



HAL
open science

Chemical advances in evolution by and changes in use of space during time

R.J.P. Williams

► **To cite this version:**

R.J.P. Williams. Chemical advances in evolution by and changes in use of space during time. Journal of Theoretical Biology, 2010, 268 (1), pp.146. 10.1016/j.jtbi.2010.09.021 . hal-00642434

HAL Id: hal-00642434

<https://hal.science/hal-00642434>

Submitted on 18 Nov 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author's Accepted Manuscript

Chemical advances in evolution by and changes in use of space during time

R.J.P. Williams

PII: S0022-5193(10)00495-9
DOI: doi:10.1016/j.jtbi.2010.09.021
Reference: YJTBI6161

To appear in: *Journal of Theoretical Biology*

Received date: 23 August 2010
Revised date: 15 September 2010
Accepted date: 15 September 2010

Cite this article as: R.J.P. Williams, Chemical advances in evolution by and changes in use of space during time, *Journal of Theoretical Biology*, doi:10.1016/j.jtbi.2010.09.021

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



www.elsevier.com/locate/jtbi

Chemical Advances in Evolution by and Changes in Use of Space during Time

R.J.P. Williams

Inorganic Chemistry Laboratory

South Parks Road

OXFORD OX1 3QR

UK

ABSTRACT

The general features of theoretical biology are outlined, starting from the beginnings of modern bioenergetics, energy transduction, in 1961. The major characteristics of the bioenergetic systems are then described and it is pointed out that especially the treatment of energised, directed chemical flow in divided space in cells is central to this activity. It is also the major theoretical problem of many other cellular activities. The development of bioenergetics until today, based on the theories of flows of electrons, protons and other chemical materials in structures, is then outlined. Great benefit has accrued from the knowledge of the structures which allow controlled flow, diffusion. In a speculative forward look at the theoretical biology of cells based on this outline the usefulness of further reductive analysis of extracted molecular units is considered as

descriptions of mobility within structures of large molecules and membranes as well as in free space. Such structures are central to catalysed control and flow. The possibility of a more holistic approach is then examined. This will require markers of flow states and spatial structures. It is shown that as the flow system is one of a combined environment and organisms the study of metal ions in both has particular advantages. Consideration is given as to how this study can assist a broad understanding of the chemistry of the evolution of organisms. The changes of the metallome, strongly linked to the proteome, the genome and the environment, is a ready source of information concerning the manner in which changes in catalysts and controls of flow of organic chemicals have evolved in the divided space of cells and of cellular systems. The connection of the code changes to the environment via the metallome changes is then mooted.

1. Introduction

It came as a very pleasant surprise for me to be asked to write an open-ended article on my views of the outlook for theoretical biology in celebration of the 50th anniversary of this journal. I shall begin with some general remarks about the underlying features of the theory of the physical chemistry which are central to the understanding of processes which have contributed to life's origin and evolution. They are all linked to energy transduction. I shall then go back to the development of ideas I

put forward in the very first article in this journal (Williams, 1961), where I explored the nature of the initial steps of energy transduction in cellular life. The elaboration of these views, largely done by others until today, will be described. It has given rise to three Nobel Prizes. I shall then look at the possibility of generalising the central points in a wide-ranging consideration of the nature of biological systems and particularly their evolution, not their origin, and the strong link between the changes of the environment and those of the code.

The essence of theoretical biology must be the understanding of the complexity that has arisen or can arise in the process of energy, light, going to heat via intermediates, Fig. 1. The intermediates here of concern are the energised chemical activities of living organisms. There are also natural processes which are deeply indirectly involved with organisms other than those we describe under life which are driven by the transduction of light as it goes to heat. They are geophysical and geochemical. The simplest to describe, though still hard to understand, is weathering due mainly to the circulation of water from the sea by evaporation followed by condensation in clouds and rainfall on land. It leads to massive irreversible erosion and movement of elements from land to sediments and the sea. Our inability to understand in detail the flow of the simple physical molecules of water, H_2O , and its consequences, described by the popular word “the weather” and erosion,

are matched by the difficulties in understanding the flows under the surface of Earth's upper crust which lead to the movement of tectonic plates, upwellings at ocean ridges and volcanoes. The overall result of these two physical flows is the redistribution of the chemical elements, especially in the seas and deposits of novel material, sedimentary land, which form much of life's more recent environments. Other sediments provide a history of the sea's chemistry. Further physical cycles have been active among which are the fluctuations of temperature which in the extreme led to periods of so-called Snowball Earth and times when the melting of ice led to huge sea level rises and floods. All of these processes are interactive with the evolution of life through chemical change, as I shall discuss.

2. The Central Points of Energy Transduction

In my paper (Williams, 1961), I presented certain points of view which I shall also refer to below in a different context. They were not all original but taken together gave a new view of bioenergetics.

1. Biological cells arose and evolved as divisions of space, two of which were observed as organelles, one using light, chloroplasts, and one using oxidation/reduction equivalents, mitochondria, to provide energy to the cells. All their important chemical

transformations were referred to as *dislocated reactions across kinetically isolated spaces*.

2. Energy transduction within an isolated region of the organelles' systems gave rise to separation of electronic charge in two parts of the divided space, Fig. 2. This, a separation of oxidative (positive charge) from reductive (negative charge) regions, originated in bacteria. The major external energy source is of light acting on chloroplasts.
3. Separation of electronic charges could produce gradients of ions across the absorbing region which divided space. The proton gradient was a special case.
4. The gradient of protons diffused to a different part of space and by passing through chemical machinery produced, while it was neutralised, energised reagents useful in transducing the energy further, notably in those chemicals able to drive condensation, e.g. ATP, Fig. 2.
5. The charge separation could also be used to produce oxidised chemicals (O_2) in one part of space and reduced chemicals (bound H) in another, from water, Fig. 2.
6. The three energised conditions of chemicals, one of separation of oxidising from reducing agents, a second of chemical gradients and

a third of condensation reagents led to the biochemistry, the energised metabolism, of cells.

7. The separation of released oxygen and of contained bound hydrogen, both from water in different parts of space could feed back to other organised reactions across different space separations to make proton gradients and then ATP in a second way in mitochondria. Such chemical reactions could have predated energisation due to light.

I need to add that the division of space can be by a membrane. All the processes were known to occur in bacterial membranes as one way of achieving kinetic isolation. At the request of Peter Mitchell I explained in letters that a membrane was a perfectly reasonable dividing device but not essential for the above energy transductions and I referred to two types of electrochemical cells. I did not do, indeed could not do, confirming experiments on the biological systems, Mitchell and many others did so later.

The exchange between Mitchell and myself is published in full (Weber and Prebble, 2006), and I shall not go into the matter here.

Mitchell in a later note in *Nature* (Mitchell, 1961) described his views (as if our exchange or my articles did not exist). In major features they do not differ though he elaborated them differently in that he described the proton gradient in terms of equilibrium charge and proton potentials and

gave a very different mechanism of ATP production. A major difference which is important in this article is that I postulated selective diffusion of chemicals, here of protons, in space of two kinds, while Mitchell believed movement of protons in anything but water was impossible. His views are really those of a microbiologist with an interest in physiological membrane potentials, while mine came from the discipline of chemistry. His mechanism of ATP formation was electrolytic field driven and is impossible. Enough. Let us look now at how the theory has developed, mostly in the hands of others, and then see how diffusion between organised structures, that is directed flow of energy and material in different points of space, is central to any theory of biological systems. Two modes of diffusion were considered in 1961, chains of catalysts for electron transfer and restricted diffusion in water and other media to reaction sites in the flow of protons.

3. Electron Transfer

In 1961 apart from myself many others had started to explore electron transfer over long distances in biological protein and model assemblies. I wish to mention in particular the work of Britton Chance to which I referred at the time (Lloyd and Chance, 1964). I advanced my thinking, using model mixed oxidation state conductors with copper and iron dilute cationic centres. The nature and the theory of these compounds was developed with one of my pupils (Culpin et al., 1965; Robin and

Day, 1967) but the real advance in the analysis of electron transfer in the relevant biological materials came with Dutton's work on electron transfer in the proteins of photo-reaction systems, Fig. 2, using Marcus tunnelling theory (Moser et al., 1992). The analysis was confirmed by the spatial distribution of the electron carriers in the membrane protein structure (Deisenhofer and Michel, 1989). Gray and Winkler (1996) extended knowledge to other model protein systems. By 2000 the way electrons moved in membranes was well understood and backed by crystal structures. A very interesting further development was the cyclic flow of electrons in Particle 2 of the electron transfer chain due to Mitchell. We brought these treatments into a general approach to electron transfer in all kinds of structured systems including those in the metallic state and those which are close to insulators with metal-containing proteins in the middle (Edwards et al., 2009). I regard the treatment of Fermion diffusion as effectively complete in contrast to that of Boson diffusion where the superconductivity electron pair theory is still not very good. It is probably not relevant to biology but we cannot be certain.

4. Proton Diffusion

The argument about proton diffusion as rate-controlling in ATP synthesis as opposed to equilibrium proton distribution in space around membranes appears to be settled in that both contribute. The theory of diffusion along a membrane surface has been well studied and

experimental evidence in favour of this mechanism has been published by several authors (Alexiev et al., 1995). Much of further understanding depends on the dynamism of the structure of water layers close to membrane surfaces and the total structure of the transducing membranes in mitochondria and chloroplasts. This is two-dimensional diffusion but we shall have to consider as well diffusion in one dimension and free diffusion in three dimensions. The diffusion of the proton through the membrane ATP-synthesis is now well-established through detailed crystal structure studies and analyses of the mechanical properties of the complex protein coupled to the uptake into it of ADP + P giving ATP (Walker, 1998), see the schematic diagram, see Fig. 3. One remaining problem is this movement of protons in the membrane (*pace* Mitchell) on the initial creation of the proton gradient and in its final neutralisation while driving ATP-release. Protons, unlike electrons, can only diffuse short distances between non-metal centres of protein side chains and water molecules in the membranes. Contribution by tunnelling may be found but has not been noted. The proton could only tunnel over distances of fractions of an Ångstrom while an electron can tunnel up to 20 Å in one step in a protein. Both diffusion steps have an associated relaxation energy. The relaxation energy associated with particle motion and the mechanical dynamics of segments of proteins on proton and ADP + P binding and ATP release, see below, return me to the central points of

my 1961 paper, the so-called “dislocated” space in cells and their connected activities in organisation by diffusion. The understanding of diffusion, broadly speaking, I take to be a major task of the future theories of biology and it includes that of changing flux as well as within steady states. Let me explain my views which depend on fixed and mobile large structures as in a modern machine.

5. Space and Time in Cellular Systems

Even the simplest cell has structure. Major structural elements are the macromolecules themselves, some isolated, some in higher structures, and some associated with membranes. The macromolecules and membranes have internal mobility of both their main chains and side chains. Moreover some have a degree of translational degrees of freedom, free mobility between very small or large distances, but all are extremely important. The membranes or proteins can be in, or can come into, contact forming higher dynamic structure and connected in activity by smaller organic molecules and metal ions both of which may become virtually permanently bound but many are in part free and mobile. In the second case protein surfaces act as selective donor and acceptor sites. In a later part of this article I shall refer to the different classes of metal ions which work in this way, much like the diffusing or non-diffusing organic molecules, to see how activity is controlled by flow of messengers between centres of activity or to create them. Now all of these motions,

both free and controlled diffusion, contribute to necessary coordinated activity and growth. Space in cells is seen to be divided or criss-crossed by membranes, with channels and pumps, and large mechanical, elastic molecules, for example by filaments of tubulin and acto-myosin which control flow. In evolution the total dynamic organisation of flow, using these dynamic structures, increased in complexity by division into many different spatial enclosures, vesicles and kinetically isolated spaces. It is important to remember the diffusion/reaction theory of Turing here (Turing, 1952). For resources of energy and materials all cells are dependent on their environment with which later cells have a more and more interactive dependence. Once again I shall illustrate these functions with different metal ions as they are an extremely important part of the total environment/organism system in evolution. In fact the environment and organisms have evolved together in very slow change over millions of years of flows in both. The environmental flows were described above but not the interdependent interactions with biological flow, see below. While we can only speculate as to the origin of this interaction we can study (and experiment) with its means of evolving.

