ISSN 1669-9106

SPECIAL ARTICLE MEDICINA (Buenos Aires) 2016; 76: 199-203

ON THE ORIGIN OF LIFE

JACK W. SZOSTAK

Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA, Massachusetts General Hospital, Boston, MA, USA

Abstract The origin of life is a very rich field, filled with possibilities and ripe for discovery. RNA replication requires chemical energy and vesicle division is easy to do with mechanical energy. These requirements point to a surface lake, perhaps at some time following the period of concentrated cyanide chemistry that gave rise to nucleotides, amino acids and (maybe) fatty acids. A second requirement follows specifically from the nature of the RNA replication cycle, which requires generally cool to moderate temperatures for the copying chemistry, punctuated by brief periods of high temperature for strand separation. Remarkably, lakes in a geothermal active area provide just such a fluctuating temperature environment, because lakes similar to Yellowstone can be generally cool (even ice covered in winter), but they contain numerous hydrothermal vents that emit streams of hot water. Protocells in such an environment would occasionally be swept into these hot water streams, where the transient high temperature exposure would cause RNA strand separation. However, the protocells would be quickly mixed with surrounding cold water, and would therefore cool quickly, before their delicate RNA molecules could be destroyed by heat. Because of the combination of favorable chemical and physical environments, this could be the most likely scenario for the early Earth environment that nurtured the origin of life.

Key words: RNA replication cycle, protocells, ribozymes, geothermal activity, concentrated cyanide

Resumen ***Sobre el origen de la vida.*** El origen de la vida es un campo lleno de posibilidades, listas para ser descubiertas. Basados en lo conocido sobre modelos de sistemas de membranas y sobre ARN, se comienza a deducir algunas características necesarias del entorno inicial. La replicación del ARN requiere energía química y la división de la vesícula es fácil de hacer con la energía mecánica. Estos requisitos apuntan a la superficie de un lago, en algún momento después del período en que la química del cianuro concentrado dio origen a los nucleótidos, aminoácidos y (tal vez) ácidos grasos. Un segundo requisito surge de la naturaleza del ciclo de replicación del ARN, que requiere temperaturas moderadas para la química de la copia, interrumpidas por breves períodos de alta temperatura para la separación en hebras. Solo lagos en una zona de actividad geotérmica proporcionan un ambiente de temperatura tan oscilante, lagos similares a Yellowstone pueden ser frescos (cubiertos de hielo en invierno), pero contienen numerosas fuentes hidrotermales que emiten chorros de agua caliente. Las protocélulas, en un ambiente así, de vez en cuando serían barridas en estas corrientes de alta temperatura, que podrían causar la separación transitoria de ARN de cadena. Pero las protocélulas serían mezcladas con rapidez en la zona de agua fría, y enfriarse antes de que sus delicadas moléculas de ARN fueran destruidas por el calor. La combinación de estos ambientes químicos y físicos favorables serían el escenario más probable del medio ambiente de la Tierra temprana que nutrió el origen de la vida.

Palabras clave: ciclo de replicación de ARN, protocélulas, ribozimas, actividad geotérmica, cianuro concentrado

There are three fundamental Origins questions – the Origin of the Universe, the Origin of Life and the Origin of the Mind and Consciousness. All of these have been debated for thousands of years, and all are now the subject of serious scientific research. To me, the first and third seem too difficult: I wouldn't know where to begin to search for a solution – but the middle question, the Origin of Life, is a very rich field, filled with possibilities and ripe for discovery. We now have the tools to study this problem, and indeed new advances are being made every year.

What exactly do we mean by an answer to the question of the Origin of Life? After all, we can't go back in time to the early Earth and watch the process unfold, so we may never know for sure precisely how life evolved here on Earth. In that sense, is this even a valid field of scientific inquiry? My answer is that what we are after, to have a scientific understanding of the Origin of Life, is understanding a reasonable series of steps that would provide a pathway going all the way from planet formation through simple and then more complex chemistry leading to the first simple life forms, and then on through the evolution of modern life. Moreover, we don't want just vague theories, we want detailed and experimentally validated mechanisms that will explain all of the steps on this long pathway. It may be that there will be many possible pathways to life – but at present, we don't have even one fully connected pathway because there are many gaps in our understanding. As scientists, this is not a bad thing or a

reason to give up, rather this is a wonderful opportunity, a rich mine of interesting puzzles to solve. The fact that there are so many interesting discoveries just waiting to be made is what makes this such an interesting field to me, my students and my colleagues.

