1
MAJOR EVENTS IN THE HISTORY OF LIFE

The Jurassic pterosaur *Pterodactylus kochi* from Solnhofen Limestone preserving impressions of the wing membranes, × 0.84. (Photograph courtesy of J.M.V. Rayner.)

1.1 Origin of Life

C. R. WOESE & G. WÄCHTERSHÄUSER

Introduction

The origins of man and his fellow creatures are concerns perhaps as old as man himself. However, before the nineteenth century these could not be given scientific form. The prescientific notions of life's beginnings were an incongruous amalgam of biblical thought, philosophy, alchemy and folk wisdom. The Bible taught that all life arose through special acts of divine creation during the first days of the Earth's existence. Commonplace experience showed, however, that life can also arise spontaneously, as maggots seemed to, for example, from rotting meat. And vitalism saw life as an ever-present non-material property of the universe.

In the nineteenth century four great scientific achievements laid the groundwork for making the origin of life a scientific problem: (1) The realization that the cell was the fundamental unit of biology introduced an enormous gulf between the living and non-living worlds; (2) Darwin's theory of evolution implied that all life came from some distant universal ancestor; (3) Pasteur painstakingly and convincingly refuted the claims of spontaneous generation (of microscopic life); so that, if life had arisen spontaneously on this planet, it must have done so under conditions no longer present and probably long gone; and (4) Mendel discovered genetics, whose origin is to this day one of biology's great mysteries.

The picture we now have of life's origin, though scientific, is based upon very few facts. It derives mainly from metaphysical assumptions, cultural images we take for granted. Consequently, it is likely to share features with the prescientific accounts of life's origin which go unrecognized and unchallenged. The present discussion of origins is framed along historical lines, a format that generally helps to reveal prejudices that impede or sidetrack the development of a scientific picture.

The conventional primitive ocean scenario

Darwin apparently gave little thought to the origin of life; he understood the problem to be intractable in his time. However, his casual 'warm little pond' allusion has had a significant impact on later thinking, undoubtedly far more than its author intended or would have liked. Oparin (1924 *in* Bernal 1967) and Haldane (1929 *in* Bernal 1967) are generally credited with formulating the issue scientifically; not because they were the first to attempt it, but because their origin scenarios were more comprehensive than those of their predecessors. The details of these theories need not concern us; for they often reflected misconceptions, for example as to the nature of genes, viruses, and protoplasm, and how they replicate. However, in their general aspects the theories are of great interest, for, remarkably, these half century old proposals still remain the foundation of our understanding of life's beginnings.

The Oparin ocean scenario has by now become almost catechismal. It begins with a primitive anoxic atmosphere, comprising gases such as carbon dioxide or methane, nitrogen or ammonia, hydrogen sulphide, water, and hydrogen. Current thinking invokes the less reduced forms of these elements, the fully reduced forms being postulated earlier, by Urey (1951) and others, on the mistaken assumption that the nascent Earth possessed an atmosphere similar to those found on the large gaseous planets. Miller (1953) was the first to put such models to scientific tests when, as a student in the nineteen-fifties, he demonstrated that electrical discharge acting on Urey's atmosphere produced a conglomeration of organic compounds that included many of the familiar amino acids. The many experiments that followed showed that not only amino acids, but also a variety of organic compounds of biological interest, can be produced by a variety of energy sources under a variety of conditions (providing that oxygen is absent). So today we believe that some anoxic, slightly reducing atmosphere, acted upon by ultraviolet light and/or electrical discharge, served as a continual source of the simple reactive organic chemicals needed to begin and sustain the evolutionary process.

The products of this atmospheric chemistry ended up in the primitive ocean, which over time became a vast repository of reactive organic chemicals. Oparin's and Haldane's primitive ocean was a 'hot

dilute soup' (Haldane's phrase); hot on the mistaken assumption (common in the nineteen-twenties) that the Earth had arisen as a fragment of the Sun. With the later acceptance of a cold accretion model for the Earth's formation (also incorrect), the oceanic soup was cool from the start. The primitive ocean, then, was a 'vast chemical laboratory', a cosmic retort in which the great alchemist Nature sought to concoct the first living cell.

The most important but weakest element in the ocean scenario is the transition from prebiotic chemistry to actual living, self-replicating entities. The early models necessarily resorted to hand waving arguments — interactions occurring among the reactive chemicals in the primitive ocean led to ever more complex structures, to more and more complicated aggregates, that ultimately somehow became self-perpetuating. Later proposals, drawing upon the structure of nucleic acid, refined this notion to that of macromolecular templating.

As Haldane (1929) put it, in the ocean 'the first precursors of life found food available in considerable quantities'. Therefore, there was no need for them to evolve the capacity to produce these metabolites, and so they did not. The aboriginal organisms were total heterotrophs. This seldom questioned assertion determines the subsequent evolutionary course. A heterotrophic life style will necessarily (and rather quickly in evolutionary terms) deplete the oceanic stores of nutrients. If organisms are then to survive, they must evolve an intermediary metabolism and eventually learn to transduce other forms of energy (light or chemical) into biochemical energy.

The current explanation for how intermediary metabolism arose in the aboriginal heterotrophs was fashioned by Horowitz (1945). When the oceanic supply of a particular amino acid, for example, became exhausted, the supply of its immediate chemical precursor, for example an hydroxy acid (for which organisms previously had had no need), still remained untapped. Were the organism to evolve an enzyme that converted this precursor to the needed amino acid at this point both the organism and its progeny would survive. When, sometime later, the supply of the precursor also became exhausted, the process would repeat — the organism evolving another enzyme to catalyse synthesis of the precursor from its own (previously unutilized) precursor; and so on. In this manner all intermediary metabolism arose, the pathways evolving 'backward', one step at a time.

Neither Oparin nor Haldane initially postulated cellular entities as a starting point for evolution, although Oparin did so later, with his coacervate model. There has been no subsequent consensus as to when cellularity arose.

Criticisms of and refinements to the standard model

The essence of the primitive ocean scenario (i.e. its metaphysics) has never been seriously questioned. However, many weaknesses in its details have come to light over the years. In each case the tendency has been to correct the problem by adding some new feature to the model. As a result today's origin scenario is a Ptolemaic hodgepodge of *ad hoc* assumptions. There is little point any longer in criticizing the standard model simply in the standard way, adding another *ad hoc* feature to remedy each new difficulty. The basic (implicit) assumptions of the model must be questioned.

Cultural roots of the primitive scenario. Although the Oparin ocean scenario was developed as a scientific alternative to the prescientific versions of the process, its similarity to the Garden of Eden myth should be worrisome. An oceanic 'paradise' is postulated in which organisms can develop safely in the midst of plenty. In addition to having 'food available in considerable quantities', the first organisms 'had no competitors in the struggle for existence' (Haldane 1929). Scientists have tended to see the first organism as arising through a series of highly unlikely events; discussions centre about improbable happenings that, given long times and enormous numbers of trials, eventually come to pass. (Remember that before the discovery of microfossils the Earth was generally thought to have been sterile for most of its existence, allowing billions of years for key events to happen.) Yet none of this is regarded as miraculous!