This is a major problem and we can look at it in two very different ways. The simpler is to examine and to develop experimental and theoretical treatments of separated units such as DNA, RNA and proteins and then attempt integration. This is the common and more conventional

reductive procedure. Structure of these molecules gives first the sequences of bases, DNA/RNA, or amino-acids, proteins. We have today complete knowledge of their coded relationship. This is in part the genetic analysis procedure but why are there just this very limited number of each? How did the connection between code and amino-acids arise? Is there a quantum mechanical answer or is it pure chance? Are the units those which will arise through their kinetic stability in an initially random set of flowing chemicals? Now the molecules have folded forms and interact to give metabolic and synthesising units. The variety of folds is considerable but limited (Lesk, 1996). The energetics of the folding process and the final states is still open to discussion and many proteins could be in unstable thermodynamic states though of long life (Dobson, 2003). In the case of proteins there are also those which can rearrange into more stable states and it is considered that this reorganisation can lead to serious human diseases, such as Alzheimer's. We need to be confident that we understand the folding process of proteins and how it is kinetically controlled by ribosomes and chaperones. Now it is not the case that all proteins fold and certainly their mobility is very variable. I was involved with the Oxford Enzyme Group in exploring mobility of proteins by NMR, both those of side chains, giving induced fit rather than lock and key matching, as well as main-chain motion (Williams, 1977 and 2005). The protein folds, especially β -sheets, are of sufficient

strength to create special catalytic centres of activity (Gray et al., 2000). It was quickly apparent from a series of studies that main-chain motion, even to the extent of some proteins remaining in an unfolded condition, was equally necessary for some biological functions. We noted in particular that helical proteins were more mobile internally, especially in membranes and signalling, but we noted too allosteric switches, and that β -sheet proteins were relatively rigid, especially defining enzyme selectivity. Fig. 3 refers to an example in bioenergetic membranes. The motions of helices are at the heart of biological mechanisms, including bioenergetics. This allows me to make an aside to ATP-synthetase, the remaining component of energy transduction from the 1961 analysis, not discussed so far.

Both Mitchell and I failed to give a good description of the ATP synthetases. The suggestion of Boyer as to the importance of energy for removal of bound ATP gave a lead to the mechanical nature of this enzyme. True insight was not available until the crystal structures of ATP synthetase (Walker, 1998), gave us a static model, which was used to describe the dynamics of a mini-machine, the general nature of which had previously been demonstrated in part. The lesson was an amplification of that learnt earlier from studies of hemoglobin and of calmodulin and of enzymes such as lysozyme (Williams, 1977). The essence lies in the relative mobility of helices, moving rods, and the immobility of sheets,

Fig. 3. The relationship to macro-mechanical machinery such as the combustion engine is clear. Helical structures can also be deformed elastically in units as in single proteins and multi-protein complexes such as acto-myosin. With the lessons from macro-machines the analysis has moved from single to multi-protein complexes and motors connected in extended networks, coordinating action and connected by chemical (messenger) flow or by electric flows (nerves) related to wirings in all kinds of mechanical and electrical mini- and macro-machines. Very frequently the flows have a feedback link included in Turing's reaction-diffusion theory, see below.

6. Grosser Motions

Unfolded, stable, mobile proteins are very rare inside the cytoplasm of cells but are found in vesicles and extracellular space. Frequently they are heavily glycosylated or phosphorylated or oxidatively modified. A different case is that of connective structure proteins which are often in very extended flexible structures, for example, collagen. Other globular proteins are organised in mobile structures by recognition of opposite surfaces to give long, also flexible, filaments such as actinomyosin and tubulins. Both are capable of generated directed transfer of other molecules, even proteins or of mechanical force. The theory of these forms of transport is that of engineering machinery. The way in which proteins come together selectively to form these and other

structures, including binding to DNA and RNA, is no different from the fitting of substrates to enzymes, except in scale. As the surfaces of proteins are more mobile and more extensive than those of substrates there are many possible very similar states. There is then the possibility that an assembly of proteins can fluctuate or can be adjusted in shape, energised, as in muscles, or that one protein can travel on another as in transport on tubules, linear diffusion. Bulk properties such as elasticity of these filaments are in principle open to theoretical analysis at the molecular level but are probably less important than their role in networks. It is very apparent that cooperative guided motion (or guided diffusion), not random diffusion, is to be seen in many cellular activities. Even whole-body or very large segment motion, macroscopic changes of state, has to be treated by analysis of the way systems behave to which we turn later. Bioenergetic organelles and muscle behaviour, probably the best studied relatively large-scale cooperative systems so far, can be approximately analysed between the molecular detail and the large scale but the molecule by molecule approach gets more and more difficult as larger units of cells are examined. For example we know that the organelles, mitochondria and chloroplasts, undergo membrane rearrangements as well as coupling between internal and external components on energisation. Now an alternative approach to the reductive analysis of separated units followed by attempted creation of

the whole by putting the pieces together, is that of the direct analysis of whole systems, referred to sometimes as holistic analysis or systems biology. This analysis will again require reference to free diffusion from donor to target and controlled motions in or by structures.

7. Systems Biology: Models and Basic Principles

How are we to come to terms with the complexity of biological systems with the very large number of mobile and structured, many relatively rigid, components in cells and their interaction with their mobile environment (Barrow et al., 2006)? The difficulty lies in the analysis of rate processes, flows, which are often in a feedback relationship, Fig. 4. Feedback control is not open to algebraic analysis and much theory has been developed by engineers, especially using computerised analysis. The magnitude of the problem can be seen by reducing observation to the analysis of the simplest non-biological systems. For myself I like to try to formulate what appears to be a very different very simple problem. I refer to the intermediate stages in the flow of energy from light to heat through evaporation of water (Williams and Fraústo da Silva, 2006). The intermediate is the formation of clouds. There are similarities to living systems in that clouds are classified by shape and that while clouds travel they maintain recognisable shape for some time after growth though ultimately they fragment or evaporate. The only molecular components in them are water molecules. The

molecules are in constant motion in temperature gradients as the droplets absorb heat (not light) at low levels in the cloud, evaporate becoming isolated molecules, and on rising to a great height the molecules condense to droplets which fall. Surprisingly this gives rise to clouds with isolated shapes with no symmetry! Compare cells. The reason is that droplets grow more rapidly on increase in size and evaporate faster as they get smaller. It is the intimate interaction with the energy source together with evaporation in an environmental temperature gradient which gives rise to this behaviour. Cooperativity is the essence of phase change much as it must be in large-scale shape changes in biology and cannot be seen in single units of the same material. Diffusion here is guided by gradients of gravity and temperature. Unfortunately the temperature gradient is difficult to quantify as in part it is altered by the size of the cloud and winds produced by distant temperature gradients. There is then a feedback character to the cloud shapes. Several kinds of shapes exist and they reappear time and time again but the puzzle remains that they can be classified as cumulus, stratus, cirrus and so on. The more we can define this dynamic interactive system mathematically the more we may be helped with the holistic approach to dynamic biological shape.

A different engineer's mathematical approach to shapes is that of rigid structures with hidden dynamics as in stress/strain analysis. Nearly a hundred years ago D'Arcy Thompson (1938) pointed to this approach in

his discussion of the rigid structures based on cones, simple and twisted as in shells, and on big scaffolds such as the Firth of Forth Bridge compared to the ribbed framework of dinosaurs. These semi-rigid systems are not very interactive with the environment. However a combination of the two systems of shaped objects can be likened to the series of organisms from amoeba to vertebrates. D'Arcy Thompson also discussed approaches to bulk dynamics such as elastic properties. A chemical system which is very simple but could also be useful to us in analysis of biological cells would be one which is clearly chemical, not just physical. Consider the formation and shape of the ozone layer which is a permanent dynamic feature of the upper atmosphere but which can be altered in shape (even destroyed) by man's injection of chemicals into the environment. It is again the production of an unstable, now chemical, intermediate, O_3 , formed by absorption of UV light, but which falls in the gravity field of life to a lower level where it decays to O_2 and rises again. It is altered in shape as are clouds by environmental factors which alter chemical rate constants. The difficulties of analysis are very like those for the treatment of clouds. The essence of both is the link between the restricted systems observed and their environment. There are no codes here as must have been true in the origin of life. With these two non-biological steady states in mind I turn now to cellular systems.

8. The Systems of Organisms and the Environment

The object of this section is to explain the problem of the changing nature of organisms as systems, Fig. 4, since we have no sure knowledge as to how organisms were first formed. I shall take it that although we do not know how life started we do know that it is a chemical system and it became reproductive, that is coded. The start however had to be from small energised environmental chemicals and it is not possible for us to reason about such chemical systems from RNA or DNA codes since neither can form membranes or undergo the essential redox changes of energy capture. Furthermore their synthesis needs much simpler chemical compounds. The essence of life is the creation of a barrier, composed of small molecules, a vesicle of some kind, to capture energy and material internally from the external environment. The understanding of energisation of the interior ions and molecules in an enclosed space has to come before that of the beginning of the synthesis of energised but relatively stable, in a kinetic sense, larger molecules. Since the basic building blocks are from C, N, O and H atoms and the environment could only provide them in a very few small molecules, mostly chemically inert, CO₂, CH₄, CO, N₂, H₂O with more reactive H₂S and NH₃, there had to be, as well as ways of capturing energy, ways of increasing reactivity. The chemistry to complex carbon compounds requires reduction of CO or CO₂ as oxidation of CH₄ is too difficult without oxygen. The initial and further reactions of these molecules virtually required catalytic metal ions

capable of redox changes coupled to energy capture. Small organic chemicals cannot give powerful enough catalysts. This is amply apparent in men's organic chemistry today. We can follow the evolution of the availability of the inorganic elements in the sea from the trace elements in sedimentary Banded Iron Formations and Shales (Vine and Tourtelot, 1996; Klein and Beukes, 1993; Siebert et al, 2005; Lehman et al., 2010).

We have no idea yet of how these primitive steps could yield proteins, RNA, DNA, polysaccharides and lipids. I shall therefore only mention those steps which must have been required before reproductive life could start and then I shall jump forward from the start of life which is coded and reproductive in the simplest prokaryote cell and ask how its chemical system could evolve. I shall stress the roles of metal ions as I consider they are the easiest markers we have, allowing us to follow chemical evolution, see below. I assume that certain chemical steps were present, including all those of bioenergetics leading to oxidation/reduction separation and to pyrophosphate compounds, as in ATP, note really Mg^{2+} .ATP, as described already so that energised condensation reactions were possible and that redox catalysts were based on the only available catalytic redox metal ion, Fe^{2+} . Condensation was required for the synthesis of all the large molecules mentioned above together with oxidation/reduction. Pyrophosphate is the major or possibly the only inorganic unit of sufficient chemical stability, yet sufficiently

reactive with the aid of the available mild catalyst, Mg^{2+} , to carry energy to the synthesis of all the compounds of a cell. (The only other simple way to transfer energised units uses thio-acyl compounds, R-CO-SR.) We wish to consider how the simple original resultant cell could lead, 3.5 billion years later, to an animal as complex as man. I shall explain why I do not consider it was due to pure chance alteration of a code. I shall claim that it is a direct consequence of a system of chemical reactions in coordinated flow and as such was inevitable as it was dependent on the inevitable original environment and its change, no matter how much the central line of progression of organisms is confused by the multitudes of varieties, so-called species, that arise at any one time. *There is a main logical chemical progression of the whole of life.* I shall not offer any alternative explanation for the origin of species but I shall assume that the mechanism of random selection is correct. The multitude may in fact represent the confusion of the survival of the fittest, where fitness allows for adjustment to obtain in small ways advantage in slightly variable environments at a given time. The central line of novelty by way of contrast is large-scale chemical change, we shall assert, where chemistry in cells follows the inevitable main change of it in the environment over 3.5 billion years. Life's general chemical progression, not its particular expression as we look around us at the confusion of forms and activities, was therefore inevitable and we still present the chemical evidence after

giving logical reasons for the progression of the environment which it is difficult to deny and which can drive organic change. Just because the environmental chemical changes are relatively extremely simple and the main advance of biological types had to follow them we have referred to evolution of chemotypes (Fraústo da Silva and Williams, 2001). Let us consider such coupled environment/organisms evolution series. We shall not refer until much later in this article to the code (DNA), which is essential for reproduction, but will explore the view that evolution is due to the back and forth flow of changing chemicals between the environment and cellular life.

