Biochemical Evolution: A Synopsis

E. N. WARDLE, MD, MRCP

Department of Medicine, Royal Victoria Infirmary, Newcastle-upon-Tyne

The age of the universe is about 12 billion years, that of the solar system 5 billion years, and the age of the earth could be $4.8 \times 10^9$ years. The earth originated, it seems, from a globe of hot gases passing into a liquid stage in which the orb became covered by a crust, a process akin to the smelting of an ore; in the case of the earth the molten nickel-iron interior extruded the rock matrix which rose to the surface to form a crust. The earth’s crust consists of a lower shell of basalt, a fine-grained basic rock found in lava flows, and on top of this a discontinuous layer of the acid igneous rock granite. Granite underlies the oceans and the continents. As a world map indicates, the continental tectonic plates have been on the move over the ages. For many ages the earth was surrounded by steam but when its surface cooled, fresh water seas formed and became saline as salts were leached out of the rocks.

The earliest time from which fossils survive is known as the Cambrian, because such rocks were found in Wales on the Harlech dome, but pre-Cambrian remains of invertebrates like hydra and annelid worms have now been found in South Australia. In that pre-Cambrian era one can imagine landscapes of hot and cold deserts, volcanoes and lava fields. The unoxidised state of these older rocks shows that the atmosphere was deficient in oxygen. 'De-gassing' from the centre of the earth by volcanic action would have produced a gaseous mixture of methane, ammonia, nitrogen, and carbon monoxide and dioxide.

Life did not then beget life. Nor can we now suppose, as did the ancients, that there was spontaneous generation from lifeless matter, because F. Redi, the Italian naturalist, and Louis Pasteur have long since disproved that theory. All we can guess is that at some stage the distinction between living and non-living must have been blurred. Like Darwin, we can only suppose that in a warm pond of ammonium and phosphoric salts organic compounds formed which led ultimately to proteins with auto-catalytic properties. Haldane in 1928 suggested that ultra-violet light would act on the 'hot dilute soup' of the primordial oceans to form organic compounds; ultra-violet light would be essential, because the simple molecules can only absorb short-wave radiation. This would supply the 'free energy' for endergonic processes. Lightning could also have played a part, for nitrogen gas is inert, being near the condition of minimum free energy, but Fritz Haber was able to begin synthetic nitrogen fixation by combining nitrogen with hydrogen to form ammonia at elevated temperature and pressure. As Schrödinger (1944) stated, 'living matter avoids the decay to equilibrium, which in an isolated
system is the state of maximum entropy. Free energy must be supplied to stop a living system fading away to the equilibrium state'. Animals do this by using energy-rich matter, namely food. That energy originates from solar radiation.

Implicit, then, in the organisation of matter into a less random state is the need for a force. Otherwise energy that is transferred between units of matter will follow the course of least resistance towards disorder. It has been argued that the time taken to transform inorganic to organic compounds in the primitive atmosphere could have been long enough to allow the transition of the improbable into the probable. The chance of being struck by lightning in a lifetime is small but over 700 million years that chance becomes a certainty! Once the amino acids were formed, it seems reasonable to suppose that the hydrophobic forms would soon lead to the formation of insoluble polypeptides; indeed, the formation of insoluble organic polymers in contact with water is yet another expression of the search for a lower free energy level.

Mechanisms of pre-biotic synthesis of organic compounds have been studied by means of the effects of electrical discharges on mixtures of gases such as methane-ammonia-nitrogen or of carbon monoxide, nitrogen and hydrogen. In this way amino acids are formed. When an electron beam, used to simulate the effect of potassium-40, was used to irradiate methane, ammonia and water in combination, the result was the formation of the nitrogen base adenine. Similarly, hydrogen cyanide can be turned by photochemical reaction into adenine and guanine. When water vapour interacts with carbides in the magma, brought to the surface of the earth by volcanic activity, organic carbon compounds are formed:

\[ 3 \text{Fe}_2\text{C}_n + 4 \text{mH}_2\text{O} = \text{Fe}_3\text{O}_4 + 3 \text{C}_n\text{H}_{8\text{m}} \]