Before jumping into the scientific story, I would like to mention another huge recent advance which has provided an important context for Origin of Life studies, and has stimulated enormous public interest in this old question. This of course is the discovery by teams of astronomers of thousands of exoplanets – planets orbiting other stars. The observational evidence that our galaxy is swarming with planets immediately raises the question of whether there is life out there, or whether our Earth is alone and unique in hosting life. The answer to this question depends upon how easy or hard it is for life to emerge from the chemistry of a young planet. If planets with the right environmental conditions are common, and the path to life is a series of steps all of which are simple or high probability, then life could be abundant in the cosmos. On the other hand, if there are several difficult, low probability events on the path to life, or even one extremely difficult step, then life could be very rare in the universe – in the extreme, our planet could be the only place in the Universe with life. At this point we simply do not know. However, this is where the astronomical studies intersect with laboratory studies: if evidence for life on a distant planet is found, then the fact that life emerged twice independently would mean that there cannot be any incredibly difficult step in the Origin of Life – implying that it might be easy for us to reconstruct a pathway to life in the laboratory. Conversely, if our work in the lab suggests that all of the steps from geochemistry to life are easy, the implication would be that life should exist on other planets, and that we should be confidant in moving forward with the search for life elsewhere in the Universe.

Since there is so much interest in the question of how life emerged, why has the problem not been solved already? I think that one reason for this is the complexity of modern life. For example all cells have a complex structure composed of many closely interacting parts, which are themselves complex. Underlying this structural complexity is an enormous biochemical complexity – hundreds or thousands of chemical reactions all catalyzed by complex protein machines called enzymes. And underlying this metabolic complexity is the machinery that guides the flow of information in cells, from the information archived in DNA, to the transcribed intermediate mRNA, to the translated proteins, and finally to the biochemical level.

The remarkable and confusing thing about this flow of information is that every partof the machinery depends upon every other part. For example, you need RNA, proteins and metabolites to replicate DNA, and you need DNA, proteins and metabolites to make RNA, and so on. How could such a self-referential system ever have evolved? Many bizarre theories were put forth in an attempt to deal with this conundrum, but little progress was made for decades. In the late 1960s, three brilliant scientists, Francis Crick, Leslie Orgel and Carl Woese, realized that a potential solution lay in the central position of RNA (between DNA and proteins). Being chemically similar to DNA, it was realized that RNA could store and transmit information, and the complex folded 3-D structure of tRNA was reminiscent of folded proteins, suggesting that RNA might also, like proteins, be able to act as an enzyme and catalyze chemical reactions. In that case, perhaps life could have started with RNA alone. However, no one took this proposal seriously until some 15 years later when Tom Cech and Sid Altman discovered examples of catalytic RNAs, or ribozymes. This revolutionary discovery led to a new conceptual model for the Origin of Life: very simple early cells, before the emergence of DNA and proteins, would have had an RNA genome and used RNA enzymes for metabolism and replication. This primitive stage in the evolution of Life, now known as the RNA World, vastly simplifies the search for an explanation of the Origin of Life, which can now be understood as a search for a pathway from prebiotic chemistry to simple RNA based cells. All of the complexity of modern life, including DNA and proteins, can be seen as emerging later, as a result of an extended period of Darwinian evolution.

How then, can we bridge the gap between the chemistry of the young Earth, and the beginnings of Life? Fortunately we can break down this over-arching question into three more specific questions, as follows: 1) what was the relevant pre-biotic chemistry that led to the availability of the basic building blocks of biology?, 2) given the right sets of molecules, how were the first cells assembled, and how did they grow, divide and evolve?, and 3) what were the geological settings for the first two processes? Fortunately each of these questions can in turn be broken down into smaller more manageable research projects, and indeed, considerable progress has been made in all three areas in the past 10-15 years.

Let's begin with a very brief review of recent advances in prebiotic chemistry, a field which has experienced something of a renaissance in recent years. There has been a new focus on chemistry that is channeled into a few products formed in high yields, as opposed to earlier processes that tended to give thousands of compounds in low yields. Ironically, the ideal starting material for this chemistry appears to be cyanide, which is relatively easy to make in the atmosphere (at least, the atmosphere of the young earth, as a result of lightning, UV or impacts, with no free oxygen). Cyanide stores a lot of energy in its carbon-nitrogen triple bond, and it is moderately reactive. However, if the cyanide formed in the atmosphere just rains out into the ocean, forming a very dilute solution, the cyanide would simply slowly hydrolyze to ammonia and formate. How then can cyanide be stored and used? A