The original environment of Earth shortly after it cooled was a reduced surface and aqueous solution with an atmosphere of CO₂, CO, (CH₄), N₂ (possibly NH₃ and HCN), H₂S and H₂O (Fraústo da Silva and Williams, 2001). Before cells formed weathering activity had removed much of the CO₂ as CaCO₃. The sea at 3.5 Ga was a reducing medium of something like today's Na⁺, K⁺, Ca²⁺, Mg²⁺, Cl⁻ and some HPO₄²⁻ ionic solution of fixed pH but with no O₂ or SO₄²⁻. Let us assume that this strong solution also inevitably contained H₂S, making it reducing, and a limited number of metal ions, plenty of Fe²⁺, Mg²⁺, some Ni²⁺ and Mn²⁺ but little Zn²⁺, no Cu⁺ (Cu²⁺) and little Mo but probably some W. For life to occur it had to use those elements which were available in catalysts and in assisting organisation of its chemistry. In the cytoplasm the only

way it could exist was to expel Ca^{2+} ions, which would have precipitated too many organic and inorganic anions, and much of the Na^+ and Cl^- ions as they, at the levels in the sea, would have caused osmotic rupture due to the presence in the cell of organic molecules of living processes. Removal of Ca^{2+} , Na^+ and Cl^- is a major, essential, very early required, energised chemical step *before* the contents of the isolated space became active in organic synthesis. It is a very early demand on energy but the ion gradients produced were from that time on possible sources of energy for transport, as were H^+ gradients. (Later by allowing these ions to flow inward on an event disturbing the permeability of the membrane the gradients became a potential source of information in message networks to cells about external conditions and events, see below). Additionally, as already indicated and without going into detail, iron became the source of redox catalysts and magnesium the catalyst for use of ATP. These two cations also became major control messengers of cytoplasmic chemistry. Of the environmental small molecules CO and CO_2 had to be reduced and hence liberated oxygen or oxidising equivalents.

The starting system of biological activity as we know it, was then unavoidably a selected electrolyte and energised solution at a fixed pH of 7.3 capable of catalysis, using Fe^{2+} and Mg^{2+} with perhaps smaller amounts of Zn^{2+} , Mn^{2+} and Ni^{2+} , and so able to give energised organic molecules. This basic system gave rise in some way to a selected set of

organic metabolic pathways including the Krebs cycle, the Calvin cycle, glycolysis, the synthesis of fats and leading to the production of amino acids, monosaccharides and nucleotide bases, Fig. 5 (Kauffman, 1993). The polymers DNA/RNA proteins and polysaccharides followed. It must be that there is a chemical selection of compounds, preferred by optimal kinetic stability, in an earlier whole system. Can we use theory to prove that no alternative existed? To believe that this system arose by unbiased chance I do not find acceptable. No matter. I wish to consider how its chemistry could evolve systematically, treated as a self-sustaining flowing network of reactions without a code.

Now the inevitability of the nature of the required polymers of cellular life, that are produced by reduction so as to generate condensation polymerisation, removal of water, from CO, CO₂, NH₃ and H₂S or H₂O, drove oxidising equivalents, sulfur or oxygen, out of cells. Let us consider that in some unknown way this system came to produce the anaerobic cells we know of. Perhaps H₂S was employed at first as an easy source of hydrogen but, later, it was the use of light acting on water as the hydrogen source which gave oxygen to the environment. It is this oxidation of the environment which is unavoidable and which drives the evolution of life by feedback, inevitably aided by increasing use of energy, Fig. 1. It took place in systematic steps, very slowly adjusting the redox potential of the sea from close to -0.2 towards +0.8 volts. At

present the potential is +0.4 volts. The slowness of the process was due to the buffering capacity of H_2S and Fe^{2+} from about 2.5 to about 0.75 billion years ago (Ga). This chemistry change of reduction to reduction-with-oxidation in cells is the main line of chemical evolution and it is decided by oxidative equilibria in the environment which has a fast rate on the timescale of millions of years. It increased the presence of certain metal ions, some possibly useful, but poisonous at first to simple cells, including Cu^{2+} , Zn^{2+} and MoO_4^{2-} . Organisms, with a reductive cytoplasm, then had to live in this changing, poisonous, environment which they created. It is then permissible to look upon the activity of generating oxygen as an unavoidable pollution of the anaerobic life's world. One route for anaerobes to survive was for them to hide in anaerobic niches, the other was to add, mainly externally, oxidative chemistry to the reductive chemistry of the cytoplasm of the original cells, creating aerobes, Fig. 6. The simplest devices were used – in large part the exploitation of compartments inside and outside cells. The greater the diversity of oxidative chemistry, the more spaces, compartments, were required, hence the title of this article. Generally speaking oxidative and reductive steps in chemical processes have to be kept separate as is clear in man's chemistry. The evolutionary sequence was anaerobic, then aerobic prokaryotes, unicellular then multicellular eukaryotes, the last three dependent on a micro- then a macro-aerobic environment, with

more and more compartments, Fig. 6. This is the main line of change to 0.4 Ga independent of species. Moreover the last three coexist as the complex cells, alone, have not sufficient separate survival strength due to their complexity. In fact the majority of unicellular organisms supports the minority of multicellular organisms. This symbiosis is just an extra way of utilising space efficiently as in modern industrial practice where different locations provide different components. We need to appreciate the details of this sequence of an overall system with flow between compartments and the environment, based on the evolution of oxidative chemistry in the environment. Underlying all is the increase in the conversion of light to heat, Fig. 1.

In summary the major new elements in the environment were copper, zinc, molybdenum and selenium while iron diminished greatly. It is relatively easy to follow the continuous changes in the environment by following changes of elements in sedimentary Banded Iron Formations and Shales, Fig. 7 (Klein and Beukes, 1993; Siebert et al., 2005; Lehmann et al., 2010; Williams and Rickaby, 2011). The elements Cu, Zn and Mo increased tenfold. They show that before 3.0 Ga little changed but from around 2.5 Ga oxidation increased slowly due to the above buffering of the environment until 0.75 Ga when oxygen overcame the buffering, Fig. 6, see also Anbar and Knoll (2002). It is then very

interesting to follow or estimate the changes of the elements Cu, Zn and Fe in organisms during evolution (Fraústo da Silva and Williams, 2001; Williams and Fraústo da Silva, 2006). Using bioinformatics analysis of DNA (Andreini et al., 2009; Dupont et al., 2006 and 2010; Zheng and Gladychyev, 2009) and insight into the evolution of organisms we find that, while molybdenum, zinc and copper increased, the iron content of cells remains approximately fixed and found new uses. The most informative changes in the biological part of the system are in the new uses of the elements Zn and Cu mainly in the increased number of compartments. The cytoplasmic chemistry of all cells is largely unchanged though it develops specialist functions, especially when cells differentiated. The dependence on nickel and cobalt in the cytoplasm of higher multicellular organisms is less than in early unicellular organisms, very probably as early substrates such as available hydrogen and the need to transfer $-\text{CH}_3$ groups are greatly diminished. The major change of cells was oxidation of organic compounds, especially in the periplasm of bacteria and in both the extracytoplasmic fluids of the outer regions and vesicles of eukaryotes. The roles of copper as oxidases in these regions dominates that of iron since Fe^{2+} centres are not stable in the high redox potential in them. Copper oxidases also made possible the cross-linking of extra cytoplasmic proteins and other polymers to make collagen, chitin and lignin ordered structure and hence biominerals in shells. To make

growth possible required the break-up of this network by zinc hydrolases. However zinc and iron also created new chemistry internal to the cytoplasm. We note particularly the zinc finger transcription factors which act as receptors of hormonal messengers. Iron, Fe^{2+} , was still required in reductive metabolism and Fe/S proteins were kept in the cytoplasm. Iron as Fe^{2+} also became increasingly valuable in oxidations by oxygen of substrates and of deleterious chemicals in cells – note the P-450 cytochromes and the 2-oxyglutamate dependent Fe^{2+} proteins. These oxidases avoided release of partially reduced oxygen which is poisonous and take oxygen to water. The point we are stressing here is that the many changes were a response to the environmental changes yet all the environmental changes were due to cellular production of oxygen. It was oxygen which also generated the disadvantageous change of environmental substrates, all carbon became CO, all nitrogen became N_2 or NO_3^- , all sulfur became SO_4^{2-} while in turn, as oxidation potentials rose, the surface sulfide minerals were oxidised to give the possibility for the new metal catalysts, Mo, Zn, Cu, to arise. All influenced cellular evolution. The system is one of a deeply interactive feedback. It is absolutely necessary to change DNA, the code and its expression too but we wish to explore the chemistry changes first, turning to code changes later. All these changes began in earnest with the later prokaryotes, increased with the eukaryotes and accelerate into the sudden Cambrian

Explosion. The Cambrian Explosion seen in fossils is due to the surge in oxidation to give cross-linked external matrices around 0.75 to 0.55 Ga. Enabling the evolution of numerous shells and skeletons.

At this point in the article I insert a general view of our holistic approach to chemical evolution which is based on observing the changes largely in the metal inorganic elements in the environment and those of organic chemicals based on them in cells (Williams and Fraústo da Silva 2006; Williams and Rickaby 2011). The reasons for using this approach are that the metal elements are common to both the environment and organisms in exactly the same form. Their concentrations in both can therefore be followed and compared, showing the mutual interaction and energisations of the different spaces, the environment and the cell compartments, including the cytoplasm. The elements of organic compounds cannot be followed in the same way. The ease of following the inorganic elements is increased by the speed of their reactions which means that they are in equilibrium with their compounds, both in bound states of any one oxidation state and in their exchange between oxidation states. Note the elements to which we refer are H^+ , Na, K, Mg, Ca, V, Mn, Fe, Co, Ni, Cu, Zn, Mo, Si and Cl, and not bound H, C, N, O, S and P. The second organic group in cells are energised but not in energised equilibrium with their environment or internally. The equilibria outside organisms in the environment are simple as binding generally is weak and

it can be linked to well understood complexity constants, redox potentials and solubility products. The binding constant changes in the Periodic Table of inorganic cations to organic molecules follow the well-understood Irving/Williams patterns, Fig. 8. The equilibria inside cells are made more complicated as the exchange of some ions is slow due to stronger binding. However the slow exchange is catalysed by internal exchangers, called carriers or chaperones, so that all bound states can be related to free ion concentrations by calculation that uses the above thermodynamic constants. The compartments, the environment and the internal solutions, that is of the cytoplasm and of the vesicles, can each be followed separately, although the elements in all cell compartments are in different energised concentrations, at different free ion equilibrated concentrations in each and different from the environment, and often in different oxidation states. Overall each cell compartment has a local homeostatic concentration of free ions in an observable calculable overall steady state. The steady states changed with external evolution, oxidation externally. These changes are reflected in the novel organic chemicals and inorganic matrices produced during evolution. The extent of their oxidation also follows the rise of oxygen and changes in metal ion concentrations.