One can even detect something akin to the biblical banishment in the scientific account: because organisms ultimately destroyed their oceanic paradise (by consuming the store of biochemicals), they were thrust into a harsh world where they had to fend for themselves by developing intermediary metabolism, autotrophy, and eventually phototrophy. This was the dog-eat-dog world of competitive existence: 'The further life progressed the less nutrient substances were available to the organisms and the more strongly and bitterly the struggle for existence was waged' (Oparin 1924). Its truth aside, the Oparin ocean scenario seems a prime example of culturally

determined imagery shaping a scientific concept. The sooner these cultural influences are recognized and understood, the sooner a proper scientific picture of the origin of life will emerge.

With a need for major restructuring in mind, let us analyse the main elements in the standard scenario in detail.

Energy sources, the multi-theatre assumption, and the ocean repository. The ultraviolet light or electrical discharge invoked by the standard scenario to create the initial simple reactive compounds are so energetic that they produce indiscriminate bond ruptures, ionizations and free radicals. These would be entirely destructive of any larger (organic) compounds, not to mention living systems. This dilemma forces the standard scenario into a 'two-theatre' assumption: the initial simple reactive compounds produced in one theatre (the atmosphere) are subsequently quenched and protected in a second theatre (the ocean), where they accumulate and further react to produce more complex structures, and ultimately living systems.

The need for two widely separated theatres seems to underlie a pernicious paradox in the standard scenario: the notion of reactive chemicals is at odds with their transport over great distances, from the high atmosphere to the ocean. The ocean must accumulate reactive chemicals over long times because distance necessarily translates into dilution. A protracted accumulation (and storage) in turn is at odds with the reactivity of these chemicals and their rapid removal by hydrolysis or sedimentation.

In retrospect it is strange that all attempts to correct the difficulties with this oceanic chemical repository–reaction pot have never questioned its underlying multi-theatre assumption. Rather, difficulties were overcome by invoking *additional* theatres. Mineral surfaces, particularly clays, were seen by Bernal (1967) as a vehicle for concentrating and reacting organic chemicals. He pictured the dilute organic compounds in the ocean becoming concentrated in the froth that forms on its surface, the froth being driven shoreward, to end up in estuaries — where the already concentrated compounds became even more so in the oozes that formed there. Organic compounds adsorbed in high concentrations on clay sediments (and perhaps oriented and/or activated in the process) might then undergo spontaneous condensations, to produce biopolymers, and so on. Considerable experimental work has been done, e.g. by Katchalsky (1973) and co-workers, on the properties of clay minerals *vis-à-vis* the adsorption and reactivity of organic compounds. We shall encounter below a fundamentally different role for surfaces in the origin of life.

Fox (1965) demonstrated experimentally that mixtures of amino acids (rich in the dicarboxylic amino acids) polymerized under hot (less than 200°C) non-aqueous conditions. Upon hydration these condensates produced 'proteinoid microspheres', which loosely resembled cells (in size, shape, and in having a few general catalytic properties). Because of this Fox argued that the high temperature conditions associated with volcanic environments were those under which the organic compounds in the ocean repository became concentrated and reacted to give the prototypes of living systems.

Invoking additional theatres is a Ptolemaic solution to the standard scenario's problems — which to a large extent are due to the multi-theatre assumption. The true remedy may lie in single-theatre scenarios, in which the energy source can be in close proximity to or within the evolving system. These are conditions under which an energy flux can constantly generate a rich spectrum of organic biochemicals that are turned over rather than stored.

The organism–environment dichotomy; heterotrophy and self-assembly. For a sufficiently primitive system, the organism–environment distinction does not exist. The dichotomy arises only when the evolving system has become sufficiently complex and physically separated from its surroundings that it can be viewed as an entity in its own right. However, the standard scenario (shaped by the properties of extant life) tends to see an organism–environment dichotomy early in the evolutionary process — certainly too early. What occurs in the 'organism' is strongly distinguished from what occurs in its 'environment'. The dichotomy (together with the Garden of Eden image) then makes of the ocean repository a pre-existing store of 'food' for the aboriginal organisms. They, in turn, come into being as heterotrophs, and go on to deplete and ultimately exhaust their store of food. In other words, a prematurely forced distinction between organism and environment tends to place the replicative aspect of the primitive system in the former, its metabolic aspect in the latter.

In such a dichotomous world the environment does not naturally, automatically, give rise to the organism. The latter has to 'strive' to bring itself into existence; it is the product of accidental self-assembly from simpler components in the

environment — an improbable and, therefore, protracted process. (Subject to the vagaries of chance in this way the evolutionary process has to pass through a stage of instability, of uncertain outcome.) A more extensive quote from Haldane shows these features rather clearly: 'When the whole sea was a vast chemical laboratory the conditions for the formation of such films [membranes, that is] must have been relatively favourable; but for all that life may have remained in the virus ... half-living chemical ... stage for many millions of years before a suitable assemblage of elementary units was brought together in the first cell. There must have been many failures, but the first successful cell had plenty of food.'

Were life to have originated in an autotrophic rather than heterotrophic manner, the scenario would have been markedly different. Autotrophic evolution focuses on autocatalytic reaction networks, on metabolic pathways — whence all other evolutionary developments stem. It is a single theatre scenario, in which energy source, production of reactive compounds, and condensations to form complex organic structures, occupy the same locale. When life begins with autotrophic metabolic pathways, one avoids the kind of dichotomous separation that exists between a heterotrophic organism and its host environment. An autotrophic system is a *source* of biochemical energy and complexity, not a sink for these (as are heterotrophs). With autotrophy the protracted, chancy trial and error period no longer seems required; the self-replicating entity (its genetics) can arise simply as a more refined and complex extension of the primary autotrophic and autocatalytic process. While some self-assembly might have to occur even in this process, that requirement too can be reduced by eliminating the constancy of the chemical conditions. In other words, major changes in the evolving system could be driven by, or be responses to, local or global changes in the state of the planet.

The origin of genetics; templating and the genotype/phenotype dichotomy. Mendel's great discovery, that the cell has a phenotypic-functional aspect that is determined by a cryptic genotypic-reproductive aspect, has dominated our view of the origin of life. With the discovery in the nineteen-forties that each gene corresponds to a unique enzyme, the central question could then be phrased: 'Which came first, the gene or the enzyme?' Geneticists such as H. J. Muller felt that the gene had to have come first; only a few physiologists disagreed. (The gene at that time was often thought of as proteinaceous and even as having its own primitive phenotype.) Watson and Crick's discovery of the double stranded structure of nucleic acid rendered the question meaningless. Since all genes appeared to have the same basic structure, they could not have unique phenotypes, could not be functional in their own right; and, proteins (i.e. enzymes) could not evolve without genes — a chicken and egg paradox.