Formaldehyde (H. CHO), which can be regarded as the simplest of the sugars, must have been present in the primitive atmosphere; by a series of condensations it will form pentose and hexose sugars (Ponnampерuma, 1971). The photochemical reduction of carbon dioxide to formaldehyde, followed by formation of glyceraldehyde and dihydroxy-acetone, is recognised in present-day microorganisms as the Calvin-Benson cycle. Formaldehyde will also combine with ammonia and hydrogen cyanide in a reaction called the Strecker synthesis to form amino acids.

The oxygenless conditions on the primitive earth would clearly favour the reduction that is required for carbon and nitrogen to become incorporated as amino acids, purines and pyrimidines. The fact that all organisms eliminate ammonia as urea bears witness to the fact that life evolved in the reducing atmosphere of primitive times. Urea could originally have been formed in large amounts from ammonia and cyanate. However, the instability of many similar compounds suggests that a low temperature must also have been necessary to achieve the stability for macromolecule formation. Although chemical syntheses would be slow, decomposition would be minimal.
The equilibrium of hydrolysis is such that with a modest concentration of amino acids in water some peptides will appear. Certain substances will act as catalysts; for example, clay-like materials. Moreover, when an aqueous solution of glycine and leucine is exposed to ultra-violet light in the presence of cyanamide, dipeptides and even tripeptides like leucyl-glycyl-glycine can be formed. Glycine, alanine, aspartic acid, glutamic acid and serine are all formed if an electrical discharge passes through ammonia together with methane in water; some 15 amino acids can be produced in this way, as might have been the case on the primitive sea shore or the dried up bed of a lagoon.

Similarly, if nucleosides (nitrogen base — pentose sugar molecules) were formed in an aqueous solution, under arid conditions these nucleosides would combine with phosphates to form nucleotides, as during evaporation by solar heat. At low temperatures inorganic orthophosphates can condense to form polyphosphates, which could then be phosphorylating and condensing agents for primitive syntheses. The alternation between high and low temperatures might either be the change from day to night, or, on an evolutionary time scale, from hot desert conditions to the ice ages.

About 800 million years ago, in the late pre-Cambrian era, there was an explosive evolution of life, coinciding with the appearance of oxygen in the atmosphere. This oxygen could have been created by the interaction of ultra-violet light with water vapour in the upper atmosphere, with a consequent release of hydrogen into upper space, but is more likely to have been the outcome of the photosynthetic processes of bacteria and the blue-green algae. These bacteria and blue-green algae are prokaryotes (pre-nuclear cells) whose DNA is not confined to an organised nucleus. Two hundred million years later, the trilobites of Shropshire were scuttling around in the mud, such was the rapidity and diversity of the expansion of life.

In this time organisms had learned to protect themselves from the effects of ‘oxygen toxicity’. The vitamin B flavin co-enzymes and the protein sulphydryl groups are particularly sensitive to oxygen and hydrogen peroxide. But the presence of oxygen in the atmosphere led to the development of oxidative phosphorylation as the means of energy production. Henceforth the oxidation of glucose to carbon dioxide and water could be used by animals to recover the energy that had been stored by plants as part of photosynthesis.

We must presume, as did Calvin (1959), that porphyrin molecules could have become available early in the evolution of life, because basically they are formed in the sequence

\[
\text{sucinate} + \text{glycine} \rightarrow \text{b-amino-laevulinic acid} \rightarrow \text{protoporphyrin (iron carrying)}.
\]

Random variations in the porphyrins must have led to the construction of chlorophyll and then to the development of other respiratory pigments, including
the cytochromes, and later, with the need to carry oxygen in the blood, to the haemoglobins. The cytochromes form the electron transport chain whereby electrons or hydrogen atoms derived from food molecules during biological oxidation eventually unite with oxygen to form water; in so doing energy is channelled off at three separate steps to give the energy-rich molecule ATP.

