Early Planetary Environments and the Origin of Life

2. Origin of Life on Earth

P V Sukumaran

P V Sukumaran took his M Tech degree in applied geology from the University of Saugar and has been with the Geological Survey of India since 1974. His interests include petrology, geochemistry, palaeoceanography and organic evolution. He is presently posted as Director in the Department at Nagpur.

The prebiotic organic synthesis occurred in a reducing or mildly oxidising atmosphere. There was no dearth of energy sources to drive this reaction. The speculation that life originated in ocean floor vent habitats is supported by the recent discovery of fossil microbes in a mid-Archaean vent ecosystem. Evidence from nucleic acid sequencing that the last common ancestor of all extant life is a hyperthermophile also lends credence to this hypothesis. The first living molecule that held heritable genetic information was probably ribonucleic acid. The course of events that nature would have followed to reach the RNA-world is fairly clear, but simulating this course in the laboratory to reach RNA remains a formidable problem and has little relevance to an early Earth setting. A prolonged period of abiotic chemical evolution undoubtedly preceded the emergence of the first living molecule.

Introduction

June 26th, 2000 was a momentous day in the history of science; it was on this day that scientists announced to the world that they have completed sequencing the human genome, an effort that began a decade earlier. The book of human genome has been with us ever since James Watson and Francis Crick unravelled the structure of DNA way back in 1953. But it is only now that the letters in the book could be identified; however, it would be a long time before the sentences in the book are annotated, syntaxes identified and the book finally read and understood. With this historic achievement man has understood the secrets of life and he is close to creating it in the lab.

[1] Part 1. Early Planetary Environments, *Resonance*, Vol.6, No.10, pp.16-28.

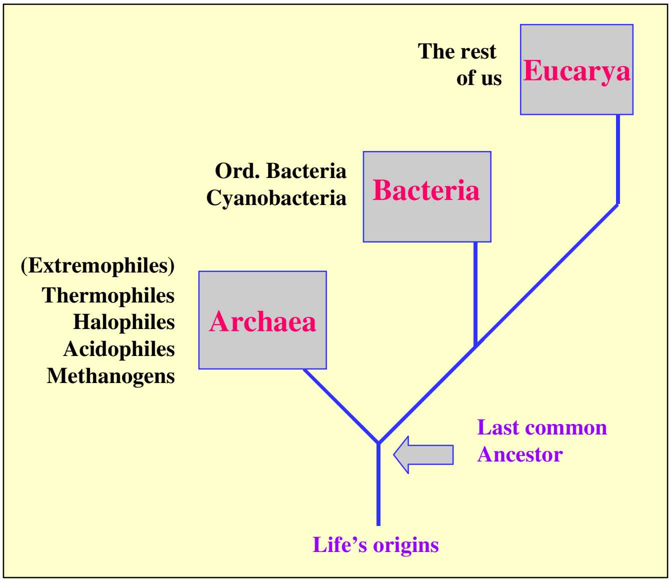


Figure 1. The phylogenetic tree of life (based on ribosomal RNA sequencing).

Though the immediate benefits of this feat are in the fields of medical sciences and biotechnology, the genome sequencing knowledge has great potential in understanding life's ancestry and thereby its origin itself. For instance, based on ribosomal RNA sequencing it has been possible to group all life on Earth into three domains, namely Archaea, Bacteria and Eucarya (*Figure* 1). The deepest roots of Archaea and bacteria are populated by thermophiles, which are heat-loving organisms, and the deepest roots of Archaea are all occupied by hyperthermophiles. Going back in this way one reaches the last common ancestor of all life on Earth, which is a hyperthermophile. From this one can conclude that in all likelihood life originated at the high temperature environments of the early Earth.

All current discussions on the origin of life on Earth are centred on two aspects, namely the sources of energy to run the prebiotic organic synthesis and the possible role of nucleic acids as the first genetic molecules with which life began. Let us now briefly look at these points.