At this point the central question should have become 'How did the genotype–phenotype relationship (i.e. translation) arise?' However, the attractive and specific mechanism for the origin of gene replication inherent in the double stranded structure for nucleic acid (plus our near total ignorance of the molecular mechanics of translation) took us in an opposite direction. The origin of the genotype (nucleic acid replication) separated completely from the origin of the phenotype (metabolism) — the former question totally eclipsing the latter (Eigen *et al.* 1981; Orgel 1973). Recently it has become popular to believe that (RNA based) 'nucleic acid life' must have preceded protein-based life; that initially nucleic acid was both the genotype *and* the phenotype. This point of view is supported by the facts: (1) that polypyrimidines can serve as templates that align complementary purine nucleotides, which (when properly activated) then go on to condense into polypurine chains; and (2) that some RNAs possess certain limited enzymatic or catalytic properties. Eigen has also reported that in the presence of a particular protein (the replicase of the virus Qβ) a certain type of RNA will spontaneously arise (in the absence of a pre-existing template).

A fascinating variation on the templating theme is Cairns-Smith's (1985) proposal that life began with replicating patterns in *clay* layers (which could adsorb organic molecules and thereby influence the course of subsequent organic evolution). While daring in one way, this proposal is conventional in another; it takes for granted the need for templating as an initial step in the origin of life.

A totally dichotomous view of the origins of the genotype (replication) and the phenotype (metabolism) is an extrapolation in the wrong direction. It has even led Dyson (1985) to propose that life arose *twice*; initially somehow as protein-based life, within which nucleic acid then separately arose as a 'disease'. The earliest life forms were almost certainly not incarnations of our dichotomies, of our attempts to define extant life. Rather, primitive living systems were undoubtedly less well de-

fined, less compartmentalized, than their modern counterparts, and so in that unusual sense more 'integrated'. It is time to reassess the genotype–phenotype dichotomy as a paradigm for the origin of life.

A proper conceptualization of translation, the process that defines the genotype–phenotype relationship, should have an integrating, unifying effect on our concept of the origin of life. Unfortunately, the translation mechanism is large and complex, and, therefore, its molecular workings and evolution are not understood. The fact that some of the proteins involved in translation are also components of certain nucleic acid replication enzymes, however, suggests primitive connections between the two processes. The facts that cells today contain transfer RNA-like molecules as essential parts of non-translational ('non-programmed') polymerizations (e.g. polypeptide antibiotic and cell wall syntheses), and that nucleotides, other heterocyclic compounds, and even transfer RNAs play important roles in intermediary metabolism, hint at still deeper evolutionary connections.

The suggestions are strong that the programmed polymerizations (translation and nucleic acid replication) have arisen out of more primitive metabolic interactions. Therefore, what seems called for at this juncture is a general view of polymerization processes, one that attempts to relate polymer formation to the full spectrum of metabolic reactions in primitive systems — e.g. the types of polymers arising under primitive conditions; the range of monomer units and chemical linkages involved; whether polymers were formed by monomer or oligomer condensations; chirality constraints; whether the sequences of the aboriginal polymers were random or (simply) ordered (e.g. homopolymers, polymers of alternating sequence, etc.); the extent to which templating is or is not involved; and oligonucleotide–amino acid interactions.

Geological and phylogenetic constraints on the primitive ocean scenario

Knowing when the evolutionary process started is crucial to understanding how it occurred. Conditions during the first few hundred million years of Earth's existence were certainly very different from those occurring 2000 million years later.

The current understanding of the geological history of the Earth, Moon, and other planets, together with recent advances in the biologist's understanding of phylogeny, substantially restrict the time interval during which the evolutionary process could have started.

The Earth's crust is now believed to have been initially quite hot, too hot to sustain liquid water. Any water present would have been partitioned between the primitive atmosphere and a semi-molten crust. There is also geological evidence to suggest that the Archaean oceans were warm. The oldest sedimentary rocks (3800 Ma), although somewhat metamorphosed, give evidence of life at that time; and the better preserved 3500 Ma sediments give clear evidence of bacterial life — showing both fossil stromatolite structures and microfossils (see also Section 1.2). In that stromatolites today are produced by photosynthetic bacteria, principally cyanobacteria (or thermophiles of the *Chloroflexus* type), photosynthetic bacteria (probably) already existed 3500 Ma.

The explosive developments in molecular phylogeny over the past decade have revealed a number of important facts: (1) the earliest phylogenetic branchings gave rise to three aboriginal lineages, the eubacteria, the archaebacteria and the eukaryotes; (2) photosynthesis appears to have arisen (early) within the eubacteria. If so (given the stromatolite evidence), eubacteria already existed at least 3500 Ma, so that the most recent ancestor common to *all* life lived at a still earlier time — probably far earlier, because of the enormous evolutionary distances that separate the three classes; (3) prokaryotic life (at least) arose in high temperature environments; (4) the ancestral environments were anaerobic; and (5) the ancestral forms of prokaryotic metabolism may have been autotrophic. Comparisons among the (sequences of the) genomes of diverse organisms will ultimately permit us to infer in some detail the nature of the most recent common ancestor of all extant life, and also certain things about still earlier evolutionary stages (see also Section 2.1). All the evidence to date, then, points to life having arisen quite early in the planet's history, and under thermophilic conditions.

Alternatives to the primitive ocean scenario

An important methodological rule of K. Popper is that a new theory should have a greater explanatory power than its predecessors, i.e. it should explain a multitude of facts with a minimum of assumptions. Clearly, today's consensus theory of the origin of life is little more than a highly amended version of the original Oparin/Haldane scenario it has replaced — which violates Popper's rule. Further

amendments to the standard scenario are not what is needed; true alternatives to it are.