The organisation of matter into structural units seems to have arisen from the inevitable fact that hydrophobic amino acids will form insoluble polypeptides and will be attracted to lipid-water interfaces and thus become associated with lipid droplets. In 1965 the Russian investigator, A. Oparin, postulated that the transition from chemical to biological evolution depended on the formation of membranes around droplets, to form microcells. The technical term is a 'co-acervate', which is an aggregate of hydrophilic particles surrounded by a shell of water molecules. Such lipid particles would rise to the surface of water and be driven to the shoreline in ever-growing numbers so that the chance for interactions would increase. The change from a co-acervate to a living complex probably occurred some two billion years ago. The microcell theory obviates the chicken and egg question of whether proteins or nucleic acids were the first template or informational molecules.

In the molecular aggregates the original peptides would have had a great variety of amino acid sequences. However, if certain sequences were more capable of producing copies of themselves, they would have a selective advantage and so would be better at producing progeny. Certain metals would act as catalysts, and particularly efficient catalytic polypeptides would evolve into enzymes. It is quite possible that acidic and basic proteinoid in the presence of magnesium ions would convert ATP into polynucleotides, and also use ATP for the formation of peptide bonds. The microcells would serve to concentrate intermediate chemicals to physico-chemically favourable concentrations. Very likely the polypeptides were the first templates but from them would soon arise nucleic acid templates. It is known that polyarginine has a particular affinity for guanosine monophosphate and a lesser one for adenosine monophosphate.

Watson and Crick's 'central dogma' stipulated that the double helical strand of DNA carries the inherited characteristics as nucleotide sequences called genes. One strand winds round the other in a helix. Adenine to thymine and guanine to cytosine are always complementary base pairs (Watson and Crick, 1953). For the purpose of protein synthesis an informational molecule called messenger RNA is 'transcribed' from one strand of the DNA helix, and this molecule leaves the nucleus to instruct the protein synthesising units in the cytoplasm (the ribosomes) as to how they should join the various amino acids together to form that particular protein. In fact the m-RNA winds round the ribosome and is there 'translated'. The complementary nucleotide sequence on the m-RNA is used to determine which anti-codons of transfer RNA molecules that bring up activated amino acid molecules shall be added next to the new peptide strand (Fig. 1).
Three nucleotide bases are required to specify the insertion of an amino acid into a polypeptide chain. A protein of 100 amino acids has $10^{130}$ possible sequences, yet the DNA → RNA → amino acid assembly organisation is capable of ensuring near constant production on the ribosomes. There are now 20 amino acids to code for and this is done by the combination of four possible nitrogen bases in triplets ($4^3$), making 64 nucleotide triplets or 'codons'. How the code arose is a tantalising speculation. It seems that there were originally only 15 amino acids; it can thus be said that the primitive code was even more 'degenerate' than at present. The word degenerate is used to mean that several codons exist for any one amino acid. Moreover, if the genetic code is examined (in any standard biochemical text) it will be seen that the most hydrophobic amino acids (easily distinguished by their higher mobility on chromatograms) are coded for up the left-hand column and along the top row. This suggests that each amino acid is found at a particular position for a definite reason; the trinucleotides could well have such a selective affinity for certain amino acids that if life were to begin all
over again certain features of the new code would turn out exactly the same on account of specific physico-chemical interactions.

The genetic code is thus based on nucleic acid-base pairing. Proteins can only achieve 'specificity instruction' when a particular spatial folding of part of a protein molecule leads to conformations that can be recognised. This is the basis of antigen-antibody interactions. However, such recognition is restricted to short sequences. In retrospect it is clear that what is required in evolutionary development is the self-instructive capacity of complementary recognition, which is inherent in the nucleic acid-base pairing system, together with the enormous functional capacity of proteins to form enzymes, antibodies, building units, and so on. It is probable that, originally, a single codon interacted directly with an amino acid. Later, the amino acid could have fitted between the double strand of a codon-anticodon sequence; still later that amino acid would be picked up by a transfer molecule that would be the originator of transfer RNA.