Prebiotic Energy Sources

Several energy sources might have existed on early Earth that drove the organic reactions to produce the precursor compounds like hydrogen cyanide (HCN), formaldehyde (HCHO) and amino acids. These include visible and UV radiations, electric discharges, shock energy of thunder and bolide impacts and chemical energy of photo oxidation of aqueous Fe^{2+} in the Archaean-Hadean oceans releasing H_2. The solar flux in the Hadean and early Archaean was 30% less than it is today, while UV energy output was more intense in the absence of atmospheric ozone shield. The solar energy is considered to have played an important role in the production of simple precursor molecules.

Hydrothermal Environments

Submarine hydrothermal vent systems (see [1]) were first discovered in the Galapagos spreading centre in the early 1970s. Many subsequent undersea explorations have brought to light totally unknown ecosystems in these vents that support myriads of life. Ever since their discovery biologists have considered hydrothermal ecosystems as possible habitats for the origin of life and those advocating this view have come to be known as 'ventists'. Attractive features of vent environments are energy sources from thermal gradients and chemical reducing powers of sulphide minerals and possible protection from annihilating impacts and exposure to harmful UV radiations.

Attempts to test experimental predictions of hydrothermal organic synthesis have been hampered by inaccessibility to submarine vents. Besides, simulating vent environments in the laboratory is a more arduous task. Shock (1990) proposed that amino acids and other organic molecules could be synthesized in vent systems, but the stability of these compounds in vent environments is transient for their eventual polymerization to peptides and polypeptides. Marshal (1994) reported synthesis of amino acids in simulated vent conditions and suggested that hydrothermal ecosystems provide the right habitat for early biosynthesis. However, these amino acids could not be resolved for their chirality (*Box* 1) and therefore contamination of bio-

Submarine hydrothermal vent systems as possible habitats for the origin of life.

Box 1. Racemicity

Many chemical compounds like amino acids and sugars occur in two optically active forms. Structurally they are mirror images of each other like the left and right hands, similar in all respects except that they rotate plane-polarized light (ppl) in opposite directions. This property of chemical compounds to exist in two 'handed' forms is called chirality or enantiomorphism and such compounds are called chiral compounds and the members enantiomers. The one that rotates ppl to the left is *L*-(laevorotatory) enantiomer and that which rotates ppl to the right is called *D*-(dextrorotatory) enantiomer. A mixture containing the two in equal proportions is called racemic and is optically inactive. Life has the unique property to incorporate only one enantiomer, as in sugars (only *D*-sugars) or amino acids (only *L*-amino acids). This property of life is called homochirality. Homochirality is thought to be inherited from primordial times when circularly polarized light, inimical to certain enantiomers, was abundant in the universe. However, Rikken and Raupach (2000) have demonstrated that static magnetic field can bias chemical process in favour of one of the enantiomers.

genic amino acids could not be ruled out from this synthesis. According to Jeffrey Bada (2000) vent environments have more sterilizing effect on organic compounds than their synthesis. In general, proteins become denatured at temperatures characteristic of vent systems and DNA becomes unstable due to hydrolysis. As a consequence, origin of life in vent habitats seems to be a remote possibility, though such ecosystems could have been the cradle for the evolution of early life after having originated elsewhere.

Though thermophilic microbes are ubiquitous in modern ocean floor hydrothermal settings, they were totally unknown until recently in similar settings in the Precambrian. Birger Rasmussen (2000) reported the first ever discovery of early microbial fossil life from the 3.25 Ga old volcanogenic massive sulphides from Australia. These are filamentous pyritic sulphides that Rasmussen considers as fossil prokaryotic microbes such as bacteria that inhabited sea floor environments. These organisms were chemotrophic hyperthermophiles, i.e. they are non-photosynthetic organisms that derive their energy from inorganic substances. If Rasmussen's interpretations are correct there would have been widespread organisms of this kind in the Archaean sea floor volcanic systems. This discovery rekindles

Origin of life in vent habitats seems to be a remote possibility, though such ecosystems could have been the cradle for the evolution of early life after having originated elsewhere.