very interesting hypothesis has been put forward by Prof. John Sutherland, to explain how cyanide might accumulate over long periods of time in a stable reservoir. The idea involves a lake, perhaps something like Yellowstone Lake now, in which sub-surface magma provides heat that drives ground water circulation. As water circulates through fractured rock, it leaches metal ions, which are brought to the surface lake waters through hydrothermal vents. The resulting metal ions, especially iron, quickly react with cyanide brought to the lake from rainwater and streams, forming a stable complex known as ferrocyanide. Certain salts of ferrocyanide are quite insoluble, and thus would precipitate, building up layers on the floor of the lake, which could accumulate over thousands of years as a reservoir of cyanide in a stable form. Subsequent drying of the lake, along with increased heat from below (or perhaps from above, from a meteorite impact), would transform these ferrocyanide salts into a series of more reactive compounds through well-known chemistry. Finally, once the environment returned to a wetter phase, these compounds would react with water to give a concentrated solution of all the key starting materials needed for the synthesis of amino acids, nucleotides, sugars and maybe even lipids. This bold new idea thus provides a geological and chemical scenario for forming a very concentrated mix of starting materials that could yield all the needed building blocks of biology.

Given a scenario in which the synthesis of the necessary building blocks looks plausible, we can now ask how these molecules might assemble into the first simple cells. Our model of a primitive cell, or protocell, is a stripped down version of a modern cell: we imagine a simple cell membrane surrounding the cell contents, which would include RNA or RNA-like molecules that can both replicate, and so transmit information from generation to generation, as well as fulfilling simple biochemical roles, such as a primitive form of metabolism. It turns out that very simple molecules such as fatty acids can spontaneously self-assemble into sheet-like membranes, similar to the membranes of modern cells but with some important differences. For example, these primitive membranes allow ions and molecules to cross the membrane without help from the complicated protein machines that control molecular movement across modern cell membranes. As these membrane sheets form, they eventually close up to form spherical shells that trap water and anything in the water, such as RNA, peptides, and small molecules, on the inside. As a result, simply forming cell like structures is pretty easy – the harder and more interesting questions concern growth, division, and the replication of the genetic molecules. How could these processes have occurred?

Because the first cells were so simple, and there was by definition no evolved cell machinery at the origin of life, we think that a rich and complex environment must have driven growth and division. All of the chemical building blocks of the cell must have been supplied by the environment, as a result of the prebiotic chemistry discussed above. In addition, the environment must have supplied the necessary energy to drive growth and division. For example, chemical energy for RNA replication could come from high energy compounds, possibly derived from cyanide. Simple mechanical energy, e.g. from waves on the lake, could have played a role in division. We would like to understand all the ways that the early environment could have provided the molecules and the energy required for primitive cell growth and division.

Before discussing our experiments with models of primitive cells, or protocells, I would like to address the question of why compartments and membranes are necessary at all? After all, wouldn't it be simpler to just have RNA molecules floating in solution, replicating with no membrane as a barrier blocking their access to environmentally supplied nucleotides? The answer is that some form of compartmentalization is necessary in order for Darwinian evolution to work. Imagine an RNA enzyme that catalyzed RNA replication: in solution, it would simply replicate unrelated RNAs, but in a replicating vesicle it would replicate closely related molecules. The flip side of this argument is that parasitic RNA molecules will be segregated away from active RNAs during the division of protocells, so they don't poison the whole system. However, in solution, parasites inevitable arise and can outcompete active molecules.

In addition to these arguments, primitive membranes are simple self-assembling systems that are remarkably similar to modern cell membranes, making it easier to see how protocells could evolve into modern cells through a series of gradual changes. Continuing on from work by Deamer, Luisi, and others, we think that primitive cell membranes were made from simple molecules such as fatty acids (one example is oleic acid, which comes from olive oil). Very beautiful spherical vesicles form simply by shaking fatty acids in water with a bit of salt, near neutral pH (not too acidic, not too basic). These vesicles grow simply by adding more fatty acids, which can be done in a variety of ways. A few years ago, Ting Zhu, then a graduate student in my lab, discovered an amazing aspect of this vesicle growth. Through careful experiments and observations, documented by video microscopy, Ting showed that fatty acid vesicles could grow into long filamentous structures, without having any of their contents leak out. The resulting filamentous vesicles were quite fragile, and could be made to divide, with no loss of contents, simply by gentle shaking. Thus, by adding new fatty acid molecules in the right way, we could drive repeated cycles of vesicle growth and division. This system mimics, in a very simplified way, the repeated growth and division of living cells – but only with respect to the cell membrane of course! The important point is that the cycle of growth and division is controlled by changes in the environment, e.g. addition of new 'food' molecules for growth, and periodic agitation for division.

Thus, the first cells would not have required any evolved biological machinery to enable growth and division.