The evolutionary development of individual molecules is a fascinating subject. The essential underlying mechanism is the hereditable variability caused by mutation and, thereafter, the influence of natural selection. Vogel (1965) has described the discovery of two different pathways for lysine biosynthesis. The diamino-pimelic acid (DAP) lysine pathway is the lysine synthetic pathway in bacteria and plants. Animals can no longer synthesise lysine, which is an essential amino acid, but the pathway of lysine synthesis via amino-adipic acid is found in fungi and euglenids and this fact suggests that it was from this off-shoot of the mainstream of plant evolution that the protozoans diverged.

One molecule that has attracted a great deal of interest is cytochrome c. Dickerson (1972) has presented its composition in 38 species; no other molecule has been studied in so many different organisms. Cytochrome c of man and the chimpanzee are identical but differ, for example, from that in the red mould Neurospora by 44 out of 104 amino acids. The cytochrome c of man and horse differ by 12 out of 104 amino acids only. These statements illustrate that a structural gene which codes for a vital polypeptide is often quite resistant to mutation. Similarly, the B chains of insulin are the same in different species but substitutions are allowed on the A chain at positions 8, 9 and 10. In haemoglobin the alpha-chain of man resembles that of the horse more closely than its own beta-chain.

Dickerson has illustrated how the study of amino acid sequences is a most powerful tool for the elucidation of evolutionary development. It is possible to draw graphs in which the number of amino acid changes in a particular type of molecule are related to the geological time since those species showed evolutionary divergence. The line indicating amino acid divergence for cytochrome c rises gently at an angle of 30 degrees, indicating that some 20 million years are required to produce a 1 per cent change in the amino acid sequence of two diverging types of this molecule. The line for haemoglobin rises at an angle of 60
degrees, showing that haemoglobin is changing at that rate in only six million years. Still more dramatic is the 80 degree rate of change for fibrinopeptides, which indicates a 1 per cent amino acid change in this molecule in only one million years. What is actually being measured is the rate of appearance of harmless mutations. Whereas a fibrinogen molecule can tolerate quite a marked change in its end-group fibrinopeptide, the function of a haemoglobin molecule is closely geared to its structure (Perutz and Lehmann, 1968) and a random mutation is five times more likely to be harmful and so not to be perpetuated in evolution. In the pig, cow and sheep cytochrome c is identical, but there are 11 differences in the amino acid sequence in fibrinopeptide A, which is only 19 amino acids long (Fraser and Mayo, 1975). The molecules that have hardly changed in 600 million years are the histones. These are the basic proteins, rich in arginine and lysine, that are bound to the major groove of the DNA helix. Even during mitosis they remain bonded to DNA. It is thought that they protect DNA from damage by radiation or nucleases, that they exert an important regulatory role on DNA transcription into the messenger RNAs, and that they function as protective blocks to maintain the integrity of chromosomes. It is therefore not surprising that the histones from calf thymus and the seedling pea still retain an almost identical structure.

Gene duplication plays an important role in evolutionary development. Haemoglobin is a tetramer of two identical types of polypeptide chain so that adult haemoglobin A can be represented as \( \alpha_2 \beta_2 \) and fetal haemoglobin F as \( \alpha_2 \gamma_2 \). Haemoglobin and myoglobin have similar functions and in contemporary species are controlled by separate genes, but it would seem that they arose by duplication of a common gene during the evolution of species (Ingram and Stretton, 1962). Figure 2 indicates that the primordial gene duplication that gave

![Figure 2](image-url)

