

# The Origin of Life: A Critique of Current Scientific Models

AW SWEE-ENG

## ABSTRACT

*Profound advances in the fields of molecular biology in recent years have enabled the elucidation of cell structure and function in detail previously unimaginable. The unexpected levels of complexity revealed at the molecular level have further strained the concept of the random assembly of a self-replicating system. At the same time, the recent discovery of fossil algae and stromatolites (primitive colonies of cyanobacteria) from as early as the Precambrian, have reduced the time for development of the first cell as much as tenfold. Together with implications of this for the oxidative state of the primitive atmosphere, these developments will force researchers to rethink many fundamental ideas pertaining to current models of the origin of life on Earth. The evidence for the nature of the primitive atmosphere is examined and the possibility of ribonucleic acid (RNA) as the first self-replicating molecule is evaluated. The focus is then on DNA, proteins and the first cells.*

---

## THE EARLY ATMOSPHERE

The nature of the atmosphere under which life arose is of great interest. The high oxygen content of the Earth's atmosphere is unique among the planets of the Solar System and could have been tied up with the composition of the core and its crust. It has to be said that none of the hypotheses of core formation of the Earth survives quantitative scrutiny. The gross features of mantle geochemistry, such as its redox state (FeO) and its iron-sulphur systems, apparently do not agree with experimental data.<sup>1,2</sup> There are outstanding questions relating to the formation and recycling of the Archaean crust.<sup>3</sup>

Interesting organic molecules such as sugars and amino acids can be formed from laboratory 'atmospheres' of different proportions of CO<sub>2</sub>, H<sub>2</sub>O, N<sub>2</sub>, NH<sub>3</sub>, H<sub>2</sub>, CH<sub>4</sub>, H<sub>2</sub>S and CO. This happens only in the absence of free O<sub>2</sub>. Oxygen is highly reactive, breaking chemical bonds by removing electrons from them. A reducing gas (H<sub>2</sub>, CH<sub>4</sub> or CO) is therefore thought to be essential for the successful

synthesis of prebiotic organic molecules.

It has been generally accepted that at about 1.5 Ga [*Giga annum* = billion years ago] the oxygen content of the air rose at least 15-fold. (Note that evolutionary/uniformitarian 'ages' are only used here for argument's sake.) Before this, the oxygen had been reduced by Fe(II) in sea water and deposited in enormous bands as oxides or hydroxides on the shallow sea floors. The source of the ferrous iron was hydrothermal vents in the company of reducing gases such as hydrogen sulphide (H<sub>2</sub>S).

In 1993 Widdel and his team cultured non-sulphur bacteria from marine and freshwater muds. These anoxygenic, photosynthetic bacteria use ferrous iron as the electron donor to drive CO<sub>2</sub> fixation. It was a signal discovery that oxygen-independent biological iron oxidation was possible before the evolution of oxygen-releasing photosynthesis. Quantitative calculations support the possibility of generating such massive iron oxide deposits dating from Archaean and Early Proterozoic times, 3.5–1.8 Ga.<sup>4</sup>

In 1992 Han and Runnegar made a discovery which impinged on discussions of oxygen evolution during the Precambrian. To everyone's surprise they reported the spiral algal fossil *Grypania* within banded iron formations (BIFs) in Michigan, USA. Algae require oxygen, so their existence at this juncture shows banded iron formations do not necessarily indicate global anoxic conditions.<sup>5</sup>

Indeed, as early as 1980 two reports appeared on the discovery of stromatolites in the 3.4–3.5 Ga Warrawoona Group sediments from the Pilbara Block, Australia.<sup>6,7</sup> Similar remains were also discovered in Zimbabwe<sup>8</sup> and South Africa.<sup>9</sup>

It is fair to conclude that the Earth's early atmosphere before 3.5 Ga could have significant quantities of oxygen. This should discourage the sort of hypothesising on abiotic monomer and polymer syntheses so often assumed to have occurred in Archaean times. Robert Riding says that the *Grypania* discovery

*'could spell the end of BIF-dominated models of oxygen build-up in the early atmosphere . . . The cat really will be put among the pigeons, however, if [further] fossil discoveries extend the eukaryote record back much beyond 2200 million years ago, into what is still widely perceived to have been an essentially anaerobic world.'*<sup>10</sup>

## SCENARIOS FOR PREBIOLOGY

A number of revised textbooks on molecular biology came out in 1994–1995 which, while conveying the standard arguments for origin-of-life hypotheses, are cautious in their affirmation. Rightly so, because advances in the field have uncovered exquisite details of intracellular processes. These challenge superficial explanations that their origin and subsequent refinement were fed by randomness. After mentioning the famous simulation by Miller and Urey of prebiotic synthesis of organic compounds (Figure 1), Voet and Voet handle the riddle of the formation of biological monomers with a caveat. They write:

*'Keep in mind, however, that there are valid scientific objections to this scenario as well as to the several others that have been seriously entertained so that we are far from certain as to how life arose.'*<sup>11</sup>

The text of **Molecular Cell Biology** in its second edition was well indexed on the evolution of cells, describing the Miller experiment in detail.<sup>12</sup> The third edition has dropped the chapter on evolution of cells found in the second edition.<sup>13</sup> Similarly, Stryer's fourth edition of his textbook on biochemistry makes no mention of the abiotic synthesis of organic molecules.<sup>14</sup>

*'Doubt has arisen because recent investigations indicate the earth's atmosphere was never as reducing as Urey and Miller presumed. I suspect that many organic compounds generated in past studies would have been produced even in an atmosphere containing*

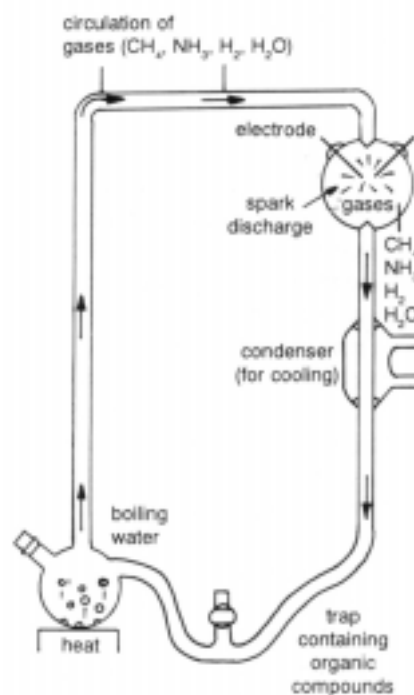
*less hydrogen, methane and ammonia. Still, it seems prudent to consider other mechanisms for the accumulation of the constituents of proteins and nucleic acids in the prebiotic soup.*

*For instance, the amino acids and nitrogen-containing bases needed for life on the earth might have been delivered by interstellar dust, meteorites and comets.'*<sup>15</sup>

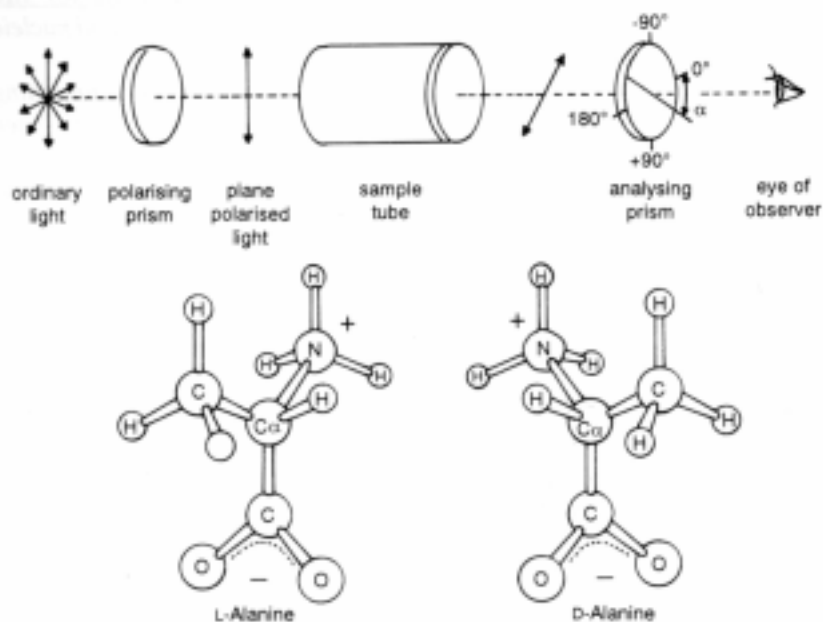
In his essay on the origin of life on Earth, Orgel quotes the experiments of Miller, and of Juan Oro' who used the Miller model to produce adenine with hydrogen cyanide and ammonia.<sup>16</sup> His conclusions overall are:

*'Since then, workers have subjected many different mixtures of simple gases to various energy sources. The results of these experiments can be summarized neatly. Under sufficiently reducing conditions, amino acids form easily. Conversely, under oxidizing conditions, they do not arise at all or do so only in small amounts.'*

Saturn's giant moon, Titan, has an atmosphere composed mainly of molecular nitrogen and up to 10 per cent methane. Carl Sagan and Bishun Khare of Cornell University simulated the pressure and composition of Titan's atmosphere and irradiated the gases with charged particles. A dark solid was formed, which on dissolving in water yielded amino acids and traces of nucleotide bases, polycyclic hydrocarbons and many other compounds. It was then assumed that from this 'wonderful brew' life would have originated.<sup>17</sup> In the text **Molecular Biology of the Cell** the authors note that experimentalists are



**Figure 1.** Simplified apparatus for abiotic synthesis of organic compounds as performed originally by Miller and Urey. By varying the mixture of gases, including using volcanic gases of today, experimenters have been able to produce many types of organic compounds.