The sure knowledge that the rates of the evolution of the states of the elements in the environment are simply dependent on the rate of

increase of oxygen and are based on simple equilibrium thermodynamics means that we can go on to consider the timetable of their changes in the light of the rise of oxygen itself and of the contents of sediments. It is largely oxygen's rate of formation from water and environmental buffering which decided this timetable, since organisms had only a minor part in oxygen removal. A further factor of considerable importance and limiting the oxidation potential changes of elements in the sea was poor mixing of the oxygenated upper regions and the reducing lower levels (Canfield, 1998; Canfield et al., 2007). For this reason, although free oxygen in the atmosphere rose rapidly in two periods around 2.5 and 0.75 Ga, the sea is not at a uniform potential until after 0.75 Ga. We have therefore represented the change of potential as a slowly changing average during this period, Fig. 6. which gave a slowly changing metal ion content, Fig. 7. Turning to the effect of these changes in the evolution of organisms our premise is that they had to adapt and so their chemistry evolved firstly as a protective measure, removal of O_2 and its half-reduced forms of O_2^- and H_2O_2 , and then by utilising the new environment largely in new oxidative organic chemistry in novel spaces. In both the old and new equilibria which arise both in the cytoplasm and outside cells, no matter which oxidation state is concerned, the binding of a metal ion to organic sites is, as stated, just that of its equilibria in model complex ion chemistry, for example in the Irving/Williams series, Fig. 8

(Fraústo da Silva and Williams, 2001). The apparent limiting factor to the amounts of metal ions, free and bound, in cells in all this description of compartmentalised thermodynamic equilibrium systems, is firstly the kinetics of energised pumping of ions into or out of a compartment and secondly it appears to be limited by the production of proteins which bind the ions' selectivity. In fact the individual element pumps to and from the cytoplasm are controlled by feedback stopping of entry by the inside free ion concentration binding. Again amongst the proteins produced there are transcription factors which bind the selected ion, so that when the free concentration reaches a certain level it prevents further m-RNA and protein formation. As all the selected proteins, that is enzymes, outward pumps and these transcription factors have very similar binding constants of a selected ion they equilibrate with the same free concentration. The whole cytoplasm comes to a steady state related to the series of free ion concentrations and their proteins, binding constants. This network can only operate if the metal ion binding constants are selected relative to one another and therefore the free ion content of the cytoplasm has to follow the inverse of the Irving/Williams series found in model complexes, Fig. 8, as seen in Fig. 9. Vesicles can be very different in that their energised membrane pumps can operate selectively and so include only ions of given kinds to give a limited range of reactions. There are high concentrations of free Ca^{2+} and Zn^{2+} in very discrete vesicles. It must be

clear that each compartment is energised differently but its energised content of metal ions and proteins is restricted by equilibria, again dependent on pumps but now on protein as well as ion pumping. Note that all the equilibria and steady states are maintained by flow, controlled diffusion. We can then add to the ways of following evolution by biological form and activity, Darwinian style (Darwin, 1859), and by changes of coded molecules, genetics (Cavalier-Smith et al., 2006), the understandable changes in basic chemistry, inorganic and organic, created in the cells from the changes of the environment. The remarkable fact appears to be that there is only one limiting set of free ion values in the cytoplasm, Table 1, in all of the four classes of aerobic cells suggesting that this particular chemical condition is a unique solution to the problem of cell cytoplasmic activity. We shall therefore need to relook at the possible ways in which the environmental changes can alter not just biological metabolism but the DNA code, since the code must represent the whole protein activity and must reflect the environment changes since it controls reproduction. In this way it is in principle possible to connect a system of inorganic thermodynamic constants and catalysed rate constants of organic reactions to a flow network of elements and compounds in both the environment and the cell compartments into one holistic steady state. It is important to realise that the major changes of the chemistry of evolution were completed by 0.4 Ga. Thereafter

evolution is largely diversity, not novelty, of chemistry (Williams and Rickaby, 2011). Now apart from their functions in catalysis the inorganic elements are active in controls other than by binding to transcription factors and maybe these controls too are common to all aerobic cells.

The increase in complexity of cells from prokaryotes to unicellular, then multicellular, organisms required more and more management. Staying with our wish to refer to metal ions, Fe^{2+} and Mg^{2+} remained as major controls of metabolic pathways in the cytoplasm in all cells, of both anaerobic and aerobic organisms. Control systems responsive to external changes became necessary for the vulnerable eukaryote cells. They evolved, using the available metal ion gradients, especially of calcium, (Carafoli and Klee, 2008), and later of the pair of ions Na^+/K^+ (Grant, 2010), for external/internal cell messages. The controls using all three ions arose from the need of the earliest prokaryotes for protection which created gradients of these ions with high concentration of Ca^{2+} and Na^+ in the sea and later in extracellular fluids to low values in the cytoplasm. The flows of the two govern many cellular mechanical and chemical metabolic switches. The essential role of calcium arose as a necessary connection of unicellular eukaryotes to communication between the environment and compartments, including the cytoplasm. The role of Na^+ and K^+ arose from the fact that they do not bind significantly and are then excellent electrolytic current carriers. There are

also some functions of other ion gradients such as those of zinc. The messenger system was extended to organic transmitters and hormones released from vesicle or the cytoplasm of cells of multicellular eukaryotes. Most transmitters, in vesicles, are again produced by copper enzymes in vesicles while iron oxidised the basic frameworks of many messenger organic molecules in the cytoplasm, e.g. sterols. In turn the sterols controlled the steady state of the minerals in the extracellular fluids, Na^+ , K^+ and Ca^{2+} . As mentioned the transcription factors for many of these hormones are zinc fingers. These chemical messenger flow systems were devised, and the nervous system and the chemistry of the brain, by 0.5 to 0.4 Ga. Once again we insist that the functioning of all the metal ions in a compartment are based on equilibrium binding although it is only Lewis acid strength not redox potentials which are important for Na^+ , K^+ , Ca^{2+} and Zn^{2+} . Messenger systems give rise only to fluctuations in steady states. While the use of calcium gradients arose with the single-cell eukaryotes, the use of Na^+/K^+ gradients, the last chemical innovation, arose soon after the evolution of multi-cell eukaryotes.

We turn finally to the earliest need of further protection from dangerous chemicals such as O^-_2 and peroxides and here increasingly Cu/Zn enzymes played a new role in eukaryotes together with that of novel selenium enzymes. Once again we observe the total dependence of

the evolution of life on the changes in the environment due to organisms' activities. It is the increased availability of the three elements, Cu, Zn and Se, which increased the protection in eukaryotes with an increased presence of oxygen.

Now complexity places great stress on the ability to organise. A remarkable feature of evolution in the broadest sense is not survival of the fittest, competition between individual organisms, but on survival of the whole, cooperative, symbiotic system. The development of unicellular eukaryotes is a major, little understood, *symbiotic* evolution (Margulis, 1999). Utilising symbiosis is a new use of space. It was greatly increased in multicellular eukaryotes, plants and animals, as they come to rely on chemicals from lower organisms living independently, attached outside or even inside these eukaryotes. Plants depend on bacteria for nitrogen and on fungi for minerals. Plants and animals depend on vitamins, including the many essential coenzymes, and recent animals require additionally essential amino acids and sugars from many sources. They are obtained by feeding. The ensemble of organisms became a large cooperative network in an environment/organism system in which the definition of species becomes difficult. However speciation is not important in main line chemical evolution. Man is an extreme example. We do not know how many organisms (species) man's existence depends upon. The total

complexity involves and is relieved by ever increasing use of biological plus environmental space.

We must now turn to the discussion of the connection to genetic information so essential for the maintenance of the whole system of supposedly selfish organisms and their reproduction. As the system depends on the environment and organisms' reproduction depends on the genetic code, DNA, the code had to change as the environment changed. How is that possible? It points not to random evolution (sometimes called Darwinism (Darwin, 1859) but to some kind of directed main-line evolution (sometimes linked to Lamarck). What do we know about DNA evolution from 3.5 Ga? We shall turn again to metallo-proteins and enzymes in the next section for they provide an easily followable link between the development of chemistry in the environment and the DNA codes of organisms. We note that the majority, if not all, these chemical changes were complete by 0.5 to 0.4 Ga, close to the Cambrian Explosion of life and after the final rise of oxygen when the oxidation potential of the sea was virtually finalised. From 0.4 Ga to today evolution is not so much about chemical changes but about variety of very similar chemistry in vast numbers of organisms permitted by complexity, based on those extant at 0.4 Ga. It is far easier to use genetics and biological comparisons after, say, 0.75 Ga than before that (Cavalier-Smith et al.,

2006). We shall not attack the problem of these varieties but concentrate on the major line of chemical evolution: prokaryotes, anaerobic then aerobic eukaryotes, unicellular and multicellular, all aerobic, and their genes before 0.4 Ga (Williams and Rickaby, 2011).

9. The Links between the Environment and Genetic Code Changes

It is generally considered that the changes in the code of DNA are random, arising from mutations. It might then appear that the organisms which arise are a linked random competitive set which meets and is biased by environmental change, no matter if it is accepted that the major line of environmental change is systematic depending on oxygen rise. In our opinion this does not fit the above observations of the chemistry of the whole system, certainly from 3.5 to 0.4 Ga. Notably we cannot doubt that symbiosis has increased at all times and this involves cooperativity not competition. In fact cooperativity can hardly be the same thing as competition and is not random. It often involves genetic code gains in one organ or loss in one partner and gain in another in symbionts to overall advantage without mutation. We consider that cooperativity has increased in some cases with or without gene swapping, and is known in early prokaryote life. The division of chemical activity in different cell types, often chemotypes, in different prokaryotes is obvious in names of groups of species, such as sulfur bacteria, nitrobacteria, photosynthetic both in eubacteria and archaea. They could then be dependent on one another in

symbiosis at a distance but gene exchange is common in these prokaryotes too. Equally clear is symbiosis in the origin of several types, not one, of unicellular eukaryotes some of which need essential chemicals produced by other organisms. Subsequently all kinds of symbiosis have arisen in multicellular organisms. Gene swapping in some and gene re-arrangement in many are common. We have related symbiosis increases particularly to the difficulty of carrying increasing complexity in single cells whenever chemistry in the environment became more complicated and useful to cells but not directly related to the original maintained chemistry of the basic cell cytoplasm. This suggests that we should look carefully for the effect of stress on organisms down to the level of DNA, the major stress being the consequences of the increase in oxygen and of an oxidised environment until 0.5 Ga. These mainly concerned the changes in marker inorganic elements, Fe, Cu, Zn and Mo and the rise of oxidised substrates difficult for the original anaerobic chemistry of cells, O_2 , N_2 , NO^- , CO_2 , SO_4^{2-} and SeO_4^{2-} . Such considerations of stress had to result not just in improvement of an enzyme by mutation but to the evolution of new proteins for novel functions. We note that major chemical changes in organisms as a whole, not separately, in species, arose at very similar times. As a consequence we have turned our attention to *multiplication of genes as the major possible way of evolving novel DNA coding* with novel catalysis while maintaining the nature of

the original code. Moreover we can see if in fact code changes are or are not random with the changes to oxidised substrates as above and using catalysts based on the changed availability of the three metal elements, Fe, Cu and Zn. If they are not random then codes limited to specific related processes of genes. needing, say Cu, will be shown by numerous *different* increases in the copper catalysts for these processes. in a comparison of genes of many types of organisms. Bioinformatics gives us the chance to explore this possibility,

