

RNA-based life forms.

When considering forms of life based upon ribonucleic acid (RNA), it is clearly important to firstly establish what exactly is meant by “life form” and secondly to ensure that the term “RNA-based” is well defined. On first thought, one might posit that “RNA-based” translates to an organism whose genome consists solely of RNA. However, if one then goes on to consider RNA’s role not only as a carrier of genetic information, but also as a catalytic agent (in the form of a “ribozyme”), this definition could be considered unnecessarily broad; why not limit the meaning to an organism whose entire being, rather than simply genetic material, consists of RNA? Or conversely, why not open up the definition, as is conventional when referring to viruses, to include organisms that merely use RNA as an intermediate, albeit a compulsory one, of their genetic material in their life cycle? To facilitate a more comprehensive discourse on the subject of RNA-based life, it would be prudent to use the latter definition, as doing otherwise would preclude discussion not only of the hypothetical self-replicating RNA molecules of the pre-biotic world, but also of the multitudinous RNA viruses.

Life is an extraordinarily difficult concept to define. Philosophers and scientists alike have struggled for thousands of years to formulate a satisfactory description of life that is neither too prescriptive nor too broad. Western philosophers from pre-Socratic times, through Aristotle’s *De Anima* to 20th century philosophers such as phenomenologist Michel Henry have all contemplated and contributed ideas to the origin and definition of life. The question also pervaded Eastern philosophy, forming an almost defining facet of the questions and ideas behind Vedic and classical Indian philosophy. Today’s widely accepted scientific definition of life specifies a list of criteria to which a species must conform if it is to be considered alive. Autonomous growth, metabolism (including consumption, transformation and storage of energy/mass and excretion of waste), motion, the ability to reproduce, and the ability to respond to stimuli are the generally accepted criteria. Once again, however, this definition would exclude RNA viruses from the discussion and so, for the purposes of the topic, the definition can be broadened to exclude the necessity for autonomous reproduction and growth. Additionally, the ability to respond to stimuli need not be restricted to one organism, but should be thought of as including adaptive heritable mutations over many generations, i.e. Darwinian evolution.

Both deoxyribonucleic acid (DNA) and RNA consist of a series of nucleotides chemically linked by phosphodiester bonds. Each nucleotide consists of one of four nitrogenous nucleobases, a pentose sugar and a phosphate group. RNA is structurally indistinguishable from

DNA, with the exception of the presence of a hydroxyl group (in place of a hydrogen) on the 2' position of the pentose sugar and a difference in one of the four possible bases. The bases in DNA are known as cytosine, adenine, guanine and thymine, whereas in RNA, thymine is replaced with uracil. These two differences are chemically very significant and are almost certainly the reason why DNA is now the predominant macromolecule for storage of genetic information. The presence of the 2'-hydroxyl group in the RNA sugar-phosphate backbone makes it considerably more susceptible to alkaline hydrolysis. This occurs via a two-step mechanism involving a reversible transesterification, resulting in the formation of a 2', 3' cyclophosphate, which then undergoes an irreversible hydrolysis. In this reaction, RNA's 2' hydroxyl is able to act as an internal nucleophile, whose intramolecular catalysis yields such vast rate enhancements that the energetically unfavourable formation of the cyclophosphate is overcome and the molecule is hydrolysed at a rate that has been estimated to be in the order of 10^6 times faster than DNA hydrolysis under the same conditions.¹

The presence of uracil in place of thymine also makes RNA less suited to long-term storage of genetic information. Uracil is relatively easily aminated to form cytosine and, conversely, cytosine may undergo either spontaneous or induced deamination to form uracil, either way resulting in a mutagenic event. In DNA, deamination of a cytosine base to uracil can be recognized by the enzyme uracil-DNA-glycosylase, which effects a process known as base excision repair, involving removal of the foreign uracil base to leave an abasic site, which is later repaired. In RNA, however, this presents a problem: neither uracil nor cytosine are foreign bases in the molecule. As such, either amination or deamination results in a mutagenic event, which goes entirely unnoticed. However, while this undetectable mutagenesis would be unacceptable for slow-growing multicellular organisms, it may actually be advantageous to populations of unicellular organisms, RNA viruses or even single RNA molecules competing for free nucleotides. An increased mutagenic rate within the population resulting from deamination may give rise to non-viable organisms, but equally may lead to, for example, increased metabolic efficiency, leading to the natural selection of the mutant organism and proliferation of the advantageous mutation throughout the population.