Wächtershäuser (1988) has proposed one such alternative, which dispenses with the multi-theatre assumption, the ocean repository, heterotrophic origin, and modular self-assembly. This theory, moreover, is sufficiently detailed to make testable assertions regarding the nature and evolution of primitive biochemical pathways. The first organisms are assumed to be truly autotrophic (not heterotrophic) — the result of *de novo* biosynthesis of organic constituents by the uptake of inorganic material (e.g. CO_2), and subsequent rearrangement reactions. They are not the products of accidental modular assembly. The theory's central idea is that life began with autocatalytic, metabolic processes occurring in an essentially two-dimensional fashion, within organic monolayers anionically bonded to positively charged surfaces of minerals, such as pyrite, and in contact with water at high temperature. Adherence to the mineral surface is *not* the result of adsorption but of an *in situ* autotrophic growth of organic constituents that acquire their anionic surface bonding *in statu nascendi*. The concentration of dissolved organic constituents in the water phase is negligible. Hence the process by which a constituent loses its surface bonding is irreversible; detachment is tantamount to disappearance. (In this respect the theory is the opposite of Bernal's clay theory, which is based upon adsorption). On these pyrite surfaces large polyanionic constituents, with ever stronger surface bonding, are automatically selected — to begin with polyanionic coenzymes, eventually nucleic acids and polypeptides. The primitive system grows by spreading onto vacant surfaces, reproduces by producing its autocatalytic coenzymes, and its evolution is driven by environmentally induced ignitions of new autocatalytic cycles. The system evolves toward higher complexity, since the thermodynamic equilibrium in a surface metabolism would favour synthesis, not degradation (as would occur with solution reactions). High energy phospho-anhydride groups are not required for the formation of covalent bonds. Phosphate groups (whose source is taken to be the mineral substrate) have the sole function of surface bonding. The energy for carbon fixation is provided by the redox process of converting ferrous ions and hydrogen sulphide into pyrite, which is not only a waste product but provides the all-important binding surface for the organic constituents.

This initial laminar organism is succeeded by two further stages. The second stage organisms are semicellular entities still supported by a pyrite surface but having an (autotrophically grown) lipid over-layer, with an internal broth of detached constituents. In this 'bleb' stage a membrane metabolism and a cytosol metabolism appear, first as a supplement to, and later as a substitute for, the aboriginal surface metabolism. Membrane-bound electron transport chains allow the tapping of other redox energy sources and ultimately of light energy. The cytosol metabolism allows the salvaging of detached constituents by catabolic processes and the development of modular modes of synthesis that rely upon energy coupling. Eventually heterotrophy appears, as a by-product of the catabolic salvage pathways. The cell's genetic machinery develops from surface-metabolic precursors. It produces self-folding enzymes which compete with the mineral surface for bonding the metabolic constituents. In this stage evolution becomes double tracked, an evolution of metabolic pathways and one of the bonding surfaces for their constituents. In the third stage the pyrite support is abandoned and true cellular organisms arise.

Since the ocean cannot reasonably function as a reaction pot in which life originated, and its role as a repository is suspect, the question is whether it played any significant role at all in the origin of life. Two types of scenarios exist that make minimal use of the ocean. One is the idea that hydrothermal vents served as the aboriginal environment. Since hydrothermal vents create chemical gradients, a single-theatre vent scenario can be developed that has no need for the ocean repository assumption. How the model would cope with the fact that vents, and so their products, are ephemeral (especially so on an evolutionary time-scale) is unspecified.

It was suggested by Woese (1979) that evolution began in the primitive atmosphere, at a time when the planet's surface was too hot to sustain liquid water. The early Earth can be pictured as surrounded by vast cloud banks, as Venus is today. The severe weather conditions that must then have existed would have caused large quantities of minerals (dust), from the dry surface, to be swept into the atmosphere. Atmospheric water vapour then condensed on the dust, dissolving it (in part). As a consequence, the primitive Earth was enshrouded in clouds of salt water. In addition to containing (possibly high concentrations of) minerals, the droplets in these clouds would accumulate organic compounds, produced by interactions among atmospheric gases and other constituents (or with

compounds produced by thermal reactions on the Earth's surface and swept into the atmosphere). These droplets are natural precursors of cells — their surfaces coated with mixtures of the larger organic compounds, their interiors solutions of reactive (organic and inorganic) compounds. The different layers of the atmosphere would each have characteristic chemistries, the whole being in effect a connected series of chemostats. Droplets (and hydrated dust) offer enormous amounts of surface, and so surface chemistry becomes all important in life's beginnings.

As the primitive Earth cooled, its surface would pass from a dry condition, through cycling damp/dry stages, to one where large bodies of (hot) water could accumulate. These major global transitions would bring about major changes in the evolutionary course (see above). The cloud setting suggests a single theatre scenario, requiring no repository assumption; it also suggests that major stages in evolution were driven by (were responses to) major changes in the state of the planet.

In one sense the origin of life problem today remains what it was in the time of Darwin — one of the great unsolved riddles of science. Yet we have made progress. Through theoretical scrutiny and experimental effort since the nineteen-twenties many of the early naive assumptions have fallen or are falling aside — and there now exist alternative theories. In short, while we do not have a solution, we now have an inkling of the magnitude of the problem.

References

Bernal, J.D. 1967. *The origin of life*. World Publishing Co., Cleveland, Ohio.

Cairns-Smith, A.G. 1985. *Seven clues to the origin of life*. Cambridge University Press, Cambridge.

Dyson, F.J. 1985. *Origins of life*. Cambridge University Press, Cambridge.

Eigen, M., Gardiner, W., Schuster, P. & Winkler-Oswatitsch, R. 1981. The origin of genetic information. *Scientific American* **244**, 88–118.

Fox, S.W. (ed.) 1965. *The origins of prebiological systems: and of their molecular structure*. Academic Press, New York.

Horowitz, N.H. 1945. On the evolution of biochemical syntheses. *Proceedings of the National Academy of Sciences, USA* **31**, 153–157.

Katchalsky, A. 1973. Prebiotic synthesis of biopolymers on inorganic templates. *Naturwissenschaften* **60**, 215–220.

Miller, S.L. 1953. A production of amino acids under possible primitive earth conditions. *Science* **117**, 528–529.

Oparin, A.I. 1938. *The origin of life*. (Translation of 1936 Russian Edition.) Macmillan, London.

Orgel, L.E. 1973. *The origins of life: molecules and natural selection*. John Wiley & Sons, New York.

Urey, H.C. 1951. The origin and development of the earth and other terrestrial planets. *Geochimica et Cosmochimica Acta* **1**, 209–277.

Wächtershäuser, G. 1988. Before enzymes and templates: theory of surface metabolism. *Microbiological reviews*. **52**, 452–484.

Wald, G. 1964. The origins of life. *Proceedings of the National Academy of Sciences, USA* **52**, 595–611.

Woese, C. R. 1979. A proposal concerning the origin of life on the planet Earth. *Journal of Molecular Evolution* **13**, 95–101.