**Figure 2.** Optical activity and chirality. Ordinary light consists of waves vibrating in all possible directions perpendicular to its path. Certain substances will selectively transmit light waves vibrating only in a specific plane — plane polarised light. Most compounds isolated from natural sources are able to rotate the plane of polarised light a characteristic number of degrees for any specific substance. The significance of this phenomenon to molecular biology and the origin of life is that stereoisomers, molecules of identical but mirror image structure, possess such 'optical activity'. For example, in the case of the stereoisomers of the amino acid alanine shown above, L-alanine will rotate the plane of polarised light in the opposite direction to D-alanine. Why biological systems utilise exclusively levorotatory (left-handed) amino acids and dextrorotatory (right-handed) sugars remains unfathomable. Mixtures of organic compounds synthesised in Urey-Miller type experiments always consist of racemic (equal amounts of left- and right-handed) mixtures.

beguiled by the 'surprisingly easy' manner in which organic molecules form.<sup>18</sup> Little store is laid for such crucial points as the lability of the organic products, or their reactivity among themselves to form mixed polymers. Indeed, the problem of spontaneously producing a simple homochiral compound, say, L-alanine, from racemic reaction systems has not been solved (see Figure 2).

Classical mechanisms generally rely on chance for the selection of L-amino and D-sugars by self-replicating systems. Mason has put forward the tantalising speculation that a weak nuclear interaction will stabilise the L-amino acids and their polypeptides over their D-forms. This electroweak advantage is considered too weak to affect the outcome of biochemical evolution. An imaginary flow reactor of a kilometre in diameter and four metres deep would be needed to autocatalyse a change of  $10^{-2}$  to  $10^{-3}$  moles of one isomer over 10,000 years if the temperature is kept at ambient. Admittedly a good thought experiment 'but it will find no popular primitive Earth scenarios'.<sup>19</sup>

The discovery of hydrothermal vents at oceanic ridge crests has spawned several origin-of-life hypotheses. It seemed an attractive suggestion that, given the dissolved

gases issuing from the vents, with hydrothermal mixing there would emerge peptides, nucleotides and even protocells of some sort. Miller and Bada, however, dispute the plausibility.

'This proposal, however, is based on a number of misunderstandings concerning the organic chemistry involved. An example is the suggestion that organic compounds were destroyed on the surface of the early Earth by the impact of asteroids and comets, but at the same time assuming that organic syntheses can occur in hydrothermal vents. The high temperatures in the vents would not allow synthesis of organic compounds, but would decompose them, unless the exposure time at vent temperatures was short. Even if the essential organic molecules were available in the hot hydrothermal waters, the subsequent steps of polymerization and the conversion of these polymers into the first organisms would not occur as the vent waters were quenched to the colder temperatures of the primitive oceans.'<sup>20</sup>

## TIME-SPAN FOR PREBIOLOGY

A pillar of 'prebiological evolution' has been the long period of time supposedly available for the emergence of 'protocells' whose development in turn profoundly altered the climate of the planet and its geology. For an estimated age of the Earth of 4.6 Ga this seemed initially to pose no problem. However, the discovery of stromatolites in Western Australia<sup>21,22</sup> and in South Africa<sup>23,24</sup> upset the timetable severely. The finding of algal filaments dated at only slightly more than 1 Ga younger than the Earth itself restricted the time required for the evolution of the living cell. *Pari passu* the list of processes thought to occur abiotically has been shrinking.<sup>25,26</sup> Even the origin of the huge banded iron formations of the Archaean can now be attributed to microorganisms,<sup>27</sup> and Raup and Valentine have suggested that bolide impacts have, at intervals of  $10^5$  to  $10^7$  years, periodically erased more than one origin of life.<sup>28</sup> According to this scenario, ten or more extinct bioclades could have preceded the Cambrian. A bioclade is a group of life forms descended from a single event of life origin. 4.2 Ga has been given as the date of the oldest rocks, which is ostensibly consistent with the cooling and degassing of an active molten Earth that is said to be 4.6 Ga old.<sup>29</sup> According to the isotopic carbon record in sedimentary rocks, 3.8 Ga would date the origin of life.<sup>30</sup>

Fred Hoyle, the Cambridge astronomer and physicist,

made some sobering calculations on the origin of the cell.<sup>31</sup> The probability of forming the 2,000 or so enzymes needed by a cell lies in the realm of 1 in  $10^{40,000}$ . This makes the conceptual leap from even the most complex 'soup' to the simplest cell in the time available (that is, about 500 Ma) so dramatic that it requires some suspension of rationality in order to accept it. Small wonder that latterly it is being touted that life may have taken far less time to appear.

Carl Sagan has opined:

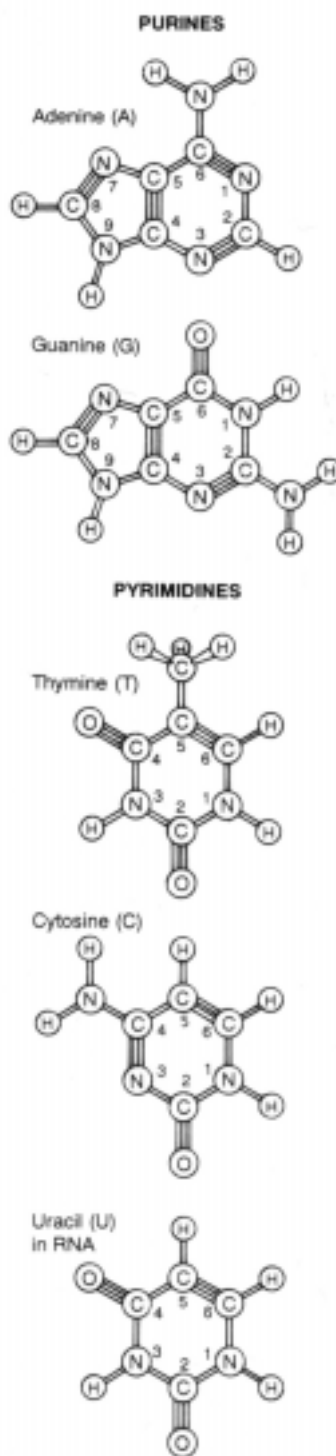
*'If 100 million years is enough for the origin of life on the earth, could 1,000 years be enough for it (to appear) on Titan?'*<sup>32</sup>

### A RIBONUCLEIC ACID (RNA) WORLD

RNA is a linear polymer of ribonucleotides, usually single stranded. Each ribonucleotide monomer contains the sugar ribose linked with a phosphate group and one of four bases: adenine, guanine, cytosine or uracil. RNA appears in both prokaryotic and eukaryotic cells as messenger RNA (mRNA), transfer RNA (tRNA) and ribosomal RNA (rRNA) which are involved in protein synthesis with DNA the source of information. Some viruses however contain genomes of RNA. The nuclei of eukaryotic cells carry two other types of RNA; heterogeneous nuclear RNA (hnRNA or pre-mRNA) and small nuclear RNA (snRNA).

In recent literature there is much excitement over the discovery that there are RNAs that can catalyse specific biochemical reactions. These are the ribozymes, that is, RNA with enzymatic functions.<sup>33</sup> RNA can do this surprising feat by folding its linear chains to appropriate secondary and tertiary structures thereby conferring 'domain' type catalytic structures as seen in protein enzymes.

That RNA can act as a template and also now exhibits catalytic activity fuelled hypotheses for the evolution of an 'RNA world'.<sup>34</sup> In this scenario RNA is the primary polymer of life that replicates itself. DNA and proteins were later refinements. So the first genes were short



**Figure 3.** The molecular structures of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are built using the nitrogenous bases adenine and guanine (purines), and thymine, cytosine and uracil (pyrimidines), which are the 'letters' of the genetic code.

strands of RNA that reproduced themselves, perhaps on clay surfaces. This conjecture is strengthened by the fact that in cells today there are segments of some eukaryotic pre-rRNAs which can cleave themselves off and join the two cut ends together to reform the mature rRNA. In 1982 Thomas Cech and his colleagues at the University of Colorado discovered this can take place in the absence of protein in the ciliated protozoan *Tetrahymena thermophila*.<sup>35</sup> Just as remarkable are the small nuclear RNAs (snRNAs), which complex with protein to form small nuclear ribonucleoproteins (snRNPs; pronounced 'snurps'). Particles called spliceosomes convert pre-mRNA to mRNA.<sup>36</sup> Other ribozymes include the hammerhead variety and RNase P, which generates the 5' ends of tRNAs. The former are found in certain plant viruses. Origin-of-life theories see prebiotic significance in these 'vestigial' post-translational mechanisms.

Though attractive, there are several serious objections to the notion that life began with RNA:—

(1) Pentose sugars, constituents of RNA and DNA, can be synthesised in the formose reaction, given the presence of formaldehyde (HCHO). The products are a melange of sugars of various carbon lengths which are optically left- and right-handed (D and L). With few exceptions sugars found in biological systems are of the D type; for instance, β-D-ribose of RNA, which is always produced in small quantities abiotically.

(2) Hydrocyanic acid (HCN) undergoes polymerisation to form diamino-maleonitrile which is on the pathway to producing adenine, hypoxanthine, guanine, xanthine and diaminopurine. These are purines: there is difficulty in producing pyrimidines (cytosine, thymine and uracil) in comparable quantities<sup>37,38</sup> (see Figure 3).

(3) Neither preformed purines nor pyrimidines have been successfully linked to ribose by organic chemists. An attempt to make purine nucleosides resulted in a 'dizzying array of related compounds'.<sup>39</sup> This is expected if sugars and bases were randomly coupled. The prebiotic production of numerous isomers and closely related molecules hinders the

likelihood of forming desirable mononucleosides. Furthermore, unless ribose and the purine bases form nucleosides rapidly they would be degraded quite quickly.

### Purine and Pyrimidine Nucleotide Biosynthesis

Purine ribonucleotides (for example, AMP, GMP) are synthesised from scratch by living systems in ways not remotely connected with the laboratory models. The purine ring system is built up stepwise from an intermediate 5'-phosphoribosyl-1-pyrophosphate (PRPP) to a larger molecule inosine monophosphate (IMP). This involves a pathway comprising 11 reactions.