Let's go back to RNA, the genetic and functional molecule in our hypothetical protocell. As noted above, when vesicles first form, they encapsulate any molecules, including RNA molecules that happen to be floating in the solution. How would such RNA molecules assemble in the first place? It turns out that it is not difficult to persuade activated nucleotides to join together into the strings of nucleotides we call RNA. This kind of polymerization can be assisted by certain mineral surfaces, or even simply by freezing the solution (this works because when a solution freezes, and pure ice crystals start to grow, the dissolved molecules become highly concentrated in between the ice crystals, which helps them to react with each other). The more difficult problem, and therefore the problem that we have been concentrating on in my lab, is how to replicate these strands of RNA without enzymes. This is a long standing and difficult problem that has been worked on by scientists such as the late Leslie Orgel and his former student Gerald Joyce since the 1970s and '80s. After an initial period of rapid advances, further progress stalled and it began to seem almost impossible that RNA strands could be replicated without enzymes. As a result, the emphasis in many labs, including my own, shifted to studies of RNA-catalyzed RNA replication, in other words, the search for an RNA enzyme or ribozyme with RNA polymerase activity. A ribozyme that could copy its own sequence would be an RNA replicase, and this concept lies at the heart of most models of primitive cells in the RNA World.

As attractive as the idea of RNA-catalyzed RNA replication is, we believe that a simpler and purely chemical process must have preceded this more sophisticated mechanism. Over the last few years, we have returned to this focus on chemical, i.e. nonenzymatic, RNA replication. The difficulties with using chemistry to copy RNA sequences, and thus to allow cycles of RNA replication within replicating vesicles, come down to a list of about eight distinct problems, that can be considered separately. Without getting too technical, I'll briefly summarize our recent progress. One problem that bothered us from the beginning was that previous efforts to chemically replicate RNA resulted in a rather messy heterogeneous backbone structure – in other words, the individual nucleotides were not all joined together in the correct manner. This seemed like a big problem, because without a uniform backbone, we thought it would be impossible for RNA to reproducibly fold into complex 3-D shapes required for enzymatic activity. Much effort had been expended over the years in trying to find conditions that would lead to a uniform backbone, without much success. However, recent work done in my lab by Matt Powner and Aaron Engelhart showed that the RNA backbone is so flexible that the messy backbone resulting from chemical copying was not actually a problem. In fact, folded RNA structures could form correctly despite the presence of a large fraction of incorrect backbone linkages. This was a considerable surprise, but we were of course delighted to realize that what had been thought to be a potentially a fatal flaw with RNA was not so much of a problem after all. Even more remarkably, it turns out that this heterogeneous RNA backbone could even be an advantage! The reason for this is that following the copying of a single stranded RNA molecule, the product is a duplex, a two-stranded double-helix. In order for the next round of copying to begin, the two strands must be separated from each other. This can be accomplished by brief heating for DNA, but not for RNA – unless the backbone contains a fraction of the 'incorrect' linkages that result from chemical copying. Thus, our thinking on this problem has completely reversed – what we thought was a terrible problem that required a solution turns out to be a big advantage of RNA!

Another problem with the chemical copying of RNA is that, so far at least, the chemistry requires high concentrations of metal ions such as magnesium. Unfortunately, these magnesium ions are very disruptive to our fatty acid membranes, causing the vesicles to break down and release their contents. As a result, we were unable to do any RNA copying chemistry on RNA molecules that were inside fatty acid vesicles. In order to assemble a complete replicating protocell, we needed to find a way to make the chemistry of RNA copying compatible with the survival of the protocell membrane. Former graduate student Kate Adamala took on this problem, and discovered a simple 'proof-of-principle' solution to the problem. We already knew that molecules that would bind to magnesium ions and completely surround them would protect membranes, but they would also block the RNA copying chemistry. What Kate found was that citric acid would also bind to magnesium, but would only cover up about half of its surface. As a result, the membranes were protected, but the magnesium ions could still interact with RNA and assist with the copying chemistry. This enabled Kate to, for the first time, demonstrate RNA copying inside fatty acid vesicles. Those experiments were a big advance towards the assembly of a complete model protocell, and they have given us the confidence to move forward, because now we know that if we solve the remaining problems with nonenzymatic RNA replication, we will be able to combine RNA and vesicle replication since the systems can be made to be compatible. I referred to this solution as a proof-of-principle solution, since we do not think that citric acid is a prebiotically realistic way to solve the problem. However, knowing that we have one solution means that we can continue to explore and search for other solutions – eventually hoping to find one that is simple and robust enough to work on the early Earth environment.