1.2 Precambrian Evolution of Prokaryotes and Protists

A. H. KNOLL

Introduction

The Phanerozoic Eon, the interval under discussion in most of this volume, encompasses the most recent 13% of our planet's history. A sedimentary record documenting more than 3000 Ma of Archaean and Proterozoic time extends below the base of the Cambrian System, and research conducted over the past three decades has demonstrated that this entire sweep of history is the proper domain of palaeontology. Stromatolites, microfossils, and geochemical markers provide fragmentary, sometimes frustrating, but critically important evidence for early evolution. Like younger invertebrate fossils, fossil

prokaryotes and protists must be studied as populations characterized by a measurable range of morphological variation, reproductive pattern, behavioural orientation, taphonomic features, and distribution within and among sedimentary environments. Unlike invertebrate fossils, significant questions of metabolism may remain after populations have been otherwise characterized. The interpretation of early metabolic diversity requires that morphological investigations be supplemented by trace fossil studies (stromatolites and oncolites being the preserved traces of microbial communities) and geochemical analyses of ancient metabolic and environmental indicators. Geological data must be integrated with information from molecular phylogeny and the comparative physiology of living organisms, and interpreted with a clear appreciation of our incomplete understanding of both living micro-organisms and their geological record.

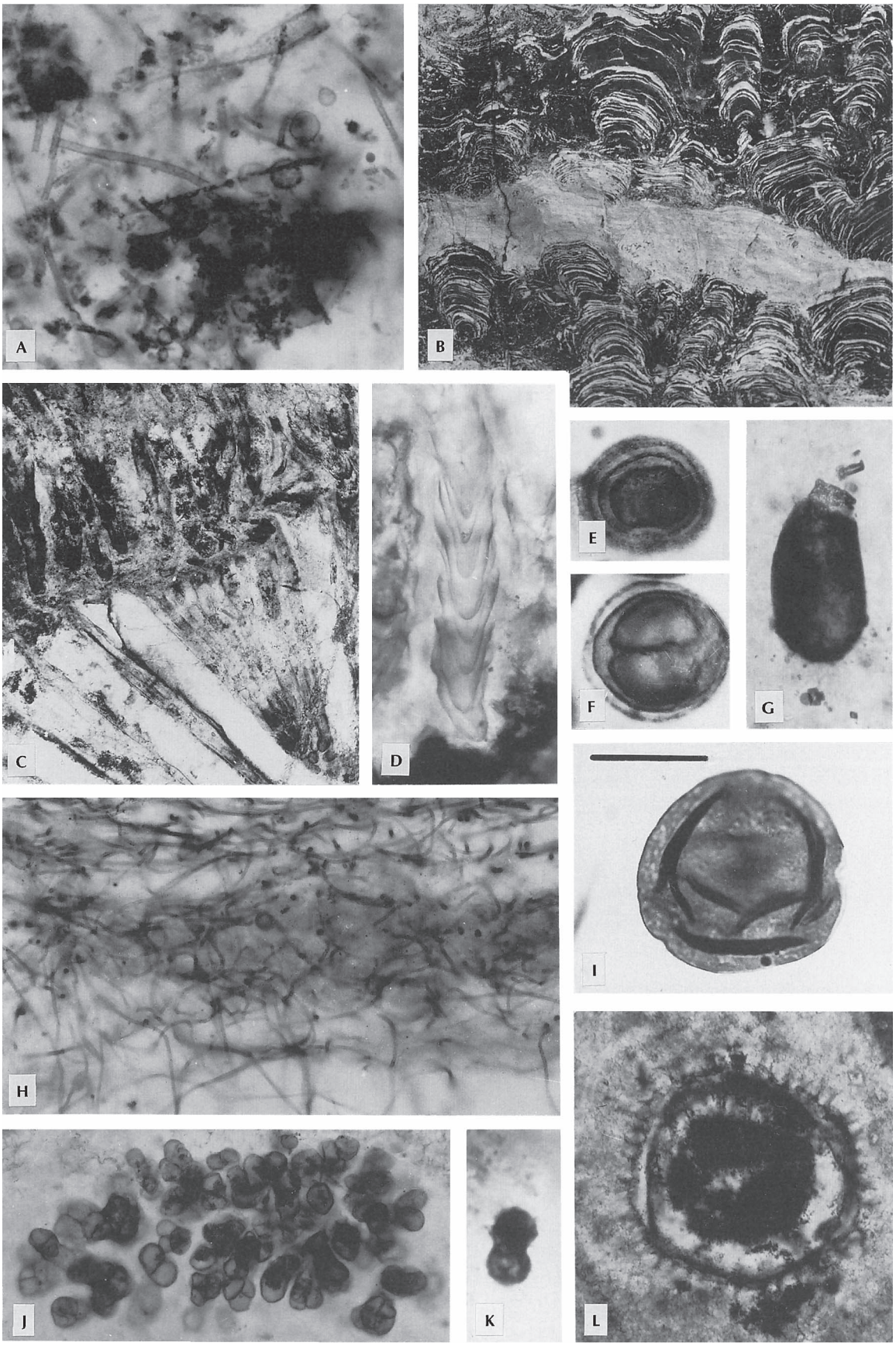
The Archaean Eon: the early diversification of micro-organisms

The age of the earliest palaeobiological record has not changed appreciably in more than 20 years, but the quality of interpretable evidence has improved significantly at decadal intervals. Palaeobiological investigations of Early Archaean rocks have concentrated on two successions, the Onverwacht Group of South Africa and the Warrawoona Group, Western Australia. Both sequences are dated at *c.* 3500 Ma. Both are little-metamorphosed greenstone belt successions characterized by thick mafic and ultramafic lavas, subordinate felsic volcanics, and intercalated sedimentary rocks. Sediments originated largely as volcaniclastics and chemical precipitates, including carbonates, but most have been extensively silicified. Stratiform, domal, and columnar to pseudocolumnar stromatolites occur locally in both areas (Byerly *et al.* 1986; Walter *in* Schopf 1983). These structures have generally been interpreted as the trace fossils of microbial communities. Although this interpretation is reasonable, no Early Archaean stromatolites are known to contain microfossils. Thus, abiological alternatives must be considered, and biogenicity defended on the basis of gross morphology and microstructure (Buick *et al.* 1981).

Microfossils have also been reported from both groups. Simple carbonaceous spheroids of varying size were reported from several horizons in the Onverwacht and overlying Fig Tree groups during the nineteen-sixties but the biogenicity of many of these structures is open to question. During the nineteen-seventies, several authors reported populations of spheroidal carbonaceous microstructures that show a number of features more consistent with a biological interpretation. These include a narrow, nearly normal size frequency distribution about a mean diameter of 2.5 μm, clear evidence for binary division, a sedimentary context comparable to that of younger, undisputed microfossils, and taphonomic features comparable to younger fossils such as flattened and wrinkled vesicles and the occasional preservation of internal carbonaceous contents (Fig. 1K). Rod-like and filamentous microstructures have also been reported from the Swaziland succession, but their antiquity and mode of origin remain subjects for debate.

