now the era is not merely to propose a key reaction of the RNA world but also to produce an RNA molecule that can perform such a reaction.

New ribozymes bring important insights into the feasibility of the ‘RNA world’ hypothesis, but the in vito selection approach has limitations. Even if it can demonstrate that there are RNA sequences that can catalyse the reaction in question, the possibility that other ribozymes with different folds could also catalyse the reaction precludes any claims that a particular fold would have been relevant in early evolution. Moreover, it cannot prove that the RNA world ever existed. Even if ribozymes for all the essential activities of an RNA world were generated and assembled into RNA-based life, this would only show that the fundamental properties of RNA are compatible with the ‘RNA world’ scenario. Perhaps most disconcerting is that the in vito selection approach cannot disprove the ‘RNA world’ hypothesis. Only a minute fraction of the possible RNA sequences can be sampled in each experiment and, therefore, a negative result does not mean that the activity is absent among all possible sequences. Despite these limitations, an increasing number of research groups are isolating new ribozymes. For some, the issue of the origin and early evolution of life is such an important question that it must be approached by all possible avenues. Even if the ‘RNA world’ hypothesis cannot be proven, recreating key features of an RNA world would make the hypothesis much more credible, and such efforts do enforce a more rigorous focus on the relevant issues and a deeper appreciation of the complexities of even the simplest imaginable life. An assembly of molecules with life-like properties, although a distant prospect, would be a fascinating tool for studying the basic processes and properties of life. Other experimentalists are primarily interested in the fundamental properties and possibilities of RNA, and how they compare with those of other biopolymers, such as protein. They want to place current biocatalysis in the context of what is possible and would be pursuing new ribozymes even if there was no ‘RNA world’ hypothesis. Others are interested in building the technology for generating enzymes with new or enhanced reaction chemistries, substrate preferences or reaction conditions. A deoxyribozyme selected in vitro has already become a useful research tool\(^1\), and enzymes selected in vitro might eventually find uses as diagnostic, therapeutic or synthetic tools.

With this combination of motives, important elements of the putative RNA world are under construction. At best, these cumulative efforts will resemble an ‘RNA world’ theme park — artificial and fragmented when compared with the real thing, but still well worth a visit. The second part of this article highlights the main attractions under construction: replication, coded peptide synthesis and several metabolic activities presumed by the ‘RNA world’ scenario.

### RNA self-replication

The ‘RNA world’ hypothesis hinges on the assumption that somewhere, among all RNA-sequence possibilities, ribozymes exist that can replicate RNA. The fact that some RNA molecules can catalyse the chemistry of translation is a testament to the potential for RNA to replicate itself. The fact that RNA can catalyse the formation of amide bonds is at least five times higher than that required for RNA polymerisation\(^3\). The ‘RNA world’ hypothesis hinges on the assumption that some of these properties of RNA are compatible with the ‘RNA world’ scenario. Perhaps most disconcerting is that the in vito selection approach cannot disprove the ‘RNA world’ hypothesis. Only a minute fraction of the possible RNA sequences can be sampled in each experiment and, therefore, a negative result does not mean that the activity is absent among all possible sequences. Despite these limitations, an increasing number of research groups are isolating new ribozymes. For some, the issue of the origin and early evolution of life is such an important question that it must be approached by all possible avenues. Even if the
For example, a ribozyme has been selected that forms a peptide linkage between a phenylalanine moiety attached to the ribozyme and any one of several biotinylated amino acids. The biotinylated amino acid was activated by adenosine in a manner analogous to activation of amino acids by tRNA. It would be interesting to explore whether this ribozyme can be a starting point for generating ribozymes capable of coded oligopeptide synthesis. Unlike replicase activity, modern protein synthesis might still harbour remnants of RNA catalysis, suggesting another approach for generating coded peptide synthesis: it might be possible to start with ribosomal RNA sequences and work back towards an RNA-only activity. It is encouraging that, when the RNA of the ribosomal large subunit is stripped of all but a few of its proteins, it still promotes a non-coded peptidyl transfer reaction, which is considered a model for the chemistry of protein synthesis.