Box 2. The Deep Hot Biosphere

Did life originate within the Earth rather than on the surface? Well, yes according to Thomas Gold, the author of the book *The Deep Hot Biosphere*. Exotic submarine life has been known in ocean vents at temperatures of 80°C or more since 1977. Bacterial life was long known in deep oil wells but was suspected as surficial contaminant. The existence of indisputable subterranean life was discovered in 1987 in deep boreholes at Savannah River nuclear site, South Carolina. Microbes were subsequently identified in South African gold mines at 3.5 km depth and also in many deep-sea cores. Many of these microbes are new and form totally unknown genera, adapted to strange conditions, completely isolated from surface and sunlight. Bacteria are the primary producers in this world like phytoplankton in surficial ocean. Life in this habitat is very primitive and is dependent on non-photosynthetic chemistry. This led Gold to postulate that life originated inside the Earth and migrated to the surface.

the long-held view of origin and early evolution of life in hydrothermal habitats and its thermophilic ancestry, supporting similar findings from genome sequencing mentioned earlier. Many metal-incorporating proteins are known and these metals may owe their source to hydrothermal exhalations enriched in Cu, Ni, Zn and Mo.

The deep crust of the Earth is home to an altogether different ecosystem, populated by some of the most primitive life forms. Could this be the starting point for the origin and diversification of life? (*Box* 2).

Emergence of RNA

The early Earth, soon after accretion, was most certainly an inhospitable and lifeless planet. But a billion years later it was a haven for microbial life. How could life arise on an inhospitable planet? This question fascinated both scientists and the laymen alike from times immemorial and many speculations and experiments have been made to understand its origin. Many of these are concerned with the possibility that the emergence of self-replicating RNA as a crucial milestone in the process that led to the origin of life.

The widely prevalent view prior to the 17th century was that all higher forms of life including human life were the result of some supernatural creation and the rest of all life originated spontaneously from decaying matter or filth. Louis Pasteur was the first to disprove the belief of spontaneous origin of life and he demonstrated with evidence that all life including microbial life arose from parents resembling themselves. His explanation of the evolution of life obviously posed the intriguing question of how did the first generation of each species come into being. The enunciation of the theory of natural selection and evolution by Charles Darwin and Alfred Russell Wallace in the mid 19th century was another milestone in our understanding of life. According to this theory, life evolved from simple ones to complex ones by natural selection, that is, by adaptation of new traits by species to suit changing environmental pressures as a result of which organisms aptly suited to the new environments stood the best chances of survival. Accordingly, through generations of adaptation, life radiated from simple organisms to complex ones. Going down along the family tree (*Figure* 1) we must then have a progenitor organism from which all extant life descended. This organism is now called the last common ancestor.

The quality defining life is that it can store and transmit genetic information required for its perpetuation.

The most fundamental quality defining life is that it has the ability to store and transmit genetic information required for the perpetuation of life. Without this ability life would have left no descendants. Obviously the last common ancestor must have possessed heritable genetic information for replication; otherwise it would not have left any progeny. Despite the complexity and diversity, certain biochemical commonalities are very obvious in all contemporary life: Firstly all known life is based on carbon compounds and utilises H, N, O and P for metabolism. A second commonality is that all life synthesises proteins out of only 20 amino acids although a hundred and odd amino acids are known to science (the 21st amino acid selenocysteine was discovered in 1986, but very few proteins incorporate this amino acid). These proteins, besides catalysing life's vital metabolic functions, also dictate its fundamental traits including whether a living organism is an ant or an elephant. The third common property is that current life holds genetic information in nucleic acids, RNA and DNA, in fact in the sequence of nucleotides that the nucleic acids are composed of. Lastly, all life uses the same genetic code for assembling proteins from amino acids. This code, a triplet of nucleotide sequence in the nucleic acid, decides which amino acids are to be linked to form the protein chain. The nucleotide sequences in the nucleic acids thus carry the instructions to synthesise all the proteins required by a living organism.