The biosynthesis of pyrimidines is less complex, but again the process is elegantly dissimilar to the *in vitro* chemistry, with some of the enzymes on the pathway exercising regulatory functions.

The purine and pyrimidine biosynthetic pathways are finely tuned, and defects such as enzyme deficiencies, their mutant forms or loss of feedback inhibition, cause diseases in man.

Suppose that we already have mononucleosides — purines (or pyrimidines) linked to ribose. Heating these in a mixture of urea, ammonium chloride and hydrated calcium phosphate has been shown to produce mono-, di- and cyclic phosphates of the mononucleoside. The subsequent chemistry would yield a rich (or untidy, depending on how it is viewed) racemic mixture of D and L-oligonucleotides in all sorts of combinations and permutations. Internal cyclisation reactions would destroy much of these oligonucleotides.<sup>40</sup>

Suppose further that we have a parent strand of RNA in a chirally-mixed pool of activated monoribonucleotides. By base-pairing the strand correctly aligns on itself the incoming monomeric units in matching sequence. Phosphodiester bonds are spontaneously forged.

The chief obstacles to efficient and faithful copying appear to be threefold.<sup>41</sup>

- D-monomer nucleotides and L-monomer nucleotides hinder each others' polymerisation on an RNA template.
- Short chains of nucleotides tend to fold back on themselves to form double helical Watson-Crick segments.
- Newly formed strands separate with difficulty from their parent RNA strands. The process grinds to a halt. Using activated monomers — both nucleotides and amino acids — Ferris and his co-workers could form oligomers up to 55 monomers long on mineral surfaces. Such surfaces bind monomers of one charge (negative in these experiments) and strength of binding increases with chain length. Desorption then becomes impossible.<sup>42</sup>

Joyce sums up the difficulties of conjuring up a hypothetical RNA world in these words.

*'The most reasonable interpretation is that life did not start with RNA . . . The transition to an RNA world, like the origins of life in general, is fraught with*

*uncertainty and is plagued by a lack of relevant experimental data. Researchers into the origins of life have grown accustomed to the level of frustration in these problems . . . It is time to go beyond talking about an RNA world and begin to put the evolution of RNA in the context of the chemistry that came before it and the biology that followed.'*<sup>43</sup>

These sentiments are shared by Orgel, a long-time, well-known prebiotic chemist. In 1994 he wrote:

*'The precise events giving rise to the RNA world remain unclear. As we have seen, investigators have proposed many hypotheses, but evidence in favour of each of them is fragmentary at best. The full details of how the RNA world, and life, emerged may not be revealed in the near future.'*<sup>44</sup>

As we have seen, the intuition that an RNA world preceded DNA and protein is based on some features found in modern cells. But it appears to be contradicted by the available experimental evidence. In fact, the extra hydroxyl of ribose renders it more reactive than deoxyribose and, in principle, makes the more stable DNA a more likely progenitor.

#### KEY POINTS

- **The presumed rise of oxygen levels in a primitive reducing atmosphere formerly attributed to the evolution of photosynthesis can be explained by oxygen-independent biological iron oxidation.**
- **Recent investigations indicate that the Earth's atmosphere was never as reducing as previously thought.**
- **Recent discovery of fossil stromatolites and algae from the Precambrian has reduced the time for evolution of the first cell ten-fold.**
- **The atmosphere of 3.5 billion years ago could have contained significant quantities of oxygen.**
- **Under oxidising conditions, the formation of organic compounds and their polymerisation do not occur.**
- **Biological homochirality of sugars and amino acids remains an enigma.**
- **Hypotheses of ribonucleic acids (RNAs) as the initial self-replicating molecule have serious unresolved difficulties.**
- **Extrapolating results of *in vitro* synthesis of purines and pyrimidines should take into account that biosynthesis utilises different reaction pathways.**

#### Other Options

Attention switched to other molecules that can carry information and replicate themselves. In 1991 a team of Danish chemists led by Egholm strung the four familiar bases of nucleic acids along a peptide (polyamide) backbone forming a peptide nucleic acid (PNA).<sup>45,46</sup> Unfortunately, PNAs bind natural DNA and RNA tightly (about 50 to 100 times stronger than the natural polymers

bind among themselves) so that it is difficult to envisage their being a prebiotic replicating system. So strong is their affinity for DNA that they would disrupt nucleotide duplexes unless they were removed from an evolving RNA milieu. Their base-specificity for natural nucleic acids of oligomers of 10 units or more, and consequently their fidelity in copying RNA or DNA, is uncertain. This militates against the co-evolution of multiple genetic systems, a suggestion raised by Böhler and his co-workers.<sup>47</sup> Using an unusual activated monomer, guanosine 5'-phosphoro (2-methyl) imidazolide, they formed 3'-5'-linked oligomers with PNA as template. In fact, because of problems of cyclisation the activated dimer rather than the monomer was used. No oligomers of more than 10 were formed, and there was present in the complex mixture short oligomers with unnatural 2'-5'-phosphodiester bonds, pyrophosphate linked oligomers and possibly cyclic oligomers.

### THE DNA STORY

Like RNA, deoxyribonucleic acid (DNA) is a linear polymer of nucleotides. Each nucleotide consists of a pentose sugar, a nitrogenous base and a phosphate group. The sugar-phosphate linkages form an external backbone with the bases sticking in and hydrogen-bonding with complementary bases of the opposite sugar-phosphate backbone, zipper-fashion, producing the famous double helix structure of DNA. The helix can take on alternate forms in which it twists to alter the compactness of its spiral and bends to change its overall shape. The packing of DNA in a microscopically visible chromosome represents a 10,000-fold shortening of its actual length. Little is known of the structure of DNA in the natural state within the cell. Clearly it is dynamic, and by assuming different forms DNA controls various biological processes such as replication, transcription and recombination. This is a fruitful area for research.

#### The Synthesis of $\beta$ -D-Ribose

The abiotic origin of DNA is beset with problems similar to those seen with RNA.<sup>48</sup> The synthesis of deoxyribose forms the nub. We have already mentioned the difficult synthesis of even small amounts of  $\beta$ -D-ribose for the *in vitro* production of RNA. Furthermore, we might have expected deoxyribonucleotides to be biosynthesised *de novo* from deoxyribose precursors. In real life, however, DNA components (the deoxyribonucleotides dADP, dCDP, dGDP and dUDP) are synthesised from their corresponding **ribonucleotides** by the reduction of the C2' position. The enzymes that do this are named ribonucleotide reductases. There are three main classes of reductases. All replace the 2'-OH group of ribose via some elegant free radical mechanisms.<sup>49,50</sup> The class III anaerobic *Escherichia coli* reductase is thought to be the most closely related to the common reductase ancestor from which the three main

classes are presumed to have evolved. It has been proposed that the pristine reductase enzyme, similar to present-day class III enzymes, arose before the advent of photosynthesis and therefore before the appearance of oxygen.

Now the *E. coli* class III enzyme mentioned above can be induced by culturing the bacteria under anaerobic conditions. This enzyme is an Fe-S protein that in its active form contains an **oxygen-sensitive glycyI** free radical.<sup>51</sup> This poses a conundrum: the survival and continual evolution of an oxygen-sensitive enzyme when oxygen appeared. On the other hand, the class I reductases **require** oxygen for free radical generation. Surely they could not have evolved and operated in the anaerobic first cell in an oxygen-free environment.<sup>52</sup> Moreover, one of the most remarkable aspects of this *E. coli* ribonucleotide class I reductase is its ability to maintain its highly reactive free radical state for a long period. Interestingly, this is achieved *in vivo* by **internally generated oxygen**. Four proteins have to be in place:—

- Flavin oxidoreductase, which releases superoxide ion ( $O_2^{\cdot-}$ ),
- Superoxide dismutase, to rapidly convert this destructive radical to  $H_2O_2$  and  $O_2$ ,
- A catalase, to disproportionate  $H_2O_2$  to  $H_2O$  and  $O_2$ , and
- A fourth protein, thioredoxin, that functions as a reductant.

The oxygen oxidises Fe II and a deeply buried tyrosyl residue (Tyr122). Herein lies a difficulty. The reductases are complex protein reaction centres acting in tandem on each other and on the 2'-OH group of ribose. These must all have co-evolved **before DNA and along with RNA**. Could this be seriously contemplated for a metabolically naive RNA 'progenote'?

The origins of deoxyribose and of DNA therefore remain unsolved mysteries.

Even if the DNA molecule were assembled abiotically, there is the instability and decay of the polymer by hydrolysis of the glycosyl bonds and the hydrolytic deamination of the bases.<sup>53</sup> Each human cell turns over 2,000–10,000 DNA purine bases every day owing to hydrolytic depurination and subsequent repair. Genetic information can be stored stably only because a battery of DNA repair enzymes scan the DNA and replace the damaged nucleotides. Without these enzymes it would be inconceivable how primitive cells kept abreast of the constant high-level damage by the environment and by endogenous reactions. If unrepaired, cell death would result. Indeed, the spontaneous errors resulting from intrinsic DNA instability are usually many times more dangerous than chance injuries from environmental causes.<sup>54</sup> The enzymes of the DNA repair system are a marvel in themselves and have been rightfully recognised as such.<sup>55</sup>

Reports of the culture of *Bacillus sphaericus* from spores preserved in amber for over '25 million years' does

not tally with what is known of the physico-chemical properties of DNA.<sup>56</sup>

### Several DNA Paradoxes

The total amount of DNA in the haploid genome is its C-value. Intuitively we would expect that there should be a relationship between the complexity of an organism and the amount of its DNA. The failure to consistently correlate the total amount of DNA in a genome with the genetic and morphological complexity of the organism is called the C-value paradox.<sup>57</sup> This paradox manifests itself in three ways.