10. Gene Duplication

For a mutation to be successful it requires either (1) an alteration in the code to change amino acids so as to improve an existing protein activity or (2) the addition of a new coded DNA section to provide a new activity. The first case is clearly the case in adjusting the binding constants of carbon dioxide against those of oxygen in the evolution of ribulose –1,5, biphosphate carboxylase/oxygenase (Tortell, 2000; Rickaby et al., 2011; Williams and Rickaby, 2011). It was also deemed to have occurred in the “perfection” of triose phosphate isomerase described by Knowles and Albery (1977). (Note this conclusion of a drive to perfection is based here on equilibrium being established between the forward and back reactions of an enzyme. Generally “perfection” is not a requirement in a step in metabolism not at equilibrium but an optimal rate in a network is desirable.) The second evolution, of a new activity, can

only occur through creation or capture of new segments of DNA. Capture can arise from input using carriers, viruses or symbionts. Creation can occur most strongly however through gene duplication, followed by mutation of one of the duplicates (Ohno, 1970; Koonin, 2009). It is especially favourable if a new substrate needs to be metabolised while a very similar metabolism of a substrate already present in the cell is maintained. The new substrate could even be an inhibitor of the original enzyme. An easy way to follow the expansion of the DNA by duplication is to follow the appearance of repeated DNA sequences for a given class of enzyme or function. It is here that bioinformatics of metalloproteins is a particularly useful tool (Zerkle et al., 2005). This is illustrated for Ca^{2+} proteins, Table 2 (Morgan et al., 2005), and as we have illustrated in some detail in two short papers on Zn, Cu and two classes of iron proteins, those of iron porphyrin and ferrous iron, 2-oxyglutamate oxidases, 2-OG, Fe^+ , Table 2 (Decaria, Bertini and Williams (a) 2010; (b) 2011). Any bioinformatic method may be used but we have used that of Bertini et al., 2009, assisted by that of Dupont et al., 2006 and 2010, and of Zhang and Gladyshev (2009). Some results are shown, dependent on metal ion proteins linked to oxygen increase in the environment, Table 2, when there are considerable duplications, increasing in evolution. The table is of particular organisms but DNA sequences show that their duplications in size are general to large groups of related organisms. The

data are for today's organisms but arose in each class of organism as it evolved. Immediately we see that multiplication of the metal proteins in the metallome is very considerably different from one protein to another and although they differ in number from one group of organisms to another, considerable multiplication of particular enzymes is common to all within unicell and multicell eukaryote groups. It is highly selective, whilst we know that some other proteins are hardly duplicated at all, see above. What are the implications? The clearest is that evolution is very much dependent on localised, not whole genome duplication and is not random, no matter that subsequent, probably random, mutations in one at least of the duplicates is required to change activity. (It is unlikely that much protection is initially present for these duplicates as in part they may start as junk DNA or pseudo-genes. Many are lost easily but clearly many remain.) The second is that there is common (but different in size) multiplication in different classes of organisms, plants, animals and fungi. Plants and animals have many duplications of oxidases (Cu, Fe), though plants have more and animals have the greater number of protein hydrolases (Zn), calcium signalling proteins, and zinc finger transcription factors (Zn). The implication is that these very different organisms have either much greater need for protection from environmental chemicals and cross-linking (plants) or for managing greater complexity (animals). Finally many large duplications in the enzymes are strikingly connected

to oxygen, related to metabolism, and to zinc and copper which correlates with the increase of all three in the environment, and all of which as they increased were initially very poisonous, and damaged DNA. Moreover the changes are not evenly spread in time but appear to occur in many organisms close to the time of rapid environmental change of oxygen and metal ion availability. These are times of very considerable stresses. They also occur with increase in the division of space in unicellular, then multicellular eukaryotes and with symbiosis. The changes in DNA are strongly indicative of an indirect environmental influence.

Provided further inspection of the DNA data bases confirms these results we conclude that environmental changes are reflected in particular gene duplication, to which we add one other supporting observation. It is observed that gene loss is rationally associated (1) with a loss of particular enzymes, for example of those proteins containing nickel and cobalt (vitamin B₁₂) which are lost in higher organisms as their substrates are very reduced molecules related to H₂ and CH₄. These substrates became virtually absent in the atmosphere, and (2) with coalescence of organisms, symbiosis. The question which follows is that if DNA changes, gains or losses, are so selectively dependent on environmental changes, how could this happen? We believe that gene duplication is connected to the stress imposed initially on the proteome which is

directly dependent on changes of the environment. Increased selective protein production is related to this stress as seen in the effect of oxygen, of oxidation of the environmental and in the immune response. These stresses at first damage proteins which must be made in new copies. The increase in selective protein production could cause DNA misfitting at specific points, namely at those at which the required m-RNA for this protein production occurs. Gene duplication could be a consequence of this misfitting. This statement is open to investigation but it leads to a concluding sentence. Major evolution of classes of organisms, seen in protein (DNA) changes is an inevitable *chemical* consequence of the cellular coupling to the changes of the environment. Major classes of organisms are subsequently increasingly decorated as seen in the random appearance of species. Expansion of activity in organisms increases complexity and, as seen in the random appearance of species, greater complexity allows greater diversity in increase in sensitivity to the environment. Complexity is reduced by symbiosis and gene loss but aided by directed gene duplication.

11. The Integrated Approach

The above section has shown in principle that a connection can be made from the environmental chemical changes to the details of steady states of organisms through metal ion, substrates and proteins to the genetic code. This has to be the case for the systematic main line of the

chemistry of the continuously increasing energisation of reduction in the cytoplasmic interior by increasing numbers of cells resulting in continuously increased, energised oxidation of the environment and of the outside of the cytoplasm of cells and in vesicles. The steady states, as explained above, apply since each compartment is in overall equilibrium and in contact with all others though they are differently energised. Their changes are a rational chemical progression and is the resultant of oxidation and reduction chemistries which forced compartmentalisation of cells in order to adapt to their increasing oxidised environment while creating more effective conversion of light to heat overall. The proof of the steady state applicable to all cells' vesicles lies in their different ion concentration and the observed universal concentration of ions in the cytoplasm, Table 1 and Fig. 8. We are now used to the idea of proteins as mini-machines maintaining the steady state by pumping and energy capture so we have to find the way to describe cells as micro-factories of many interlocking machines linked by the flow of small molecules and metal ions and so to the environment. The approach does not require us to ask about the steps of creation of life but only of the changing state of chemical flow from new input to new product once life began. That is new machinery added to old machinery. There are limits as to how the machinery of cells has changed over time with incorporation of extra compounds which can operate only if they network with the maintenance

of the earlier machines in cells. The way forward to understanding may well come from theory taken from engineering sciences where reductive approaches are replaced by whole-system thinking and control theory and industrial activities are coordinated in space. We may even have to go first to the “simplest” systems, such as the weather, to see how well-defined systems between cause and effect with feedback are created.

Our contention is that full use of bioinformatics will allow and assist the analysis of the conventional molecular biology of the code with that of the expression of proteins, the proteome and the metallome. From the understanding of the link of the resultant observed change of steady states of flow with the change of the environment, I suggest that a beginning can be made by an analysis of metallo-proteins which provide the one easy link from DNA to the environment. The metallo-proteins and their changes connect in resolvable ways to metabolism and controls and to the code. These data will allow us to see evolution of the system as in a flow diagram which is complicated by symbiosis and is not “tree-like”.

New Environment Metal Ions → Binding to existing proteins →
 additional activities of DNA to replace these proteins →
 stress on DNA reproduction → duplication of stressed region →
 new proteins using the new metal ions →
 new metabolic products in new reproductive organisms → symbiosis.

All organisms at a given time will suffer from environmental change and would be expected to make similar, not the same, adjustments as the experience of the inorganic environment is somewhat different for different organisms. We observe in particular the consequences in the evolution caused by oxygen and the oxidised environment, especially by oxidation of metal compounds, as a progression of steps. Examination of the multiplication of sections of the code and activity of particular new metabolic paths linked to environmental change will allow us then to treat advances not as due to random mutations but to the chemical change of the environment which “caused” the particular duplication which gave rise to a general new activity. The evolutionary timetable can be related to the existing environment, the sea, in any particular period through the geochemistry of sediments. The code and its changes are then a necessary component in the activities. There is the possibility that such a study will reveal the mechanistic steps by which a new process is introduced and coded and then even the origin of the system. While we have noted the major effect of new inorganic chemicals in causing stress on cells we believe this phase of evolution was virtually complete by 0.4 Ga. We do not think that physical stress can be disregarded since it may play an important part, for example in evolution around 0.3 Ga when there were extreme periods of volcanic activity and of extinction. Especially after the Cambrian Explosion in 0.54 Ga evolution has been largely of extension

and variety and follows the conventional Darwinian random process of selection. It is hoped that this admittedly speculative account of evolution of the organisms/environment system will lead to further inquiry into the mystery of life.

Acknowledgment Much of this article is based on a book by the author and Dr Rosalind Rickaby to be published shortly. I thank Dr C. Dupont for bioinformatic information and an exchange of views.

Accepted manuscript

Fig. 1 The fundamental physical chemical system which we have to analyse. The entropy gain lies in the conversion of few large high-energy (light) quanta to many low-energy (heat) quanta. All living processes are cyclical and so are many, not all, environment steps driven by light energy. The two are intimately mixed in spatial compartments, the topic of this article. Here we connected the changes in the environment by energy from cells, e.g. oxygen release, which acts to supply elements to cells.

Fig. 2 A very simplified diagram of the scheme of the transduction of light energy via electron and proton flow through two parts of space separate by an impenetrable region as described first in Williams, 1961. The fact that the impenetrable region between the two is a membrane is not essential. The initial charge separation is indicated by the electron flow.

Fig. 3(a) An outline diagram of either an ATP-ase pump as shown or in reverse, of a proton energised synthesis of ATP, see Fig.

2. The stress here is on the construction of the dynamics using energy controlled α -helical channels for flow through the membrane and hinged binding zone of β -sheets, Fig. 3(b) to move substrates. The easiest hinge to picture is of two α -helices. X^1 is the site of applied energy to create an active site by coupled motion.

Fig. 4. A very simple diagram of the relationship between energy and materials from the environment and metabolism in cells linked to environmental waste (oxygen) to change the environment and by feedback change organisms. Apart from the cycling of material the conversion of light to heat (increased entropy) drives the system.

Fig.5 The outline scheme of metabolism modified from the work of Kauffman, 1993.

Fig. 6 The rise of oxygen, and therefore of oxidation, and the increase in the complexity of some cells. The increase seen as increasing numbers of compartments is not at the expense of simple cells which act with complex cells in a symbiotic system. The rise in oxygen was buffered by Fe^{2+} and H_2S between 2.5 and 0.75 Ga, and again from about 0.5 Ga but uncertainly by oxygen and reduced carbon.

- Fig. 7 The calculated continuous increase in copper and zinc and the fall of iron in the environment due to the rise of oxygen, (Fraústo da Silva and Williams, 2001; Saito *et al.*, 2003).
The rise of oxygen in the atmosphere may have been in two sharp steps at 2.5 and 0.75 Ga but this would not be reflected in the sea as mixing in this period was poor. Any outline of changes is very approximate and dependent on locality.
- Fig. 8 The Irving/Williams series of stability constants. The solubility product of sulfides follows the top line closely while that of oxides follows the bottom line. We take the switch to indicate the changes in the state of the sea from the sulfide period around 3.5 Ga to the oxide period of today (Fraústo da Silva and Williams, 2001; Saito *et al.*, 2003).
- Fig. 9 The concentrations of free ions in cells in part from experimental data and in part by calculation from known cell inorganic and organic contents and stability constants in model systems, (Fraústo da Silva and Williams, 2001).
- Fig. 10 The close relationship between geological observations, the appearances of different classes of organisms, the rise of oxygen in the atmosphere and the availability of elements in

the atmosphere or the sea. The detailed evidence will be given in a forthcoming book (Williams and Rickaby, 2011).

Accepted manuscript

Alexiev, U., Mollaaghababa, R., Scherrer, P., Khorana, H.G., Heyn, M.P., 1998. Rapid long-range proton diffusion along the surface of the purple membrane and delayed proton transfer into the bulk. *Proceedings of the National Academy of Sciences* 92, 372-376.