Undoubted filamentous microfossils have recently been described from cherts of the Warra-

Fig 1 Representative Archaean and Proterozoic fossils. A, *Gunflintia* (filaments) and *Huroniospora* (spheroids) in stromatolitic chert from the Lower Proterozoic Gunflint Iron Formation, Ontario. B, Stromatolites from the Upper Proterozoic Backlundtoppen Formation, Spitsbergen. C, D, Low and high magnification views of a surface-encrusting cyanobacterial population from the Upper Proterozoic Limestone–Dolomite 'Series', central East Greenland — the nested cups are successive extracellular envelopes produced by coccoidal cyanobacteria that jetted upward from the sediment surface, much as morphologically similar populations in peritidal environments of the Bahama Banks do today. E, F, Chroococcalean cyanobacteria from silicified playa lake carbonates of the Upper Proterozoic Bitter Springs Formation, Australia. G, Vase-shaped protist from the Upper Proterozoic Elbobreen Formation, Spitsbergen. H, Low magnification view of oscillatorian cyanobacteria from the Upper Proterozoic Backlundtoppen Formation, Spitsbergen, showing the alternation of vertical and horizontal orientations characteristic of many mat-building populations. I, Acritarch isolated from shales of the Upper Proterozoic Chuar Group, Arizona. J, Endolithic hyellacean cyanobacterium in silicified ooids from the Upper Proterozoic Limestone–Dolomite 'Series', central East Greenland — ooid surface is toward the top of the photograph. K, Spheroidal microstructure from a population showing various stages of binary division, Early Archaean Onverwacht Group, South Africa. L, Large, process-bearing acritarch preserved in chert nodules within a moderately metamorphosed succession of latest Proterozoic age, Prins Karls Forland, Svalbard. Bar = 30 μm for A, 10 cm for B, 400 μm for C, 100 μm for D, 20 μm for E, F and I, 50 μm for G, H and J, and 75 μm for L.


A
B
C
D
E
F
G
H
I
J
K
L

woona Group (Schopf & Packer 1987), where they occur in association with clusters of spheroidal unicells encased in multiple extracellular envelopes. These microfossils are morphologically similar to extant cyanobacteria, and may be early representatives of this group; however, that interpretation is by no means assured. Even if the fossils do represent early cyanobacterial ancestors, there is no assurance that they were oxygenic photoautotrophs using two photosystems. In the presence of H_2S, many living blue-greens photosynthesize anoxygenically using only photosystem I, i.e. H_2S, H_2, or organic molecules donate electrons, and no O_2 is produced. Comparative biochemistry indicates that this photosynthetic system evolved earlier than the cyanobacterial (and higher plant) pathway in which water donates electrons.

The apparent low morphological diversity of described Early Archaean microfossils cannot be taken too literally. Studies of Early Proterozoic assemblages from Western Australia have demonstrated that, as morphologically varied assemblages of fossils undergo increasing diagenetic and incipient metamorphic alteration, they become 'archaeanized' — i.e. they appear to converge morphologically on the simple microstructures found in weakly metamorphosed Early Archaean cherts (Knoll *et al.* 1988).

The biological fixation of CO_2 is accompanied by a marked fractionation of the stable isotopes of carbon, ^{12}C and ^{13}C. Carbon isotopic ratios in Onverwacht and Warrawoona carbonates and kerogens indicate significant fractionation between oxidized and reduced species, suggesting an Early Archaean carbon cycle fuelled by photosynthesis, possibly under conditions of elevated Pco_2. Sulphur isotopes are likewise fractionated during dissimilatory sulphate reduction, but in contrast to carbon, Early Archaean sulphur-bearing samples show little fractionation between sulphides and sulphates. At the same time, sedimentological evidence indicates that sulphate was an important anion in the water bodies beneath which both the Onverwacht and Warrawoona beds accumulated. This apparent paradox has several possible explanations: (1) it is possible that Early Archaean oceans contained negligible sulphate concentrations, and that rocks containing evidence for sulphates in both the Onverwacht and the Warrawoona groups accumulated under non-marine conditions — an explanation that is unsatisfactory to many geologists familiar with the rocks; (2) it is possible that significant concentrations of sulphate existed in oceans for several hundred million years before prokaryotes learned to use it — an explanation unsatisfactory to microbiologists, who note that bacteria evolve rapidly to exploit novel substrates; or (3) perhaps almost all sulphate in pore fluids was reduced biologically to sulphide in an essentially closed system with little fractionation because of high ambient temperatures (70°C or more) — a theory for which the geological record provides little supporting evidence. A generally acceptable solution to this problem has not yet been proposed.

Despite outstanding problems of palaeobiological interpretation, it seems clear that 3500 Ma the Earth supported complex prokaryotic ecosystems driven by photosynthesis. Oxygen may have been generated by Early Archaean cyanobacteria, but geochemical evidence indicates that any O_2 produced was largely consumed by the oxidation of organic matter, ferrous iron, and sulphides. Ambient Po_2 appears to have been low and physiological pathways, consequently, anaerobic. Oxide facies iron formation is found in Early Archaean basinal facies, but not in shallow volcanic platform sequences, prompting speculation that oxygenic photosynthesis may have originated in 'mid-gyre' environments far from sites of volcanic or sedimentary H_2S generation. Comparisons of informational macromolecules in extant micro-organisms independently suggest rapid metabolic diversification early in evolutionary history. Early branching groups in both the eubacteria and archaebacteria are predominantly anaerobic, thermophilic, and sulphur-dependent; several are autotrophic (Woese 1984).

The search for older biological records is limited by the paucity of pre-3500 Ma sedimentary sequences. 3800 Ma rocks from Isua, southwestern Greenland, contain reduced carbon that is isotopically fractionated relative to carbonates in the same succession, but the metamorphism of these rocks to amphibolite grade has obliterated any unambiguous indications of biological activity. Later Archaean successions in Australia, Africa, and North America contain diverse stromatolites, rare microfossils of cyanobacterial aspect, and local evidence of unusually strong carbon isotope fractionation. Most of the isotopically light kerogens come from non-marine deposits, so their interpretation in terms of global conditions is not straightforward; however, it has been suggested that isotopically light kerogens fix a minimum age for the evolution of aerobic methylotrophy (the metabolic oxidation of methane or other one-carbon compounds; Hayes *in* Schopf 1983).

The Early Proterozoic Eon: the diversification of aerobes

The modern era of Precambrian palaeontology began in 1954 with the brief description by S. Tyler & E.S. Barghoorn of microfossils preserved in cherts from the 2000 Ma Gunflint Iron Formation, Canada. Subsequent research has demonstrated that several discrete microfossil assemblages occur in Gunflint rocks. Stromatolitic cherts near the base of the formation contain abundant microfossils preserved as organic, haematitic, or pyritic structures. Although more than a dozen valid species have been described from this facies, two taxa together comprise more than 99% of all individuals (Fig. 1A).