The aforementioned concept of competition between RNA molecules is one of the many ideas that have contributed to the "RNA world hypothesis", a term first used by Nobel laureate Walter Gilbert in 1986 to describe the concept of independent RNA life existing in the pre-cellular world, a world preceding proteins and DNA. The idea itself, however, was formulated earlier than this, being discussed in Carl Woese's book *The Genetic Code* in 1967. Additionally, James

Watson heavily praised Sir Francis Crick's thinking on the matter in 1968, when Crick predicted that RNA could act not only as a carrier of genetic information, but also as a catalyst. This hypothesis was not conclusively proven until the 1980s when Sidney Altman's work on ribonuclease P (RNase P)^{2,3} and Thomas Cech's work on RNA from *T. thermophila*⁴ showed that RNA could catalyse both self-splicing and splicing of other RNA molecules – work that was to win the pair a joint share of the 1989 Nobel Prize in chemistry. The discovery that RNA alone could perform the functions performed by both DNA *and* protein in modern organisms came as a complete surprise to many scientists and was such a major discovery that, according to the Nobel press release in 1989, “Many chapters in our textbooks [will] have to be revised.”

In order for the RNA world hypothesis to gain acceptance as a plausible description of the events that occurred in the period between the solidification of the earth's crust and the appearance of “life”, another question had to first be answered: how did the first RNA molecule arise? Aleksandr Oparin and John Haldane had both independently asked a similar question and devised a model to answer it in the 1920s. The Oparin-Haldane model for chemical evolution hypothesised that organic molecules (amino acids, sugars etc.) may have been spontaneously formed from inorganic molecules under the reducing conditions thought to have been present in the early earth's atmosphere. In 1953, Stanley Miller and Harold Urey conducted experiments that attempted to simulate these conditions based upon the then current understanding of the earth's atmosphere around 4 billion years ago. The now famous Miller-Urey experiment involved placing ammonia, methane, water and hydrogen into a sealed sterile system consisting of an array of glass tubes and flasks organised in a loop. One of the flasks was half full of water, while another contained two electrodes that simulated lightning. The system was cyclically heated and cooled to induce multiple rounds of evaporation and condensation of the water and any solutes dissolved in it.

After one week of continuously running the Urey-Miller system, the results provided evidence that was broadly supportive of the Oparin-Haldane hypothesis. While no nucleotides were found in the resulting mixture, between ten and fifteen percent of the carbon originally in the methane had been incorporated into organic compounds and two percent of the carbon had been incorporated into amino acids, the monomers that polymerise to form proteins. Indeed, 13 of the 20 proteinogenic amino acids were found in the resulting mixture. With regard to RNA, various later experiments, starting with those conducted by Joan Oró in the 1960s, showed that the essential building blocks of RNA including ribose, phosphate, purines (adenine and guanine) and pyrimidines (cytosine, thymine and uracil) may well have been available in the prebiotic

chemical mix. Adenine, for example, can be synthesised abiotically from ammonia and hydrogen cyanide⁵ and cytosine may be synthesised in surprisingly high yields from cyanoacetaldehyde in a concentrated urea solution.⁶