The two biopolymers responsible for all the mysteries and miracles of life are nucleic acids and proteins; nucleic acids carry the blue print of life as they store genetic information while proteins carry out all cellular work.

These intricate commonalities obvious in all modern life must also have been present in the last common ancestor because it is considered unlikely that such universal features evolved independently. This means that the last common ancestor stored genetic information in nucleic acids, whose nucleotide triplets specified the amino acids needed to assemble all the proteins it required. These proteins, in the last common ancestor, must also have directed the metabolic functions such as synthesis of new strands of nucleic acids needed for its material continuity.

The two biopolymers responsible for all the mysteries and miracles of life are nucleic acids and proteins; nucleic acids carry the blue print of life as they store genetic information while proteins carry out all cellular work. In all modern organisms nucleic acids are synthesised only with the help of proteins and proteins are synthesised with the information held by nucleic acids, and neither of them can do anything independent of the other. Both proteins and nucleic acids are chemically very complex and are considered unlikely to have come into being spontaneously, simultaneously and at the same place on a barren early Earth. Thus the most basic problem in understanding the origin of life is how did such an interdependent system of nucleic acids and proteins evolve and which of these appeared first on the evolutionary path. If only one of them appeared first how did this early life carry out the functions of the other. In other words, if nucleic acid appeared first how did it replicate without protein enzymes or if protein appeared first how did it store genetic information? This chicken-egg dilemma in biochemistry is called the nucleic acid-protein paradox.

What appeared first on earth: nucleic acids or proteins?

An apparent way out of this dilemma emerged with the discovery in the early 1980s, by Cech and Altman, of certain RNAs having enzymatic powers. Subsequent research confirmed that these RNA complexes called ribozymes have catalytic powers and could have served as enzymes, besides being information carriers, in primitive metabolism including their own replication. This discovery led Walter Gilbert (1986) to conceive of an 'RNA-world' as an early stage in organic evolution when RNA functioned both as genetic material and enzymes (catalyst) without protein enzymes. In other words, this implies that life began with RNA, though a protracted period of chemical evolution necessarily preceded the advent of RNA. The problem of origin of life thus becomes explaining how RNA came into being.

Nucleic acids form the chemical basis of heredity; they have template properties to store and replicate information and to make any number of their own copies.

But Why RNA?

Nucleic acids form the chemical basis of heredity; they have template properties to store and replicate information and to make any number of their own copies. Other information storage systems may have existed in the primitive world but we know little about their chemical nature. For instance, Cairns Smith (1966) postulated that life began with self-replicating clays. Replicating systems other than nucleic acids have also been suggested. But experimental evidences are lacking to confirm their existence on the early Earth. Some experts advocate that the first genetic molecule did not have template properties and replicated in some other way. Oparin for instance, proposed in 1965 that life began with self-replicating proteins. This is supported by the catalytic properties of some polypeptides and it is likely that they were produced abiotically on the primitive Earth.

Did life begin with RNA?

According to Jerald Joyce (1989) [2] the presumption that life was based on RNA at some stage in its evolution is circumstantial. He drew a number of evidences from contemporary biochemistry that attests to this presumption. For example, RNA plays an important though subsidiary role in modern metabo-

Dyson argued that proteins are the hardware and nucleic acids are the software of life.

lism and participates in all cellular processes some of which are among the most ancient such as priming DNA synthesis during DNA replication. Besides, DNA can validate the RNA-before-DNA hypothesis further if one considers the selective pressures that led to the evolution of DNA as in all contemporary life with the exception of RNA viruses. The reduced stability of RNA vis-a-vis DNA, the absence of proof reading ability in RNA machinery resulting in higher mutation rates in RNA genomes, the tendency of genetic information to degrade in RNA genomes owing to hydrolysis and the photochemical changes that the RNA molecules are prone to under the influence of UV radiation are considered to have made RNA inefficient as a genetic information storage device that led to its take over by the more efficient DNA. However, some of the objections that go against RNA as the first living molecule are that it is an inept catalyst unlike proteins and that its availability in sufficient quantities in the prebiotic world is doubtful.