Anbar, A.D., Knoll, A.H., 2002. Proterozoic ocean chemistry and evolution. A bioinorganic bridge. *Science* 297, 1137-1142.

Andreini, C., Bertini, I., Rosato, A., 2009. A bioinformatic approach. Metalloproteins. *Accounts of Chemical Research* 42, 1471-1479.

Barrow, J.D., Freeland, S.J., Harper, C.L., Morris, S.C. (eds.), 2006. *Fitness of the Cosmos for Life: Biochemistry and Fine-Tuning*. Cambridge University Press, Cambridge.

Canfield, D.E., 1998. A new model for proterozoic ocean chemistry. *Nature* 396, 450-453.

Canfield, D.E., Poulton, S.W., Narbonne, G.M., 2007. Late neoproterozoic deep ocean oxygenation and the rise of animal life. *Science* 315, 92-95.

- Carafoli, E., Klee, C. (eds), 2008. Calcium as a Cell Regulator. Oxford University Press, New York.
- Cavalier-Smith, T., Brasier, M., Empey, T.M. (eds.), 2006. Major steps in cell evolution. Philosophical Transactions of the Royal Society B361, 848-1083.
- Culpin, D., Day, P., Edwards, P.R., Williams, R.J.P., 1965. Charge transfer in solid chlorocuprates. Chemical Communications No. 19.
- D'Arcy Thompson, W., 1938. On Growth and Form, second ed. Oxford University Press, Oxford.
- Darwin, C., 1859. On the Origin of Species by Natural Selection. J. Murray, London.
- Decaria, L., Bertini, I., Williams, R.J.P., 2010. Zinc proteomes, phylogenetics and evolution. Metallomics (to be completed)
- Decaria, L., Bertini, I., Williams, R.J.P., 2011. Copper proteomes, phylogenetics and evolution. Metallomics in press.
- Deisenhofer, J., Epp, O., Miki, K., Huber, R., Michel, H., 1985. The structure of the photosynthetic reaction centre. Nature 318, 618-624.

- Dobson, C.M., 2003. Protein folding and misfolding. *Nature* 426, 884-890.
- Dupont, C.L., Yang, S., Palenisk, B., Bourne, P.E., 2006. Modern proteomes contain imprints of ancient shifts in ocean chemistry. *Proceedings of the National Academy of Sciences*, 103, 17822-17827.
- Edwards, P., Gray, H.B., Lodge, M., Williams, R.J.P., 2008. Electron transfer through an intervening medium. *Angewandte Chemie International Edition* 36, 6758-6765.
- Fraústo da Silva, J.J.P., Williams, R.J.P., 2001. *The Biological Chemistry of the Elements*. Oxford University Press, Oxford.
- Grant, S.G.N., 2010. Computing behaviour in complex synapses. *The Biochemist* 32, April.
- Gray, H.B., Malmström, B, Williams, R.J.P., 2000. Copper coordination in blue proteins. *Journal of Biological Inorganic Chemistry* 5, 551-560 and references therein.
- Gray, H.B., Winkler, J.R., 1996. Electron transfer in proteins. *Annual Review of Biochemistry* 65, 537-561.

Kauffman, S., 1993. *The Origins of Order*, Oxford University Press, New York.

Klein, C., Beukes, N.J., 1993. Sedimentology and geochemistry of the glaciogenic late proterozoic Rapitan iron formation in Canada. *Economic Geology* 88, 542-565.

Knowles, J.R., Albery, W.J., 1977. The perfect enzyme. *Accounts of Chemical Research* 10, 105-121.

Koonin, E.V., 2009. Darwinian evolution in the light of genetics. *Nucleic Acid Research* 37, 1011-1034 and references therein.

Lehmann, B., Nagler, T.F., Holland, H.D., Willie, M., Mao, J., Pan, J., Ma, D., Dulski, P., 2010. Highly metalliferous carbonaceous shale and Early Cambrian seawater. *Geology* 35, 403-406.

Lesk, A.M., 1996. *Protein Architecture*. Oxford University Press, Oxford.

Lloyd, D., Chance, B., 1968. Electron transport in mitochondria. *The Biochemical Journal* 107, 829-837.

Margulis, L., 1999. *Symbiotic Planet*. Basic Books, New York.

- Mitchell, P., 1961 Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic mechanism. *Nature* 191, 144-146.
- Morgan, R., Martin-Almedina, S., Gonzalez, M., Fernandez, M.P., 2002. Evolutionary perspectives on annexin calcium-binding proteins. *Biochimica Biophysica Acta* 1742, 133-140.
- Moser, C., Keske, J.H., Warneke, K., Farid, R.S., Dutton, P.L., 1992. Nature of biological electron transfer. *Nature*, 355, 796-802.
- Ohno, S., 1970. *Evolution by Gene Duplication*, Springer Verlag, Berlin.
- Rickaby, R.E.M, 2011. to be published.
- Robin, M.B., Day, P., 1967. Mixed valence chemistry. *Advances in Inorganic Chemistry and Radiochemistry* 10, 247-422.
- Saito, M.A., Sigman, D.M., Morel, F.M.M., 2003. The bioinorganic chemistry of the ancient ocean: the evolution of cyanobacterial metal requirements and biochemical cycles at the archaean proterozoic boundary. *Inorganic Chimica Acta* 356, 308-318.
- Siebert, C., Kramer, J.D., Meisel, T.R., Morel, P.H., Nagler, Th.F., 2005. PGE, Re-Os and Mo isotope systematics in archaean

sedimentary systems as proxies for redox conditions of the early earth. *Geochimique et Cosmochimique Acta* 69, 1787-1801.

Tortell, P.D., 2000. Evolutionary and ecological perspectives in carbon requisition in phytoplankton. *Limnol Oceanography* 45, 744-760.

Turing, A., 1952. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society B237*, 37-72.

Vine, J.D., Touzelot, E.B., 1970. Geochemistry of black shale deposits – a summary report. *Economic Geology* 63, 253-272.

Walker, J.E., 1998. ATP-synthesis by rotary catalyst. *Angewandte Chemie, International Edition* 37, 2309-2319.

Webber, B.H., Prebble, J.N., 2006. An issue of originality and priority. *Journal of the History of Biology* 39, 125-163.

Williams, R.J.P., 1961. The functions of chains of catalysts. *Journal of Theoretical Biology* 1, 1-13.

Williams, R.J.P., 1977. Flexible drug molecules and dynamic receptors. *Angewandte Chemie, International Edition* 11, 805-816.

Williams, R.J.P., 2005. NMR studies of mobility within protein structure.

European Journal of Biochemistry 183, 479-497.

Williams, R.J.P., Fraústo da Silva, J.J.R., 2006. The Chemistry of

Evolution. Elsevier. Chichester, UK.

Williams, R.J.P., Rickaby, R.E.M., 2011. Evolution's Destiny: the

Integration of the Chemistry of the Environment and of

Organisms (to be completed).

Zhang, Y., Gladyshev, V.N., 2009. Comparative genetics of trace

elements. Chemical Reviews 109, 4828-4861.

Accepted manuscript

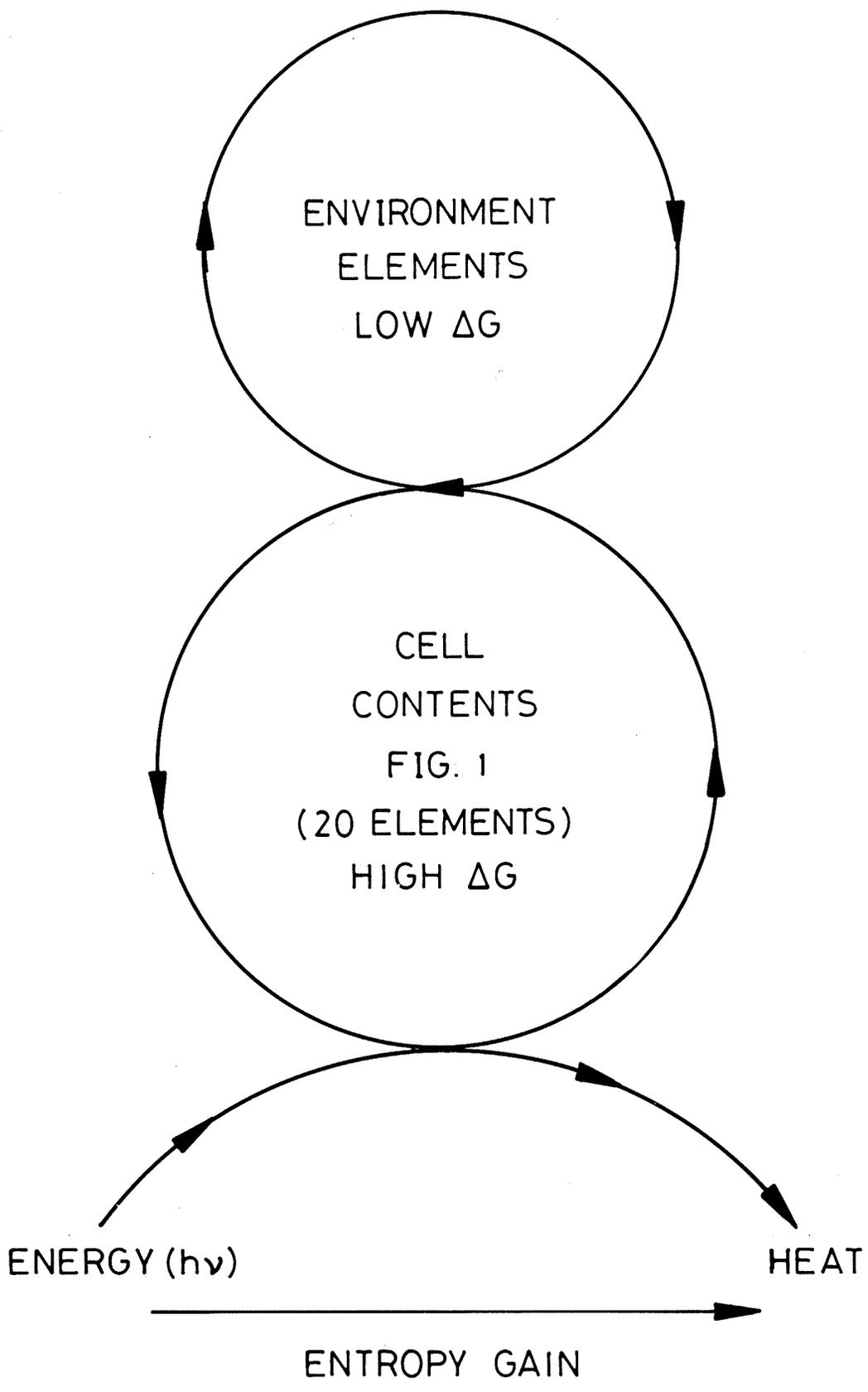
Table 1 Composition of Cytoplasmic Fluid of All Aerobic Cells

| Element | Form | Concentration (M) |
|----------------|--------------------------------|--------------------------|
| H | H ⁺ | 10 ⁻⁷ |
| Na | Na ⁺ | <10 ⁻³ |
| K | K ⁺ | 10 ⁻¹ |
| Cl | Cl ⁻ | <10 ⁻³ |
| Mg | Mg ²⁺ | 10 ⁻³ |
| Ca | Ca ²⁺ | <10 ⁻⁶ |
| Mn | Mn ²⁺ | ~10 ⁻⁶ |
| Fe | Fe ²⁺ | ~10 ⁻⁷ |
| Co | Co ²⁺ | <10 ⁻⁹ |
| Ni | Ni ²⁺ | <10 ⁻¹⁰ |
| Cu | Cu ²⁺ | <10 ⁻¹⁵ |
| Zn | Zn ²⁺ | <10 ⁻¹¹ |
| Mo | MoO ₄ ²⁻ | ~10 ⁻⁷ |

Table 2 Local DNA (Gene) Duplication

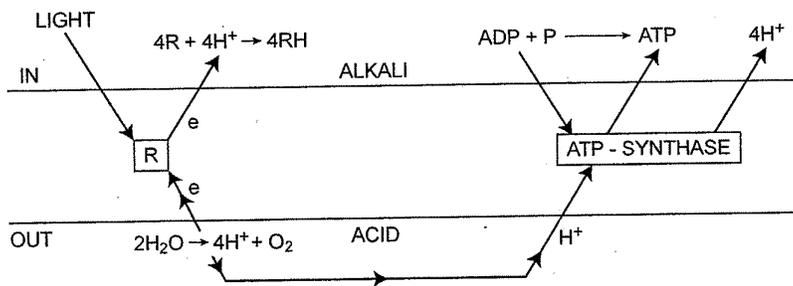
| Organism | | Numbers of very Similar Domains | | | | |
|--------------|------------|---------------------------------|-------------|-------------|------------|------------|
| | Peroxidase | P450 | Cu oxidases | Fe oxidases | Zn fingers | Calmodulin |
| | | Heme | EC:1 | 20G-Fe(II) | | |
| Eubacteria | 1 | 1 | 3 | 1? | 5 | 0 |
| Yeast | 1 | 3 | 9 | 1 | 373 | 4 |
| Arabidopsis | 95 | 268 | 111 | 116 | 1631 | 80 |
| Homo Sapiens | 0 | 70 | 21 | 9 | 2984 | 160 |

Note average proteome size for the organisms from top ~3,500, 6,000, and then > 35,000 for the other two organisms.