Despite the promising results of the various abiogenesis (i.e. from non-biological origins) experiments, it is almost certainly not the case that RNA molecules as we know them today appeared spontaneously from inorganic molecules. It has been proposed that a pre-RNA world existed before the RNA world but after the world of pre-biotic chemistry, which is where the results of experiments in the Miller-Urey genre are most applicable. Even if it were the case that all of the necessary RNA precursors were present at adequate concentrations to form nucleotides and subsequently polynucleotides, it would be highly unlikely that there would be any regular structure to the resulting complexes. The problem arises from the vast number of structural permutations that could result from a mixture of the components that constitute RNA. For example, there could be multiple phosphate groups joining the riboses together, Dextro- or Levo-stereoisomers and α or β mutarotative forms of ribose and 2', 5'-, 3', 5'- or 5', 5'- phosphodiester linkages between the riboses. The problem is further exaggerated by the much larger array of various nucleotide and ribose analogues that would be present in the prebiotic mix. The scope for variability (and hence complexity) within such RNA makes it difficult to envisage any mechanism for replication that could arise by chance and yet maintain any possible sequence information stored within the heterogeneous structure.

Armed with this knowledge, there are two possible schools of thought as to what occurred following the formation of the various organic compounds. The first is that a series of conditions arose that favoured the production of "standard" nucleotides (i.e. β -D-nucleotides) from the prebiotic mixture, or at least favoured their polymerisation should they be formed at low concentration. The second hypothesis is that a genetic system prior to RNA effectively brought some order to the prebiotic chemicals and then later led onto the RNA world via a mechanism possibly similar to today's transcription. Both of these ideas are very much valid and each one has its merits, but it is the second hypothesis that has shown most promise in recent years, despite relatively little chemical knowledge of potential RNA precursors from metabolic pathways. A whole series of potential nucleic acid analogues have been investigated⁷ including various sugars and linkages to form the backbone. Promising candidates have included threose nucleic acid (TNA), peptide nucleic acid (PNA), pyranosyl RNA and glycerol-derived nucleic acid analog.⁸

Of the possible RNA precursors, TNA is probably the most promising in terms of both the relative likelihood of its formation from the prebiotic mixture and the ease with which a transition to RNA might have been achieved. TNA consists of α -L-threofuranosyl sugars linked by 3', 2'-phosphodiester bonds.⁹ The probability of TNA spontaneously forming from its components is much greater than that of RNA, as threose is one of just two aldotetroses, each of which has only four stereoisomers and may only be linked at the 2' and 3' positions. Single-stranded (ss) TNA may form a double strand with either another ssTNA molecule or with RNA via standard Watson-Crick base pairing. The possibility for a TNA:RNA hybrid double helix makes it easy to envisage a transcription-like mechanism for retention of genetic information in the hypothetical transition between TNA and RNA. One problem with such a conversion would be that any catalytic activity present in the RNA precursor would not be expected to carry over to the RNA. However, the RNA may retain some catalytic activity due to features of its secondary structure conserved across the transition or alternatively, the RNA precursor may have been able to catalyse reactions involving the RNA, resulting in a molecular symbiosis.

Assuming that the death of any given organism is inevitable, reproduction, whether autonomous or not, is a necessary capability of the species to which it belongs if that species is to persist. Given this criterion, it must follow that in order for the first RNA molecules to persist, they must either have themselves possessed the ability to replicate or been replicated by a "third-party", as proposed above with the possibility of a putative RNA precursor providing replicative ability to the RNA. Before further investigating the catalytic capabilities of RNA, it would be advantageous to first lay down a formal definition of self-replication. Barry McMullin proposed such a definition in 2000 with regard to von Neumann machines¹⁰, but it is both straightforward and convenient to adapt it for use in a biological context. If we take r to be an RNA molecule, $O(r)$ to be a set of RNA molecules that r is capable of producing (the "offspring" of r), r to be a member of the set ψ ($r \in \psi$) and $O(r)$ to be a proper subset of ψ ($O(r) \subset \psi$), then r is capable of self-reproduction if $O(r) \neq \emptyset$ and $r \in O(r)$. This last condition is really the defining aspect of self-replication – the set that comprises the offspring of r contains r itself.