Freeman Dyson, in his book *Origins of Life*, argues for proteins as the hardware and nucleic acids as the software of life and the logical precedence of hardware before software. Just as hardware development triggered software revolution in computer technology, the advent of proteins must have triggered organic evolution on our planet. According to Dyson, proteins evolved first and once this hardware was available the software, that is, the nucleic acids evolved subsequently. Miller's classic experimental results (see [3]) that amino acids are readily synthesized in simulated prebiotic settings but not nucleic acids support Dyson's arguments. But Miller's experiments fail badly when considering that the amino acids so synthesized are racemic compounds (see below) irrelevant in current life.

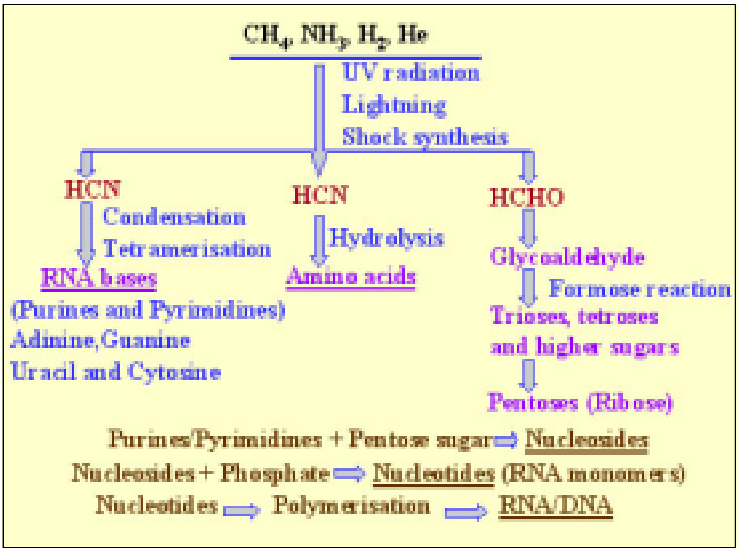
Assuming that life began with RNA let us now consider the course of events that nature would have followed to reach the RNA-world and how difficult it is to synthesise RNA in the laboratory. To understand the discussion further it is necessary to have some understanding of the structure of RNA (*Figure* 2). RNA is a polymer, whose monomer units are called nucleotides.

Nucleotides are constructed of a backbone of pentose sugar (ribose in RNA and deoxyribose in DNA), linked on one side to a nitrogenous base and on the other to a phosphate (PO_4^-) ion. The bases are of four kinds namely, adenine and guanine (purines), and cytosine and uracil (pyrimidines; thymine replaces uracil in DNA). These bases shortened as A, G, C and U are called the alphabets of the genetic language of RNA (it is ATGC in DNA) and it is the sequence in which these letters are arranged in the RNA chain that makes sense in the genetic language. Just as the letters in a language can be arranged in different ways to convey different meanings, so also the letters of the genetic language are arranged in a variety of ways to mean different things. Take for instance the word LIFE; the very same letters in this word can be rearranged to get another word FILE with altogether different meaning. In a similar way the alphabets of the RNA genetic language are linked in different ways to signal coining of different amino acids in the protein chain.

Figure 2. The general structure of RNA.

Therefore, for RNA to evolve as the first genetic material there must be sufficient quantities of the bases, the ribose sugar and phosphate ion on the primitive Earth and these must be capable of combining, utilizing the available energy sources to produce RNA. However, the bases and sugar are not ready-made compounds themselves but are the end products of long series of chemical reactions. The most probable sequence of events leading to the synthesis of RNA is given in *Figure* 3. From the CH_4-NH_3-H_2-He primary atmosphere it is relatively easy to synthesise hydrogen cyanide (HCN) or formaldehyde (HCHO) and has been shown to form under spark experiments [1]. Both are considered as good starting materials for more complex organic compounds. HCN could undergo a series of tetramerisation reactions to produce RNA bases like adenine and guanine through a chain of intermediate products. Hydrolysis of HCN has been demonstrated to produce a number of amino acids of prebiotic relevance. However, a major drawback of laboratory synthesis of purines and pyrimidines is the low yield and instability of the products to make the reaction progress to the RNA stage. Be-

Figure 3. The path to RNA.



sides, laboratory synthesis involves use of several chemical compounds and catalysts the availability of which is doubtful on the primitive Earth.