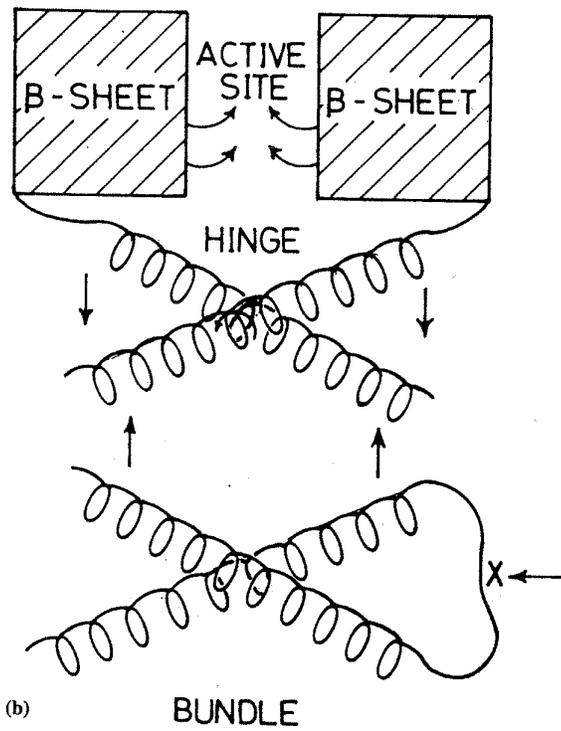
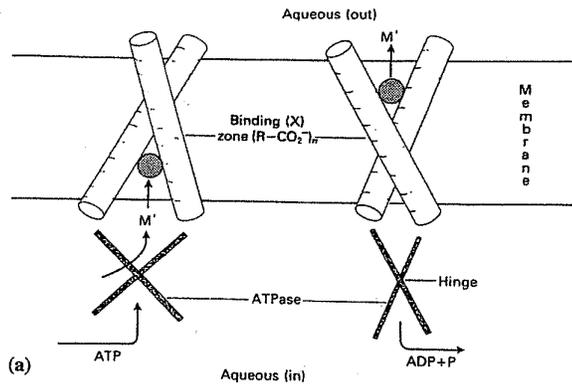


R ^{semper 1} metal in cortex

Fig. 4.2 (2)



4.10 (12)
(3)



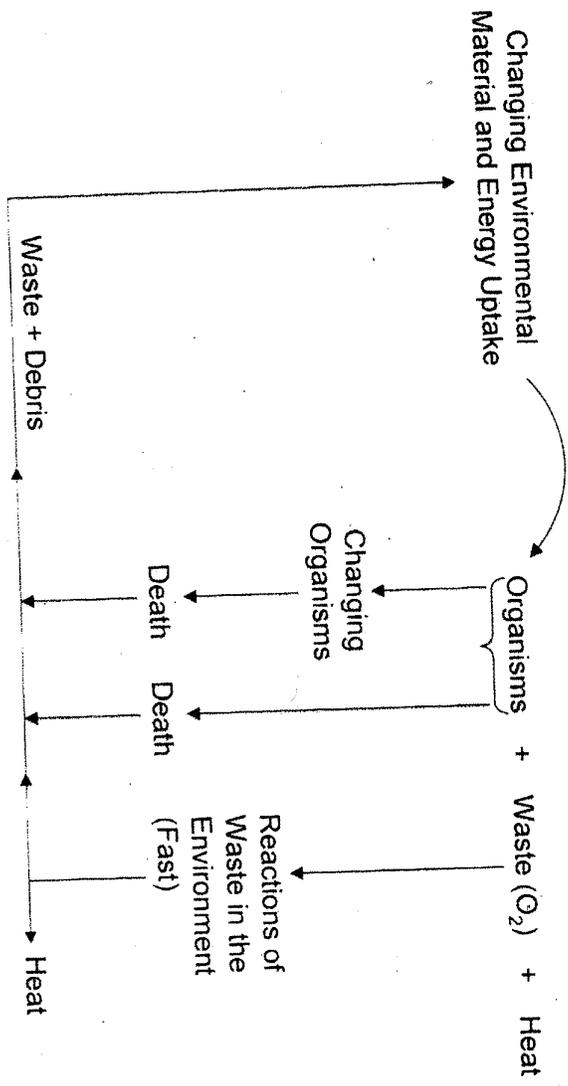


Fig 4.2

10?

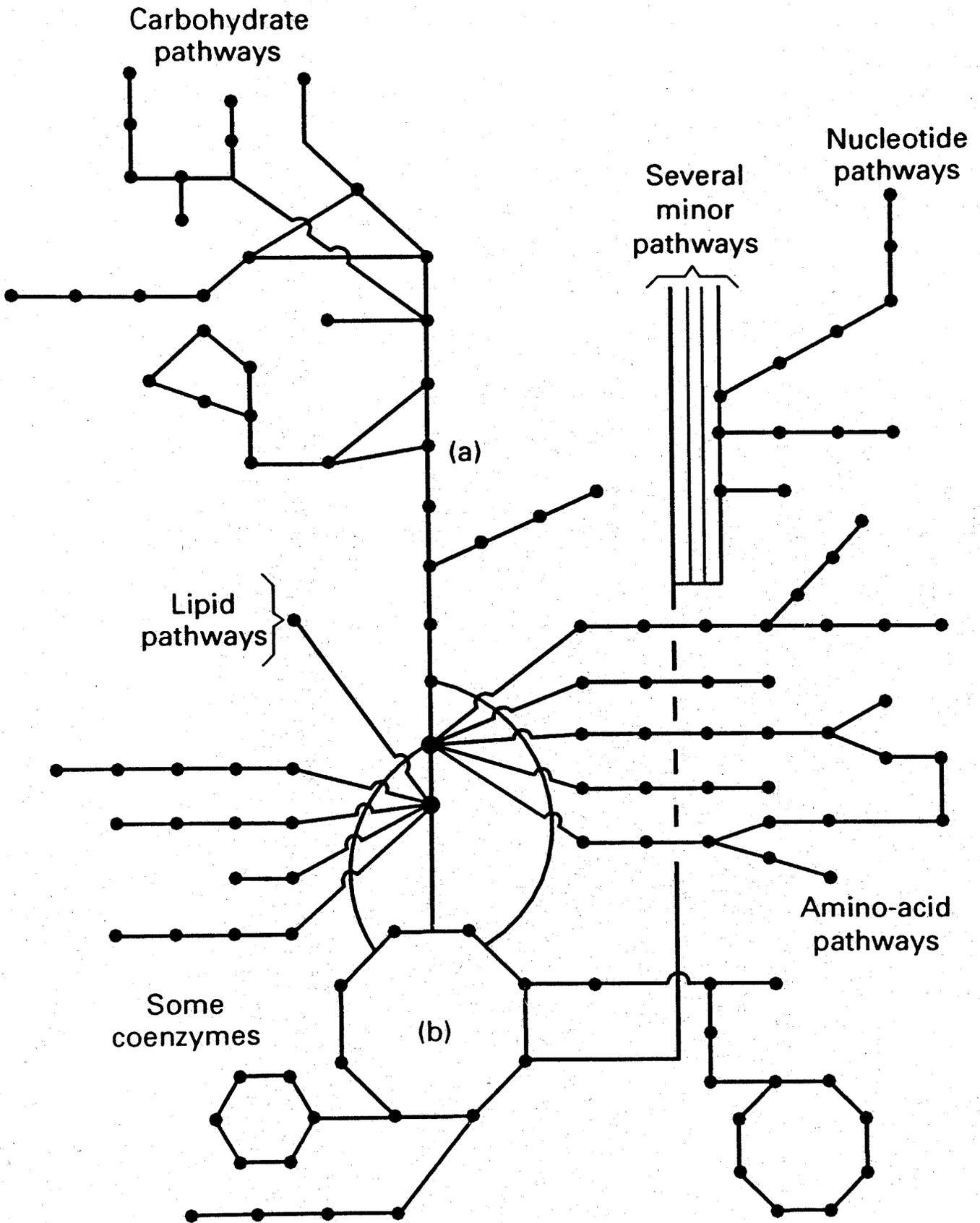
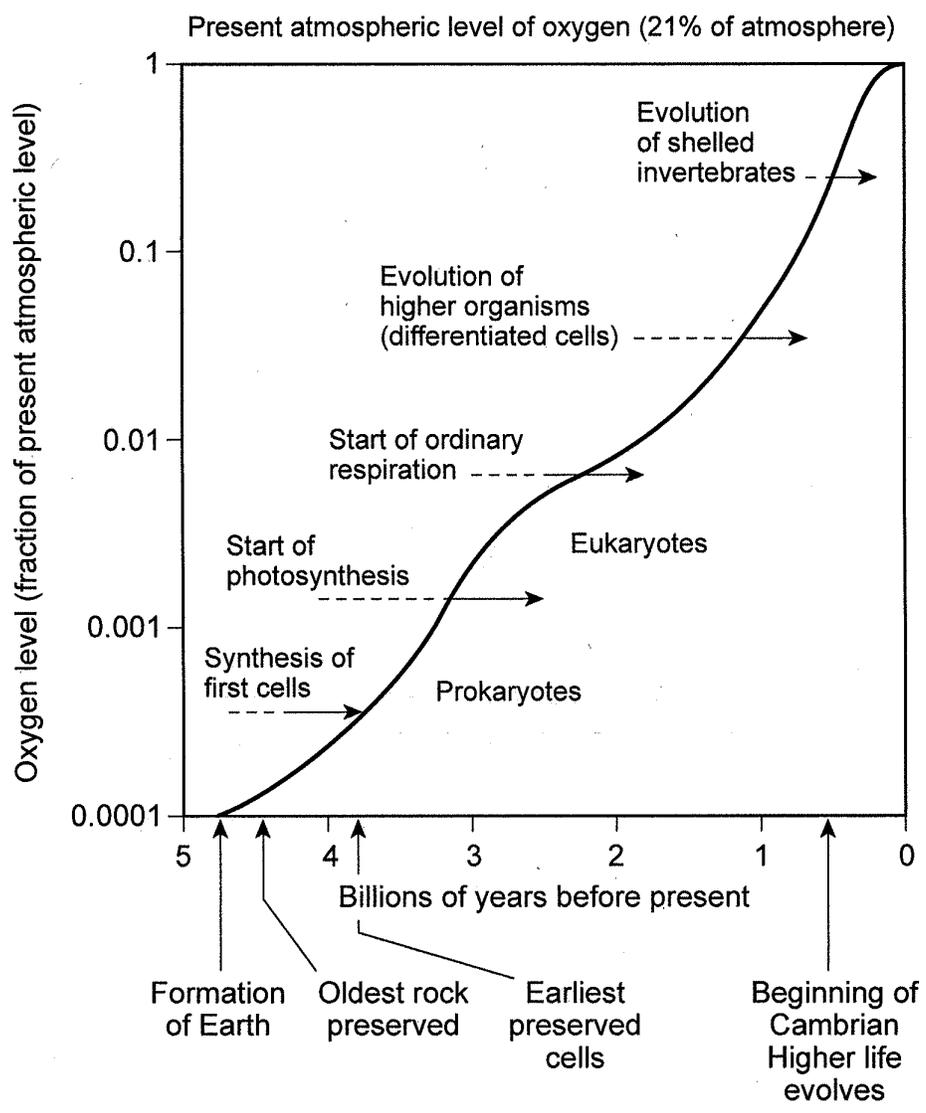
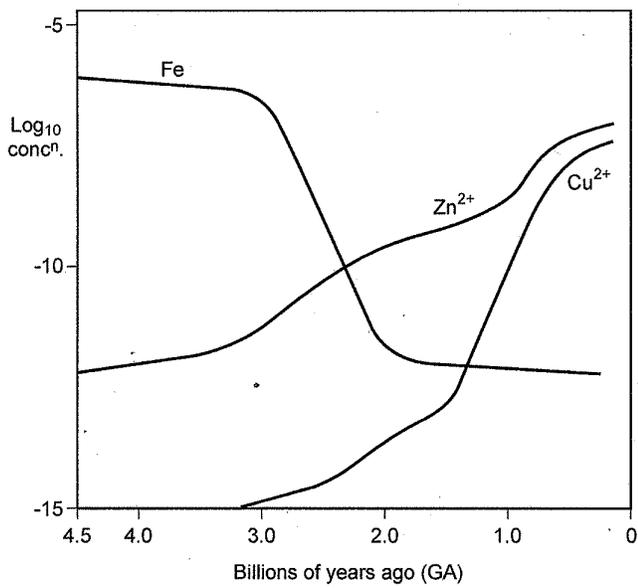