RNA-based metabolism

Significant metabolic complexity would have been required for RNA-based life to develop coded protein synthesis sufficient for the transition to protein-based metabolism. Although these metabolic reactions do not present the coding and translocation intricacies of RNA polymerization and translation, they offer a unique set of challenges, typically involving more difficult chemical transformations as well as recognition of small-molecule substrates. The argument that the nucleotides of enzyme cofactors are remnants of the RNA world implies that key reactions involving these cofactors were once catalyzed by RNA. Specifically, ribozymes are presumed to have promoted oxidation and reduction reactions, aldol and Claisen condensations, transmethylations and porphyrin biosynthesis. At least, ribozymes would have had to synthesize any nucleotides, lipids, amino acids and cofactors that could not be scavenged from the environment. Ribozymes would also have had to activate nucleotide and amino acid monomers for polymerization, presumably by taking advantage of energy metabolism based on nucleotide phosphates. Ribozymes promoting several of these classes of chemical transformations have been isolated from random sequences. Ribozymes have been found that aminoacylate themselves. Aminoacylation of tRNA is the second step of the reaction needed to activate amino acids for translation. Other RNAs promote the chemistry required for glycosidic bond formation, a key reaction in nucleotide synthesis. Hydroxyl phosphorylation and acyl transfer reactions are also among the catalytic repertoire of RNA.
as a cofactor during RNA cleavage. A further step would be to show that the use of organic cofactors can extend the scope of RNA catalysis beyond that seen with metal cofactors.

Although some key reactions of the RNA world hypothesis, most notably RNA polymerization, ammoxysylation and translation, involve RNA or derivatized RNA, typical metabolic reactions involve small molecules that are not attached to RNA. Demonstrating that RNA can promote an interesting reaction using a reduced substrate is important but is only the first step in addressing the question of whether RNA can catalyze a reaction involving relevant small molecules. For example, some RNA sequences can phosphorylate an RNA oligonucleotide, raising the question of whether any RNAs could phosphorylate free ribose. Addressing this is challenging because the most productive approach for isolating ribozymes requires that one of the substrates be attached to RNA. In a few cases, the molecule originally attached to the ribozyme can be removed and used as a substrate when added in solution.

Another approach might be to isolate RNAs based on their ability to bind transition-state analogues, but this strategy has not yet generated ribozymes able to form new covalent bonds.

As RNA-based metabolic pathways bifurcated and formed networks, regulation would have become increasingly important in the RNA world. In this regard, it is interesting that allosteric ribozymes have been generated with activities modulated dramatically by the presence of organic small-molecule effectors, such as adenosine and thymine. This evokes a scenario with ribo-organisms able to respond rapidly to environmental changes or internal needs.

References

5. Green, R. and Noller, H.F. (1997) Ribosomes to forms that use free small-molecule substrates might require the development of selection protocols that involve endowment within membranes in order to discriminate between ribozyme variants based on colocalization with their small-molecule products. Another approach might be to isolate RNAs based on their ability to bind transition-state analogues, but this strategy has not yet generated ribozymes able to form new covalent bonds.

As RNA-based metabolic pathways bifurcated and formed networks, regulation would have become increasingly important in the RNA world. In this regard, it is interesting that allosteric ribozymes have been generated with activities modulated dramatically by the presence of organic small-molecule effectors, such as adenosine and thymine. This evokes a scenario with ribo-organisms able to respond rapidly to environmental changes or internal needs.

Millennium issue

David P. Bartel and Peter J. Unrau Early RNA-based life?

work was supported

Acknowledgments

We thank R. Lawrence, E. Schulten, S. Bassendine and other members of the lab for helpful comments on this manuscript. This work was supported by grants from the NIH and NIGMS to D.P.B. and a postdoctoral fellowship from the Medical Research Council (Canada) to P.J.U.
The origin, history, and singularity of our species has fascinated storytellers, philosophers and scientists throughout, and doubtless before, recorded history. Anthropology, the modern-era discipline that deals with these issues, is a notoriously contentious field, perhaps because the topic at hand — the nature of our own species — is one that is difficult or impossible to approach in an unblurred way. Recently, molecular genetics has increasingly contributed to this field. Here, I briefly discuss three areas where I believe molecular studies are likely to be of decisive importance.