- (1) Many plant species have from two to ten times more DNA per cell than the human cell. Among the vertebrates with the greatest amount of DNA are the amphibians. Salamander cells contain 10–100 times more DNA than mammalian cells. It is hard to make sense of the existence of such major redundancies in organisms evolutionarily less complex than man.
- (2) There is also considerable intragroup variation in DNA content where morphology does not vary much. For example, the broad bean contains about three to four times as much DNA per cell as the kidney bean. Variations of up to 100 times are found among insects and among amphibians. In other words, cellular DNA content does not correlate with phylogeny.
- (3) Large stretches of DNA in the genome, say, of humans, appear to have no demonstrable function. This will be discussed later.

### INTRONS AND EXONS

Once the genes of unrelated cells were studied it became clear that the molecular genetics of higher organisms are different from those of bacteria. The principles uncovered in prokaryotes cannot simply be applied to eukaryotes. For one thing, the precursor RNA found in the nucleus, called heterogeneous nuclear RNA (hnRNA), was far greater in amount than the mRNA that emerged from the nucleus into the cytoplasm. It was discovered that the linear hnRNA molecule contained excess RNA which was cut out, and the mRNA was then constructed from splicing together the in-between pieces. An editing process had taken place.<sup>58</sup> The logical inference from this finding was that the genomic DNA from which the hnRNA was transcribed must be similarly constructed. The notion of the co-linear relationship between a segment of DNA and the protein for which it codes is not true, at least for higher organisms.

The word 'intron' was used to describe such a non-coding region of a structural gene. They separate the 'exons', which encode the amino acids of the protein.<sup>59</sup> For instance, the human  $\beta$ -globin gene comprises, in linear sequence, three exons separated by two introns within a total length of 1,600 nucleotides. Introns are abundant in higher eukaryotes, uncommon in lower eukaryotes, and

rare in prokaryotic structural genes. Variations in the length of the genes are primarily determined by the lengths of the introns. Since the discovery of introns/exons the intricate processes of nuclear mRNA splicing have been elegantly elucidated. Among these are the remarkable self-splicing introns<sup>60</sup> and the equally revolutionary finding that individual nucleotides can be inserted into RNA after transcription altering them remarkably.<sup>61</sup>

The inevitable questions emerged. What role does having genes in pieces serve? How have such interrupted genes 'evolved' over time?

One hypothesis points out that exons usually encode for a part of the protein that folds to form a domain. What constitutes a domain has been a matter of controversy. By dispersing individual exons of a protein among introns it is reasoned that breaking DNA and rejoining and recombining different exons is that much easier. This process of shuffling exons/domains is presumed to have created new proteins with multi-domain structures. This is thought to be a more efficient way for a cell to create proteins rather than through random DNA mutations. Here is a means of duplicating, modifying, assembling and reassembling units with modular functions into larger structures. According to this hypothesis this is the reason why introns have survived through time. Several queries may be raised. First, exon shuffling as a device to speed up evolution is logically tied up with a subsidiary assumption that possessing similar domains qualifies proteins for biochemical kinship, which is to say, these proteins are alleged to bear the marks of descent from a common ancestral protein.<sup>62</sup> But the construction of phylogenetic trees relies on unstable molecular clocks and other genetic mechanisms largely unknown<sup>63</sup> and, as discussed below, should be approached with caution.

Biochemical kinship aside, would not domains exercising similar function be structurally alike such as we see between, say, the catalytic domains of the two serine proteases chymotrypsin and tissue plasminogen activator?

Second, RNA splicing is an accurate and complex procedure comparable in complexity to protein synthesis and initiation of transcription. It is carried out by a 50S to 60S ribonucleoprotein made up of small nuclear ribonucleoproteins (snRNPs) as well as other proteins. Just as the ribosome is built up in the process of translation, the spliceosome components assemble in an orderly manner on the intron to be spliced before the initial cleavage of the 5' splice site. The splicing must be carried out precisely, joining the 5' end of the preceding exon to the 3' end of that following. A frameshift of even one nucleotide would change the resulting mRNA message. The inescapable conclusion is that these interlocking components must have 'evolved' together, as an imperfect splicing mechanism is worse than none.

Third, were the original protein-coding units seamless, that is, uninterrupted by introns? And were the introns bits of 'selfish DNA' that later insinuated themselves into

the hosts' structural genes? What purpose then the subsequent evolution of a multi-step complicated splicing machinery to remove the introns?<sup>64-69</sup> Would not simply eliminating the introns make better sense for selective advantage?

Fourth, and most importantly, transport of mRNA from the nucleus to the cytoplasm is coupled to splicing and **does not occur until all the splicing is complete**. How does the RNA enter the cytoplasm for translation during the evolution of the splicing mechanism? This would have disrupted protein synthesis and would be powerfully selected against.<sup>70-72</sup> Why is splicing in all its variants so rampant today?

The problem would arise too were introns abundant in cells without nuclear membranes — the prokaryotes. Mattick wrote:

*'If introns were introduced into a procaryotic cell's genes, there would be no opportunity to remove them before protein is made, and the result would be "nonsense" non-functional proteins.'*<sup>73</sup>

This is essentially correct because spliceosomes would be needed for their removal, but again begs the question on the viability of the transitional phases.

The relationships between exons and protein domains remain to be worked out. Where introns came from and how they were integrated into the genome is a mystery to evolutionists.<sup>74</sup>

## THOSE OVERLAPPING CODES

Messenger RNAs generally contain only one reading frame which is dictated by the position of the initiation codon. This correct reading frame translates the nucleotide code into a functional protein. Starting at an AUG codon, translation continues in triplets to a termination codon. The starting point can be altered by a mutation, usually resulting from insertion or deletion of a single nucleotide to give an alternate reading frame. A frameshift error results in the synthesis of a polypeptide that does not resemble the normal product. Typically, it will be inactive and, because stop codons are abundant in the alternative frames, shorter than the native protein.

Some organisms store information in their DNA in the form of overlapping codes. The overlapping codes are still triplet but have different initiation points. In other words, the same stretch of DNA carries the information for producing two proteins of entirely different amino acid sequence. This discovery is truly startling, because the possibility that genes might overlap in different reading frames imposes severe evolutionary constraints. A favourable mutation in one frame must be favourable in the other. A termination codon in the second frame would be fatal to the organism as a whole. So the two overlapping genes have to evolve in parallel. Yockey considered the problem from the point of view of information theory applied to biology, itself a venture fraught with caveats.<sup>75</sup>

In his opinion information theory shows that transcription from two or even three reading frames in a DNA or RNA sequence is possible, provided the total informational content to be transcribed does not exceed the full informational capacity of the DNA or RNA sequence. This interesting bit of information is a necessary but not a sufficient explanation for the origin of overlapping codes. The packing of information for synthesising additional essential proteins through weaving such information into a pre-existing nucleotide sequence is little short of miraculous, assuming that chance is the author.

Most of the known examples of such programmed frameshifts occur in viral genes.<sup>76,77</sup> The notorious hepatitis B virus has four open reading frames on the long strand of its DNA to produce four different proteins. In a striking demonstration of sheer economy it turns out that each reading frame overlaps **at least one other frame**. And the code for the polymerase enzyme overlaps the other three.<sup>78</sup> It is true that programmed frameshifts are not common, but they have been found across a wide spectrum of organisms. Yeast and *E. coli* also practise frameshifting.<sup>79,80</sup> The mechanisms by which they work seem to involve 'shifty' messages in the mRNA, where the ribosomes may read four nucleotides as one amino acid and then continue reading triplets. Or it may back up one base before reading triplets in the new frame. 'Shifty tRNAs' are also implicated.<sup>81-83</sup>

## THE NON-UNIVERSAL CODE

Even the code's universality — a strong argument for the hypothesis that life on Earth evolved only once — has a large number of 'exceptions'. These are usually credited to later evolutionary developments, as the following quote from a paper by Jukes and his colleagues shows. Commenting on the dearth of molecular studies on

*'the more than 10 million species of organisms now living on Earth, all of which are derived from a single pool of the ancestor',*

they continue:

*'... nonuniversal codes have been discovered at a relatively high incidence. Codon UGA Trp has been found in seven Mycoplasma species and related bacteria; at least two kinds of nonuniversal code are independently used in ciliated protozoans; the same code change was found in two different organismic lines, ciliated protozoans, and unicellular green algae; a yeast line uses a still different code. All nonplant mitochondria that have been examined use nonuniversal codes, which are more or less characteristic for each line. It is remarkable that mitochondria from one species use more than two nonuniversal codons; six in yeasts, four or five in many invertebrates, and four in vertebrates. Thus, nonuniversal codes are widely distributed in various groups of organisms and organelles. . . . The*



*nonuniversal codes are not randomly produced, but are derived from the universal code as a result of a series of nondisruptive changes.*<sup>84</sup>

All this just means that hypotheses of the origin of the genetic code based on our understanding of the nature of the DNA, its transcription and translation have to be substantially revised.

### THE SILENT MAJORITY

It is now agreed that any theory on the origin of DNA must take into account that the genomes of multicellular organisms are characterised by high intron content. Mattick has proposed that introns having a high sequence complexity be regarded as informational RNA (iRNA).<sup>85</sup> Each chromosome is increasingly being viewed as a complex 'informational organelle'. At least some now regard the idea that there is 'junk' or 'useless' DNA as untenable,<sup>86</sup> but the logical extensions are not usually followed through.

An unanswered question concerns the enormous amount of DNA in most eukaryotic genomes which appears to serve no useful purpose. Introns contribute to this excess. The highly conserved nature of the sequences in introns points to the possibility that they have served important function(s) from the time of their first appearance in their hosts' genomes. For instance, mouse and human T-cell receptor genes show 71 per cent homology over their entire 100 kb length even though less than six per cent of that length encodes the receptor protein.<sup>87</sup> Recent studies describe finding a RNA regulator of gene expression originating from the introns of another mRNA.<sup>88,89</sup> This small RNA binds to the so-called 3' untranslated region (3'UTR) which lies at the end of each gene's mRNA, once again confounding the notion of 'functionless' RNA.