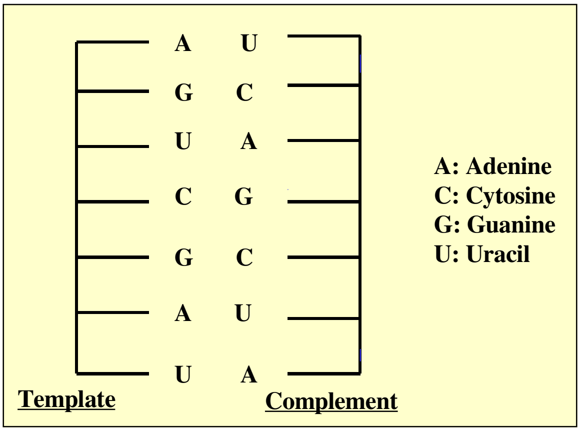
Condensation of HCHO yields glycoaldehyde and then a series of trioses, tetroses pentoses and higher sugars. This reaction, leading to the generation of sugars, is called formose reaction and is considered to be of great prebiotic significance to produce the ribose backbone of RNA. But ribose is only one of the 5-carbon sugars in the complex sugar products of HCHO-hydrolysis. Further, the ribose so formed is a racemic (*Box* 1) mixture of both *L*-ribose and *D*-ribose whereas life utilizes only the *D*-enantiomer. It remains a mystery how nature could first isolate the ribose sugar from a virtual sugar forest (Chyba and McDonald, 1995) [4] and then the *D*-enantiomer from the racemic mixture.

It remains a mystery how nature could first isolate the ribose sugar from a virtual sugar forest and then the *D*-enantiomer from the racemic mixture.

Having produced ribose sugar and the bases the next step on the road to RNA is joining the sugar with the bases to form nucleosides (base + ribose = nucleoside). This is a very formidable reaction. While purine nucleosides have been synthesized in the laboratory with relative ease, synthesis of pyrimidine nucleosides has been met with more serious difficulties. The next step namely phosphorylating the nucleosides (i.e. joining the nucleosides with phosphate ions to produce the RNA monomers) is a very discouraging task. Phosphate ions were available on the early oceans (Chyba and McDonald, 1995) [4] but predominantly as orthophosphate without high-energy bonds. For the reaction to proceed further, high-energy phosphates like metaphosphates or polyphosphates such as pyrophosphate are required. Pyrophosphate has been synthesised in the laboratory with the help of several other chemical compounds, but are again of doubtful prebiotic significance.

Assuming that all the constituent components of RNA were available on the primitive Earth, polymerizing these components to produce an RNA strand in a prebiotic setting is a difficult task. Polymerisation has been achieved in the laboratory by dry heating the constituents but this course has little significance to the origin of RNA. Polymerisation is a dehydration reaction during which water molecules are eliminated and it is difficult to visualise how polymerization could occur on an ocean-covered Hadean or early Archean Earth when continents were either absent or were very limited in extent. The most important step leading to RNA is to use the polymerised strands as templates to synthesise complementary strands (*Figure* 4), that is, as genetic material to store and replicate information. In contemporary life protein enzymes catalyse this vital process and in the absence of proteins replication of the RNA strand in the prebiotic world remains a mystery. Laboratory attempts to synthesise RNA strands and further use of these strands to produce complementary strands have been successful through various permutations and combinations of reactants and catalysts. Efforts to synthesise the original strands, on the other hand, making use of the complementary strands as templates, have proved unsuccessful till today. In addition the various chemical reagents employed in RNA synthesis in controlled experiments are unlikely to have

Figure 4. Template and complementary strands of RNA.