Fig. 15.6

NO need to learn



18



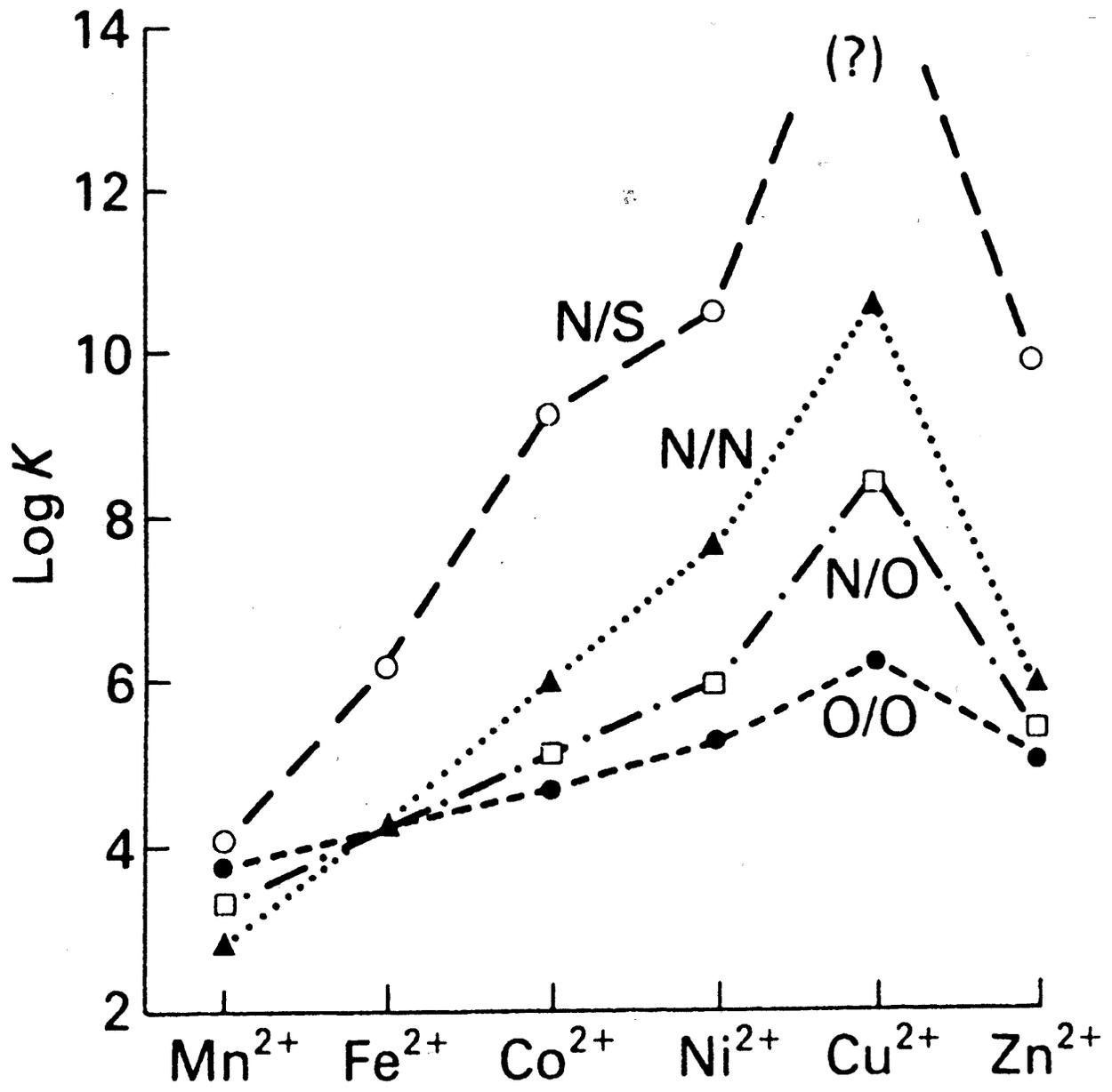
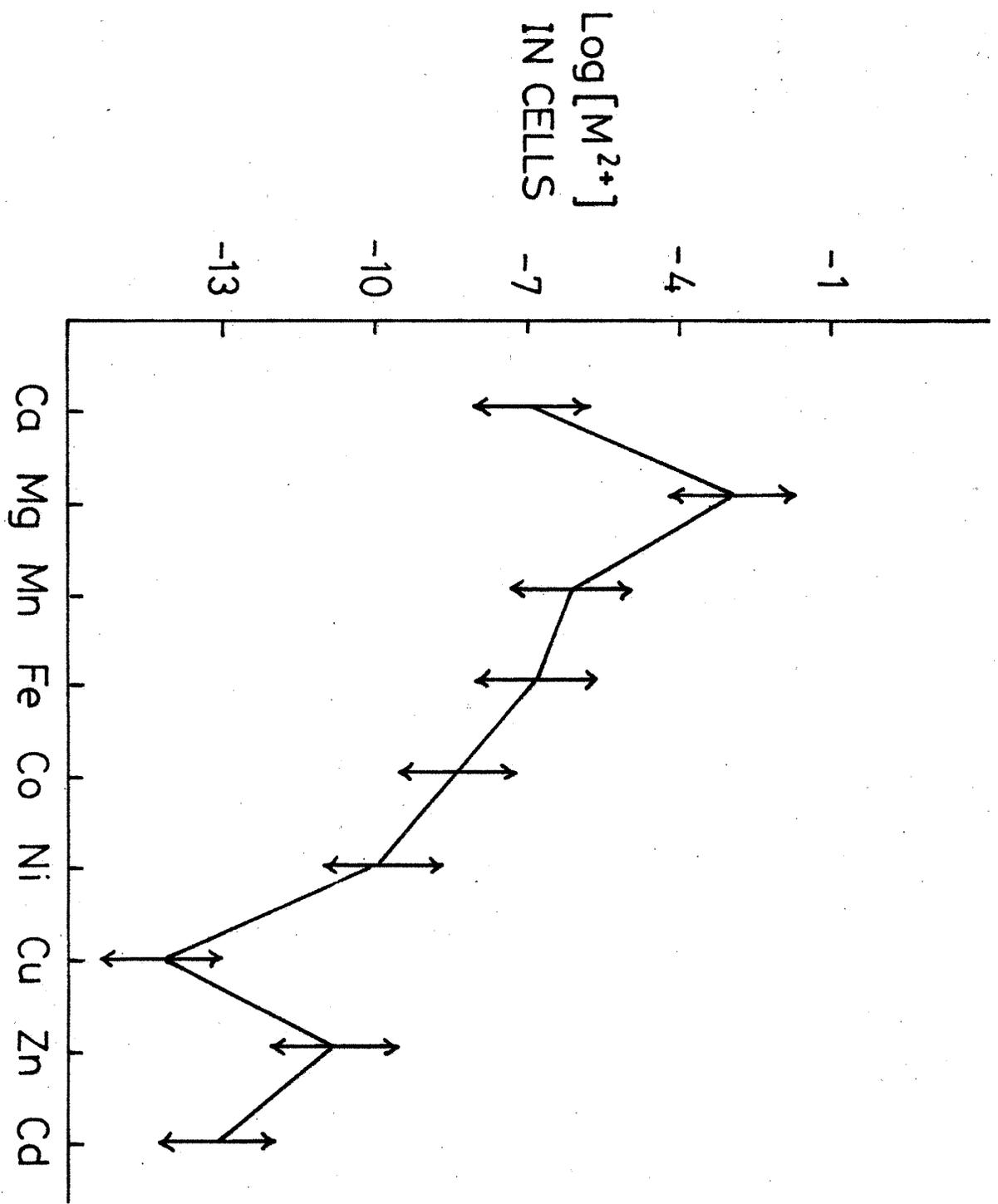


Fig. 109



| Time (MYr ago) | Biological evidence | Features | Glucose use | Geological evidence | Oxygen/ PAL* | Loss of element | Gain of element | Little change of element |
|----------------|--|--|--------------|--------------------------------|--------------|---|---|--|
| 400 | Large fishes, first land plants | | Respiration | Shells and bones | 100 | Fe ²⁺ , S ²⁻ , Se ²⁻ , H ₂ S ²⁻ , MoS ₄ , NH ₃ , CO ₂ | Cu ²⁺ , Zn ²⁺ , Cd ²⁺ , (Fe ³⁺), MoO ₄ ²⁻ , N/O, SO ₄ ²⁻ , SeO ₄ ²⁻ , I ₂ | |
| 550 | Cambrian fauna | Shelly metazoans, absorption through external shell | | | 10 | | | |
| 670 | Ediacarian fauna | Metazoans, collagen | | Red beds [Fe(III), Fe(II/III)] | 7 | | | Mn ²⁺ , Cs ²⁺ , Mg ²⁺ , Si(OH) ₄ , HPO ₄ ²⁻ , Cl ⁻ , Na ⁺ , K ⁺ |
| 1400 | Cells larger in diameter | Eukaryotic cells, mitosis uses actomyosin | Respiration | | >1 | | | |
| 2000 | Enlarged, thick-walled cells at intervals on algal filaments | Oxygen tolerating blue-green algae, protection against photo-oxidation | Fermentation | | 1 | | | |
| 2800 | Stromatolites, filamentous chains | Resemble living blue-green algae | | Uraninite | 0.1 | | | |
| >3500 | Stromatolites, depletion of ¹³ C | Precursors of blue-green algae active | | | <0.01 | | | |
| 3800 | Rhythmically banded rocks, depletion of ¹³ C | Microbial organisms (?) Biological activity (?) | | Banded iron [Fe(II)] | <0.01 | | | |

*PAL = present atmospheric level