Taking this definition of replication, a self-replicating RNA molecule, known as R3C, has in fact been created in the laboratory through a process of directed nucleic acid evolution.¹¹ However, the mechanism of replication is not, as one might first expect, a template directed nucleotide-by-nucleotide polymerisation, but is rather a (considerably simpler) ligation of two substrate RNA molecules. Before more closely investigating the implications and mechanism of this intriguing molecule, it would be instructive to first briefly explore the catalytic arsenal of

ribonucleic acids. In eukaryotic cells, RNA has three separate roles in the process of translation alone (the production of protein from an mRNA transcript); it acts as an encoder (mRNA), a translator (tRNA) and as a facilitator of the translation (rRNA).¹² Each of these scenarios involves some RNA catalysis, but *in vivo*, this has thus far been found to be limited to phosphoester transfer, phosphoester hydrolysis and peptide bond formation. However, *in vitro* directed evolution experiments, similar to those used to create R3C, have created ribozymes capable of polynucleotide phosphorylation and ligation, mononucleotide polymerisation and amide bond formation and cleavage, amongst others.

Such catalytic diversity is impressive for a polymer that contains only four subunits which are, when compared with the diversity found in amino acid side chains, chemically quite similar. Ribonucleic acid monomers lack, for example, a sulphhydryl group (as found in cysteine), a carboxylate group (as in aspartate) and even a general acid base with a neutral pK_a (as found in histidine). Indeed, the functional groups of ribonucleosides don't ionise to any significant degree between pH 5 and pH 9, making them unlikely to accept or donate protons at a neutral pH – one of the basic mechanisms of catalysis. Furthermore, the RNA monomers have so far proved to be incapable of providing charge stabilisation to an electronegative transition state, as none of the functional groups carries a positive charge at neutral pH.¹³

Despite these apparent shortcomings, research into protein metalloenzymes has shown that metallic cation cofactors could more than compensate for the lack of catalytic competence found in the ribonucleosides. Metal ions are able to withdraw electrons from ribose or water (facilitating their deprotonation) or, for example, act as Lewis acids, stabilising negatively charged leaving groups after separation from phosphorus in the RNA backbone. Hexacoordinated metal ion complexes (such as $[Mg(H_2O)_6]^{2+}$) are able to act either as general acids, donating a proton to an oxyanion leaving group or general bases, accepting a proton from an oxygen nucleophile, for example. Because of this clear potential for enhancing catalysis and their known role in stabilising RNA tertiary structure, it was thought for a long time that metallic cofactors were required by all ribozymes, although this has now been shown not to be the case. The hairpin ribozyme, for example, maintains activity whether in the presence of Mg^{2+} or the “exchange-inert” metal complex cobalt hexamine – $[Co(NH_3)_6]^{3+}$.¹⁴ Also, directed evolution has conferred dependence upon adenosine triphosphate (ATP) and flavin mononucleotide (FMN) to the class I ligase ribozyme. However, it is still true to say that the vast majority of naturally occurring ribozymes require (mostly divalent) metal cations such as Mg^{2+} or Mn^{2+} . Dependence upon such cofactors is even more pronounced if the ribozyme in

question is composed of unaltered nucleotides with no functional substituents.

RNA, when accompanied by these metal cofactors, clearly has the potential to be a competent catalyst for many reactions, even if those catalysed by the naturally occurring ribozymes are limited to peptide bond formation and two variations of phosphodiester transesterification. Considering this catalytic versatility and, in some cases, ability to self-replicate, would it be accurate to say that a ribozyme, like R3C, is a living molecule? If we refer back to the list of criteria an entity must satisfy in order to be considered “alive”, R3C, thus far probably the most promising candidate for a single “living” molecule, doesn't meet all the requirements. Its replicative strategy of ligating two shorter molecules rather than addition of single nucleotides means that at no stage can the ribozyme said to be growing. Nor does R3C metabolise or undergo Darwinian evolution. While any of the five domains that comprise the ribozyme can be experimentally mutated with retention of function, the progeny molecules do not then give rise exclusively to additional copies of themselves. The necessity for addition of mutagenised ribozyme components, evolved separately from the self-replicating RNA reactions, also somewhat nullifies any conclusions that might otherwise be drawn about Darwinian evolution of the ribozyme.