Intron-containing genes have yet another intriguing property, uncovered in 1992 by Peng and his co-workers in Boston. They introduced a new quantitative method to display correlations in the sequence of nucleotides. To their surprise they discovered that the nucleotide sequence in intron-containing genes is correlated over remarkable ranges of thousands of base pairs apart. Their results are based on a statistical assessment of 24 viral, bacterial, yeast and mammalian sequences. This means that a particular nucleotide at one site would somehow influence which nucleotide would locate at a remotely distant site. This long-range dependence indicates an intricate self-similarity that is reminiscent of fractal dynamics.<sup>90</sup> In addition there are hints of a **language structure**, akin to that seen with ordinary languages, in the lengths of non-protein coding DNA. Their findings support the possibility that non-coding regions of DNA may carry biological information. The two standard linguistic tests applied were those of George Zipf and Claude Shannon. The coding regions of the genes returned negative results for both tests.<sup>91</sup>

Distinctive and previously unsuspected features of genomic DNA are beginning to be revealed. What is surprising is the tiers of immense complexity which are buried in its structure. An analogy will not be out of place. Viewing from a great height a road traversing the length of a continent, a being from outer space might at first wonder what purpose such a structure could serve. Unfamiliar with the ways of man, the alien realises that the ribbon-like structure actually links areas that are intensely bright at night, which are, of course, our cities and towns.

Further study by the alien is even more revealing. The night-bright entities seem to correlate with the lie of the land, its mountain ranges, rivers and underground mineral resources. The alien may even be momentarily distracted by the question of whether the link or the entities came first! What he can conclude, however, is that the structure he had examined is neither random in design nor intention over its whole length, but serves to link entities which themselves evince design and purpose.

What is increasingly seen as the DNA story unfolds is *prima facie* evidence of intelligent design extending over the **whole** molecule. What used to be thought of as a prodigious 95 per cent excess of repetitive and useless DNA turns out to be an interactive regulatory network controlling gene expression in the remaining five per cent. Even the humble trinucleotide repeat sequence CAG has been implicated in the pathogenesis of a number of serious neurological diseases.<sup>92</sup> This illustrates the complicity of the simplest codes in the intricate regulatory network, and puts further strain on ideas of the code's abiotic origin. In summing up, let me quote the editor of *Nature*, who wrote in 1994:

*'The problem of the genetic code has several facets, of which the most compelling is that of why it is why it is . . . it was natural that people should look for an explanation, both for its own sake and because an understanding of how the code evolved must certainly be a pointer to the origin of life itself . . . It was already clear that the genetic code is not merely an abstraction but the embodiment of life's mechanisms; the consecutive triplets of nucleotides in DNA (called codons) are inherited but they also guide the construction of proteins.*

*So it is disappointing, but not surprising, that the origin of the genetic code is still as obscure as the origin of life itself.'*<sup>93</sup>

### THE ORIGIN OF PROTEINS

As with the D-sugars of carbohydrates, so with the amino acids from which proteins are made. They are typically L (left-rotating) in optical activity. D-amino acids are found in bacterial products and peptide antibiotics, but they are not incorporated into proteins via the ribosomal protein synthesising system.

The almost total dominance of one chiral form in present life forms is an enigma. Vital processes such as protein biosynthesis, ligand-receptor activity, substrate binding, enzymatic catalysis and antigen-antibody interaction depend on the present chemical-handedness. Fisun and Savin have provided another example of monochiral utility by examining proton transfer along the hydrogen-bonded chain formed by amino acids.<sup>94</sup> After all, membrane proteins are structured to enable such transfers to take place as a means of regulating proton concentrations. The amino acids they examined were L-tyrosine, L-serine and L-threonine. What would happen, they asked, if a long sequence of such OH-bearing acids were interrupted by an unnatural D isomer? Their analyses revealed that it suppressed transfer through the hydrogen-bonded network. The authors point out the generally disruptive effects that deforming natural polymers with D-amino acids would have on diverse biological phenomena, such as information, charge, energy and mass exchanges.

The evolutionary explanation for left-handed amino acids is simply that a common ancestor, by sheer coincidence, happened to have this mirror image. Well-worn explanations, such as the anisotropic effects of refracted light, are convincing only to those who propose them. 'Chiral fields' that could effect a critical prebiotic transition to one chiral species have been worked out on paper.<sup>95</sup> The trouble is that, so far, there has been no success for the apparently simple problem of tipping the experimental scales to favour one of two isomers.

The problem of chirality is crucial to the origin of life. For Darwinian evolution involves selection, a winnowing process that separates the 'fit' from the 'unfit'. The 'fit' are then amplified to ensure a progeny. The 'fit' are those able to do one of two things, depending on the school of thought. The 'genes first' school envisages primitive replicons that later surrounded themselves with metabolic cycles.<sup>96,97</sup> The 'cells first' school pictures primitive cells covered with primitive membranes engaged in a sort of metabolic exchange with the environment. These propagated themselves by simple expansion followed by division. Genetic mechanisms of inheritance developed gradually.<sup>98,99</sup> Both schools founder on the unsettling and unsettled question — which came first, homochirality or life?<sup>100</sup> If one holds that homochirality came first, it is an admission that without 'left-handed' amino acids and 'right-handed' sugars life's structures and processes would have been impossible. One then has to account for the origin of homochirality. If one assumes that life came first, then one is saying that chirality was not important to the origin of life's structures and processes as we now know them. One has to enter a special pleading for a vastly different metabolism in the 'protobiont', ignoring, for instance, the pivotal role of polypeptide homopolymers in hydrogen-bonded networks for proton and electron transport.<sup>101</sup> One has also to account for the successful transition to homochirality as we have it today.

The logical conclusion from these considerations is a

simple and parsimonious one, **that homochirality and life came together**. But evolutionary lore forbids such a notion. It claims to explain how life began, but on the profound issue of life's 'handedness' there is no selective mechanism that it can plausibly endorse.

## FOLDING PROTEINS

Much thought has been given to suggesting pathways as to how a polypeptide chain, freshly made, folds into its unique shape.<sup>102</sup> But biological systems are inherently complicated and so are their components. Today the concept that proteins can self-assemble has been modified to incorporate the astonishing part played by accessory proteins called **chaperones**, first identified in *E. coli*.<sup>103-107</sup> Chaperones are found in all types of cells and in every cellular compartment. They bind to target proteins to facilitate proper folding, prevent or reverse improper associations, and protect their accidental degradation. Of special interest are a subset of chaperones called chaperonins. They are large, barrel-shaped, polymeric proteins present in bacteria, mitochondria, chloroplasts and eukaryotes. They enfold protein chains in a cavity, a protected micro-environment to allow their guest molecules opportunity to fold correctly. Chaperones utilise the energy of ATP hydrolysis to bind and release their charges. They are also involved in many macromolecular assembly processes, including the assembly of nucleosomes, protein transport in bacteria, assembly of bacterial pili, binding of transcription factors, and ribosome assembly in eukaryotes. A subset of molecular chaperones has even been implicated in signal transduction. This follows upon the discovery that steroid hormone receptors, which are cytoplasmic proteins, combine not only with their respective hormones, but also require chaperones in order to form functioning recycling complexes.<sup>108</sup> Such structural arrangements must be highly conserved, seeing that these chaperones are found in similar macromolecular complexes in organisms as diverse as mammals and yeasts.<sup>109</sup> This is supposed to attest to their great antiquity (if evolution is true), because properly folded proteins are absolutely essential for a cell's viability.

Lodish and his co-authors express their opinion:—

*'Folding of proteins in vitro is inefficient; only a minority undergo complete folding within a few minutes. Clearly, proteins must fold correctly and efficiently in vivo, otherwise cells would waste much energy in the synthesis of non-functional proteins and in the degradation of misfolded and unfolded proteins.'*<sup>110</sup>

How did cellular proteins avoid being tied up into kinks individually and aggregates corporately before chaperones came on the scene? If chaperones help other proteins fold, what mechanism helps chaperones to fold? And chaperones are themselves complex proteins. A well-studied chaperonin, Cpn60, has a unique structure, consisting of fourteen identical subunits of a 60 kDa protein arranged

in two stacked rings of seven.<sup>111,112</sup> It interacts with another conserved protein chaperonin Cpn10, itself a complex of seven subunits.<sup>113</sup> The answers to these questions would indeed be illuminating.

## THE ANCIENT CELLS

### Prokaryotes and Eukaryotes

The existence of chaperones influences the endosymbiont hypothesis of the origin of eukaryotes. This hypothesis proposes that chloroplasts and mitochondria began as free-living aerobic prokaryote ancestors which were engulfed by, and formed, a mutually advantageous relationship with an ancient large anaerobic prokaryote with a nucleus.<sup>114,115</sup> These endosymbionts became the organelles mentioned, which then apparently lost many of their own genes to the nuclei of their hosts. Now, the time-frame of oxygen levels in the primitive Earth is extremely controversial in the face of conflicting palaeobiological evidence.<sup>116</sup> Nevertheless, how a stable relationship between ingested aerobic invaders and an anaerobic, or aerotolerant, host was possible, and why some genes and not others should be transferred to the host's nucleus is not clear.

An idea of how many genes were 'lost' to the host nucleus may be gleaned from the fact that the cytosol synthesises for the mitochondria the following proteins: ribosomal proteins, DNA replication enzymes, aminoacyl-tRNA synthases, RNA polymerase, soluble enzymes of the citric acid cycle and so on.<sup>117</sup> It is clear that, since proteins are made at two separate sites, nuclear-coded proteins must be imported into mitochondria and chloroplasts. This is not made easy by the fact that imported proteins have to cross subcompartments to get into both organelles as the organelles possess double membranes: two subcompartments in the case of mitochondria, three for chloroplasts because of the thylakoid membrane.