**Fig. 2.** The formation of present forms of haemoglobin by gene duplication in relation to geological time.
rise to the alpha and beta chains occurred some 375 million years ago, before the ancestral divergence of man, horse and cattle. Before further gene duplication gave rise to separate adult (beta) and fetal (gamma) chains the ancestral haemoglobin could have been a functional compromise between intra- and extra-uterine demands. There must have been a later translocation of genes, since the alpha and beta chain genes are now on separate chromosomes. However, the existence of the Lepore haemoglobins, that produce a form of thalassaemia in which the molecule consists of beta chain residues and variable N-terminal d-chain residues, indicates that there has been no translocation since the duplication of the b-d genes; the b and d chains of haemoglobin differ by only ten amino acid substitutions. The d gene found in haemoglobin A2 (α2 d2) is actually a poor polypeptide chain producer accounting for only 2 per cent of adult haemoglobin.

The haemoglobin amino acid sequences of at least 20 species are known. Curiously enough, there are nine positions that contain the same amino acid in nearly all 20 species and these 'conserved' positions are important for oxygen binding. The reversible oxygenation of haemoglobin depends on the haem being located in a non-polar niche where it is protected from water, otherwise oxidation would occur. The haem with the iron at its centre sits in a crevice at the centre of the haemoglobin molecule and electrons are shuttled in and out. The molecule thus alternates from an oxygenated ferricytochrome to a reduced ferrocyanochrome.

The considerable evolutionary significance of gene duplication will be clear if one recalls, for example, that it is the basis of 'heterozygous advantage' (Allison, 1954) by which sickle cell heterozygotes (β−βs) have survived the ravages of Falciparum malaria, so accounting for that unexpected 'balanced polymorphism'. Similarly, the fact that pyruvate kinase has two separate gene loci explains why a human homozygote for pyruvate kinase deficiency suffers only a non-spherocytic haemolytic anaemia and not a lethal condition.

What then are the mechanisms for gene duplication? It is known that unequal crossing over of chromosomes during meiosis is common; this means that sometimes there is gene duplication on one chromatid and deletion of the gene locus on another. This is how the haptoglobin variants came about. There may also be unequal exchange of material between two chromatids of the same chromosome during the synthetic phase of the mitotic cycle. Usually, the exchange is equal and of no consequence but, if it is unequal, two genes for a particular protein may be left on one chromatid and none on the other. Gene duplication in the era of pre-vertebrates could also have occurred by 'polyploidisation', the simultaneous duplication of genes in a process that placed the original gene and the new genes on separate chromosomes. In such a process there is an increase of the DNA, which is one reason why it would have occurred a long time ago, although even now there is evidence that the mouse genome is increasing in size.

Much study has been directed to determination of the differences in the nuclei
of the DNA of related species in relation to the geological time since those species became separated. Such differences can be estimated by using the principle that hybrid molecules can be formed from the DNA of two species, since it is easy to dissociate and re-associate DNA (Kohne, 1970). For every 1.5 per cent nucleotide pair mismatch the thermal stability will be lowered by 1°C. Therefore one will expect that the longer the period since divergence of the species, the greater will be the species differences in the nucleotide base pairs, and the greater the reduction in thermal stability. Table 1 gives approximate values.

Table 1

| nucleotide divergence % | geological time |
|-------------------------|-----------------|
| man/chimpanzee          | 2.5             |
| man/gibbon              | 6.2             |
| man/rhesus monkey       | 10.1            |

Clearly the effect of a mutation will depend on which part of the protein synthesising apparatus is involved, what its importance is, and whether the part concerned can be mended or merely ignored because there is reserve gene material. Mutations in non-functional DNA will not matter because they are neutral. Mutations in genes for ribosomal RNA or transfer RNA are not directly read as changes in protein synthesis but in due course will alter the synthesis of proteins. In fact, the genes for ribosomal RNA are highly conserved ‘repeater sequences’ which change little if at all in evolution. When a mutation affects the DNA that is acting as a messenger RNA template, any sequence change will be significant; many are lethal and cannot be passed on. Indeed, induced mutations in the DNA, together with exact biochemical and genetical analyses, have been used to confirm the genetic code assignments. In general, there is a ‘one gene-one polypeptide’ rule, so inborn errors of metabolism due to an anomaly in one enzyme might ultimately be traced back to one abnormal nucleotide triplet in the codon.