There are a number of other problems that need to be solved before we can demonstrate the efficient and

general copying of RNA sequences within replicating protocells. However, we and others continue to make progress in removing these roadblocks, which makes it interesting and fun to speculate about particular environments on the early Earth that could have fostered the growth of the first populations of protocells. Based on what we know about our model membrane systems, and our work with RNA, we can begin to deduce some of the necessary features of such an environment. We think that the environment must have been able to provide a concentrated mixture of the right starting materials, as well as multiple sources of energy. The assembly of membranes, and polymerization of RNA, both require high concentrations of the corresponding building blocks, while RNA replication requires chemical energy and vesicle division is easy to do with mechanical energy. These requirements point to a surface lake, perhaps at some time following the period of concentrated cyanide chemistry that gave rise to nucleotides, amino acids and (maybe) fatty acids. A second requirement follows specifically from the nature of the RNA replication cycle, which requires generally cool to moderate temperatures for the copying chemistry, punctuated by brief periods of high temperature for strand separation. Remarkably, lakes in a geothermal active area provide just such a fluctuating temperature environment, because lakes similar to Yellowstone can be generally cool (even ice covered in winter), but they contain numerous hydrothermal vents that emit streams of hot water. Protocells in such an environment would occasionally be swept into these hot water streams, where the transient high temperature exposure would cause RNA strand separation. However, the protocells would be quickly mixed with surrounding cold water, and would therefore cool quickly, before their delicate RNA molecules could be destroyed by heat. Because of the combination of favorable chemical and physical environments, we think this is the most likely scenario for the early Earth environment that nurtured the Origin of Life.

Working in this field is a great pleasure for many reasons, not least because of the frequent stimulating discussions with colleagues both across the world and within my own laboratory. We all enjoy the thrill of discovery, and the opportunity to learn about new areas of science from astronomy and planetary science to chemistry and biology. I look forward with great optimism to continued scientific advances in the study of the Origin of Life, and hope in the future to be able to describe a comprehensive pathway leading to the appearance of the first cells on the early Earth.

Conferencia: El Origen de La Vida

Jack W. Szostak. Premio Nobel de Medicina 2009.

El 12 y 13 de mayo de 2016, se realizó en el Aula Magna de la Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina, el Simposio Internacional Programa RAICES Red de Científicos Argentinos en el Noreste de EE.UU. “Ganando la guerra contra el cáncer”. En esa reunión se le concedió el Título *Honoris Causa* de la Universidad de Buenos Aires al Dr. Jack W. Szostak, Premio Nobel de Medicina 2009 y el expositor de la Conferencia inaugural del simposio.

En este número MEDICINA brinda la versión de la conferencia editada por el Dr. Szoszak y además una lista de las referencias citadas en la conferencia y otras que pueden orientar a los lectores.

Medicina (Buenos Aires) agradece la gentileza del Dr. Jack Szostak por brindarnos el manuscrito con su conferencia y la ayuda prestada por los coorganizadores del simposio Raúl Mostoslavsky, del *Massachusetts General Hospital Cancer Center*, *Harvard Medical School*, Boston, EE.UU. y Gabriel Rabinovich del Laboratorio de Inmunopatología, Instituto de Biología y Medicina Experimental (IBYME-CONICET) y Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina.

Referencias

Adamala KP, Engelhart AE, Szostak JW. Collaboration between primitive cell membranes and soluble catalysts. *Nat Commun* 2016 Mar 21; 7:11041. doi: 10.1038/ncomms11041.

Engelhart AE, Powner MW, Szostak JW. Functional RNAs exhibit tolerance for non-heritable 2’-5’ versus 3’-5’ backbone heterogeneity. *Nat Chem* 2013; 5: 390-4.

Gangadhara KL, Srivastava P, Rozenski J, et al. Design and synthesis of nucleolipids as possible activated precursors for oligomer formation via intramolecular catalysis: stability study and supramolecular organization. *J Syst Chem* 2014; 5: 5. Epub 2014 Dec 9.

Cech TR. Nobel lecture. Self-splicing and enzymatic activity of an intervening sequence RNA from Tetrahymena. *Biosci Rep* 1990; 10: 239-61.

Adamala KP, Engelhart AE, Szostak JW. Collaboration between primitive cell membranes and soluble catalysts. *Nat Commun* 2016 Mar 21; 7: 11041. doi: 10.1038/ncomms11041.

Altman S. The road to RNase P. *Nat Struct Biol* 2000; 7: 827-8.

Zhu TF, Szostak JW. Coupled growth and division of model protocell membranes. *J Am Chem Soc* 2009; 131: 5705-13.

Zhu TF, Budin I, Szostak JW. Preparation of fatty acid micelles. *Methods Enzymol* 2013; 533: 283-8.