Here is where chaperones are needed to bind the polypeptide chains just as they emerge through special pores into the mitochondrial matrix. Assistance with protein folding is given by yet other chaperones near at hand.<sup>118</sup> A similar process operates in the importing of proteins into the chloroplast. As plant cells have both chloroplasts and mitochondria, two different kinds of signal peptides are also required to send proteins to the correct addresses.<sup>119</sup> The very complicated transport arrangements described force us to query how they arose and what selective advantages there could be for original endosymbionts to share genomes with the nucleus of the host cell. As if this is not difficult enough, a further logical and logistical problem is created by the fact that all of the host cell's fatty acids and a number of amino acids are made by enzymes in the chloroplast stroma. We have now a transfer in reverse.<sup>120</sup>

## THE MOST ANCIENT CELL

We are running ahead somewhat because endosymbiosis could only take place when cells with well-developed metabolism were in existence. These were the three prokaryotic lines — the Archaeobacteria, the Eubacteria and those nuclei-bearing prokaryotes destined to initiate the eukaryotic line by acquiring organelles.<sup>121,122</sup> Antedating these three in time was their hypothetical universal ancestor, at the very root of the phylogenetic tree — an anaerobic prokaryote shrouded in mystery, barely surviving on the simplest molecules diffusing in from the surroundings. How simple was its metabolism? A recent textbook suggests that it must be glycolysis.

*'If metabolic pathways evolved by the sequential addition of new enzymatic reactions to existing ones, the most ancient reactions should, like the oldest rings in a tree trunk, be closest to the center of the "metabolic tree", where the most fundamental of the basic molecular building blocks are synthesized. This position in metabolism is firmly occupied by the chemical processes that involve sugar phosphates, among which the most central of all is probably the sequence of reactions known as glycolysis, by which glucose can be degraded in the absence of oxygen (that is, anaerobically). The oldest metabolic pathways would have had to be anaerobic because there was no free oxygen in the atmosphere of the primitive earth.'*<sup>123</sup>

It is extremely unlikely that the earliest cell was such a heterotroph 'feeding' on organic compounds such as acids and sugars. Many strictly anaerobic bacteria today break down glucose through the Entner-Doudoroff pathway. This pathway comprises more than six enzymes acting in sequence and is therefore rather advanced for the rudimentary first cell.

If the specific qualities of the ancestor are to reflect the geothermal environment it occupied it should be a thermophilic autotroph, that is, a heat-tolerant cell subsisting on the simplest compounds. It happens that the Archaeobacteria of today inhabit environments of extreme heat or salinity or acidity. They can utilise (fix) CO<sub>2</sub>, although not by the Calvin cycle, as in most photosynthetic organisms. Indeed, current belief is that the closest to a prototype of the earliest cell are those Archaeobacteria that are completely anaerobic, with inorganic electron acceptors, and which use H<sub>2</sub> and CO<sub>2</sub> as sole reductant and carbon source, respectively.<sup>124</sup> These cells called chemolithotrophs are (were) able to extract energy and synthesise their cellular constituents from simple molecules such as SO<sub>4</sub><sup>2-</sup>, S<sub>2</sub>, H<sub>2</sub> and CO<sub>2</sub>. For most anaerobic Archaeobacteria, CO<sub>2</sub> can be used as the sole carbon source for growth, and acetyl-CoA is the central biosynthetic intermediate or 'building block' for other molecules. The formation of acetyl-CoA requires two molecules of CO<sub>2</sub>, a nickel enzyme complex and other cofactors. Furthermore, pyruvate obtained from the breakdown of

glucose is converted to acetyl-CoA by a thiamine-pyrophosphate (TPP) enzyme called pyruvate oxidoreductase.<sup>125</sup>

The recruitment of coenzymes such as TPP so early in evolution is puzzling. Recently, Keefe and his colleagues attempted the successful synthesis of pantetheine, a precursor to coenzyme A, presuming the abundance of the precursor molecules on the primitive Earth. Heating pantetheine with ATP or ADP failed to produce the dephosphocoenzyme A.<sup>126,127</sup> All things considered, a chemolithotroph, whether ancient or modern, is anything but simple for the kinds of enzymes and metabolic pathways it possesses.

### KEY POINTS

- How deoxyribonucleic acid (DNA) sequence integrity could have been maintained in the absence of the many enzymes which continually scan and replace missing, incorrect and damaged nucleotides has not been satisfactorily explained.
- The amount of DNA in species does not correlate consistently with organism complexity.
- Exon shuffling creates problems in molecular phylogeny.
- The numerous components involved in RNA splicing must have all appeared simultaneously to be advantageous because a partially complete mechanism would function detrimentally.
- Introns introduced into a prokaryotic cell's genes would have no opportunity to be removed before protein is made, resulting in 'nonsense' non-functional proteins.
- The weaving of information coding for one polypeptide into an existing nucleotide sequence coding for another imposes severe evolutionary constraints.
- The universality of the genetic code — a strong argument that all organisms are derived from a single ancestor — in fact has many exceptions.
- Intron sequences correlate over remarkable ranges of thousands of base pairs, strongly suggesting they are functional.
- It has not been explained how proteins could have managed to fold correctly in the absence of chaperones — themselves complex proteins.
- In hypotheses involving the incorporation of a prokaryote to account for organelles such as mitochondria, it is not clear how a stable relationship between anaerobic invaders and an aerobic or aerotolerant host was possible or why some genes and not others should be transferred to the host's nucleus.
- Current attempts to root the phylogenetic tree of life are based on relatively simple and therefore unrealistic models of evolution.
- Accidental assembly of a self-replicating molecule now has so many qualifications that its scientific integrity is questionable.

### REPRISE

Evolution is **biology as a historical science**.<sup>128</sup> Evolutionists seek to unravel the tangled strands of hypothetical ancient life forms assumed to have developed over billions of years. In so doing they hope to learn the secret of that most profound of scientific enigmas, namely, the origin of life.

The driving forces for the enterprise are two: the fossil record of cellular structures, and the reasonable inference that nucleotide and protein molecular changes over time should enable their ancestral lineages to be traced.

Of the first, there is the hard evidence for the presence of Precambrian stromatolites. This indicates that cells identical to modern cyanobacteria were thriving at 3.5 Ga.<sup>129-132</sup> This and the discovery of the algal fossil *Grypania*<sup>133</sup> support the most ancient dates for the origin of fully-developed cells and have skewed the current opinion on the oxygen content of the primitive atmosphere towards higher values.<sup>134</sup> Strong support also comes from the studies of Schidlowski on the fractionation of the carbon isotopes in the waxy carbon polymers (kerogens) of Archaean sediments. In photosynthesis, somewhat more of the lighter <sup>12</sup>CO<sub>2</sub> is fixed in slight preference to the heavier <sup>13</sup>CO<sub>2</sub>. Enrichment of <sup>12</sup>C with respect to <sup>13</sup>C in kerogens extracted from 3.8 Ga rocks is evidence that photosynthetic life must have been around for almost 4 Ga.<sup>135</sup>

The time available for the origin of the cell has shrunk to one-tenth or less than has been assumed.<sup>136,137</sup> There now seems to be little or no time for the genesis of the anaerobic first cell — the progenote of the RNA world.<sup>138</sup>

Turning now to rooting the phylogenetic tree of life, investigators in the field have voiced concern over attempts to do this and plead for greater understanding of phylogenetic methods. Only recently, Hillis and Huelsenbeck caution that

*'current phylogenetic implementations of maximum likelihood are limited to relatively simple and therefore unrealistic models of evolution.'*<sup>139</sup>

At the same time workers in Canada and Switzerland have commented on uncertainties of trying to work out phylogenies using both parsimony and maximum-likelihood methods.<sup>140,141</sup>

The current belief that life's ancestral lineage is through the Archaeobacteria also faces major unsolved problems with rooting the tree, as witness the following opinions:

*'However, using protein phylogeny to root the tree of life is not safe; besides the possibility of lateral gene transfer, one cannot be sure that proteins compared in an individual tree descend from a single gene in the common ancestor, or from already duplicated genes.'*<sup>142</sup>

Doolittle laments the fact that there is *'still profound disagreement among different kinds of biologists about what a phylogenetic taxonomy is.'*<sup>143</sup>

In conclusion, molecular biology in recent years has

revealed previously unimagined levels of sophistication in the details of subcellular organisation and function.<sup>144-149</sup> The available evidence from the field and the laboratory is not amicable to the theory that life began with the accidental assembly of a self-replicating molecule. It is now accepted with so many qualifications that its scientific integrity, even as a heuristic device, is questionable.