Gunflintia minuta is a thin (usually 1–2 μm) filamentous sheath that has been compared to both nostocalean cyanobacteria and iron bacteria. Its affinities remain uncertain; locally inflated areas along filaments interpreted as akinetes and heterocysts (distinct cell types produced by nostocalean blue-greens) are probably diagenetic in origin.

Small (2–15 μm) spheroidal fossils assigned to the genus *Huroniospora* occur in the same beds. The phylogenetic relationships of these populations are also unclear, but their recent interpretation as bacterial spores merits serious consideration. Other microfossils in the Gunflint stromatolitic assemblage are uncommon; they include probable iron-oxidizing bacteria, possible cyanobacteria, and problematica, but no strong candidates for eukaryotic assignment.

Although these fossils occur within laminated stromatolitic structures, *Gunflintia* and *Huroniospora* populations do not display the orientations characteristic of mat-building micro-organisms in younger rocks. Thus, like their phylogenetic relationships, their ecological interpretation as mat-builders is open to question.

Non-stromatolitic Gunflint assemblages include microbenthos preserved in silicified muds and probable planktic populations. The mud microbenthos is dominated by stellate microfossils interpreted as iron and manganese oxidizing bacteria, while the apparent planktic forms are 6–31 μm diameter spheroids of uncertain systematic position. Whatever the taxonomic affinities of Gunflint microfossils, it is clear that generally similar assemblages were widely distributed 2000 Ma. Assemblages comparable to Gunflint mud, mat, and plankton florules occur in Labrador, the Canadian Northwest Territories, and two areas in Western Australia (references in Knoll *et al.* 1988). Not all of these occur in iron formations, and several contain microfossils not found in the Gunflint Formation itself. For example, silicified carbonate muds of the Duck Creek Dolomite, Western Australia, contain septate filaments as much as 63 μm in diameter — among the largest such fossils known from any Proterozoic formation. Although Gunflint-like assemblages are widely distributed in Lower Proterozoic formations, they are not the only fossils in rocks of this age. Assemblages from hypersaline peritidal rocks of the Belcher Supergroup, Hudson Bay, Canada, contain populations that are indistinguishable from cyanobacteria found today in comparable environments (Hofmann 1976).

Stromatolites are abundant and morphologically diverse in Lower Proterozoic platform carbonates (Walter 1976). It is not certain whether the observed increase in stromatolite diversity between the Late Archaean and Early Proterozoic eras reflects a radiation in mat-building prokaryotes, a preservational consequence of Late Archaean continental crustal growth and stabilization, or both.

What may have been the most profound evolutionary changes of the Early Proterozoic Era are events that must be inferred from sedimentological and geochemical data. During the Early Proterozoic, the degree of isotopic fractionation recorded in sulphur-bearing minerals increased substantially. Detrital uraninite ceased to be a significant constituent of fluviatile and deltaic sediments, while red beds became widespread. Limited data suggest that iron retention in palaeosols developed on mafic parent materials decreased by the end of this interval.

Beginning with Preston Cloud, numerous commentators have suggested that these phenomena reflect a significant increase in the partial pressure of oxygen in the Earth's atmosphere. This has sometimes been interpreted as meaning that the Early Proterozoic atmosphere shifted from reducing to a composition comparable to the present; however, such a black-and-white view no longer seems tenable. The Archaean (especially the late Archaean) atmosphere undoubtedly contained some molecular oxygen, albeit in low concentrations. At the end of the Early Proterozoic Era, the atmosphere probably contained only one to a few per cent of present day O_2 levels. The difference, however, is metabolically significant; aerobic respiration is possible in the latter atmosphere, but not in the former. Some palaeontological evidence supports the idea of Early Proterozoic aerobic prokaryotes, but clearer insights come from molecular phylogeny and comparative

physiology. In many aerobic physiological pathways, oxygen-requiring steps are appended to an otherwise anaerobic series of reactions (Chapman & Schopf *in* Schopf 1983). Molecular data, specifically comparisons of nucleotide sequence in 16S ribosomal RNA molecules among different living microorganisms, suggest that aerobic respiration evolved independently in a number of groups, most of which are fundamentally photoautotrophic (Woese 1984). If one accepts that broad constraints on the timing of evolutionary events can be gleaned from molecular data, then it can be inferred further that the polyphyletic evolution of aerobic prokaryotes occurred during a relatively brief period following a long period of anaerobic evolution (Fig. 2).

The later Proterozoic Eon: the emergence of protists

Although treated last in this chronological account, the later Proterozoic Eon might have justifiably been discussed first, because its palaeobiological record, especially for the period 900–600 Ma, is far more extensive and better preserved than that of earlier epochs. Nearly 200 Late Proterozoic microfossil biotas are known from seven continents (Knoll 1985). Environmental sampling is far better than for earlier eras. Thus, it is in later Proterozoic sequences — where the record is clearest — that principles of palaeoecological, palaeogeographical, taphonomic, systematic and, hence, evolutionary interpretation can best be established.

Late Proterozoic microfossil assemblages have been reported from silicified carbonates representing a variety of peritidal depositional environments. *In situ* microbenthic populations occur in stratiform stromatolites and, much less frequently, in conoidal, domal, or columnar forms (Fig. 1C–F, H). Microbenthos can also be found in silicified micrites, oncoids, and ooids, as well as in shales and, rarely, in unsilicified carbonates. There is a strong correlation between facies and assemblage composition. Many populations are convincingly interpreted as cyanobacteria, although under exceptional circumstances bacterial heterotrophs can be recognized. Less amenable to interpretation are populations of unornamented 10–20 μm spheroids that are distributed sporadically throughout most fossiliferous rocks. Although their simple morphology precludes confident systematic