Although a single self-replicating ribozyme might not be described as living, there is still much scope for finding an RNA-based entity that can. For example, a hypothetical system consisting of multiple ribozymes interacting in a series of cross-catalytic reactions could be described as living. Such a system could occur in a heterogeneous mix of ribozymes with different functions (polymerisation, phosphorylation, ligation etc.) or it could be more complex, with each RNA molecule hydrogen-bonding to others through specific binding motifs. Under certain conditions, the ribozymes could even accumulate within coacervates – spherical aggregates of lipids held together by hydrophobic forces, which Oparin proposed may have been present in the primordial soup. However, if we leave the requirement of independent growth and replication behind, there is a whole world of RNA-based life in existence today in the form of viroids¹⁵ and RNA viruses.

Viroids are a fascinating group of small, single-stranded, circular (although covalently closed) RNA molecules, which have no form of encapsidation or protective coating, as would be found surrounding an RNA virus. They are obligate parasites responsible for causing infectious diseases in certain plants, in which the viroids are capable of autonomous replication. Incredibly, the viroids possess no functional open reading frames (ORFs) – that is to say, they

encode no protein and, as such, are classified as non-coding RNAs. However, non-coding does not mean the genome is non-functional. Quite the opposite: the viroid genome (between 246 and 401 bases long) performs an extremely large range of functions, including the ability to subvert host polymerases for self-replication and to induce specific pathogenic effects in the host.¹⁶ Studies from the past two decades have, in fact, revealed that almost every nucleotide in a viroid is functional and under active selection. Functionality beside that conferred by the primary sequence of viroids is found in the secondary structure which, in the *Avsunvioidae* family, folds to form a Hammerhead ribozyme with the ability to self-cleave replicative RNA intermediates of both polarities.¹⁷

Viroids have been described as “RNA parasites that have eliminated all of the dispensable ‘extravagance’ of viral pathogens to make use of the host in a more sophisticated manner.” Whether or not this statement accurately represents the origin of viroids, it is definitely overly dismissive of the capabilities of RNA viruses, which are, in fact, considerably more diverse than viroids – particularly of note is the ability of RNA retroviruses, such as HIV, to integrate into the host genome. Viruses can most simply be defined as obligatory intracellular parasites consisting of viral nucleic acid (either RNA or DNA) surrounded by a protein coat known as the virion. There are five major classes of RNA virus: dsRNA, positive strand ssRNA, negative strand ssRNA, positive strand ssRNA with DNA intermediate and dsDNA with RNA intermediate. These various subtypes differ not only in the molecule used to store their genetic information, but also in the replicative apparatus that they subvert in the host. The debate as to whether or not viruses are living has raged since they were first discovered in 1892 by Dimitri Ivanovski. While the argument continues, the emerging consensus is that viruses are somewhere between living and non-living. They have heritable genetic material which, upon infection of a host, is reproduced using parts of the host’s replication machinery. Those viruses that encode their own polymerase could be said to “metabolise” nucleotides, and all viruses are effectively able to respond to the stimulus of host cell attachment, at which point the virus injects its genetic material into the cytosol. Additionally, viruses are able to exert considerable change in the host cell by, for example, shutting down protein synthesis, stimulating host cell division, promoting cell fusion or interfering with the host’s immune response.

In conclusion, we have seen that RNA plays a large and diverse range of roles in modern day prokaryotes, eukaryotes, viruses and viroids, with substantial evidence to suggest that it once also took on the significant and impressive role of bridging the pre-biotic world and the world dominated by protein and DNA that we know today. The recurring theme in tackling the

question of the existence of RNA-based life is the problem of defining life itself. If we look at the essential constituents (notably protein and DNA) of an organism whose status as “living” is undisputed, and compare the chemical and structural capabilities of these constituents with the various abilities of RNA, we find them to be roughly comparable. With everything from self-replicating RNA molecules through RNA viruses up to eukaryotic cells relying upon RNA as a genetic intermediate and catalyst, it would be unreasonable to conclude by saying that RNA-based life is purely hypothetical.