## REFERENCES

- Jones, J. H. and Drake, M. J., 1986. Geochemical constraints on core formation in the Earth. *Nature*, **322**:221–228.
- Towe, K. M., 1990. Aerobic respiration in the Archaean? *Nature*, **348**:54–56.
- Bowring, S. A. and Housh, T., 1995. The Earth's early evolution. *Science*, **269**:1535–1540.
- Widdel, F., 1993. Ferrous iron oxidation of anoxygenic phototrophic bacteria. *Nature*, **362**:834–836.
- Han, T.-M. and Runnegar, B., 1992. Megascopic eukaryotic algae from the 2.1-billion-year-old Negaunee Iron-Formation, Michigan. *Science*, **257**:232–235.
- Lowe, D. R., 1980. Stromatolites 3,400-Myr old from the Archaean of Western Australia. *Nature*, **284**:441–443.
- Walter, M. R., Buick, R. and Dunlop, J. S. R., 1980. Stromatolites 3,400-3,500 Myr old from the North Pole area, Western Australia. *Nature*, **284**:443–445.
- Orpen, J. L. and Wilson, J. F., 1981. Stromatolites at 3,500 Myr and a greenstone-granite unconformity in the Zimbabwean Archaean. *Nature*, **291**:218–220.
- Byerly, G. R., Lowe, D. R. and Walsh, M. M., 1986. Stromatolites from the 3,300-3,500 Myr Swaziland Group, Barberton Mountain Land, South Africa. *Nature*, **319**:489–491.
- Riding, R., 1992. The algal breath of life. *Nature*, **359**:13–14.
- Voet, D. and Voet, J. G., 1995. *Biochemistry*, John Wiley and Sons, Inc., New York, p. 21.
- Darnell, J., Lodish, H. and Baltimore, D., 1990. *Molecular Cell Biology*, second edition, Scientific American Books, distributed by W. H. Freeman and Co., New York, pp. 1049–1071.
- Lodish, H., Baltimore, D., Berk, A., Zipursky, S. L., Matsudaira, P. and Darnell, J., 1995. *Molecular Cell Biology*, third edition, Scientific American Books, distributed by W. H. Freeman and Co., New York, p. 9.
- Stryer, L., 1995. *Biochemistry*, fourth edition, W. H. Freeman and Co., New York.
- Orgel, L. E., 1994. The origin of life on the Earth. *In: Life in the Universe*, *Scientific American*, **271**(4):53–61.
- Orgel, Ref. 15, p. 56.
- Sagan, C., 1994. The search for extraterrestrial life. *In: Life in the Universe*, *Scientific American*, **271**(4):71–77.
- Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J. D., 1994. *Molecular Biology of the Cell*, third edition, Garland Publishing Inc., New York and London, p. 4.
- Mason, S., 1985. Origin of biomolecular chirality. *Nature*, **314**:400–401.
- Miller, S. L. and Bada, J. L., 1988. Submarine hot springs and the origin of life. *Nature*, **334**:609–611.
- Lowe, Ref. 6.
- Walter *et al.*, Ref. 7.
- Orpen and Wilson, Ref. 8.
- Byerly *et al.*, Ref. 9.
- Han and Runnegar, Ref. 5.
- Riding, Ref. 10.
- Widdel, Ref. 4.
- Raup, D. M. and Valentine, J. W., 1983. Multiple origins of life. *Proceedings of the National Academy of Science USA*, **80**:2981–2984.
- Compton, W. and Pidgeon, R. T., 1986. Jack Hills, evidence of more very old detrital zircons in Western Australia. *Nature*, **321**:766–769.
- Schidlowski, M., 1988. A 3,800-million-year isotopic record of life from carbon in sedimentary rocks. *Nature*, **333**:313–318.
- Hoyle, F., 1983. *The Intelligent Universe*, Michael Joseph, London, p. 16.
- Sagan, Ref. 17, p. 75.
- Sharp, P. A., 1994. Ribozymes. *Current Opinions in Structural Biology*, **4**:322–330.
- Lewin, R., 1986. RNA catalysis gives fresh perspective on the origin of life. *Science*, **231**:545–546.
- Cech, T. R., 1990. Self-splicing of group I introns. *Annual Review of Biochemistry*, **59**:543–568.
- Steitz, J. A., 1988. 'Snurps'. *Scientific American*, **258**(6):56–63.
- Joyce, G. F., 1989. RNA evolution and the origins of life. *Nature*, **338**:217–224.
- Robertson, M. P. and Miller, S. L., 1995. An efficient prebiotic synthesis of cytosine and uracil. *Nature*, **375**:772–774.
- Joyce, Ref. 37, p. 221.
- Lohrmann, R. and Orgel, L. E., 1971. Urea-inorganic phosphate mixtures as prebiotic phosphorylating agents. *Science*, **171**:490–494.
- Orgel, L. E., 1992. Molecular replication. *Nature*, **358**:203–209.
- Ferris, J. P., Hill Jr., A. R., Liu, R. and Orgel, L. E., 1996. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature*, **381**:59–61.
- Joyce, Ref. 37, pp. 222, 223.
- Orgel, Ref. 15, p. 61.
- Nielson, P. E., Egholm, M., Berg, R. H. *et al.*, 1991. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science*, **254**:1497–1500.
- Egholm, M., Buchardt, O., Christensen, L. *et al.*, 1993. PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen bonding rules. *Nature*, **365**:566–568.
- Böhler, C., Nielsen, P. E. and Orgel, L. E., 1995. Template switching between PNA and RNA oligonucleotides. *Nature*, **376**:578–581.
- Joyce, Ref. 37.
- Nordlund, P. and Eklund, H., 1993. Structure and function of the *Escherichia coli* ribonucleotide reductase protein R2. *Journal of Molecular Biology*, **232**:123–164.
- Sjöberg, B. M., 1994. The ribonucleotide reductase jigsaw puzzle: a large piece falls into place. *Structure*, **2**:793–796.
- Mao, S. S., Holler, T. P., Yu, G. X. *et al.*, 1992. A model for the role of multiple cysteine residues involved in ribonucleotide reduction: amazing and still confusing. *Biochemistry*, **31**:9733–9743.
- Reichard, P., 1993. From RNA to DNA, why so many ribonucleotide reductases? *Science*, **260**:1773–1777.
- Lindahl, T. and Karlström, O., 1973. Heat-induced depyrimidation of deoxyribonucleic acid in neutral solution. *Biochemistry*, **12**:5151–5154.
- Lindahl, T., 1993. Instability and decay of the primary structure of DNA. *Nature*, **362**:709–715.
- Koshland, D. E., Jr., 1994. Molecule of the year: the DNA repair enzyme. *Science*, **266**:1925–1926.
- Fischman, J., 1995. Have 25-million-year-old bacteria returned to life? *Science*, **268**:977.
- Lodish *et al.*, Ref. 13, p. 312.
- Dreyfuss, G., Matunis, K. J., Piñol-Roma, S. and Burd, C. G., 1993. hnRNP proteins and the biogenesis of mRNA. *Annual Review of Biochemistry*, **62**:289–321.
- Gilbert, W., 1985. Genes-in-pieces revisited. *Science*, **228**:823–824.
- Cech, T. R., 1986. RNA as an enzyme. *Scientific American*, **255**:64–75.
- Benne, R., 1996. RNA editing: how a message is changed. *Current Opinion in Genetics and Development*, **6**:221–231.
- Patterson, C., 1987. *Molecules and Morphology in Evolution: Conflict or Compromise?*, Cambridge University Press, Cambridge, UK.
- Doolittle, R. F., 1986. *Of Urfs and Orfs*, A primer on how to analyze derived amino acid sequences, University Science Books, California, p. 38.
- Green, M. R., 1991. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. *Annual Review of Cell Biology*, **7**:559–599.
- Moore, M. J. and Sharp, P. A., 1993. Evidence for two active sites in spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature*, **365**:364–368.
- McKeown, M., 1993. The role of small nuclear RNAs in RNA splicing. *Current Opinions in Cell Biology*, **5**:448–454.
- Nilsen, T. W., 1994. RNA-RNA interactions in the spliceosome: unravelling the ties that bind. *Cell*, **78**:1–4.
- Horowitz, D. S. and Krainer, A. R., 1994. Mechanisms for selecting 5' splice sites in mammalian pre-mRNA splicing. *Trends in Genetics*, **10**:100–105.
- Sharp, P. A., 1994. Nobel lecture: split genes and RNA splicing. *Cell*, **77**:805–815.
- Green, M. R., 1989. Pre-mRNA processing and mRNA nuclear export. *Current Opinions in Cell Biology*, **1**:519–525.