Only one peptide comes to mind that does not depend for its synthesis on ribosomes or on messenger RNA. This is the antibiotic gramicidin, which is a decapptide incorporating some D-amino acids. Its synthesis is rather like fatty acid synthesis since the two enzymes involved use amino acids activated as thio-esters (amino-acyl AMP). These enzymes may represent a relic of a primitive mechanism used in early evolution. This particular molecule has endotoxin neutralising properties. It is, however, the product of a bacillus.

Stability in the genome of a species is imparted by ‘repeater DNA’. The genomes of mouse, monkey, guinea-pig and man contain families of sequences that are very similar. When a new sub-species appears it seems that there has been a production of similar but non-identical nucleotide sequences by a process called
‘saltatory replication’ (Britten and Kohne, 1970). Whenever such a group of DNA sequences produces a useful metabolic potential it may become the basis for a new evolutionary event resulting in the reorganisation of an organ, emergence of a new organ or the development of a new species. Many copies of the DNA sequence are produced, so that there is flexibility and stability together in the new line. This is akin to the production of antibody molecules which are functional proteins of great similarity, yet with significant differences.

Presumably, the replication process of new repeater sequences affects the germ cell line and copies are integrated into normal genetic apparatus to be transmitted to the progeny. If they give rise to a favourable genetic element the new family will spread through the species by the process of natural selection. This whole process could be rather like a virus infection in which many copies of the virus are made within a cell and, in the ‘latent state’, chromosome-integrated material is being transmitted to the progeny of the species.

As a result of background radiation there is a certain spontaneous mutation rate. However, in natural evolution the introduction of new mutations is allowed to occur at a rate that normally allows the species to incorporate the few that are advantageous and eliminate the deleterious mutants. In medicine today we are not only allowing certain undesirable recessive characteristics to be propagated but we also tend to forget that the genetic effect of small doses of radiation to a large section of the population will have cumulative effects. This story is a tribute to man’s inquisitiveness, acquisitiveness and intellectual capacity, but how he evolves in the future will largely depend on how he can learn to control his own personal and group behaviour.

References
Allison, A. C. (1954) Transactions of the Royal Society of Tropical Medicine and Hygiene, 48, 312.
Black, S. (1973) Advances in Enzymology, 38, 193.
Britten, R. J. and Kohne, D. E. (1970) Science, 224, 24.
Calvin, M. (1959) Evolution, 13, 362.
Dickerson, R. E. (1972) Scientific American, 226 (2), 58.
Fraser, G. R. and Mayo, O. (1975) Textbook of Human Genetics, p. 139. Oxford: Blackwell.
Ingram, V. M. and Stretton, A. O. W. (1962) Biochimica et Biophysica Acta, 62, 456.
Kohne, D. E. (1970) Quarterly Review of Biophysics, 3, 327.
Oparin, A. I. (1965) Advances in Enzymology, 27, 347.
Orgel, L. E. (1968) Journal of Molecular Biology, 38, 381.
Perutz, M. F. and Lehmann, H. (1968) Nature, 219, 902.
Ponnamperuma, C. (1971) Quarterly Review of Biophysics, 4, 77.
Schrödinger, E. (1944) What is Life? London: Cambridge University Press.
Vogel, H. J. (1965) in Evolving Genes and Proteins. (Ed. V. Bryson and H. J. Vogel). New York: Academic Press.
Watson, J. D. and Crick, F. H. C. (1953) Nature, 171, 964.