On tackling the question of living matter as opposed to non-living matter, is it fair or, more importantly, correct to say that there is a well-defined line between living and non-living or is the boundary perhaps more “fuzzy”? Is one collection of matter wholly “alive” and another one “dead” or is there a graduated scale between the two extremes? I am inclined to say that the latter is the more accurate representation, especially when we consider the scope of entities that possess qualities readily identified as living which, despite this, are still classified as non-living. If we work up from fundamental biological units say, for example, adenosine through a short polynucleotide strand, the R3C self-replicating ligase all the way up to the HIV retrovirus it is clear that each of these entities has varying degrees of characteristics associated with life. I therefore consider it unnecessary and perhaps even fallacious to pigeonhole any one of these entities into the blunt categories of living or non-living. If we utilise such a graded system in defining life, then, based on the presented evidence of entities that are wholly or in part constructed from RNA, it would be judicious to conclude that RNA-based life forms have been in existence for millions of years. Some of them more alive than others.

¹ Loverix S. *et al.* Reconsidering the energetics of ribonuclease catalysed RNA hydrolysis. **Eur. J. Biochem** 257, 286-290 (July 8, 1998).

² Guerrier-Takada C, Gardiner K, Marsh T, Pace NR, Altman S. The RNA moiety of RNase P is the catalytic subunit of the enzyme. **Cell** 35, 849-857 (1983).

³ Stark BC, Kole R, Bowman EJ, Altman S. Ribonuclease P: an enzyme with an essential RNA component. **Proc. Natl. Acad. Sci.** 75, 3717-3721 (1978).

⁴ Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. **Cell** 31, 147-157 (1982).

⁵ Oró, J. Mechanism of synthesis of adenine from hydrogen cyanide under possible primitive earth conditions. **Nature** 191, 1193-1194 (Sep 16, 1961).

⁶ Robertson MP, Miller SL. An efficient prebiotic synthesis of cytosine and uracil. **Nature** 377, 257 (Sep 21, 1995).

⁷ Eschenmoser, A. Chemical etiology of nucleic acid structure. **Science** 284, 2118-2124 (1999).

⁸ Joyce, GF. The antiquity of RNA-based evolution. **Nature** 418, 214-221 (July 11, 2002).

⁹ Schöning, K.-U. *et al.* Chemical etiology of nucleic acid structure: the α -threofuranosyl-(3' \rightarrow 2') oligonucleotide system. **Science** 290, 1347-1351 (2000).

¹⁰ McMullin, B. John von Neumann and the Evolutionary Growth of Complexity: Looking Backwards, Looking Forwards... **Proceedings of Artificial Life** Vol. 6, 347-361 (2000).

¹¹ Paul N, Joyce GF. A self-replicating ligase ribozyme. **Proc. Natl. Acad. Sci.** 99, 12733-12740 (October 2002)

¹² Darnell JE, Doolittle WF. Speculations on the early course of evolution. **Proc. Natl. Acad. Sci.** 83, 1271-1275 (March 1986).

¹³ Cech TR and Bass BL. Biological Catalysis by RNA. **Ann. Rev. Biochem.** 55, 599-629 (1986).

¹⁴ Hampel A and Cowan JA. A unique mechanism for RNA catalysis: the role of metal cofactors in hairpin ribozyme cleavage. **Chem. Biol.** 4, 513-517 (1997).

¹⁵ Diener TO. Potato spindle tuber 'virus'. IV. A replicating, low molecular weight RNA. **Virology.** 45, 411-428 (1971).

¹⁶ Tabler M and Tsagris M. Viroids: petite RNA pathogens with distinguished talents. **Trends in Plant Science.** Vol. 9 No. 7, 339-348 (July 2004).

¹⁷ Fadda P, Daròs JA, Fagoaga C, Flores R and Durán-Vila N. Eggplant Latent Viroid, the Candidate Type Species for a New Genus within the Family Avsunviroidae (Hammerhead Viroids). **J. Virol.** 77, 6528-6532 (2003).