**Human evolution**

Svante Pääbo

The origin, history, and singularity of our species has fascinated storytellers, philosophers and scientists throughout, and doubtless before, recorded history. Anthropology, the modern-era discipline that deals with these issues, is a notoriously contentious field, perhaps because the topic at hand — the nature of our own species — is one that is difficult or impossible to approach in an unblurred way. Recently, molecular genetics has increasingly contributed to this field. Here, I briefly discuss three areas where I believe molecular studies are likely to be of decisive importance in the future. These concern the questions of where and when our species originated, what the genetic background for characters that differ between us and apes is, and how the phenotypic traits that vary among human groups have evolved.

Studies of the genetic variation of humans, the concern of the field of molecular anthropology, attempt to produce objective data with which to arrive at new insights about human history. These insights can be of great practical importance, as in the data with which to arrive at new insights about human history. These insights can be of great practical importance, as in the case where we have recently been able to use DNA sequences derived from individuals who lived 10000 years ago to trace the spread of farming from the Near East to Europe. We have also been able to use DNA sequences from individuals who lived 50000 years ago to trace the spread of modern humans throughout the world. These insights have been of great practical importance, as in the case where we have recently been able to use DNA sequences derived from individuals who lived 10000 years ago to trace the spread of farming from the Near East to Europe. We have also been able to use DNA sequences from individuals who lived 50000 years ago to trace the spread of modern humans throughout the world.

Origins of human genetic variation

The questions of where and when our species originated might seem quite straightforward, but, in fact, the definition of the origin of a species is not trivial. However, from a molecular-genetic perspective, it is clear that the DNA sequences found in contemporary individuals have been passed down to them from previous generations. It is also clear that, in every generation, some DNA sequences are not passed on because some individuals have no children or the sequence fails to be transmitted during meiosis. Therefore, the genealogy of a DNA sequence will trace back to fewer and fewer ancestors until it comes together in one common ancestor. To reconstruct this genealogy, the most straightforward approach is to determine DNA sequences from individuals that are distributed such that they represent the entire species. We can then use mathematical techniques to estimate the age of the most recent common ancestor of this collection of contemporary DNA sequences. However, because the genealogy of sequences at different locations in the genome differs owing to recombination and segregation, the age and place of the origin will be different for each genetic locus. Thus, from a genetic perspective, there will not be a single answer to the question of when and where our species emerged. Only if many loci show the same or a similar pattern can one infer that some kind of a population phenomenon occurred, as such an event would affect several parts of the genome.

The mitochondrial genome is the locus for which the most information on DNA sequence diversity in humans is currently available. The great majority of estimates of the age of the deepest divergence among human mitochondrial genomes fall between 100 000 and 200 000 years ago (for a review, see Ref. 1). When a DNA sequence is not trivial. However, from a molecular-genetic perspective, it is clear that the DNA sequences found in contemporary individuals have been passed down to them from previous generations. It is also clear that, in every generation, some DNA sequences are not passed on because some individuals have no children or the sequence fails to be transmitted during meiosis. Therefore, the genealogy of a DNA sequence will trace back to fewer and fewer ancestors until it comes together in one common ancestor. To reconstruct this genealogy, the most straightforward approach is to determine DNA sequences from individuals that are distributed such that they represent the entire species. We can then use mathematical techniques to estimate the age of the most recent common ancestor of this collection of contemporary DNA sequences. However, because the genealogy of sequences at different locations in the genome differs owing to recombination and segregation, the age and place of the origin will be different for each genetic locus. Thus, from a genetic perspective, there will not be a single answer to the question of when and where our species emerged. Only if many loci show the same or a similar pattern can one infer that some kind of a population phenomenon occurred, as such an event would affect several parts of the genome.

The mitochondrial genome is the locus for which the most information on DNA sequence diversity in humans is currently available. The great majority of estimates of the age of the deepest divergence among human mitochondrial genomes fall between 100 000 and 200 000 years ago (for a review, see Ref. 1). When a DNA sequence is