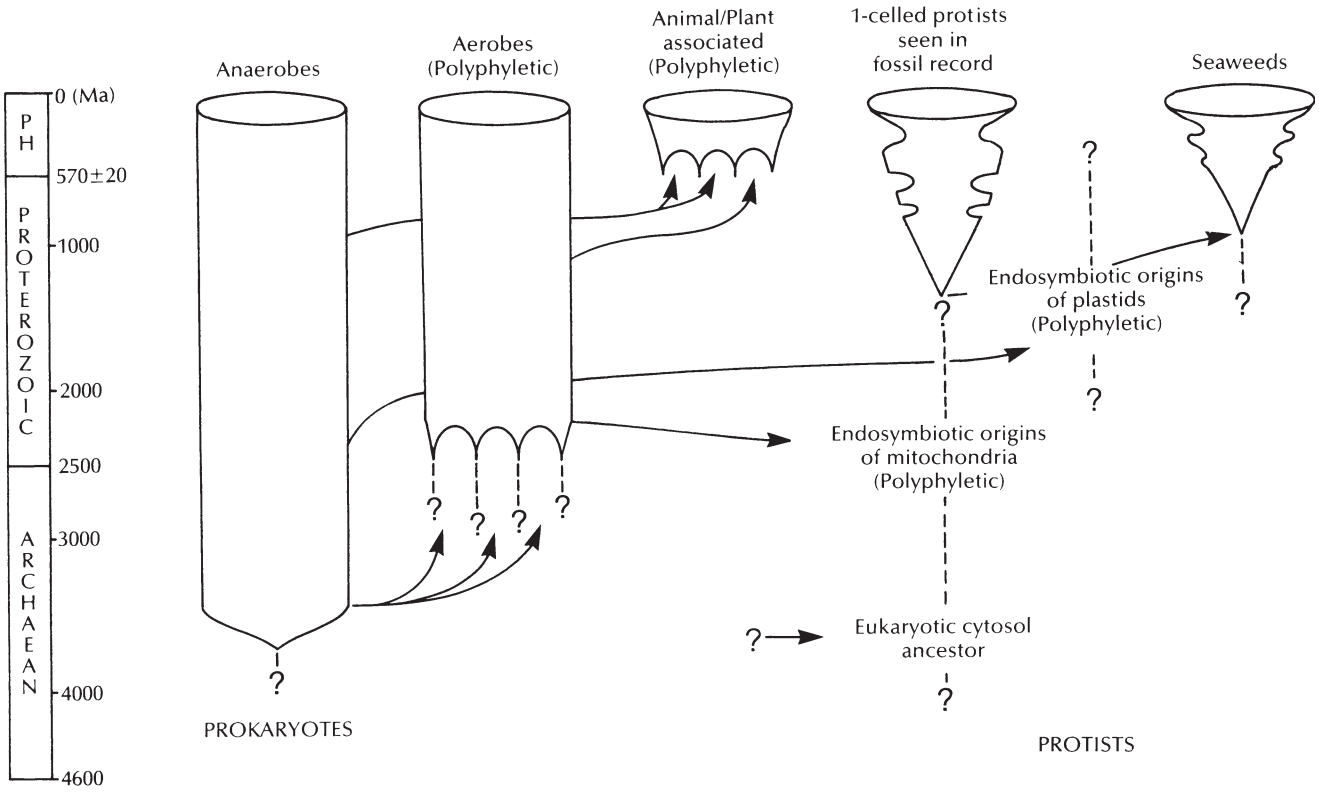


Fig. 2 Summary chart illustrating generalized patterns of prokaryotic and protistan evolution.

classification, some of these fossils resemble the cells and cysts of green algae and protozoans that occur in modern microbial communities of peritidal and hypersaline lake environments. Judging from their spatial distribution within and among facies, other spheroid populations appear to be allochthonous, probably planktic, elements.

Many Late Proterozoic prokaryotes differ little in morphology, development, or behaviour from living cyanobacterial populations found in physical environments like those inferred for the fossils. For example, endolithic microfossil assemblages found in silicified ooids from the 700–800 Ma Eleonore Bay Group, East Greenland, contain half a dozen discrete populations which have close modern counterparts in present day Bahamian ooid shoals (Fig. 1J). Late Proterozoic cyanobacteria appear to be essentially modern in their diversity and environmental distribution. One can hypothesize that the apparent increase in cyanobacterial diversity recorded in the Proterozoic as a whole is mainly a function of more complete sampling in younger successions; that is, the major features of cyanobacterial diversity were established during the Early Proterozoic Era or earlier. This hypothesis cannot be rejected on the basis of currently available data. The record of other prokaryotes is less clear, although the presence of Late Proterozoic sulphate reducers, methanogens, methylotrophs and other bacteria can be established or inferred on the basis of geochemical evidence.

Stromatolites provide sedimentary evidence for the continued wide distribution of microbial mat communities in later Proterozoic environments (Fig. 1B). It has been suggested that Proterozoic stromatolites changed systematically as a function of age, and that this provides indirect evidence for Proterozoic cyanobacterial evolution. Several objections can be raised against this view: (1) the debate over the stratigraphic distribution of stromatolite forms continues unresolved — hindered by the failure of many reports to place stromatolites in their proper sedimentological perspective and by the absence of a rational, internationally accepted system of nomenclature; and (2) it may well be true that certain stromatolites characterize particular time intervals, but this does not necessarily say anything about cyanobacterial evolution. Differences between Early and Late Proterozoic stromatolites may as easily reflect the addition of eukaryotic algae to mat-building communities, temporal changes in features of the physical environment (such as $CaCO_3$ supersaturation), the evolution of uncalcified metaphytes that outcompeted microorganisms for space in certain environments, or the evolution of meiofaunal grade metazoans.

Undisputed protistan fossils are abundant in Upper Proterozoic rocks. Large (up to 2 mm) acritarchs occur in both silicified carbonates and shales (Fig. 1I); some of these may represent the phycomata of planktic prasinophyte algae, but the systematic relationships of most are uncertain. Latest Proterozoic cherts and finely laminated shales from China, Australia, and Svalbard contain particularly complex forms, including spiny and process-bearing populations (Fig. 1L). In their general level of morphological complexity, these resemble younger Palaeozoic acritarchs, but the Proterozoic forms are invariably much larger and are certainly distinct at the specific and, usually, the generic level. Recent discoveries in Spitsbergen and Arctic Canada demonstrate that the record of spinose and process-bearing acritarchs goes back at least to 800 Ma. Vase-shaped microfossils of uncertain systematic position also occur in Upper Proterozoic shales and carbonates (Fig. 1G); in some successions, they are among the most abundant fossils preserved.

Like fossil prokaryotes, protistan microfossils reflect palaeoenvironments in their distribution, but unlike prokaryotes, they change systematically through time. Therefore, acritarchs have proved useful in at least Late Proterozoic biostratigraphy (Vidal & Knoll 1983; Hofmann 1987).

The record of eukaryotes can be traced though time at least back to 1700 Ma, when both the morphological and molecular geochemical records of protists begin (Jackson *et al.* 1986). The record of metaphytes may be almost as long. Diverse multicellular algae occur in Upper Proterozoic rocks (Hofmann 1985); with somewhat less confidence, both carbonaceous and trace fossil remains in 1300–1400 Ma rocks can be interpreted as seaweeds. No unequivocal remains of metazoans have been described from pre-Ediacaran deposits. Thus, either seaweeds and animals originated at strikingly different times or, for the first half of their history, animals must have been tiny, meiofaunal grade organisms unlikely to survive as fossils or produce recognizable traces.

While the palaeobiological trail of early eukaryotes currently turns cold at about 1700 Ma, it must be admitted that nucleated cells that were incapable of fossilization or, at least, unlikely to be recognized as eukaryotic, almost certainly existed earlier. How much earlier is unclear. The ancestors of the eukaryotic cytosol (nucleus and cytoplasm) appear to