71. Spector, D. L., 1993. Nuclear organization of pre-mRNA processing. **Current Opinions in Cell Biology**, 5:442–447.
72. Jimenez-Garcia, L. F. and Spector, D. L., 1993. *In vivo* evidence that transcription and splicing are co-ordinated by a recruiting mechanism. **Cell**, 73:47–59.
73. Mattick, J. S., 1994. Introns: evolution and function. **Current Opinions in Genetic Developments**, 4:823–831.
74. Hurst, L. D., 1994. The uncertain origin of introns. **Nature**, 371:381–382.
75. Yockey, H. P., 1979. Do overlapping genes violate molecular biology and the theory of evolution? **Journal of Theoretical Biology**, 80:21–26.
76. Moore, R., Dixon, M., Smith, R., Peters, G. and Dickson, C., 1987. Complete nucleotide sequence of a milk-transmitted mouse mammary tumour virus: two frameshift suppression events are required for translation of gag and pol. **Journal of Virology**, 61:480–490.
77. Jacks, T., Madhani, H. D., Masiarz, F. R. and Varmus, H. E., 1988. Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. **Cell**, 55:447–458.
78. Tiollais, P., Purcell, D. and Dejean, A., 1985. The hepatitis B virus. **Nature**, 317:489–495.
79. Wilson, W., Malim, M. H., Mellor, J., Kingsman, A. J. and Kingsman, S. M., 1986. Expression strategies of the yeast retrotransposon Ty: a short sequence directs ribosomal frameshifting. **Nucleic Acid Research**, 14:7001–7016.
80. Craigen, W. J. and Caskey, C. T., 1986. Expression of peptide chain release factor 2 requires high-efficiency frameshift. **Nature**, 322:273–275.
81. Hatfield, D. and Oroszlan, S., 1990. The where, what and how of ribosomal frameshifting in retroviral protein synthesis. **Trends in Biochemical Science**, 15:186–190.
82. Weiss, R. B., 1991. Ribosomal frameshifting, jumping and read through. **Current Opinions in Cell Biology**, 3:1051–1055.
83. Gesteland, R. F., Weiss, R. B. and Atkins, J. F., 1992. Recoding: reprogrammed genetic coding. **Science**, 257:1640–1641.
84. Osawa, S., Jukes, T. H., Watanabe, K. and Muto, A., 1992. Recent evidence for the evolution of the genetic code. **Microbiology Review**, 56:229–264.
85. Mattick, Ref. 73, p. 827.
86. Nowak, R., 1994. Mining treasures from ‘junk DNA’. **Science**, 263:608–610.
87. Koop, B. F. and Hood, L., 1994. Striking sequence similarity over almost 100 kilobases of human and mouse T-cell receptor DNA. **Nature Genetics**, 7:48–53.
88. Lee, R. C., Feinbaum, R. L. and Ambros, V., 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. **Cell**, 75:843–854.
89. Wightman, B., Ha, I. and Ruvkin, G., 1993. Post-transcriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal formation in *C. elegans*. **Cell**, 75:855–862.
90. Peng, C.-K., Buldyrev, S. V., Goldberger, A. L., Havlin, S., Sciortino, F., Simons, M. and Stanley, H. E., 1992. Long-range correlations in nucleotide sequences. **Nature**, 356:168–170.
91. Mantegna, R. N., Buldyrev, S. V., Goldberger, A. L., Havlin, S., Peng, C.-K., Simons, M. and Stanley, H. E., 1994. **Physical Review Letters**, 73:3169–3172.
92. Warren, S. T., 1996. The expanding world of trinucleotide repeats. **Science**, 271:1374–1375.
93. Maddox, J., 1994. The genetic code by numbers. **Nature**, 367:111.
94. Fisun, O. I. and Savin, A. V., 1992. Homochirality and long-range transfer in biological systems. **BioSystems**, 27:129–135.
95. Avetisov, V. A. and Goldanskii, V. I., 1993. Chirality and the equation of the ‘biological big bang’. **Physical Letters**, A172:407–410.
96. Schuster, P., 1991. Molecular evolution as a complex optimization problem. *In: Self-Organization, Emerging Properties, and Learning*, A. Babloyantz (ed.), NATO ASI Series B260, Plenum, New York, pp. 241–254.
97. Eigen, M., 1992. **Steps Towards Life: A Perspective on Evolution** (with R. Winkler-Oswatitch), Oxford University Press, Oxford.
98. Fox, S. W., 1986. Molecular selection and natural selection. **Quarterly Reviews in Biology**, 61:375–386.
99. Fleischaker, G. R., 1990. Origins of life: an operational definition. **Origins of Life, Evolution and the Biosphere**, 20:127–137.
100. Cohen, J., 1995. Getting all turned around over the origins of life on earth. Report on ‘The origin of homochirality in life’, February 15–17, 1995. **Science**, 267:1265–1266.
101. Fisun and Savin, Ref. 94.
102. Weissman, J. S., 1995. All roads lead to Rome? The multiple pathways of protein folding. **Chemistry and Biology**, 2:255–260.
103. Ellis, R. J. and Hemmingsen, S. M., 1989. Molecular chaperones: proteins essential for the biogenesis of some macromolecular structures. **Trends in Biochemical Science**, 14:339–342.
104. Ellis, R. J., 1990. Molecular chaperones: the plant connection. **Science**, 250:954–959.
105. Matthews, C. R., 1993. Pathways of protein folding. **Annual Review of Biochemistry**, 62:653–684.
106. Rassow, J. and Pfanner, N., 1996. Protein biogenesis: chaperones for nascent polypeptides. **Current Biology**, 6:115–118.
107. Hartl, F.-U., Hlodan, R. and Langer, T., 1994. Molecular chaperones in protein folding: the art of avoiding sticky situations. **Trends in Biochemical Science**, 19:20–25.
108. Kimura, Y., Yahara, I. and Lindquist, S., 1995. Role of the protein chaperone YDJ1 in establishing Hsp90-mediated signal transduction pathways. **Science**, 268:1362–1365.
109. Smith, D. F. and Toft, D. D., 1993. Steroid receptors and their associated proteins (review). **Molecular Endocrinology**, 7:4–11.
110. Lodish *et al.*, Ref. 13, pp. 74–75.
111. Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A. L. and Hartl, F.-U., 1991. Chaperonin-mediated protein folding at the surface of GroEL through a ‘molten-globule’-like intermediate. **Nature**, 352:36–42.
112. Brag, K., Hainfield, J., Simon, M., Furuya, F. and Harowich, A. L., 1993. Polypeptide bound to the chaperonin GroEL within a central cavity. **Proceedings of the National Academy of Science, USA**, 90:3978–3982.
113. Gray, T. E. and Fersht, A. R., 1991. Co-operativity in ATP hydrolysis by GroEL increased by GroES. **FEBS Letters**, 292:254–258.
114. Margulis, L., 1981. **Symbiosis in Cell Evolution**, W. H. Freeman and Co.
115. Perna, N. T. and Kocher, T. D., 1996. Mitochondrial DNA molecular fossils in the nucleus. **Current Biology**, 6:128–129.
116. Knoll, A. H., 1992. The early evolution of eukaryotes: a geological perspective. **Science**, 256:622–627.
117. Alberts *et al.*, Ref. 18, p. 716.
118. Pfanner, N. and Meijer, M., 1995. Pulling in the proteins. **Current Biology**, 5:132–135.
119. Watson, M. D. and Murphy, D. J., 1995. Genome organisation, protein synthesis and processing in plants. *In: Plant Biochemistry and Molecular Biology*, J. L. Peter and R. C. Leegood (eds), John Wiley and Sons, pp. 197–219.
120. Alberts *et al.*, Ref. 18, p. 696.
121. Woese, C. R., 1987. Bacterial evolution. **Microbiology Reviews**, 51:221–271.
122. Woese, C. R., Kandler, O. and Wheelis, M. L., 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eucarya. **Proceedings of the National Academy of Sciences, USA**, 87:4576–4579.
123. Alberts *et al.*, Ref. 18, pp. 13, 14.
124. Fuchs, G., Ecker, A. and Strauss, G., 1992. Bioenergetics and autotrophic carbon metabolism of chemolithotrophic archaeobacteria. *In: The Archaeobacteria: Biochemistry and Biotechnology*, M. J. Danson, D. W. Hough and G. G. Lunt (eds), Biochemical Society Symposium No. 58, Portland Press, London and Chapel Hill, pp. 23–39.
125. Danson, M. J. and Hough, D. W., 1992. The enzymology of archaeobacterial pathways of central metabolism. *In: The Archaeobacteria: Biochemistry and Biotechnology*, M. J. Danson, D. W. Hough and G. G. Lunt (eds), Biochemical Society Symposium No. 58, Portland Press, London and Chapel Hill, pp. 7–21.
126. Keefe, A. D., Newton, G. L. and Miller, S. L., 1995. A possible prebiotic synthesis of pantetheine, a precursor to coenzyme A. **Nature**, 373:683–685.
127. Ferris, J. P., 1995. Life at the margins. **Nature**, 373:659.
128. Lodish *et al.*, Ref. 13, p. 4.
129. Lowe, Ref. 6.
130. Walter *et al.*, Ref. 7.
131. Orpen and Wilson, Ref. 8.
132. Byerly *et al.*, Ref. 9.
133. Han and Runnegar, Ref. 5.
134. Riding, Ref. 10.
135. Schidlowski, Ref. 30.
136. Quastler, H., 1964. **The Emergence of Biological Organization**, Yale University Press, New Haven, London.
137. Shklovskii, I. S. and Sagan, C., 1966. **Intelligent Life in the Universe**, Dell

- Publications, New York.
138. Woese, Ref. 121.
139. Hillis, D. M. and Huelsenbeck, J. P., 1995. Assessing molecular phylogenies. **Science**, **267**:255–256.
140. Schluter, D., 1995. Uncertainty in ancient phylogenies. **Nature**, **377**:108–109.
141. Benner, S. A., Jermann, T. M., Opitz, J. G., Stackhouse, J., Knecht, L. J. and Gonnet, G. H., 1995. Uncertainty in ancient phylogenies. **Nature**, **377**:109–110.
142. Forterre, P., Charbonnier, F., Marguet, E., Harper, F. and Henckes, G., 1992. Chromosome structure and DNA topology in extremely thermophilic archaeobacteria. In: **The Archaeobacteria: Biochemistry and Biotechnology**, M. J. Danson, D. W. Hough and G. G. Lunts (eds), Biochemical Society Symposium No. 58, Portland Press, London and Chapel Hill, pp. 99–112.
143. Doolittle, W. F., 1992. What are the archaeobacteria and why are they important? In: **The Archaeobacteria: Biochemistry and Biotechnology**, M. J. Danson, D. W. Hough and G. G. Lunt (eds), Biochemical Society Symposium No. 58, Portland Press, London and Chapel Hill, pp. 1–6.
144. Riding, Ref. 10.
145. Darnell *et al.*, Ref. 12.
146. Lodish *et al.*, Ref. 13.
147. Stryer, Ref. 14.
148. Orgel, Ref. 15.
149. Watson and Murphy, Ref. 119.

---

**Dr Aw Swee-Eng**, M.B., B.S., Ph.D. (London), FRC Path., MIBiol. (London), was Associate Professor of Biochemistry at the University of Singapore until 1978, and is now head of the Department of Nuclear Medicine and Director of Clinical Research at Singapore General Hospital. Author of around 30 technical papers in his field of biochemistry and nuclear medicine, he has authored a critique of origin-of-life theories entitled **Chemical Evolution — An Examination of Current Ideas**.

---

#### QUOTABLE QUOTE: The Origin of the Universe

*‘What is a big deal — the biggest deal of all — is how you get something out of nothing.  
Don’t let the cosmologists try to kid you on this one. They have not got a clue either — despite the fact that they are doing a pretty good job of convincing themselves and others that this is really not a problem. “In the beginning,” they will say, “there was nothing — no time, space, matter or energy. Then there was a quantum fluctuation from which . . .” Whoa! Stop right there. You see what I mean? First there is nothing, then there is something. And the cosmologists try to bridge the two with a quantum flutter; a tremor of uncertainty that sparks it all off. Then they are away and before you know it, they have pulled a hundred billion galaxies out of their quantum hats.’*

Darling, David, 1996. On creating something from nothing. **New Scientist**, **151**(2047):49.