Problems with hypothesis of life beginning with RNA.

been available on the early Earth. Besides, simulation experiments always result in production of racemic mixtures of the nucleotides (containing L- and D-ribose) and the use of racemic nucleotides in laboratory systems curtails polymerisation reaction. According to Chyba and McDonald (1995) [4] inhibition of homochirality in experimental systems is the most daunting obstacle to use RNA as the first genetic molecule.

The nucleotide chemistry of RNA synthesis, without the aid of enzymes in particular, is thus a mystery. This led Orgel (1989) to propose that life did not begin with RNA and RNA-world came to the arena only after many of the hurdles associated with the prebiotic synthesis and replication of RNA were overcome. This implies that there were simpler genetic systems that preceded RNA. Unfortunately we are totally ignorant of the chemical nature of the precursor genetic systems. But it is certain that a protracted period of abiotic chemical synthesis preceded the advent of the first genetic molecule. Christian de Duve (1995) [5] denotes this phase as 'protometabolism' to designate the set of unknown chemical reactions that led to the generation of RNA. Duve argued further that a 'thioester-world' preceded the RNA-world in which thioesters played the role that ATP plays in modern biochemistry. Thioesters are formed by the reaction of thiols with carboxylic acids; the former would readily form in an early volcanic setting, while the latter could be produced spontaneously in a Miller–Urey setting.

It is certain that a protracted period of abiotic chemical synthesis preceded the advent of the first genetic molecule.

To conclude, the subject of the origin of life is a gray area of frustrating research, often leading to blind alleys. The galloping pace of genome research, resulting in the successful sequencing of human genome, is certain to widen our understanding of life and its origin. But the greatest problem facing researchers today is how to develop genetic coding systems. Experimental polymerisation has, until now, been able to produce strands containing not more than about 100 base-pairs, insufficient to store genetic information. Glen Evans and coworkers (2000) claimed to have developed a technique of first producing short DNA chains and then joining them in a controlled way to achieve up to 100,000 base-pairs long strands, sufficient to produce simpler forms of life. If this claim is correct we are close to producing life in the lab and thereby understanding the most intricate and fundamental problem in science, namely the origin of life. If not we would continue to fumble while trying to answer the question of how life originated on our planet.

Suggested Reading

[1] P V Sukumaran, *Resonance*, Vol. 5, No. 9, p. 8, 2000.

[2] F Jerald Joyce, RNA evolution and the origin of life, *Nature,* Vol. 338, pp. 217-224, 1989.

[3] P V Sukumaran, *Resonance*, Vol. 4, No. 12, p. 4, 1999.

[4] F Christopher Chyba and D Gene McDonald, The origin of life in the Solar System, *Annual Review of Earth and Planetary Sciences,* Vol. 23, pp. 215-249, 1995.

[5] Christian de Duve, Beginnings of life on Earth, *American Scientist*, September-October, 1995.

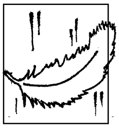
[6] J Oro, L Miller Stanley and A Lazcano, The origin and early evolution of life on Earth, *Annual Review of Earth and Planetary Sciences*, Vol. 18, pp. 317-356, 1990.

[7] L E Orgel, Molecular replication, *Nature*, Vol. 358, pp. 203-209, 1992.

[8] C Chyba and C Sagan, Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origins of life, *Nature*, Vol. 355, pp. 125-132, 1992.

[9] H John McClendon, The origin of life, *Earth Science Review,* Vol. 47, pp. 71-93, 1999.

Address for Correspondence
P V Sukumaran
Director(Geol)
Geological Survey of India
Seminary Hills
Nagpur 440006, India.
Email:pvs34@yahoo.co.uk



“And what was he? Forsooth, a great arithmetician.”

Othello